PSL UNIVERSITÉ PARIS

HABILITATION À DIRIGER DES RECHERCHES

DE L'UNIVERSITÉ PSL

Présentée à l'École normale supérieure

Decision and control in networks with stochastic demand and supply

Présentation des travaux par

Ana BUSIC

Le 21 mai 2024

Discipline

Informatique

Composition du jury :

Alain JEAN-MARIE	Directeur de recherche, Inria	*Rapporteur*
Rhonda RIGHTER	Professeur, UC Berkeley	*Rapporteuse*
Rayadurgam SRIKANT	Professeur, UIUC	*Rapporteur*
François BACCELLI	Directeur de recherche, Inria	*Examinateur*
Vivek BORKAR	Professeur émérite, IIT Bombay	*Examinateur*
Ramesh JOHARI	Professeur, Stanford University	*Examinateur*
Roland MALHAME	Professeur, Polytechnique Montréal	*Examinateur*
Catherine ROSENBERG	Professeur, University of Waterloo	*Examinatrice*

ENS | PSL

Contents

Chapter 1

Introduction

The objective of my research is to develop stochastic models and algorithms for performance analysis, optimization and control of complex networks. My contributions are on the intersection of applied probability, operations research, computer science and electrical engineering, and range from coupling techniques for Markov chains and perfect simulation algorithms, stochastic networks and probabilistic cellular automata, to reinforcement learning, and optimization and control in power networks.

This manuscript summarizes one part of my research. The focus is on the decision and control in networks with stochastic demand and supply that have to be balanced by a central entity, or using a distributed control design. Two different settings are considered: stochastic matching systems and real-time balancing of stochastic demand and supply in power grids. The objective is similar, yet the models differ considerably and use techniques from various fields including queueing systems and network flows from graph theory for stochastic matching, and mean-field approximations and control theory for real-time demand-supply balancing. Both settings use Markov decision processes.

The remaining of the introduction contains an overview of my scientific contributions, the organization of the manuscript, the list of publications since my PhD defense in 2007, and my curriculum vitae.

1.1 Scientific contributions

During my PhD, I worked on stochastic comparison of Markov chains. The main focus was on algorithmic construction of bounds for large Markov chains, with applications in reliability and performance evaluation. One of key properties I was investigating was stochastic monotonicity, under various stochastic orders (e.g. strong or usual stochastic order, convex, level-crossing ...). My interests in Markov processes evolved over the years in two directions:

- perfect sampling methods based on the coupling arguments for Markov chains,
- Markov reward and decision processes, and lately reinforcement learning.

I am interested in stochastic modeling and analysis of networks, and two particular types of networks that I focused on the most are stochastic matching systems and power networks. The overview of my contributions is organized in themes that are chronologically ordered according to the first time I started working in that area. The list of my publications since

my PhD defense is given in Section 1.3. The references starting with a letter refer to this list publications and the letter itself denotes the type of publication (B - book chapter, C - international conference, J - journal, N - national conference, P - patent, T - tool paper, W - preprint). The reference that do not start with a letter refer to the reference list at the end of this manuscript.

1.1.1 Markov (decision) processes

Bounds for Markov chains

I continued research on bounding methods for Markov chains based on stochastic comparison. I list here only the contributions that were not issued from the work in my PhD thesis. However, they use similar techniques and are issued from a collaboration with my PhD advisor, Jean-Michel Fourneau.

We developed stochastic bound algorithms for censored Markov chains in [C62, C59, C52]. Censored Markov chains can be used to study the conditional behavior of a system within a subset of observed states. They can also provide a theoretical framework to study the truncation of a discrete-time Markov chain when the generation of the state-space is too hard or when the number of states is too large [141].

In [C60, T3, J27] we proposed iterative algorithms to compute component-wise bounds of the steady-state distribution of an irreducible and aperiodic Markov chain. The proposed bounds are based on very simple properties of (max, +) and (min, +) sequences. Under some assumptions on the Markov chain, these bounds converge to the exact solution. In that case we have a clear tradeoff between computation and the tightness of bounds. Furthermore, at every step we know that the exact solution is within an interval, which provides a more effective convergence test than usual iterative methods for solving Markov chains.

In [J26] we considered two different applications of stochastic monotonicity in performance evaluation of networks. In the first one, we assume that a Markov chain of the model depends on a parameter that can be estimated only up to a certain level and we have only an interval that contains the exact value of the parameter. Instead of taking an approximated value for the unknown parameter, we show how we can use the monotonicity properties of the Markov chain to take into account the measurement error bounds. In the second application, we consider a well known approximation method: the decomposition into submodels. In such an approach, models of complex networks are decomposed into submodels whose results are then used as parameters for the next submodel in an iterative computation. This leads to a fixed point system which is solved numerically. We use stochastic monotonicity to obtain the existence proof of the solution of the fixed point system and a convergence proof of the iterative algorithm.

In [J18], we bound a discrete time Markov chain by a new chain with transition matrix that has a low rank decomposition. We show how the complexity of the analysis for steady-state and transient distributions can be simplified when we take into account this decomposition.

Structural properties in Markov decision processes

Right after my PhD, during a summer visit to CMU, I had the opportunity to meet Ingrid Vliegen, also visiting CMU for the summer. Ingrid was at that time a PhD student at TU/e, supervised by Geert-Jan van Houtum and Ton de Kok. Ingrid developed several heuristics for a complex inventory system, that were empirically always a lower and an upper bound.

Together with Ingrid and with Alan Sheller-Wolf, we developed in [J24] a new comparison technique for Markov reward processes. Our method constructs bounds for a Markov reward process by redirecting selected sets of transitions, facilitating an intuitive interpretation of the modifications on the original system. Redirecting sets of transitions is based on an extension of precedence relations to sets of states by van Houtum et al. [71], and allows to design more accurate bounds (e.g. bounds having the same mean behavior). We show that our method is compatible with strong aggregation of Markov chains; thus we can obtain bounds for the initial chain by analyzing a much smaller chain. We apply the precedence relations on set of states combined with aggregation to prove the bounds of order fill rates for an inventory system with joint demands and returns of items.

Precedence relation method, and our extension to sets of transitions, is closely related to the techniques based on structural properties of dynamic programming operator for Markov reward and decision processes, that have been largely used in queuing and inventory systems to obtain structural results for optimal policies.

Within a master internship of A. Wieczorek, co-supervised by Emmanuel Hyon, we considered lost sales inventory models with several classes of customers and investigated the optimality of critical level policies (i.e. policies defined by a set of thresholds) [C58]. We also studied in [C27, J5] structural properties and threshold-type policies in stochastic matching, which will be described further in Subsection 1.1.4.

Controlled Markov processes with stationary inputs

Within PhD thesis of Yue Chen, supervised by Sean Meyn, we studied the impact of a mean-field control on the individual quality of service in power grid applications (detailed in Subsection 1.1.5). This motivated the following new ergodic theory results for controlled Markov processes that are driven by a stationary input process [J12]: Consider a stochastic process $\boldsymbol{X}$ on a finite state space $\mathsf{X} = \{1, \ldots, d\}$, assumed to be conditionally Markov, given a real-valued 'input process' $\boldsymbol{\zeta}$. This input process is assumed to be small, which is modeled through the scaling,

$$\zeta_t = \varepsilon \zeta^1_t, \qquad 0 \le \varepsilon \le 1\,,$$

where $\boldsymbol{\zeta}^1$ is a bounded stationary process. Subject to smoothness assumptions on the controlled transition matrix and a mixing condition on $\boldsymbol{\zeta}$:

(i) A stationary version of the process is constructed, that is coupled with a stationary version of the Markov chain $\boldsymbol{X}^\bullet$ obtained with $\boldsymbol{\zeta} \equiv 0$. The triple $(\boldsymbol{X}, \boldsymbol{X}^\bullet, \boldsymbol{\zeta})$ is a jointly stationary process satisfying

$$\mathsf{P}\{X(t) \neq X^\bullet(t)\} = O(\varepsilon)$$

Moreover, a second-order Taylor-series approximation is obtained:

$$\mathsf{P}\{X(t) = i\} = \mathsf{P}\{X^\bullet(t) = i\} + \varepsilon^2 \pi^{(2)}(i) + o(\varepsilon^2), \quad 1 \le i \le d,$$

with an explicit formula for the vector $\pi^{(2)} \in \mathbb{R}^d$.

(ii) For any $m \ge 1$ and any function $f\colon \{1, \ldots, d\} \times \mathbb{R} \to \mathbb{R}^m$, the stationary stochastic process $Y(t) = f(X(t), \zeta(t))$ has a power spectral density S_f that admits a second order Taylor series expansion: A function $\mathrm{S}^{(2)}_f\colon [-\pi, \pi] \to \mathbb{C}^{m \times m}$ is constructed such that

$$\mathrm{S}_f(\theta) = \mathrm{S}^\bullet_f(\theta) + \varepsilon^2 \mathrm{S}^{(2)}_f(\theta) + o(\varepsilon^2), \quad \theta \in [-\pi, \pi]$$

in which the first term is the power spectral density obtained with $\varepsilon = 0$. An explicit formula for the function $\mathrm{S}^{(2)}_f$ is obtained, based in part on the bounds in (i).

ODE for MDPs

In tracking problems for power grid, described more in detail in Subsection 1.1.5, we encountered a new computational challenge: how to efficiently solve not one, but an entire family of Markov decision processes (MDP), parameterized by a scalar ζ that appears in the one-step reward function? In collaboration with Sean Meyn, we proposed in [J10] a new approach to computation of optimal policies for parameterized families of average-cost MDPs. For an MDP with d states, the family of relative value functions $\{h^*_\zeta : \zeta \in \mathbb{R}\}$ is the solution to an ordinary differential equation (ODE),

$$\tfrac{d}{d\zeta} h^*_\zeta = \mathcal{V}(h^*_\zeta)$$

where the vector field $\mathcal{V}\colon \mathbb{R}^d \to \mathbb{R}^d$ has a simple form, based on a matrix inverse.

Two general applications are presented: Brockett's quadratic-cost MDP model, and a generalization of the "linearly solvable" MDP framework of Todorov in which the one-step reward function is defined by Kullback-Leibler divergence. The latter was introduced in [132], where it was shown under general conditions that the solution to the average-reward optimality equations reduce to a simple eigenvector problem. Since then many authors have sought to apply this technique to control problems and models of bounded rationality in economics. A crucial assumption is that the input process is essentially unconstrained. For example, if the nominal dynamics include randomness from nature (eg, the impact of wind on a moving vehicle), then the optimal control solution does not respect the exogenous nature of this disturbance. In [C33] we introduce a technique to solve a more general class of action-constrained MDPs.

1.1.2 Perfect simulation of Markov chains

My interests in algorithmic construction of bounding chains lead to a post-doc at Inria Grenoble (2007-08) where I had an opportunity to work with Bruno Gaujal and Jean-Marc Vincent on perfect simulation of Markov chains, a simulation technique that is based on coupling constructions for Markov chains.

It is well known since the pioneering works of Loynes [93] and then Borovkov [17], that backwards schemes and specifically strong backwards coupling convergence, can lead to an explicit construction of the stationary state of the system, within its stability region. One can then use pathwise representations to compare systems in steady state (see Chapter 4 of [7] on such comparison results for queues).

Propp and Wilson [116] introduced a coupling-from-the-past algorithm (CFTP) (which essentially uses backwards coupling convergence), a powerful tool for simulating the steady state of the system, even whenever the latter distribution is not know in closed form. For an ergodic Markov chain with finite state space, CFTP provides an unbiased sample from the stationary distribution in finite expected time. In the general case, the algorithm starts trajectories from all states at some time in the past until time 0. If the final state is the same for all trajectories, then the chain has coupled and the final state has the stationary distribution of the Markov chain. Otherwise, the simulations are started further in the past. This algorithm is efficient under monotonicity assumptions that allows reducing the number of trajectories considered in the coupling from the past procedure only to extremal initial conditions.

In the non-monotone case, the original CFTP algorithm requires to generate one trajectory per state of the Markov chain, which limits its application only to chains with a state space of very small cardinality. Two general ideas of bounding processes, sandwiching all the trajectories of the original chain, have been proposed in the literature, based on a particular structure of the process: [84] assumes a partially ordered state space structure, while [75] constructs bounding chains evolving on the power set of the state space, and it is suitable for dynamics on graphs with local interactions.

Envelope perfect sampling

Motivated by the general bounding idea in [84], with Bruno Gaujal and Jean-Marc Vincent, we proposed in [C61] an algorithm to construct envelope bounding chains, for the case of a Markov chain on a lattice state space. The envelope technique has been implemented in a software tool PSI2 (Perfect SImulator) [T2].

More precisely, this envelope technique amounts to replace the initial equation $X_{n+1} = \Phi(X_n, U_{n+1})$ by a couple of equations:

$$\begin{aligned} M(t+1) &= \sup_{m(t)\leq x\leq M(t)} \Phi(x, U_{t+1}), \\ m(t+1) &= \inf_{m(t)\leq x\leq M(t)} \Phi(x, U_{t+1}). \end{aligned}$$

Starting from the extreme states in the state space, the couple $(m(t), M(t))$ always provide lower and upper bounds on the state $X(t)$. Therefore, whenever m and M meet, all trajectories of the original chain have coalesced and the algorithm returns a sample from a stationary distribution.

When the cardinality of the state space makes challenging even storing the state of the Markov chain, we proposed to combine the ideas of bounding processes and the aggregation of Markov chains [C55]. We illustrated the proposed approach of aggregated enveloppe bounding chains on assemble to order systems.

The enveloppe approach has two weak points:

- It requires the computation of the supremum and the infimum of the transition function Φ over all states in some lattice interval I. These computations should be done in sublinear time in $|I|$, in order to gain over the original CFTP.
- The coupling time of the envelope chain is larger than that of the original chain, and may even be infinite.

The former issue was addressed in [J23] where we showed that the enveloppe approach is particularly effective when the state space can be partitioned into pieces where envelopes can be easily computed. Most markovian queueing networks have this property. For the latter issue, in [C61] we propose a splitting technique: when the number of states contained in the interval of the enveloppe chain drops below a certain threshold, it is possible to "split" the interval into the remaining trajectories that are considered from that point on, until time 0.

Perfect sampling for queueing networks

In [C53,J17], we consider open Jackson queueing networks with mixed finite and infinite buffers and analyze the efficiency of sampling from their exact stationary distribution. We

show that perfect sampling is possible even when the underlying Markov chain has infinite state space. The main idea is to use a Jackson network with infinite buffers (that has a product form stationary distribution) to bound the number of initial conditions to be considered in the coupling from the past scheme. We also provide bounds on the sampling time of this new perfect sampling algorithm under hyper-stability conditions (defined in [J17]) for each queue. These bounds show that the new algorithm is considerably more efficient than existing perfect samplers even in the case where all queues are finite.

Within the PhD thesis of Cristelle Rovetta, that I co-advised with Anne Bouillard, our goal was to propose new efficient perfect sampling algorithms that are not necessarily based on a notion of a partially ordered state space. Our motivating example was a closed queueing network. When the queue capacity is finite, the stationary distribution has a product form only in a very limited number of particular cases and numerical algorithms are intractable due to the cardinality of the state space. Closed networks do not exhibit any monotonicity property and the global constraint on the total number of packets in the network prevents to directly use previously mentioned bounding chain approaches. In [J19], we derived a new bounding chain for closed queueing networks that is based on a compact representation of sets of states, reducing the complexity of the one-step transition in the CFTP algorithm to $\mathcal{O}(KM^2)$, where K is the number of queues and M the total number of customers (while the cardinality of the state space is exponential in the number of queues). The coupling time of the bounding chain is almost surely finite. In [C47] these results are extended to the multiclass case. CLONES (Closed Queueing Networks Exact Sampling), a Matlab toolbox for exact sampling of closed queueing networks developed by Christelle Rovetta, received the best-tool paper award at Valuetools 2015 [T1]. In [J15] we proposed a new representation, that leads to one-step transition complexity of the CFTP algorithm that is in $\mathcal{O}(KM)$.

Besides queueing networks, I used perfect sampling and bounding chains to study probabilistic cellular automata [C57, J20], described more in detail in Subsection 1.1.3.

Acceleration of perfect sampling

Within a master internship of Furcy Pin, co-supervised with Bruno Gaujal, we proposed a new method to speed up perfect sampling of Markov chains by skipping the passive events, i.e. an adaptive method for random event generation that considers only the events that can change the state of the envelope bounding chain. We proved that this can be done without altering the distribution of the samples [C56]. This technique is particularly efficient for the simulation of Markov chains with different time scales such as queueing networks where certain servers are much faster than others. In such cases, the coupling time of the Markov chain can be arbitrarily large while the runtime of the skipping algorithm remains bounded.

Within the PhD thesis of Rémi Varloot, we used perfect simulation for random generation of independent sets [C46]. The maximum independent set (MIS) problem is a well-studied combinatorial optimization problem that naturally arises in many applications, such as wireless communication, information theory and statistical mechanics. MIS problem is NP-hard, thus many results in the literature focus on fast generation of maximal independent sets of high cardinality. One possibility is to combine Gibbs sampling with CFTP algorithms. This results in a sampling procedure with time complexity that depends on the mixing time of the Glauber dynamics Markov chain. We proposed an adaptive method for random event generation for the Glauber dynamics that considers only the events that are effective in the CFTP scheme, accelerating the convergence time of the Gibbs sampling algorithm.

1.1.3 Probabilistic cellular automata

Dynamical systems with local interactions provide theoretical models for problems in distributed computing: gathering a global information by exchanging only local information. The challenge is two-fold: first, it is impossible to centralize the information (cells are indistinguishable); second, the cells contain only a limited information (represented by a finite alphabet). There are two natural instantiations of dynamical systems with local interactions: one with synchronous updates of the cells, and one with asynchronous updates. In the first case, time is discrete, all cells are updated at each time step, and the model is known as a Probabilistic Cellular Automaton (PCA) [134] The applications of PCA range from the designing of fault-tolerant computational models [60] to to statistical physics and life sciences. In the second case, time is continuous, cells are updated at random instants, at most one cell is updated at any given time, and the model is known as a (finite range) Interacting Particle System (IPS) [131]. IPS have wide range of applications, including sensor and wireless networks.

Within the PhD thesis of Irène Marcovici, supervised by Jean Mairesse, we investigated ergodicity of PCA. I initially joined the project to guide Irène in developing of a perfect simulation algorithm for PCA. With time, I got interested in the research area itself, so we continued collaborating during most of Irène's PhD.

A PCA on $\mathbb{Z}^d$ can be seen as Markov chains, with uncountable state space. The cells are updated synchronously and independently, according to a distribution depending on a finite neighborhood. In [C57, J20], we investigate the ergodicity of this Markov chain. We show that ergodicity of PCA is undecidable even in one dimensional case with deterministic update rule. We then propose an efficient perfect sampling algorithm for the invariant measure of an ergodic PCA. Our algorithm does not assume any monotonicity property of the local rule. It is based on a bounding process which is shown to be also a PCA.

Most of the results in the Markov chain literature are limited to the ergodic case. In many problems related to PCA, we are interested exactly in the opposite: in design of fault-tolerant systems it is important that the automaton keeps the knowledge of certain initial configurations in spite of the noise (faults) in the dynamics of the model. In [C54, J21], we address the density classification problem. Consider an infinite graph with nodes initially labeled by independent Bernoulli random variables of parameter p. Density classification problem consist in designing a (probabilistic or deterministic) cellular automaton or a finite-range interacting particle system that evolves on this graph and decides whether p is smaller or larger than $1/2$. Precisely, the trajectories should converge to the uniform configuration with only 0's if $p < 1/2$, and only 1's if $p > 1/2$. We present solutions to the problem on the regular grids of dimension d, for any $d > 1$, and on the regular infinite trees. For the bi-infinite line, we propose some candidates that we back up with numerical simulations, using our PCA perfect sampling algorithm.

1.1.4 Stochastic matching systems

The theory of matching has a long history in economics, mathematics, and computer science, with applications found in many other fields, such as health, ridesharing, power grid, or pattern recognition. Most of the research on matching considers the static setting. With the increased popularity of online advertising, various online bipartite matching models have been proposed that consider random arrivals of one population, while the other is static. Two

sided stochastic matching model was first proposed in Caldentey, Kaplan and Weiss [31]. They introduced FCFM (First Come First Matched) bipartite matching model with two infinite sequences of "customers" (also called demand in this manuscript) and "servers" (also called supply), motivated by the Boston area social housing problem and the PhD thesis of Kaplan [80]. In this bipartite matching formulation customers and servers play completely symmetric roles. Customers and servers of several types arrive to the system and wait for a compatible match. Compatibility is determined by a bipartite graph. The system is controlled by a matching policy that determines which of the available compatible items are matched together. FCFM policy matches a new arrival to the longest waiting compatible item in the system. Other systems where two sided models seem appropriate include ride sharing systems, organ transplants, assigning users to servers in distributed computing systems, assigning questions to experts in question-and-answer websites, job markets, and online advertising.

I started working on stochastic matching systems during my post-doc at LIAFA with Jean Mairesse, following our discussion with Gideon Weiss during Stochastic Networks conference in Paris in spring 2008. The initial idea was to study the FCFM model and show it admits a product form stationary distribution. This turned out to be a highly non-trivial task.

The first examples, with simple compatibility graphs were analyzed already in [31] where the authors also conjectured the matching rates, i.e. the fraction of customers of each type served by each type of server, for any bipartite compatibility graphs. The necessary and sufficient condition for stability and a product form stationary distribution for the bipartite matching model were derived by Adan and Weiss in [3]. The product form stationary distribution was derived by partial balance, similar to [137]. This stationary distribution was then used to derive expressions for the matching rates.

I describe here briefly my contributions on stochastic matching systems. This part will be covered in more detail in Part I of the manuscript.

FCFM stochastic matching

In a collaboration with Ivo Adan, Jean Mairesse et Gideon Weiss, we studied in depth the FCFM bipartite matching model in [J11], and developed product form stationary distribution results for various state descriptions. This also lead to a reversibility result for a well chosen state description, and thus a fine understanding of the product form results for this type of models. In addition to the matching rates, another important performance measure was derived, the distributions of link lags, i.e. the lags between matched customer server pairs.

The extension to the non-bipartite compatibility graphs, called the general stochastic matching model, was initially proposed by Mairesse and Moyal in [97]: there is a single i.i.d. sequence of items of several types, and a non-bipartite compatibility graph, and each item in the sequence is matched to an earlier compatible item if such exists, or it remains unmatched until it can be matched to a later item in the sequence. In a collaborative work [J7] with Pascal Moyal and Jean Mairesse we extended the reversibility and product form results from [J11] to non-bipartite compatibility graphs, under the FCFM policy.

In [C6], we extended this product form result to the case of compatibility graphs with self-loops. Similar result has been derived in parallel by Begeot et al. in [11] using a different proof technique.

These reversibility and product form results for FCFM policy greatly simplify the analysis, and can be also used to guide a system design with target performance properties [C51]. Also, FCFM policy does not favor any particular class, so it may be considered as fair by the users.

All these are very strong and appealing arguments to use FCFM policy in practice. However, in [C10] we show that FCFM policy can also exhibit an interesting performance paradox: increasing the matching possibilities by adding new edges to the compatibility graph may lead to a larger mean waiting times to find a compatible match. One may see adding an edge as increasing the flexibility of the system, for example asking a family registering for social housing to list less requirements in order to be compatible with more housing units. Therefore it may be natural to think that adding edges to the matching graph will lead to a decrease of the expected number of items in the system and the waiting time to be matched. We provide sufficient conditions for this performance paradox. These sufficient conditions are related to the heavy-traffic assumptions in queueing systems. The intuition behind is that the performance paradox occurs when the added edge in the compatibility graph disrupts the draining of a bottleneck, i.e. a set of nodes for which the arrival rates are very close to the arrival rates of their neighbors in the compatibility graph. This performance paradox is not fully surprising when analyzing a fixed policy. What is really intriguing is that the same paradox occurs even if we consider the whole family of greedy policies, i.e. the policies that always match a new arrival to a waiting item if there are any compatible items waiting in the system. An example is given in [W2]. In stochastic matching model, greedy matching policies can be interpreted as selfish behavior of new arrivals, so this performance paradox is to some extent similar to a Braess paradox observed in transportation networks [20]. Braess paradox states that, when the agents can take self-interested decisions, the travelling times of the agents can increase if we add a new road. The idea behind this phenomenon is that the extension of the network might cause a redistribution of the traffic that increases the congestion and, as a result, the delay of agents. More precisely, the Braess paradox shows that the travel time in the Nash equilibrium (the set of strategies such as no agent has incentive to deviate unilaterally) can increase if we add a shortcut in the network. This result reflects that the selfish behavior of agents in a network might lead to a situation whose performance is sub-optimal. The existence of a Braess paradox has been explored in several contexts related to queueing networks [10, 33, 46, 47, 79].

Stability and Loynes-type constructions

A drawback of FCFM policy is the need to consider the order of arrival of items into the system, which leads to a state space that consists of all finite words of item classes. In [J22] we extended the bipartite stochastic matching model to the class of admissible policies that can be described as markovian greedy policies, i.e. the policies that only depend on the current state of the system and that always match new arrivals with compatible waiting items. Also, we extended the arrival process to include possible correlations between demand and supply items, while the previously mentioned results for FCFM policy always assumed independence between the arrival supply and demand item sequences. We call this setting the Extended Bipartite Matching (EBM) model. We considered stability properties of EBM model under various greedy matching policies, including ML (match the longest), MS (match the shortest), FCFM (match the oldest), RANDOM (match uniformly), and PRIORITY. We identified necessary conditions for stability (independent of the matching policy) defining the maximal possible stability region. For some bipartite graphs, we prove that the stability region is indeed maximal for any admissible matching policy. For ML policy, we prove that the stability region is maximal for any bipartite graph. For MS and PRIORITY policies, we exhibit a bipartite graph with a non-maximal stability region. An extension of these

stability results to non-bipartite compatibility graphs, called *general matching model* (GM) was proposed by Mairesse and Moyal in [97].

In [W7], we propose an explicit construction of the stationary state of EBM model. We use a Loynes-type backwards scheme, allowing to show the existence and uniqueness of a bi-infinite perfect matching under various conditions, for a large class of matching policies and of bipartite matching structures. The key algebraic element of our construction is the sub-additivity of a suitable stochastic recursive representation of the model, satisfied under most usual matching policies. By doing so, we also derive stability conditions for the system under general stationary ergodic assumptions, subsuming the classical markovian settings. The extension to GM is studied in [W4]. We prove that most common matching policies (including FCFM, priorities and random) satisfy a particular sub-additive property, which we exploit to show in many cases, the coupling-from-the-past to the steady state, using a backwards scheme *à la* Loynes. We then use these results to explicitly construct perfect bi-infinite matchings, and to build a perfect simulation algorithm in the case where the buffer of the system is finite.

Optimization

Within the PhD thesis of Arnaud Cadas, we considered holding costs for the items that are waiting to be matched. We model this problem as an MDP (Markov decision process) and study the discounted cost and the average cost case. In [J11], we first consider a model with two types of supply and two types of demand items with an N-shaped matching graph. For linear cost function, we prove that an optimal matching policy gives priority to the pendant edges of the matching graph and is of threshold type for the diagonal edge. In addition, for the average cost problem, we compute the optimal threshold value. We then show how the obtained results can be used to characterize the structure of an optimal matching control for a quasi-complete graph with an arbitrary number of nodes. For arbitrary bipartite graphs, we show that, when the cost of the pendant edges is larger than the one of the neighbors, an optimal matching policy prioritizes the items in the pendant edges. We also provide an example that shows that it is not optimal to prioritize items in the pendant edges when the cost of the pendant edges is smaller than the one of the neighbors. Conference version [C27] obtained best paper award at the conference VALUETOOLS 2019.

In a collaborative work with Sean Meyn, in [C45] we considered the infinite-horizon average-cost optimal control problem and proposed a relaxation of the stochastic control problem, which is found to be a special case of an inventory model, as treated in the classical theory of Clark and Scarf [45]. The optimal policy for the relaxation admits a closed-form expression. Based on the policy for this relaxation, a new h-MaxWeight with threshold policy is proposed (described more in detail in Section 5.3). For a parameterized family of models in which the network load approaches capacity, this policy is shown to be approximately optimal, with bounded regret, even though the average cost grows without bound.

Matching rates

Applying a combination of a graph-theory and linear-algebra approach, in a collaborative work with Céline Comte and Fabien Mathieu, in [N1,W3] we analyze the efficiency of matching policies, not only in terms of system stability, but also in terms of matching rates between different classes. The matching model we consider is essentially a GM model, but relaxing

the assumptions on greedy policies. More precisely, we consider a matching problem in which items of different classes arrive according to independent Poisson processes. Unmatched items are stored in a queue, and compatibility constraints are described by a simple graph on the classes, so that two items can be matched if their classes are neighbors in the graph. Our results rely on the observation that, under any stable policy, the matching rates satisfy a conservation equation that equates the arrival and departure rates of each item class. We first introduce a mapping between the dimension of the solution set of this conservation equation, the structure of the compatibility graph, and the existence of a stable policy. In particular, this allows us to derive a necessary and sufficient stability condition that is verifiable in polynomial time. We describe the convex polytope of non-negative solutions of the conservation equation. When this polytope is reduced to a single point, we give a closed-form expression of the solution; in general, we characterize the vertices of this polytope using again the graph structure. Finally, we study which vectors of the polytope can be achieved by a stable policy. We show that the set of vectors reached by stable greedy policies is included in the interior of the polytope, and that the inclusion is strict in general. In contrast, we conjecture that non-greedy policies can reach any point of the interior of the polytope; whether they can also reach the boundary of the polytope depends on a simple condition on the vertices.

1.1.5 Optimization and control in power networks

The power system transformation brings new challenges and opportunities due to changes and uncertainties in electricity consumption and generation. In power networks, it is necessary that the electricity production is equal to the demand at all times. In addition to ensuring sufficient electricity production, there is also a need for flexible resources that can quickly adapt their production / consumption to compensate for demand forecasting errors and ensure real-time balancing between production and demand. This service is an example of the system services (also called ancillary services) essential to the proper functioning of power grids.

Matching electricity supply and demand used to be relatively straightforward, with large and controllable power plants on the one hand, and demand that was relatively easy to predict on the other. Slowly ramping cheaper generators were committed in advance to follow the predicted demand. The real time balancing was done by ramping up or down the most responsive power plants, such as gas turbines or hydro, when available. They were operating at lower capacity, leaving the possibility to ramp up or down their generation. This was providing balancing reserves used to correct the forecasting errors and follow the demand in real time. In recent years, there has been a significant increase in participation of intermittent renewable generation. Balancing service from traditional power plants is becoming very expensive due to the need to compensate for the missed opportunity cost the power plants are facing while operating at a lower set-point to be able to ramp up and down more aggressively than in the past. At the same time, the rapid development of "smart technologies" (e.g. Linky meter and the connected appliances) has opened new possibilities for innovation on the demand side, as well as new control solutions on the grid level.

Energy storage is one possible solution to facilitate this power network transformation. Within the PhD thesis of Md Umar Hashmi, we considered energy storage control problems both at the level of individual consumers minimizing the cost of electricity and at the grid level for increasing reliability and stability of the power network.

My main contributions in the area of control for power networks is a new distributed control approach for balancing the power grid using flexible loads. The proposed approach

relies on new smart technologies allowing for automatic control of devices. The objective is to control a great amount of devices to provide services to the system (load shaping or ancillary services) while: i) maintaining the quality of service for the users; ii) minimizing communications between controllable devices and the central controller. The proposed approach combines controlled Markov processes and mean-field models. This is a collaborative work with colleagues from the University of Florida, started during my sabbatical at the University of Florida in 2014, and then greatly accelerated by the Inria Associate Team PARIS (2015-18) and Inria International Chair of Sean Meyn (2019-25). Significant part of this research was done through collaborations involving PhD students:

- Inria: Md Umar Hashmi, Thomas Le Corre;
- University of Florida: Yue Chen, Austin R. Coffman, Joel Mathias, Neil Cammardella.

Electricity markets are also expected to undergo a significant transformation, evolving from centralized to decentralized structures, driven by digitalization, large-scale integration of renewable energy sources (RES) and distributed energy resources (DER), and active prosumer engagement. These changes motivate new market models incorporating decentralized structures, large-scale RES and DER inclusion, and the strategic behavior of prosumers. In PhD thesis of Ilia Shilov, co-advised by Hélène Le Cadre (Inria Lille), Gonçalo de Almeida Terça and Anibal Sanjab (VITO, Belgium), we studied peer-to-peer (P2P) electricity markets.

Optimization and control of batteries

Within the PhD thesis of Md Umar Hashmi, we considered energy storage control problems both at the level of individual consumers minimizing the cost of electricity and at the grid level for increasing reliability and stability of the power network.

Electricity consumers with local renewable generation, such a rooftop solar, can use a battery to minimize their electricity bills. We considered storage optimization problems under time-varying electricity prices with different net-metering policies. Using convex optimization tools, an optimal storage control policy is developed with threshold-based structure. The proposed algorithm is computationally efficient with quadratic worst-case complexity with respect to the horizon length [C38]. The extension of the battery control problem that considers the health of the battery taking into account its cycle life was studied in [C34].

Due to their high cost, batteries may need to be used for more than one dedicated application to be financially viable. In the co-optimization formulations, we considered storage performing energy arbitrage under net metering along with power factor correction, peak demand shaving, and energy backup for power outages [C18, C26, J9]. These formulations are evaluated on case studies using real data for low voltage consumers in Madeira [C19, C28] and the proposed control policies for batteries are designed taking into account the consumer contracts proposed by the local utility.

Large-scale storage applications for ancillary services were considered in [C32]. A case study is provided for energy storage revenue estimation, essential for analyzing financial feasibility of investment in batteries. This case study considers a battery that is used for electricity price based arbitrage and ancillary services for load balancing in real time. Using PJM's (a regional transmission organization in the United States) real data, we estimate short and long term financial potential for batteries. We take into account battery degradation, based on the operational cycles of the battery with various depths of discharge (DoD) and the calendar

life of the battery. A simple control mechanism is proposed to control cycles of operation in order to increase the long term gains of a battery performing arbitrage and ancillary services, under this battery degradation model. This case study suggests that participating in ancillary services is more beneficial for storage owners compared to energy arbitrage.

Control of distributed power demand

Providing new flexibility resources is crucial to integrate renewable energies into the power grid. There is an enormous flexibility potential in the power consumption of the majority of electric loads (e.g., thermal loads such as water heaters, air-conditioners and refrigerators; electric vehicles, etc.). Their power consumption can be shifted in time to some extent without any significant impact to the consumer needs. This flexibility can be exploited to create "virtual batteries". The best example of this is the heating, ventilation, and air conditioning (HVAC) system of a building: There is no perceptible change to the indoor climate if the airflow rate is increased by 10% for 20 minutes, and decreased by 10% for the next 20 minutes. Power consumption deviations follow the airflow deviations closely, but indoor temperature will be essentially constant.

The major issue lies in the distributed nature of this flexibility resource: piloting the flexible demand in real time requires a design of (simple) incentives for millions of devices. Moreover, many residential devices are on-off (e.g. water-heater or air-conditioner). In order to provide valuable balancing service, the aggregate must provide a predictable response. The future power grids will contain millions of smart components, which completely prohibits centralized decision making using standard stochastic optimization techniques, such as stochastic dynamic programming and Markov decision processes (MDP), as they do not scale well with the number of different components in the system (both the state space and the control space of the model grow exponentially with the number of components).

The answer proposed in [C50, J16] is based on probabilistic distributed demand control. Our approach combines the techniques from the theory of controlled Markov processes, mean-field theory, and automatic control. The objective is to control the average consumption of a population of N devices to track the reference signal (R_t), which is progressively revealed by the grid at discrete time steps $t = 1, ..., T$ (online reference tracking problem). Through load-level and grid-level control design, high-quality ancillary service for the grid is obtained without impacting quality of service delivered to the consumer. This approach to grid regulation is called demand dispatch: loads are providing service continuously and automatically, without consumer interference.

A starting point in our research was the fact that many of the ancillary services needed today are defined by a power deviation reference signal that has zero mean. Examples are PJM's RegD signal, or BPA's balancing reserves. We have demonstrated in [C44] that loads can be classified based on the frequency bandwidth of ancillary service that they can offer. In [C48] we focused on the issue of individual risk and in [C49] on the stability properties on the grid level for our local control architecture. The survey of the approach can be found in the book chapter [B2] and will be also summarized in Chapter 6 of this manuscript. Our patent [P1] has been exploited by an international company working on smart thermostats.

A theoretical contribution that I would like to highlight in particular, and that will be presented more in detail in Chapter 7, was motivated by the need to extend the initial distributed control approach to include the randomness that cannot be controlled (e.g. hot water usage or the weather conditions) [C41]. This lead to a new ODE method for solving

a parametrized family of Markov Decision Processes [J10]. Besides power applications, this new technique also has its potential applications in machine learning and robotics [C33].

Four application questions were explored through collaborations involving PhD students:

- Within the PhD thesis of Yue Chen (University of Florida, advisor S. Meyn), we investigated the QoS aspects for the individual consumers [C43, J14, J13] by developing new tools for state estimation for the individual and the population in the mean-field control setting, based on the ergodic theory for controlled Markov chains with stationary inputs [J12].
- Within the PhD thesis of Austin R. Coffman (University of Florida, advisor P. Barooah), the main focus was on the application of the distributed control to the Thermostatically Controlled Loads (such as fridges or air-conditioners) and the estimation of their flexibility capacity under time-varying weather conditions and cycling constraints [C35, C30, C21, C11, J2].
- Within the PhD thesis of Umar Hashmi an extension of distributed stochastic control is proposed for a fleet of batteries for tracking fast timescale supply and demand imbalance [C37].
- Within the PhD thesis of Joel Mathias (University of Florida, advisor S. Meyn), the extensions of the approach to the heterogeneous population of loads have been proposed [C42, C40, C31, J3].

Demand dispatch has many advantages:

- *minimal communication:* a unique control signal is sent from the central entity to the loads, without the communication from the loads to the centralized entity;
- local control design enables strict guarantees for the quality of service for the users;
- randomized control limits the synchronization of the response of the loads.

However, in this online reference tracking formulation of the problem, the target consumption is revealed in real time (there is no anticipation of the target by probabilistic forecasts). The fact of not allowing any anticipation of the target consumption does not make it possible to fully integrate the constraints of the different devices in terms of energy consumed over a given period. To overcome this limitation, we have proposed an offline reference tracking approach that takes into account a deterministic forecast of the target consumption over a period of anticipation (e.g. day ahead) and solves directly the tracking problem at the population level, formalized as a Kullback-Leibler-Quadratic (KLQ) optimal control problem in discrete [C20], or continuous time [C22]. This new Kullback-Leibler-Quadratic (KLQ) control approach can be seen as a special case of a finite horizon stochastic optimal control problem with the objective function that is composed of two terms: quadratic tracking error cost and a relative entropy control cost that penalizes the deviation from the nominal behavior of the load. An overview for the discrete time case is provided in Chapter 8. Within the PhD thesis of Neil Cammardella (University of Florida, advisor S. Meyn), we investigated in particular the discrete KLQ problem and its implementations based on different information architectures for the reference tracking problem by a distributed population of flexible loads [C20]. In [C8, J1] we considered an extension of KLQ to include the randomness that cannot be controlled (e.g. hot water usage or the weather conditions). Simultaneous allocation and control problem for distributed energy resources based on KLQ was considered in [C13].

P2P electricity market design and analysis

Within a PhD thesis of Ilia Shilov, co-advised with Hélène Le Cadre (Inria Lille), Gonçalo de Almeida Terça and Anibal Sanjab (VITO, Belgium), we consider different designs for peer-to-peer electricity markets, with main focus on related equilibrium problems and their properties. The PhD thesis of Ilia started while Hélène was at VITO Belgium. I met Hélène through PGMO (Gaspard Monge Program for optimization, operations research and their interactions with data sciences) community, and I invited her in 2016 to participate to my ANR JCJC PARI project. We planned to collaborate on new market designs that can accommodate demand dispatch. Due to various circumstances, our collaboration got postponed and finally started in 2019, on a topic of peer to peer (P2P) electricity markets. Our collaboration initiated me to game theory. So far, I was only contributing by applying known concepts and properties to energy network problems. In future, I hope to be able to spend some more time studying (and hopefully also contribute to) the algorithmic game theory.

In [P5], we consider a financial P2P market in which prosumers optimize their demand, generation, and bilateral trades in order to minimize their costs subject to local constraints and bilateral trading reciprocity coupling constraints, leading to Generalized Nash Equilibrium Problem (GNEP) formulations. This financial market is further analyzed in [C9], where this initial model was generalized to include the interaction with the physical level of the distribution grid, managed by the distribution system operator. We model the interaction problem between the distribution grid level and the financial level as a noncooperative GNEP. We compare two designs of the financial level prosumer market: a centralized design and a peer-to-peer fully distributed design. We prove the Pareto efficiency of the equilibria under homogeneity of the trading cost preferences. The study demonstrates the Pareto efficiency of market equilibrium and how the proposed pricing structure limits free-riding behavior on the financial level.

In [J8], we consider impact of a privacy mechanism in a peer-to-peer electricity market. The problem is modeled as a noncooperative communication game, which takes the form of a GNEP where the agents determine their randomized reports to share with the other market players, while anticipating the form of the peer-to-peer market equilibrium. We characterize the equilibrium of the game, prove the uniqueness of the variational equilibrium and provide a closed form expression of the privacy price.

In [W6], we investigate equilibrium problems arising in a decentralized electricity market involving risk-averse prosumers, having a possibility to hedge their risks through financial contracts that they can purchase from an insurance company or trade directly with their peers. We formulate the problem as a Stackelberg game where the insurance company acts as the leader while the prosumers behave as followers. We show that the Stackelberg game pessimistic formulation might have no solution. We propose an equivalent reformulation as a parametrized generalized Nash equilibrium problem, and characterize the set of equilibria. We prove that the insurance company can design price incentives that guarantee the existence of a solution of the pessimistic formulation, which is ϵ-close to the optimistic one. We then derive economic properties of the Stackelberg equilibria such as fairness, equity, and economic efficiency. We also quantify the impact of the insurance company incomplete information on the prosumers' risk-aversion levels on its individual cost and social cost.

Imprecise forecasts of RES-generation introduce additional uncertainty for the agents, which impacts their trading in P2P markets, consequently affecting the market's efficiency. However, the advent of data markets, which facilitate the exchange of data to improve forecast

accuracy, can mitigate this issue. In [C5], we explore the potential coupling between forecast and P2P electricity markets to mitigate the uncertainty from imprecise RES-generation forecasts, allowing agents to acquire a forecast of their RES-based generation. The electricity market is modeled as a two-stage peer-to-peer market, cast in the form of a GNEP. Conditions for the efficiency of the P2P market are identified, with a key condition being prosumers' participation in the forecast market. Furthermore, conditions on the probability distributions of the forecasts that assure this property are discussed.

1.1.6 Reinforcement learning

In spring 2018, I was a long term participant of the Real-time decision making semester at Simons Institute, UC Berkeley. In this highly vibrant environment, I participated to a reading group on reinforcement learning (RL) and developed a strong interest in this field, both in theoretical aspects of RL as well as its applications.

Along with the sharp increase in visibility of the field, the rate at which new RL algorithms are being proposed is at a new peak. While the surge in activity is creating excitement and opportunities, there is a gap in understanding of two basic principles that these algorithms need to satisfy for any successful application. One has to do with guarantees for convergence, and the other concerns the convergence rate.

Many RL algorithms belong to a class of learning algorithms known as stochastic approximation (SA). SA algorithms are recursive techniques used to obtain the roots of functions that can be expressed as expectations of a noisy parameterized family of functions. Book chapter [B1] provides a survey on SA approach to reinforcement learning. I briefly summarize in the following my contributions in this area, with a highlight on algorithm design and analysis using SA approach.

My main application domain for RL is energy. Two projects I have been working on recently are wind-farm production optimization and optimal control of a battery for demand charge minimization.

Stochastic approximation and reinforcement learning

Acceleration is an increasingly common theme in the stochastic optimization literature. The two most common examples are Nesterov's method, and Polyak's momentum technique. In [C23] two new SA algorithms are introduced: 1) PolSA is a root finding algorithm with specially designed matrix momentum, and 2) NeSA can be regarded as a variant of Nesterov's algorithm, or a simplification of PolSA. The PolSA algorithm is new even in the context of optimization (when cast as a root finding problem). It is well known that most variants of TD- and Q-learning may be cast as SA algorithms, and the tools from general SA theory can be used to investigate convergence and bounds on convergence rate. In particular, the asymptotic variance is a common metric of performance for SA algorithms, and is also one among many metrics used in assessing the performance of stochastic optimization algorithms. There are two well known SA techniques that are known to have optimal asymptotic variance: the Ruppert-Polyak averaging technique, and stochastic Newton-Raphson (SNR). The former algorithm can have extremely bad transient performance, and the latter can be computationally expensive. It is demonstrated here that parameter estimates from the new PolSA algorithm couple with those of the ideal (but more complex) SNR algorithm. The new algorithm is thus a third approach to obtain optimal asymptotic covariance. These strong results require

assumptions on the model. A linearized model is considered, and the noise is assumed to be a martingale difference sequence. Numerical results are obtained in a non-linear setting: In PolSA implementations of Q-learning it is observed that coupling occurs with SNR.

Motivated by applications in Markov chain Monte Carlo (MCMC) and RL, in [C14] we studied error bounds for recursive equations subject to Markovian disturbances. Many of MCMC and RL algorithms can be interpreted as special cases of SA. It is argued that it is not possible in general to obtain a Hoeffding bound on the error sequence, even when the underlying Markov chain is reversible and geometrically ergodic, such as the M/M/1 queue. This is motivation for the focus on mean square error bounds for parameter estimates. It is shown that mean square error achieves the optimal rate of $\mathcal{O}(1/n)$, subject to conditions on the step-size sequence. Moreover, the exact constants in the rate are obtained, which is of great value in algorithm design.

Zap Q-learning [52] is a recent class of reinforcement learning algorithms, motivated primarily as a means to accelerate convergence. It is based on a two time-scale stochastic approximation algorithm, constructed so that the matrix gain tracks the gain that would be used in a deterministic Newton-Raphson method. Stability theory was only provided for the tabular setting. A tutorial on Zap Q-learning can be found in [C29]. In [C16] we prove the convergence of the Zap-Q-learning algorithm for optimal stopping problem, under the assumption of linear function approximation setting. We use ODE analysis for the proof, and the optimal asymptotic variance property of the algorithm is reflected via fast convergence in a finance example. In [C15], a new framework is proposed for analysis of a broad class of SA algorithms. A special case of this new class of SA algorithms leads to a significant generalization of Zap Q-learning, for which convergence theory is obtained even in a nonlinear function approximation setting. The reliability in neural network function approximation architectures is tested through simulations.

Reinforcement learning for wind farm control

In the PhD thesis of Claire Bizon Monroc, that I co-advise with Donatien Dubuc and Jiamin Zhu (IFP Energies Nouvelles) within a joint IFPEN - Inria lab project, we investigate the possibility of using multi-agent RL algorithms for wind farm production optimization.

The power output of a wind farm is influenced by wake effects, a phenomenon in which upstream turbines facing the wind create sub-optimal conditions for turbines located downstream. Misaligning the yaw, defined as the angle between the rotor and the wind direction, is an efficient strategy to deflect the wake away from downstream turbines. This technique is known as wake steering, and has been shown to increase total production compared to the naive greedy strategy where all turbines face the wind [56, 72].

Designing efficient methods to find optimal yaw angles is a challenging task. Several classical model-based optimization methods have been proposed [85], but they are subject to model inaccuracies, ignore wake dynamics and lack adaptability. Furthermore, the complexity of this optimization problem increases with the number of turbines in the wind farm, making centralized control strategies quickly intractable for real time optimization. Deployment of any control method for real-time optimization on wind farms requires accounting for the dynamic propagation of the wind inflow.

This motivated the use of model-free decentralized RL methods. A multi-agent RL approach was proposed in [85], with static simulations, i.e. without taking into account the dynamic wake propagation.

To account for wake propagation times, in [C7] we propose a delay-aware decentralized Q-learning algorithm for yaw control on wind farms. We build on algorithm proposed in [85], but introduce a strategy to handle delayed cost collection, and show that our method significantly increases power production in simulations with realistic wake dynamics, using mid-fidelity wind farm simulator FAST.Farm [77], developed by National Renewable Energy Laboratory (NREL).

This tabular algorithm however relies on state and control space discretization. This comes with an important cost in terms of parameters to learn, and raises issues regarding the algorithm's ability to scale to larger wind farms. In [C4] we show that more efficient learning agents can be designed under the same principles of decentralization and delay-awareness. We employ actor critic agents learning in parallel with function approximation, a method that we call Delay-Aware Fourier Actor Critics (DFAC). We test this method in WFSim, a control-oriented wind farm simulator developed by TUDelft, that takes into account wake propagation dynamics which has been validated against large eddy simulations [16]. Numerical experiments on WFSim for wind farms with up to 32 turbines show that this method has great scaling potential and achieves faster coordination and convergence than the previous approach, leading to more important increases in energy production.

Reinforcement learning based demand charge minimization

A high peak demand causes high electricity costs for both the utility and end-users. Utilities have introduced peak-demand charges to encourage customers to reduce their peak demand. Within PhD thesis of Lucas Weber, co-advised by Jiamin Zhu (IFP Energies Nouvelles), we considered customers who are equipped with an energy storage device and renewable energy production. We proposed a reinforcement learning approach to reduce their electricity bills, which are composed of both energy and demand charges [C3]. We validated our approach on real data from an office building of IFPEN Solaize site.

1.2 Organization of the manuscript

The manuscript is organized in two parts: Part I covers stochastic matching systems and Part II balancing of stochastic demand and supply in power grids. The two parts can be read independently of each other.

Part I is organized in 4 chapters. Chapter 2 introduces stochastic matching models. Chapter 3 provides an overview of stability results, and it is based on [J22]. Chapter 4 is devoted to the First Come First Matched discipline, based on [J11, J7, C10]. Chapter 5 summarizes the optimization results for bipartite stochastic matching models, based on [J5, C45].

Part II is organized in 3 chapters: Chapter 6 introduces the problem of balancing in the power grid and the need for flexible resources, and introduces our distributed control approach for demand dispatch, and it is based on [B2]. Chapter 7 summarizes the ODE method for Markov decision processes, used for demand dispatch in the case of stochastic disturbances (e.g. weather conditions), based on [C33, J10]. Chapter 8 surveys the new Kullback-Leibler-Quadratic optimal control problem and its applications to demand dispatch, based on [C8, J1].

1.3 List of publications (since PhD defense)

The names of PhD or master students at the time of the work that lead to the publications are underlined. The names of PhD students under my formal supervision at the time of the work that lead to the publications are highlighted in bold.

Journals

[J1] N. Cammardella, A. Busic, S. Meyn. *Kullback–Leibler-Quadratic Optimal Control.* SIAM Journal on Control and Optimization 61(5): 3234-3258, 2023.

[J2] A. Coffman, A. Busic, P. Barooah. *A unified framework for coordination of thermostatically controlled loads.* Automatica 152, 111002, 2023.

[J3] J Mathias, A Busic, S. Meyn. *Load-Level Control Design for Demand Dispatch With Heterogeneous Flexible Loads.* IEEE Transactions on Control Systems Technology. 31(4): 1830 - 1843, 2023.

[J4] Md U. Hashmi, D. Deka, A. Busic, D. Van Hertem. *Can locational disparity of prosumer energy optimization due to inverter rules be limited?* IEEE Transactions on Power Systems. 38(6): 5726 - 5739, 2023.

[J5] **A Cadas**, J Doncel, A Busic. *Analysis of an optimal policy in dynamic bipartite matching models.* Performance Evaluation 154, 102286, 2022.

[J6] **S. Samain**, J. Doncel, A. Busic, J.-M. Fourneau. *Multiclass Energy Packet Networks with finite capacity energy queues.* Perform. Evaluation 152: 102228, 2021.

[J7] P. Moyal, A. Busic, J. Mairesse. *A product form for the general stochastic matching model.* J. Appl. Probab. 58(2): 449-468, 2021.

[J8] **I. Shilov**, H. Le Cadre, A. Busic. *Privacy impact on generalized Nash equilibrium in peer-to-peer electricity market.* Oper. Res. Lett. 49(5): 759-766, 2021.

[J9] **Md U. Hashmi**, D. Deka, A. Busic, L. Pereira, S. Backhaus. *Arbitrage with power factor correction using energy storage.* IEEE Transactions on Power Systems 35(4):2693-2703, 2020.

[J10] A. Busic, S. Meyn. *Ordinary Differential Equation Methods For Markov Decision Processes and Application to Kullback-Leibler Control Cost.* SIAM Journal on Control and Optimization, 56 (1): 343-366, 2018.

[J11] I. Adan, A. Busic, J. Mairesse, G. Weiss. *Reversibility and further properties of FCFS infinite bipartite matching.* Mathematics of Operations Research, 43(2): 598-621, 2018.

[J12] Y. Chen, A. Busic, S. Meyn. *Ergodic theory for controlled Markov chains with stationary inputs.* The Annals of Applied Probability, 28(1):79-111, 2018.

[J13] Y. Chen, A. Busic, S. Meyn. *Estimation and Control of Quality of Service in Demand Dispatch.* IEEE Transactions on Smart Grid, 9(5): 5348-5356, 2018.

[J14] Y. Chen, A. Busic, S. Meyn. *State Estimation for the Individual and the Population in Mean Field Control with Application to Demand Dispatch.* IEEE Transactions on Automatic Control, 62(3): 1138-1149, 2017.

[J15] A. Bouillard, A. Busic, **C. Rovetta**. *Low complexity state space representation and algorithms for closed queueing networks exact sampling.* Performance Evaluation, 103: 2-22, 2016.

[J16] S. Meyn, P. Barooah, A. Busic, Y. Chen, J. Ehren. *Ancillary Service to the Grid Using Intelligent Deferrable Loads.* IEEE Transactions on Automatic Control, 60(11):2847 - 2862, 2015.

[J17] A. Busic, S. Durand, B. Gaujal, F. Perronnin. *Perfect sampling of Jackson queueing networks.* Queueing Systems, 80(3):223-260, 2015.

[J18] A. Busic, J.-M. Fourneau, M. Ben Mamoun. *Stochastic Bounds with a Low Rank Decomposition.* Stochastic Models, 30(4): 494-520, 2014.

[J19] A. Bouillard, A. Busic, **C. Rovetta**. *Perfect sampling for closed queueing networks.* Performance Evaluation, 79:146-159, 2014.

[J20] A. Busic, N. Fates, J. Mairesse, I. Marcovici. *Density classification on infinite lattices and trees.* Electronic Journal of Probability, 18(51):1-22, 2013.

[J21] A. Busic, J. Mairesse, I. Marcovici. *Probabilistic cellular automata, invariant measures, and perfect sampling.* Advances in Applied Probability, 45(4):960-980, 2013.

[J22] A. Busic, V. Gupta, J. Mairesse. *Stability of the bipartite matching model.* Advances in Applied Probability, 45(2):351-378, 2013.

[J23] A. Busic, B. Gaujal, F. Pin. *Perfect Sampling of Markov Chains with Piecewise Homogeneous Events.* Performance Evaluation, 69(6):247-266, 2012.

[J24] A. Busic, I. Vliegen, A. Scheller-Wolf. *Comparing Markov Chains: Aggregation and Precedence Relations Applied to Sets of States, with Applications to Assemble-to-Order Systems.* Mathematics of Operations Research, 37(2):259-287, 2012.

[J25] R. Nair, E. Miller-Hooks, R. C. Hampshire, A. Busic. *Large-Scale Vehicle Sharing Systems: Analysis of Vélib'.* International Journal of Sustainable Transportation, 7(1):85-106, 2012.

[J26] A. Busic, J.-M. Fourneau. *Monotonicity and performance evaluation: applications to high speed and mobile networks.* Cluster Computing, 15(4): 401-414, 2012.

[J27] A. Busic, J.-M. Fourneau. *Iterative component-wise bounds for the steady-state distribution of a Markov chain.* Numerical Linear Algebra with Applications, 18(6):1031-1049, 2011.

Book chapters

[B1] A. M. Devraj, A. Busic, S. Meyn. Fundamental Design Principles for Reinforcement Learning Algorithms. In: Vamvoudakis, K.G., Wan, Y., Lewis, F.L., Cansever, D. (eds) Handbook of Reinforcement Learning and Control. Studies in Systems, Decision and Control, vol 325, Springer, Cham. 2021.

[B2] Y. Chen, **Md Umar Hashmi**, J. Mathias, A. Busic, and S. Meyn. *Distributed Control Design for Balancing the Grid Using Flexible Loads.* In: Energy Markets and Responsive Grids, pp. 383-411. S. Meyn, T. Samad, I. Hiskens, J. Stoustrup (eds). The IMA Volumes in Mathematics and its Applications, vol 162. Springer, New York, NY. 2018.

International conferences

[C1] A. Busic, J.-M. Fourneau. *Stochastic Matching Model with Returning Items.* European Workshop on Performance Engineering (EPEW 2023), 186-200, 2023.

[C2] A. Busic, S. Meyn, N. Cammardella. *Learning Optimal Policies in Mean Field Models with Kullback-Leibler Regularization.* The 62nd IEEE Conference on Decision and Control (CDC 2023), Marina Bay Sands, Singapore, 38-45, 2023.

[C3] **L. Weber**, A. Busic, J. Zhu. *Reinforcement learning based demand charge minimization using energy storage.* The 62nd IEEE Conference on Decision and Control (CDC 2023), Marina Bay Sands, Singapore, 4351-4357, 2023.

[C4] **C. Bizon Monroc**, A. Busic, D. Dubuc, J. Zhu. *Actor Critic Agents for Wind Farm Control.* The 2023 American Control Conference (ACC), May 31 - June 2, 2023 — San Diego, CA, USA, 177-183, 2023.

[C5] **I. Shilov**, H. Le Cadre, A. Busic, A. Sanjab, P. Pinson. *Towards Forecast Markets For Enhanced Peer-to-Peer Electricity Trading.* 2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Glasgow, Scotland, 2023.

[C6] A. Busic, **A. Cadas**, J. Doncel, J.-M. Fourneau. *Product form solution for the steady-state distribution of a Markov chain associated with a general matching model with self-loops.* EPEW 2022: 18th European Performance Engineering Workshop, Sep 2022, Santa Pola, Alicante, Spain.

[C7] **C. Bizon Monroc**, E. Bouba, A. Busic, D. Dubuc, J. Zhu. *Delay-Aware Decentralized Q-learning for Wind Farm Control.* 61st IEEE Conference on Decision and Control (CDC 2022), Cancun, Mexico, pp. 807-813, 2022.

[C8] N. Cammardella, A. Busic and S. Meyn. *Kullback-Leibler-Quadratic Optimal Control in a Stochastic Environment*, 60th IEEE Conference on Decision and Control (CDC 2021), Austin, TX, USA, pp. 158-165, 2021.

[C9] **I. Shilov**, H. L. Cadre and A. Busic. *A Generalized Nash Equilibrium analysis of the interaction between a peer-to-peer financial market and the distribution grid*, 2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Aachen, Germany, pp. 21-26, 2021.

[C10] **A Cadas**, J. Doncel, J.-M. Fourneau, A. Busic. *Flexibility can Hurt Dynamic Matching System Performance*. Short paper at 39th International Symposium on Computer Performance, Modeling, Measurements and Evaluation 2021. Proceedings: SIGMETRICS Perform. Evaluation Rev. 49(3): 37-42 (2022). Long version: CoRR abs/2009.10009, `https://arxiv.org/abs/2009.10009`.

[C11] A. R. Coffman, A. Busic and P. Barooah. *Control-oriented modeling of TCLs*, 2021 American Control Conference (ACC), New Orleans, LA, USA, pp. 4148-4154, 2021.

[C12] D. Kiedanski, A. Busic, D. Kofman and A. Orda. *Efficient distributed solutions for sharing energy resources at the local level: a cooperative game approach*, 2020 59th IEEE Conference on Decision and Control (CDC), Jeju, Korea (South), pp. 2634-2641, 2020.

[C13] N. Cammardella, A. Busic and S. Meyn. *Simultaneous Allocation and Control of Distributed Energy Resources via Kullback-Leibler-Quadratic Optimal Control*, 2020 American Control Conference (ACC), Denver, CO, USA, pp. 514-520, 2020.

[C14] S. Chen, A. M. Devraj, A. Busic, S. Meyn. *Explicit Mean-Square Error Bounds for Monte-Carlo and Linear Stochastic Approximation*. The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020), 4173-4183, 2020.

[C15] S. Chen, A. M. Devraj, F. Lu, A. Busic , Sean P. Meyn. *Zap Q-Learning With Nonlinear Function Approximation.* 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

[C16] S. Chen, A.M. Devraj, A. Busic, S. Meyn. *Zap Q-Learning for optimal stopping*. 2020 American Control Conference (ACC 2020), 3920-3925, 2020.

[C17] **S. Samain**, J. Doncel, A. Busic, J-M. Fourneau. *Energy Packet Networks with Finite Capacity Energy Queues*. 13th EAI International Conference on Performance Evaluation Methodologies and Tools (Valuetools 2020), 142-149, 2020.

[C18] **Md U. Hashmi**, A. Busic, D. Deka, L. Pereira. *Energy Storage Optimization for Grid Reliability*. Proceedings of the Eleventh ACM International Conference on Future Energy Systems. (e-Energy 2020), 516-522, 2020.

[C19] **Md U. Hashmi**, J. Cavaleiro, L. Pereira, A. Busic. *Sizing and Profitability of Energy Storage for Prosumers in Madeira, Portugal*. IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT 2020), Washington DC, USA, February 17-20, 2020.

[C20] N. Cammardella, A. Busic, Y. Ji, S. P. Meyn. *Kullback-Leibler-Quadratic Optimal Control of Flexible Power Demand*. 58th Conference on Decision and Control (CDC 2019). Nice, France, December 11-13, 2019.

[C21] A.R. Coffman, A. Busic, P. Barooah. *Aggregate capacity for TCLs providing virtual energy storage with cycling constraints*. 58th Conference on Decision and Control (CDC 2019). Nice, France, December 11-13, 2019.

[C22] A. Busic, S. P. Meyn. *Distributed Control of Thermostatically Controlled Loads: Kullback-Leibler Optimal Control in Continuous Time*. 58th Conference on Decision and Control (CDC 2019). Nice, France, December 11-13, 2019.

[C23] A. M. Devraj, A. Busic, S. P. Meyn. *On Matrix Momentum Stochastic Approximation and Applications to Q-learning.* 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton 2019). Monticello, IL, USA, September 24-27, 2019.

[C24] **Md U. Hashmi**, A. Mukhopadhyay, A. Busic, J. Elias, D. Kiedanski. *Optimal Storage Arbitrage under Net Metering using Linear Programming.* IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (IEEE SmartGridComm 2019). Beijing, China, 21-23 October, 2019.

[C25] D. Kiedanski, **Md U. Hashmi**, A. Busic, D. Kofman. *Sensitivity to Forecast Errors in Energy Storage Arbitrage for Residential Consumers.* IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (IEEE SmartGridComm 2019). Beijing, China, 21-23 October, 2019.

[C26] **Md U. Hashmi**, D. Deka, A. Busic, L. Pereira, S. Backhaus. *Co-optimizing energy storage for prosumers using convex relaxations.* 20th International Conference on Intelligent Systems Applications to Power Systems (ISAP 2019). New Delhi, India, December 10-14, 2019.

[C27] **A. Cadas**, A. Busic, J. Doncel. *Optimal Control of Dynamic Bipartite Matching Models.* 12th EAI International Conference on Performance Evaluation Methodologies and Tools (Valuetools 2019), Palma de Mallorca, Spain, March 13-15, 2019. **Best paper award.**

[C28] **Md U. Hashmi**, L. Pereira, A Busic. *Energy storage in Madeira, Portugal: Co-optimizing for arbitrage, self-sufficiency, peak shaving and energy backup.* 13th IEEE PowerTech 2019, Milano, Italy, 2019.

[C29] A. M. Devraj, A. Busic, S. Meyn. *Zap Q-Learning – A User's Guide.* The 5th Indian Control Conference (ICC 2019). IIT Delhi, January 9-11, 2019.

[C30] A. R. Coffman, A. Busic, and P. Barooah. *Virtual Energy Storage from TCLs using QoS persevering local randomized control.* The 5th ACM International Conference on Systems for Built Environments (BuildSys 2018). Shenzen, China, November 7-8, 2018.

[C31] N. Cammardella, J. Mathias, M. Kiener, A. Busic, S. Meyn. *Balancing California's Grid Without Batteries.* 57th IEEE Conference on Decision and Control (CDC 2018). Miami Beach, FL, December 17-19, 2018.

[C32] **Md U. Hashmi**, W. Labidi, A. Busic, S-E. Elayoubi, and T. Chahed. *Long-Term Revenue Estimation for Battery Performing Arbitrage and Ancillary Services.* IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (IEEE SmartGridComm 2018). Aalborg, Denmark, October 29-31, 2018.

[C33] A. Busic, S. Meyn. *Action-Constrained Markov Decision Processes With Kullback-Leibler Cost.* Proceedings of the 31st Conference On Learning Theory (COLT 2018). Stockholm, Sweden, July 6–9, 2018. PMLR 75:1431-1444, 2018.

[C34] **Md U. Hashmi**, A. Busic. *Limiting Energy Storage Cycles of Operation.* IEEE Green Technologies Conference (GreenTech 2018). Austin, TX, USA, April 4 - 6, 2018.

[C35] A. R. Coffman, A. Busic, and P. Barooah. *A Study of Virtual Energy Storage From Thermostatically Controlled Loads Under Time-Varying Weather Conditions.* 5th International High Performance Building Conference at Purdue. July 9-12, 2018.

[C36] **Md U. Hashmi**, D. Muthirayan, A. Busic. *Effect of Real-Time Electricity Pricing on Ancillary Service Requirements.* 1st International Workshop on Energy Market Engineering. Proceedings of the Ninth International Conference on Future Energy Systems (ACM e-Energy 2018 Workshops). Karlsruhe, Germany, June 12, 2018.

[C37] A. Busic, **Md U. Hashmi**, S. Meyn. *Distributed control of a fleet of batteries.* The 2017 American Control Conference (ACC 2017). Seattle, WA, USA, May 24-26, 2017.

[C38] **Md U. Hashmi**, A. Mukhopadhyay, A. Busic, J. Elias. *Optimal control of storage under time varying electricity prices.* IEEE International Conference on Smart Grid Communications (IEEE SmartGridComm 2017). Dresden, Germany, 23-26 October 2017.

[C39] **S. Samain**, A. Busic. *Exact Computation and Bounds for the Coupling Time in Queueing Systems.* The 11th EAI International Conference on Performance Evaluation Methodologies and Tools (Valuetools 2017). Venice, Italy. December 5-7, 2017.

[C40] J. Mathias, A. Busic, S. Meyn. *Demand Dispatch with Heterogeneous Intelligent Loads.* 50th Annual Hawaii International Conference on System Sciences (HICSS 2017). Waikoloa, HI, USA, January 4-7, 2017.

[C41] A. Busic, S. Meyn. *Distributed Randomized Control for Demand Dispatch.* 55th IEEE Conference on Decision and Control (CDC 2016). Las Vegas, NV, USA, December 12-14, 2016.

[C42] J. Mathias, R. Kaddah, A. Busic, S. Meyn. *Smart Fridge / Dumb Grid? Demand Dispatch for the Power Grid of 2020.* 49th Annual Hawaii International Conference on System Sciences (HICSS 2016). Koloa, HI, USA, January 5-8, 2016.

[C43] Y. Chen, A. Busic, S. Meyn. *State Estimation for the Individual and the Population in Mean Field Control with Application to Demand Dispatch.* 54th IEEE Conference on Decision and Control (CDC 2015). Osaka, Japan, December 15-18, 2015.

[C44] P. Barooah, A. Busic, S. Meyn. *Spectral Decomposition of Demand-Side Flexibility for Reliable Ancillary Services in a Smart Grid.* 48th Annual Hawaii International Conference on System Sciences (HICSS 2015). Kauai, HI, USA, January 5-8, 2015.

[C45] A. Busic, S. P. Meyn. *Approximate optimality with bounded regret in dynamic matching models.* SIGMETRICS 2015. Proceedings in SIGMETRICS Perform. Evaluation Rev. 43(2): 75-77, 2015. Long version: CoRR abs/1411.1044, `https://arxiv.org/abs/1411.1044`.

[C46] **R. Varloot**, A. Busic, A. Bouillard. *Speeding up Glauber Dynamics for Random Generation of Independent Sets.* SIGMETRICS 2015: 461-462. 2015. Long version: CoRR abs/1504.04517, `https://arxiv.org/abs/1504.04517`.

[C47] A. Bouillard, A. Busic, **C. Rovetta**. *Perfect Sampling for Multiclass Closed Queueing Networks.* 12th International Conference on Quantitative Evaluation of SysTems (QEST 2015). Madrid, Spain, September 1-3, 2015.

[C48] Y. Chen, A. Busic, S. Meyn. *Individual risk in mean-field control models for decentralized control, with application to automated demand response.* 53rd IEEE Conference on Decision and Control (CDC 2014). Los Angeles, CA, USA, December 15-17, 2014.

[C49] A. Busic, S. Meyn. *Passive Dynamics in Mean Field Control.* 53rd IEEE Conference on Decision and Control (CDC 2014). Los Angeles, CA, USA, December 15-17, 2014.

[C50] S. Meyn, P. Barooah, A. Busic, and J. Ehren. *Ancillary service to the grid from deferrable loads: the case for intelligent pool pumps in Florida.* 52nd IEEE Conference on Decision and Control (CDC 2013). Florence, Italy, December 10-13, 2013.

[C51] I. J. B. F. Adan, M. A. A. Boon, A. Busic, J. Mairesse, G. Weiss. *Queues with skill based parallel servers and a FCFS infinite matching model.* The Workshop on MAthematical performance Modeling and Analysis (MAMA 2013). Pittsburgh, PA, USA, June 21, 2013. ACM SIGMETRICS Performance Evaluation Review 41(3): 22-24, 2013.

[C52] A. Busic, H. Djafri, J.-M. Fourneau. *Bounded state space truncation and censored Markov chains.* 51st IEEE Conference on Decision and Control (CDC 2012). Maui, HI, USA, December 10-13, 2012.

[C53] A. Busic, B. Gaujal, F. Perronnin. *Perfect Sampling of Networks with Finite and Infinite Capacity Queues.* 19th International Conference on Analytic and Stochastic Modelling Techniques and Applications (ASMTA 2012). LNCS 7314, Springer-Verlag, pp. 136-149. Grenoble, France, June 4-6, 2012.

[C54] A. Busic, N. Fates, J. Mairesse, I. Marcovici. *Density Classification on Infinite Lattices and Trees.* 10th Latin American Symposium on Theoretical Informatics (LATIN 2012). LNCS 7256, Springer-Verlag, pp. 109-120. Arequipa, Peru, April 16-20, 2012.

[C55] A. Busic, E. Coupechoux. *Perfect Sampling with Aggregated Envelopes.* 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton 2011). Monticello, IL, USA, August 28-30, 2011.

[C56] F. Pin, A. Busic, B. Gaujal. *Acceleration of Perfect Sampling by Skipping Events.* 5th International ICST Conference on Performance Evaluation Methodologies and Tools (Valuetools 2011). Cachan, France, May 16-20, 2011.

[C57] A. Busic, J. Mairesse, I. Marcovici. *Probabilistic cellular automata, invariant measures, and perfect sampling.* 28th International Symposium on Theoretical Aspects of Computer Science (STACS 2011). Dortmund, Germany, March 10-12, 2011.

[C58] A. Wieczorek, A. Busic, E. Hyon. *Critical Level Policies in Lost Sales Inventory Systems with Different Demand Classes.* 8th European Performance Engineering Workshop (EPEW 2011), LNCS 6977, Springer-Verlag, pp. 204-218. Borrowdale, The English Lake District, UK, October 12-13, 2011.

[C59] A. Busic, H. Djafri, J.-M. Fourneau. *Stochastic bounds for censored Markov chains.* 6th International Workshop on the Numerical Solution of Markov Chains (NSMC 2010), Williamsburg, VA, USA, September 16-17, 2010.

[C60] A. Busic, J.-M. Fourneau. *Iterative component-wise bounds for the steady-state distribution of a Markov chain.* 6th International Workshop on the Numerical Solution of Markov Chains (NSMC 2010), Williamsburg, VA, USA, September 16-17, 2010.

[C61] A. Busic, B. Gaujal, J.-M. Vincent. *Perfect Simulation and Non-monotone Markovian Systems.* 3rd International Conference on Performance Evaluation Methodologies and Tools (Valuetools 2008). Athens, Greece, October 20-24, 2008.

[C62] A. Busic, J.-M. Fourneau. *Stochastic Complement and Strong Stochastic Bounds Based on Algebraic Properties.* 5th European Performance Engineering Workshop (EPEW 2008). LNCS 5261, Springer-Verlag, pp. 227-241. Palma de Mallorca, Spain, September 24-25, 2008.

International conferences (tool papers)

[T1] A. Bouillard, A. Busic and **C. Rovetta**. *Clones: CLOsed queueing Networks Exact Sampling.* 8th International Conference on Performance Evaluation Methodologies and Tools (Valuetools 2014). Bratislava, Slovakia, December 9-11, 2014. **Best tool paper award.**

[T2] A. Busic, B. Gaujal, G. Gorgo, J.-M. Vincent. *PSI2 : Envelope Perfect Sampling of Non Monotone Systems.* 7th International Conference on Quantitative Evaluation of Systems (QEST 2010). Williamsburg, VA, USA, September 15 - 18, 2010.

[T3] A. Busic, J.-M. Fourneau. *A toolbox for component-wise bounds of the steady-state distribution of a DTMC.* 7th International Conference on Quantitative Evaluation of Systems (QEST 2010). Williamsburg, VA, USA, September 15 - 18, 2010.

National conferences with proceedings

[N1] C. Comte, F. Mathieu, A. Busic. *Appariements, polytopes et dons d'organes.* AlgoTel 2022 - 24‘emes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications. 2022. `https://hal.science/hal-03656130`.
Long version: Stochastic dynamic matching: A mixed graph-theory and linear-algebra approach. CoRR abs/2112.14457, 2021. `https://arxiv.org/abs/2112.14457`.

Preprints

[W1] T. Le Corre, A. Busic, S. Meyn. *Feature Projection for Optimal Transport.* HAL preprint, hal-03923065v1, 2023. `https://hal.science/hal-03923065v1`.

[W2] A. Busic, A. Cadas, J. Doncel, Jean-Michel Fourneau. *Performance Paradox of Dynamic Matching Models under Greedy Policies.* HAL preprint, hal-04137823v1, 2023. `https://hal.science/hal-04137823v1`.

[W3] C. Comte, F. Mathieu, A. Busic. *Stochastic dynamic matching: A mixed graph-theory and linear-algebra approach.* HAL preprint, hal-03502084v4 , 2023. `https://hal.science/hal-03502084v4`.

[W4] P. Moyal, A. Busic, J. Mairesse. *On the sub-additivity of stochastic matching.* ArXiv:2305.00187, 2023. `https://arxiv.org/abs/2305.00187`.

[W5] I. Shilov, H. Le Cadre, A. Busic. *Risk-Averse Equilibrium Analysis and Computation.* CoRR abs/2004.02470, 2020. `https://arxiv.org/abs/2004.02470`.

[W6] I. Shilov, H. Le Cadre, A. Busic, G. de Almeida Terça. *A Stackelberg Game Analysis of Risk-Hedging Strategies in Decentralized Electricity Markets.* HAL preprint, hal-03674562, 2022. `https://hal.science/hal-03674562v2`.

[W7] P. Moyal, A. Busic, J. Mairesse. *Loynes construction for the extended bipartite matching.* ArXiv:1803.02788, 2018. `https://arxiv.org/abs/1803.02788`.

Patents

[P1] S. Meyn, A. Busic. Using loads with discrete finite states of power to provide ancillary services for a power grid. US Patent 10692158, 2020.

1.4 Curriculum vitae

Ana BUŠIĆ
Born in Zagreb (Croatia), December 1977
Citizenship: Croatian, French
http://www.di.ens.fr/~busic/

2 Rue Simone Iff, 75012 Paris, France
Email: ana.busic@inria.fr
Phone: +33 (0)1 80 49 43 35

Education

- **2007: PhD in Computer Science**, Université de Versailles Saint-Quentin.
  Thesis: Stochastic comparison of Markov models: an algorithmic approach and its applications in reliability and performance evaluation. Advisor: Prof. Jean-Michel FOURNEAU. Defended: July 2007.
- **2003: Master Degree in Applied Mathematics**, Université de Versailles St-Quentin.
  Thesis: Qualitative simulation of metabolic networks.
- **2002: B.S. in Mathematics**, Department of mathematics, University of Zagreb, Croatia.

Current position

- **Since 2009: Research Scientist (CR) Inria**, Inria Paris and Computer Science Department of Ecole normale supérieure (DI ENS, UMR 8548), Paris, PSL University, France.
- **Since 2022: Adjunct Professor for AI** at PSL University.
- **Since 2013:** Member of the Laboratory of Information, Networking and Communication Sciences (LINCS); https://www.lincs.fr/

Previous positions

- **2008 - 2009: Postdoc**, CNRS and University Paris Diderot - Paris 7.
- **2007 - 2008: Postdoc**, INRIA Grenoble - Rhône-Alpes.

Career break

- **February - June 2020:** maternity leave.

International research visits (1 month or longer)

- **August - December 2020:** Long-term participant of Theory of Reinforcement Learning program, Simons Institute, UC Berkeley (online due to COVID-19).
- **Spring 2019:** Program participant (5 weeks), The mathematics of energy systems, Isaac Newton Institute for Mathematical Sciences, Cambridge, UK. https://www.newton.ac.uk/event/mes
- **March - Mai 2018:** Visiting Scientist, Real-Time Decision Making program, Simons Institute, UC Berkeley.
- **March - August 2014:** University of Florida (4 months) and MIT (2 months).
- **June 2009:** EURANDOM, Technische Universiteit Eindhoven.

– **August - September 2007:** Carnegie Mellon University.

Supervision of graduate students

– **PhD students:**
  – Shu Li (ENS, PSL). Thesis topic: *Learning in dynamic matching models.* Started: November 2023.
  – Thomas Le Corre (ENS, PSL). Thesis topic: *Distributed control of flexible demand in power networks.* Started: November 2021.
  – Claire Bizon-Monroc (ENS, PSL). Co-supervised with Jiamin Zhu (IFPEN) and Donatien Dubuc (IFPEN). Thesis topic: *Deep reinforcement learning with constraints.* Started: November 2021.
  – Lucas Weber (ENS, PSL). Co-supervised with Jiamin Zhu (IFPEN). Thesis topic: *Connections between reinforcement learning and control.* Started: October 2021.
  – Ilia Shilov (ENS, PSL). Co-supervised with Hélène Le Cadre (Inria Lille) and Gonçalo de Almeida Terça (VITO, Belgium). Thesis topic: *Algorithmic Game and Distributed Learning for Peer-to- Peer Energy Trading.* Defended: September 2023.
  – Arnaud Cadas (ENS, PSL). Thesis title: *Stochastic matching models and their applications to demand-supply balancing.* Defended: November 2021.
  – Umar Hashmi (ENS, PSL). Thesis title: *Optimization and control of storage in smart grids.* Defended: December 2019.
  – Rémi Varloot (ENS Paris). Co-supervised with Laurent Massoulié. Thesis title: *Dynamic network formation.* Defended: June 2018.
  – Christelle Rovetta (ENS Paris). Co-supervised with Anne Bouillard. Thesis title: *Applications of perfect sampling to queuing networks and random generation of combinatorial objects.* Defended: June 2017.
– **Master or equivalent:** Thomas Le Corre (stage EDF; M2 MDA Paris Saclay and CentraleSupélec, 2021); Eva Bouba (M2 IASD Paris-Dauphine, PSL, 2021); Victor Vermès (M2 PMA, Sorbonne University and ENPC, 2021); Arnaud Cadas (M2 Statistics, UPMC, 2017); Augustin Pauchon (M2 MPRO, CNAM, 2017); Rémi Varloot (ENS Paris, France, 2013); Julieta Bollati (National Univ. of Rosario, Argentina, 2012); Aleksander Wieczorek (Univ. of Poznan, Poland, 2011); Furcy Pin (ENS Paris, France, 2010-2011).

Teaching Activities:

- At **Paris Dauphine, PSL - Mathematics, Machine Learning, Sciences, and Humanities (M2 MASH) and Mathematics of Insurance, Economics and Finance (M2 MASEF)**: *Reinforcement Learning* (master level, 2nd year), 2023.
- At **Parisian Master of Research in Computer Science (M2 MPRI)**: *Foundations of network models* (master level, 2nd year), 2023, 2022, 2021, 2020, 2018, 2017, 2016, 2014, 2013.

- At **Ecole normale supérieure, PSL**: *Network models and algorithms* (master level, 1st year), 2023, 2022, 2021, 2020, 2019, 2018, 2017, 2016, 2015, 2014; *Communication Networks* (undergraduate, 3rd year), 2013, 2012, 2011; *Random structures and algorithms* (undergraduate, 3rd year), 2023, 2022, 2021, 2019, 2018, 2017.
- At **UPMC Sorbonne Universités**: *Conception of algorithms and applications* (undergraduate, 3rd year), 2013, 2012, 2011, 2010.
- At **Université de Versailles (UVSQ) - M2 AMIS**: *Simulation* (master level, 2nd year), 2016, 2015, 2012, 2011.

Responsibilities

- **Scientific lead** of ARGO (Apprentissage, graphes et optimisation distribuée) common project-team between Inria and ENS, PSL; since October 2023.
- **Scientific lead** of Inria Associate Team PARIS (Probabilistic Algorithms for Renewable Integration in Smart grid) with University of Florida; `http://www.inria.fr/en/associate-team/paris`, 2015 - 2017.
- Since 2014: **Co-lead** of the working group COSMOS (Stochastic Control and Optimization, Modeling, and Simulation) of CNRS research network GdR RO; `http://gdrro.lip6.fr/?q=node/78`; scientific organization of one workshop per year (30-50 participants).
- Since 2017: Member of DI ENS Laboratory Board (Conseil du laboratoire).

Juries and service to the scientific community

- **Member of the hiring committee** for Junior research positions at Inria (2015 at Inria Saclay; 2016 at Inria Paris).
- **Member of the hiring committee** for Assistant Professor at Ecole Normale Supérieure (2024 and 2017) and IUT Orsay (2017).
- **Co-president of the hiring committee** for PhD, post-doc and visiting professor positions at Inria Paris (2016 and 2017); served as a committee member 2012 - 2017, 2021 - present.
- **Member of the Committee for the Technological Development** at Inria Paris (funding for software development projects), 2015 - 2022.
- **Examiner for PhD:** A. Ben Ameur (Télécom SudParis), 2023; K. Khun (Université de Grenoble), 2023; E. Anton (Toulouse, INPT), 2021; J. Horta (Télécom ParisTech, France), 2018, A. Ugolnikova, LIPN (Université Paris Nord, France), 2016; R. Kaddah (Télécom ParisTech, France), 2016; P.-A. Brameret (ENS Cachan, France), 2015.
- **Organizing Committee:** Flexible operation and advanced control for energy systems, INI Cambridge `https://www.newton.ac.uk/event/mesw01/` (2019); ALEA Days, CIRM, `https://conferences.cirm-math.fr/1776.html` (2018).
- **TPC co-chair:** Valuetools 2015. **TPC member:** ACM SIGMETRICS/Performance 2012; IFIP Performance 2013, 2014; ACM SIGMETRICS 2023, 2015; Valuetools 2013, 2014, 2016; IEEE SmartGridComm 2014, 2015; WiOpt 2015; QEST 2016, 2017.
- **Associate Editor** for Operations Research journal, Data Science and Machine Learning Area

- **Reviewing for journals:** Annals of App. Probability; Manag. Science; IEEE Trans. Automat. Control.; IEEE Trans. on Parallel and Distrib. Syst.; Perf. Evaluation; The Comp. Journal; Probab. in the Eng. and Inform. Sciences; Queueing systems; Stochastics.
- **Reviewer for ANR** (The French National Research Agency).

Awards and fellowships

- 2017: PEDR Inria (confirmed level)
  2015: Google Faculty Research Award
- 2014: Prime d'excellence scientifique Inria (junior level)
- 2002: French Governement Scholarship (BGF Master 2) A selective, merit-based scholarship for one year of graduate studies in France (5 fellowships in 2002 for Croatia)

Projects and grants

- **PI for partner Inria Paris** of the project AI-NRGY (Distributed AI-based architecture of future energy systems integrating very large amounts of distributed sources) of PEPR TASE (Technologies Avancées des Systèmes Energétiques), within the France 2030 Program, 2023 - 2017.
- Research grant by VITO, Belgium. Topic: Algorithmic Game and Distributed Learning for Peer-to- Peer Energy Trading, 2020-2023.
- Research grant by EDF. Topic: demand dispatch of flexible loads, 2019-2021.
- **PI** of the national project ANR JCJC PARI (Probabilistic Approach for Renewable Energy Integration: Virtual Storage from Flexible Loads); 2017 - 2021; 48 months.
- **PI** of PGMO research project *Decentralized control for renewable integration in smartgrids*, 2015 - 2017.
- **PI for partner Inria Paris** of the national project ANR MARMOTE (MARkovian MOdeling Tools and Environments). Started: Jan 2013 - Jan 2017. `https://wiki.inria.fr/MARMOTE/Welcome.`
- **PI** of INRIA ARC OCOQS (Optimal threshold policies in COntrolled Queuing Systems); 2011-13. `http://www.di.ens.fr/~busic/OCOQS`.
- **Participant:** National projects: ANR SMS (2005-08); ANR MAGNUM (2010–14), `http://www-apr.lip6.fr/anrMagnum/`; European project: *EuroNGI (Next Generation Internet).* `http://eurongi.enst.fr`.

Part I

Stochastic matching systems

Chapter 2

Stochastic matching models

Stochastic matching systems combine graph theoretic matching models with queueing systems. We consider two main variants: non-bipartite and bipartite stochastic matching. In the non-bipartite case, the items arrive one by one into the system and wait until they find a compatible match. Once matched, the items leave the system. The main difference in the bipartite case is that the items arrive two by two. Although the bipartite stochastic matching was introduced earlier than the non-bipartite model, we will start with the non-bipartite case that is notationally easier to describe and that will motivate the arrivals by pairs in the bipartite stochastic matching model.

This chapter sets the notation and the definitions of both models. Chapter 3 describes the stability results for the bipartite stochastic matching model and the connections with the non-bipartite case. Chapter 4 is devoted to the First Come First Matched discipline, under which the stochastic matching model has product form result, both in bipartite and non-bipartite case. This result allows to study analytically the performance of the matching system and in particular, the expected number of unmatched items in the steady state. We will also show that FCFM policy suffers from the following performance paradox: adding an edge to the matching graph (and therefore increasing the potential matches) does not necessarily decrease the mean number of waiting items in the steady-state. In Chapter 5 we present the results on the optimal control in the stochastic bipartite matching systems. We start with analyzing in detail the N-graph and then move to the asymptotic results for any bipartite graphs. We will introduce a new family of matching policies, called h-Max-Weight policies with threshold and show that there is a policy in this class that has bounded regret for the average number of waiting items.

This part is based on the following publications: [J22] (Chapter 3), [J7, J11, C10] (Chapter 4), [J5, C45] (Chapter 5).

General notation. Denote by $\mathbb{R}$ the set of reals, by $\mathbb{N}$ the set of non-negative integers and by $\mathbb{N}_+ = \mathbb{N}\backslash\{0\}$, the subset of positive integers. For any two integers m and n, denote by $[\![m,n]\!] = [m,n] \cap \mathbb{N}$. For any $n \in \mathbb{N}_+$, let $\mathfrak{S}_n$ be the group of permutations of $[\![1,n]\!]$. Let A^* (respectively, $A^{\mathbb{N}}$) be the set of finite (resp., infinite) words over the alphabet A. Denote by $\emptyset$ the empty word of A^*. For any word $w \in A^*$ and any subset B of A, we let $|w|_B$ be the number of occurrences of elements of B in w. For any letter $a \in A$, we denote $|w|_a := |w|_{\{a\}}$, and for any finite word w we let $|w| = \sum_{a\in A} |w|_a$ be the *length* of w. Finally, for any $w \in A^*$, let $[w] := (|w|_a)_{a\in A} \in \mathbb{N}^A$ be the commutative image of w.

Consider a simple undirected graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes the set of nodes, and

$\mathcal{E} \subset \mathcal{V}^2 \setminus \Delta$ is the set of edges, where $\Delta = \{(i,i) : i \in \mathcal{V}\}$, i.e. without self-loops. We use the notation $u-v$ for $(u,v) \in \mathcal{E}$ and $u \not- v$ for $(u,v) \notin \mathcal{E}$. For $U \subset \mathcal{V}$, we define $U^c = \mathcal{V} \setminus U$ and

$$\Gamma(U) = \{v \in \mathcal{V} : \exists u \in U,\ u - v\}.$$

An *independent set* of G is a non-empty subset $\mathcal{I} \subset \mathcal{V}$ which does not include any pair of neighbors, *i.e.* $\big(\forall i,j \in \mathcal{V},\ i-j \Rightarrow i \notin \mathcal{I} \text{ or } j \notin \mathcal{I}\big)$. Let $\mathbb{I}(G)$ be the set of independent sets of G. An independent set $\mathcal{I}$ is said to be *maximal* if $\mathcal{I} \cup \{j\} \notin \mathbb{I}(G)$ for any $j \notin \mathcal{I}$.

2.1 Non-bipartite stochastic matching model

Non-bipartite stochastic matching model was first introduced in [97] under the name of *stochastic matching model*. It is also know in the literature under the name *general stochastic matching model* (GM) [110]. Items arrive one by one in the system and depart by pairs, upon being matched. There is a finite set of item classes, denoted by $\mathcal{V}$, and identified with $[\![1, |\mathcal{V}|]\!]$. The compatibility between item classes is given by the *matching graph*, a connected simple undirected graph $G = (\mathcal{V}, \mathcal{E})$, where the set of edges $\mathcal{E}$ are the allowed matchings between classes. Two examples of matching graphs are given in Figure 2.1.

Upon arrival, any incoming item of class $i \in \mathcal{V}$ can be either matched with an item present in the buffer, of a class j such that $i-j$, or it is stored in the buffer to wait for its match. Whenever several possible matches are possible for an incoming item i, it is the role of the matching policy to decide the match of the latter item without ambiguity. Each matched pair departs the system instantenously.

We assume that the successive classes of arriving items are random. We fix an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, on which all random variables are defined. For any $n \in \mathbb{N}$, let $V_n \in \mathcal{V}$ denote the class of the n-th incoming item. We assume that the sequence $(V_n)_{n\in\mathbb{N}}$ is i.i.d., from the distribution μ on $\mathcal{V}$. Without loss of generality, we assume that μ has full support $\mathcal{V}$, i.e. $\mu \in \mathcal{M}^+(\mathcal{V})$, which we will write shortly as $\mu \in \mathcal{M}(\mathcal{V})$.

A *matching model* is a triple $(G, \mu, \textsc{Pol})$, where:

- $G = (\mathcal{V}, \mathcal{E})$ is the matching graph;
- $\mu \in \mathcal{M}(\mathcal{V})$ is the arrival distribution;
- Pol is a Markovian matching policy which defines the new buffer-content after the arrival of a new item and the possible matches with the old buffer-content;

The sequence of buffer-content forms a Markov chain. The first natural question is to find the conditions on $(G, \mu, \textsc{Pol})$ for stability, i.e. positive recurrence of this Markov chain. In [97], the authors give the necessary stability conditions:

$$\textsc{Ncond}(G) : \{\mu \in \mathcal{M}(\mathcal{V}) : \text{for any } \mathcal{I} \in \mathbb{I}(G), \mu(\mathcal{I}) < \mu(\Gamma(\mathcal{I}))\} \tag{2.1}$$

which, from Theorem 1 in [97], is non-empty if and only if G is non-bipartite.

Remark 1. *This Markov chain is periodic with period* 2*. This is not a problem for defining stability, however, if we wish to have convergence to the stationary distribution from any initial condition, it is possible to modify the model by adding a small probability of having zero arrivals at each given time step. This does not change the stationary measure of the*

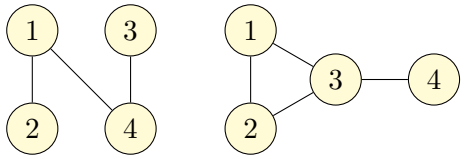
Markov chain, but it solves the issue of periodicity. Also sometimes it is more convenient to consider a related model in continuous time with items of different classes arriving to the system according to independent Poisson processes, with rates $\lambda_i > 0$, $i = 1, \ldots, n$.

After applying standard uniformization technique, with a uniformization constant $\Lambda \geq \sum_{i=1}^{n} \lambda_i$, *we obtain the following matching model in discrete time. In each time slot* $t \in \mathbb{N}^*$, *there are no arrivals with probability*

$$\alpha_0 = \frac{\Lambda - \sum_{i=1}^{n} \lambda_i}{\Lambda} \geq 0.$$

Otherwise, with probability $1 - \alpha_0$, *one item arrives to the system. This item belongs to a class within the set of item classes* $\mathcal{V}$, *sampled from a conditional probability distribution* $\mu = (\mu_1, \cdots, \mu_n)$, $\mu_i = \frac{\lambda_i}{\Lambda}$, *over* V *given the event that there is an arrival. Note that this discrete time system has average stationary queue lengths equal to the original continuous time model.*

The discrete general matching model, as defined in [97], is obtained for a particular choice of $\alpha_0 = 0$.



(a) Bipartite. (b) Non-bipartite.

Figure 2.1: Examples of matching graphs.

2.2 Bipartite stochastic matching model

In the bipartite stochastic matching model there are two populations of items, that we will call demand and supply. There are finite sets D and S, respectively of demand and supply classes. Authorized *matchings* between demand and supply items are given by a fixed bipartite graph $(\mathcal{D}, \mathcal{S}, E \subset \mathcal{D} \times \mathcal{S})$. Upon being matched, the demand and the supply item depart simultaneously.

To model arrivals, we have a priori more flexibility, but there is basically one non-trivial choice that can lead to stability (in the sense of positive recurrence for the Markov chain of the model), which is to assume that time is discrete and that demand and supply items arrive in pairs.

Consider indeed the simplest possible model with continuous-time arrivals: (i) there is only one demand class and one supply class; (ii) demand, resp. supply items, arrive according to a Poisson process of rate λ, resp. μ; (iii) matching is instantaneous. Let us describe the state by $Z = X - Y$, where X is the number of unmatched demand and Y the number of unmatched supply items. The process Z is a birth-and-death continuous-time Markov process on $\mathbb{Z}$ with drift $\lambda - \mu$. It is either transient (if $\lambda \neq \mu$) or null recurrent (if $\lambda = \mu$), but it is never positive recurrent. Let us switch to discrete-time i.i.d. arrivals. At each time step, a batch of demand and a batch of supply items arrive into the system. If the size of the batches are allowed to

be different for demand and supply, then we are back to the continuous-time situation, and even the simplest model is never positive recurrent.

We assume that the time is discrete and that at each time step, one demand and one supply item arrive in the system according to a joint probability measure μ on $\mathcal{D} \times \mathcal{S}$, independently of the past. Also, at each time step, pairs of *matched* demand and supply, if they exist, depart from the system. A *matching policy* decides how to match when there are several possibilities. Demand/supply items that are not matched are stored in a buffer. For simplicity, we assume that the matching is instantaneous. So the model is specified by: (i) the finite set $\mathcal{D}$ of demand classes and the finite set $\mathcal{S}$ of supply classes; (ii) the probability law μ on $\mathcal{D} \times \mathcal{S}$ for the arrivals in pairs; (iii) the bipartite graph $(\mathcal{D}, \mathcal{S}, E \subset \mathcal{D} \times \mathcal{S})$ giving the possible matchings between demand and supply classes (hence the possible departures in pairs); (iv) the matching policy to decide how to match when several choices are possible. We consider Markovian policies which depend only on the current state of the system. Under these assumptions, the buffer content evolves as a discrete-time Markov chain.

We call this model the *extended bipartite matching* (EBM) model. The particular case of product probability measure μ for the arrivals, i.e. under an additional assumption of independence between arriving demand and supply items ($\forall d, s,\ \mu(d, s) = \mu(d, \mathcal{S})\mu(\mathcal{D}, s)$) will be called *bipartite matching* (BM) model.

We now proceed to a more formal definition of the model.

Definition 2.2.1. *A* bipartite matching structure *is a quadruple* $(\mathcal{D}, \mathcal{S}, E, F)$ *where*

- $\mathcal{D}$ *is the non-empty and finite set of demand types;*
- $\mathcal{S}$ *is the non-empty and finite set of supply types;*
- $E \subset \mathcal{D} \times \mathcal{S}$ *is the set of possible matchings;*
- $F \subset \mathcal{D} \times \mathcal{S}$ *is the set of possible arrivals.*

The bipartite graph $(\mathcal{D}, \mathcal{S}, E)$ *is called the* matching graph. *It is assumed to be connected (otherwise we can decompose the model into connected components and treat them separately). The bipartite graph* $(\mathcal{D}, \mathcal{S}, F)$ *is called the* arrival graph. *It is assumed to have no isolated vertices (otherwise we can consider a new model without such demand or supply classes).*

In Figure 2.1, the graph on the left is an example of a bipartite matching graph for $\mathcal{D} = \{1, 3\}$ and $\mathcal{S} = \{2, 4\}$.

Demand and supply play symmetrical roles in the model. Also E and F play dual roles. The graph $(\mathcal{D}, \mathcal{S}, E)$ defines the pairs that may depart from the system, while the graph $(\mathcal{D}, \mathcal{S}, F)$ defines the pairs that may arrive into the system.

Definition 2.2.2. *A* bipartite matching model *is a triple* $[(\mathcal{D}, \mathcal{S}, E, F), \mu, \textsc{Pol}]$, *where*

- $(\mathcal{D}, \mathcal{S}, E, F)$ *is a bipartite matching structure;*
- μ *is a probability measure on* $\mathcal{D} \times \mathcal{S}$, *and* $\mu_{\mathcal{D}}$ *and* $\mu_{\mathcal{S}}$ *are the* $\mathcal{D}$ *and* $\mathcal{S}$ *marginals of* μ, *with supports satisfying*

$$supp(\mu) = F,\ supp(\mu_{\mathcal{D}}) = \mathcal{D},\ supp(\mu_{\mathcal{S}}) = \mathcal{S}\,, \tag{2.2}$$

  where $\mu_{\mathcal{D}}$ *and* $\mu_{\mathcal{S}}$ *are the* $\mathcal{D}$ *and* $\mathcal{S}$ *marginals of* μ.

- Pol *is a Markovian matching policy (defined more formally in Section 3.1) which defines the new buffer-content after the arrival of the new demand and supply items and the possible matches with the old buffer-content.*

Observe that we can simplify the notation to $[(\mathcal{D}, \mathcal{S}, E), \mu, \textsc{Pol}]$. We say that the model $[(\mathcal{D}, \mathcal{S}, E), \mu, \textsc{Pol}]$ is *associated* with the structure $(\mathcal{D}, \mathcal{S}, E, F)$.

A realization of the model is as follows. Consider an i.i.d. sequence of random variables of law μ, representing the arrival stream of pairs of demand/supply. A state of the buffer consists of an equal number of demand and supply items with no possible matchings between the classes. Upon arrival of a new ordered pair (d, s), two situations may occur: if neither d nor s match with the supply/demand already present in the buffer, then d and s are simply added to the buffer; if d, resp. s, can be matched then it departs the buffer with its match. If several matchings are possible for d, resp. s, then it is the role of the matching policy to select one. A Markovian matching policy selects according to the current state of the buffer (and not according to the whole history of the buffer contents). The resulting evolution of the buffer is described by a discrete-time Markov chain.

The bipartite stochastic matching model was first studied in Caldentey, Kaplan and Weiss [31]. They introduced FCFM (First Come First Matched) bipartite matching model with two infinite sequences of "customers" (also called demand in this manuscript) and "servers" (also called supply), motivated by the Boston area social housing problem and the PhD thesis of Kaplan [80]. FCFM policy matches a new arrival to the longest waiting compatible item in the system. The first examples, with simple compatibility graphs were analyzed already in [31] where the authors also conjectured the matching rates, i.e. the fraction of customers of each type served by each type of server, for any bipartite compatibility graphs. The necessary and sufficient condition for stability and a product form stationary distribution for the bipartite matching model were derived by Adan and Weiss in [3]. The product form stationary distribution was derived by partial balance, similar to [137]. This stationary distribution was then used to derive expressions for the matching rates.

The EBM model, as presented here, was first introduced in [J22], motivated by the question of finding a simple matching policy that has maximal stability region. We discuss the stability of stochastic matching models in the next chapter. FCFM policy will be discussed more in detail in Chapter 4.

Chapter 3

Stability of stochastic matching systems

This chapter summarizes the results on the stability of the EBM model. It is based on the publication [J22], that contains more details and the proofs of all the results summarized in this chapter. We start by giving the necessary conditions for stability. For some bipartite graphs, we prove that the stability region is maximal (i.e. coincides with the necessary conditions) for any admissible greedy matching policy. For the Match the Longest (ML) policy, we prove that the stability region is maximal for any bipartite graph. For the Match the Shortest (MS) and priority policies, we exhibit a bipartite graph with a non-maximal stability region.

At the end of the chapter, we provide the discussion on the related results in the literature for the non-bipartite matching model.

3.1 Matching policies

For a matching graph $(\mathcal{D}, \mathcal{S}, E)$, let $\mathcal{D}(s)$ denote the set of demand classes that can be matched with an s-supply; and $\mathcal{S}(d)$ the set of demand classes that can be matched with a d-demand:

$$\mathcal{S}(d) = \{s \in S \ : \ (d,s) \in E\}, \quad \mathcal{D}(s) = \{d \in D \ : \ (d,s) \in E\}.$$

For any subsets $A \subset D$, and $B \subset S$, we define

$$S(A) = \cup_{d \in A} S(d), \quad D(B) = \cup_{s \in B} D(s).$$

A matching policy is Markovian if only the current state of the buffer is taken into account, i.e. there exists a state space $\mathcal{E}$ and a mapping $\odot : \mathcal{E} \times (D \times S) \to \mathcal{E}$ which returns the new state of the system after an arrival.

A matching policy is called *greedy* if the buffer content never contains any compatible items. A greedy matching policy satisfies *buffer-first* assumption if priority is given to items that are already present in the buffer: if the state is (u, v) and the new arrival is (d, s), then d and s are matched together iff there are no servers from $S(d)$ in v and no customers from $D(s)$ in u. This is not a real restriction: a matching policy that gives priority to new arrivals can be seen as a special case of the above with an arrival probability μ such that $\mu(E) = 0$.

We will focus in this chapter on Markovian greedy buffer-first policies that can be defined on the following commutative and non-commutative state spaces. Non-greedy Markovian matching policies will be considered in Chapter 5.

3.1.1 Commutative state space

A state of the system is given by (x, y), $x = (x_d)_{d\in D}$ and $y = (y_s)_{s\in S}$, where x_d denotes the number of demand of type d and y_s the number of supply of type s. The *commutative state space* for greedy policies is:

$$\mathcal{E} = \Big\{(x, y) \in \mathbb{N}^D \times \mathbb{N}^S \ : \ \sum_{d\in D} x_d = \sum_{s\in S} y_s;\ \forall (d, s) \in E,\ x_d y_s = 0\Big\}. \tag{3.1}$$

For $d \in D$, let $e_d \in \mathbb{N}^D$ be defined by $(e_d)_d = 1$ and $(e_d)_c = 0, c \neq d$. For $s \in S$, let e_s be defined accordingly.

Definition 3.1.1. *A matching policy is* admissible *greedy policy if there are functions* Φ *and* Ψ *such that:*

$$(x, y) \odot (d, s) = \begin{cases} (x + e_d, y + e_s), & \text{if } x_{D(s)} = 0,\ y_{S(d)} = 0,\ (d, s) \notin E \\ (x, y), & \text{if } x_{D(s)} = 0,\ y_{S(d)} = 0,\ (d, s) \in E \\ (x - e_{\Phi(x,s)}, y - e_{\Psi(y,d)}), & \text{if } x_{D(s)} \neq 0,\ y_{S(d)} \neq 0 \\ (x - e_{\Phi(x,s)} + e_d, y), & \text{if } x_{D(s)} \neq 0,\ y_{S(d)} = 0 \\ (x, y - e_{\Psi(y,d)} + e_s), & \text{if } x_{D(s)} = 0,\ y_{S(d)} \neq 0 \end{cases}$$

The following commutative matching policies are admissible greedy policies (for RANDOM, ML, and MS policies $\Phi(u, s)$ and $\Psi(v, d)$ are random variables):

- PR (Priorities). For each demand type $d \in D$, we define a priority function $\alpha_d : S(d) \to \{1, \ldots, |S(d)|\}$. Similarly, for each supply type $s \in S$, we define $\beta_s : D(s) \to \{1, \ldots, |D(s)|\}$. In the case of several matching options, a demand/supply is matched with the supply/demand that has the highest priority: $\Phi(x, s) = \arg\max\{\beta_s(d) \ : \ d \in D(s), x_d > 0\}$ and $\Psi(y, d) = \arg\max\{\alpha_d(s) \ : \ s \in S(d), y_s > 0\}$.
- RANDOM : $\Phi(x, s)$, resp. $\Psi(y, d)$, is a random variable valued in $D(s)$, resp. $S(d)$, and distributed as $\left(x_i/\sum_{j\in D(s)} x_j\right)_{i\in D(s)}$, resp. $\left(y_i/\sum_{j\in S(d)} y_j\right)_{i\in S(d)}$. Intuitively, the match is chosen uniformly among all possible ones.
- ML : $\Phi(x, s)$, resp. $\Psi(y, d)$, is a random variable uniformly distributed on $\arg\max\{x_i : i \in D(s)\}$, resp. $\arg\max\{y_i : i \in S(d)\}$.
- MS : $\Phi(x, s)$, resp. $\Psi(y, d)$, is a random variable uniformly distributed on $\arg\min\{x_i > 0 : i \in D(s)\}$, resp. $\arg\min\{y_i > 0 : i \in S(d)\}$.

3.1.2 Non-commutative state space

A state of the system is given by two finite words of the same size $k \geq 0$, respectively on the alphabets D and S, describing unmatched demand and supply. The (greedy) *non-commutative state space* for greedy policies is:

$$\mathcal{E} = \Big\{(u, v) \in \cup_{k\geq 0}(D^k \times S^k) \ : \ ([u], [v]) \text{ belongs to } (3.1)\Big\}. \tag{3.2}$$

For a word $w \in A^k$ and $i \in \{1, \dots, k\}$, we denote by $w_{[i]} := w_1 \dots w_{i-1} w_{i+1} \dots w_k$ the subword of w obtained by deleting w_i.

Definition 3.1.2. *A matching policy is* admissible *greedy policy if there are functions* Φ *and* Ψ *such that:*

$$(u,v) \odot (d,s) = \begin{cases} (ud, vs), & \text{if } |u|_{D(s)} = 0,\ |v|_{S(d)} = 0,\ (d,s) \notin E \\ (u,v), & \text{if } |u|_{D(s)} = 0,\ |v|_{S(d)} = 0,\ (d,s) \in E \\ (u_{[\Phi(u,s)]}, v_{[\Psi(v,d)]}), & \text{if } |u|_{D(s)} \neq 0,\ |v|_{S(c)} \neq 0 \\ (u_{[\Phi(u,s)]} d, v), & \text{if } |u|_{D(s)} \neq 0,\ |v|_{S(d)} = 0 \\ (u, v_{[\Psi(v,d)]} s), & \text{if } |u|_{D(s)} = 0,\ |v|_{S(d)} \neq 0 \end{cases}$$

The First Come First Matched *FCFM* and Last Come First Matched *LCFM* policies are admissible greedy matching policies with:

- FIFO : $\Phi(u,s) = \arg\min\{u_k \in D(s)\}$, $\Psi(v,d) = \arg\min\{v_k \in S(d)\}$.
- LIFO : $\Phi(u,s) = \arg\max\{u_k \in D(s)\}$, $\Psi(v,d) = \arg\max\{v_k \in S(d)\}$.

3.2 Necessary conditions for stability

Consider first a simpler finite and deterministic problem. Let (D, S, E) be a matching graph. Consider a batch of demand $x \in \mathbb{N}^D$ and a batch of supply $y \in \mathbb{N}^S$ of equal size: $\sum_d x_d = \sum_s y_s$. A *perfect matching* of x and y is a tuple $m \in \mathbb{N}^E$ such that:

$$\forall d \in D,\ x_d = \sum_{s \in S(d)} m_{ds}, \qquad \forall s \in S,\ y_s = \sum_{d \in D(s)} m_{ds}\,.$$

By Hall's Theorem, there exists a perfect matching if and only if:

$$\begin{array}{l} \sum_{d \in U} x_d \leq \sum_{s \in S(U)} y_s, \quad \forall U \subset D \\ \sum_{s \in V} y_s \leq \sum_{d \in D(V)} x_d, \quad \forall V \subset S \end{array} \tag{3.3}$$

A perfect matching, if there is one, can be obtained by restating the model as a flow network and by solving the maximum flow problem [58, 54].

The bipartite matching model is more complicated: first it is random, and second the matchings have to be performed online. However the two ingredients of the simpler model play an instrumental role in the analysis: (i) the conditions NCond, to be defined in (3.4), are related to (3.3); (ii) the restatement as a flow problem is used in most of the proofs.

3.2.1 Stability definition

Consider a bipartite matching model $[(D, S, E), \mu, \text{Pol}]$. We identify the model with the Markov chain on the state space $\mathcal{E}$ describing the evolution of the buffer content.

Let P be the transition matrix of the Markov chain. A probability measure π on $\mathcal{E}$ is *stationary* if $\pi P = \pi$. It is *attractive* if for any probability measure ν on $\mathcal{E}$, the sequence of Cesaro averages of νP^n converges weakly to π.

Definition 3.2.1. *The model is said to be* stable *if the Markov chain has a unique and attractive stationary probability measure.*

It implies in particular that the graph of the Markov chain has a unique terminal strongly connected component with all states leading to it.

3.2.2 Necessary conditions

Let μ_D be a probability measure on D and μ_S a probability measure on S. Define the following conditions on (μ_D, μ_S):

$$\text{NCOND}: \quad \begin{cases} \mu_D(U) < \mu_S(S(U)), & \forall U \subsetneq D \\ \mu_S(V) < \mu_D(D(V)), & \forall V \subsetneq S \end{cases} \tag{3.4}$$

Lemma 3.2.1. *The conditions* NCOND *are necessary stability conditions: if the Markov chain is stable then the conditions* NCOND *are satisfied by the marginals of* μ.

The conditions NCOND appear already in [31] for the BM model. They have a natural interpretation. Demand items from U need to be matched with supply from $S(U)$. The first line in NCOND asks for strictly more supply items in average from $S(U)$ than demand from U. The second line has a dual interpretation.

Definition 3.2.2. *Consider a bipartite graph* (D, S, E) *and an admissible matching policy* POL. *The* stability region *is the set of values of* μ *for which the bipartite matching model* $[(D, S, E), \mu, \text{POL}]$ *is stable.*

The stability region is included in the polyhedron defined by NCOND. The stability region is *maximal* if it is equal to this polyhedron.

3.2.3 Complexity

The number of inequalities in NCOND is exponential in $|D| + |S|$. So checking directly if all the inequalities are satisfied is a method whose time complexity is exponential in $|D|+|S|$. A polynomial algorithm to check if the conditions NCOND are satisfied can be obtained using network flow arguments.

We use the standard terminology of network flow theory, see for instance [58]. Consider the directed graph

$$\mathcal{N} = \left(D \cup S \cup \{i, f\}, E \cup \{(i, d), c \in D\} \cup \{(s, f), s \in S\}\right). \tag{3.5}$$

Endow the arcs of E with infinite capacity, an arc of type (i, d) with capacity $\mu_D(d)$, and an arc of type (s, f) with capacity $\mu_S(s)$. Recall that a *cut* is a subset of the arcs whose removal disconnects i and f. The *capacity* of a cut is the sum of the capacities of the arcs. Set $A = E \cup \{(i, d), d \in D\} \cup \{(s, f), s \in S\}$. Recall that $T : A \to \mathbb{R}_+$ is a *flow* if: (i) $\forall d, T(i, d) = \sum_{s \in S(d)} T(d, s),\ \forall s, \sum_{d \in D(s)} T(d, s) = T(s, f)$; (ii) $\forall (x, y) \in E$, $T(x, y)$ is less or equal to the capacity of (x, y). The *value* of T is $\sum_d T(i, d) = \sum_s T(s, f)$.

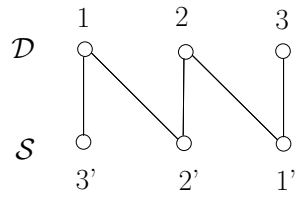


Figure 3.1: NN graph.

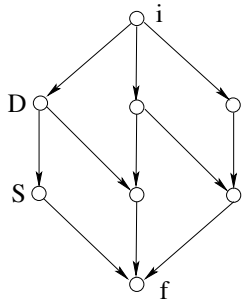


Figure 3.2: The graph $\mathcal{N}$ associated with the NN model of Figure 3.1.

Let NCOND$_{\leq}$ be the set of inequalities obtained from NCOND by replacing the strict inequalities by $\leq$.

Lemma 3.2.2. *There exists a flow of value 1 in $\mathcal{N}$ iff (μ_D, μ_S) satisfies* NCOND$_{\leq}$. *There exists a flow T of value 1 such that $T(d,s) > 0$ for all $(d,s) \in E$ iff (μ_D, μ_S) satisfies* NCOND.

The first part of Lemma 3.2.2 is proved in [31, Prop. 3.7].

There exist algorithms to find the maximal flow which are polynomial in the size of the underlying graph, independent of the arc capacities. For instance, the classical "augmenting path algorithm" of Edmonds & Karp [54] has time complexity in $O((|D|+|S|)|E|^2)$, and there exist more sophisticated algorithms with time complexity in $O((|D|+|S|)^3)$.

We take one of these polynomial algorithms, call it MAXFLOW and consider it as a black-box. We build on this to design a polynomial algorithm to check NCOND.

Lemma 3.2.3. *Fix η such that $0 < \eta < 1/|E|$. Define $(\widetilde{\mu}_D, \widetilde{\mu}_S)$ as*

$$\widetilde{\mu}_D(d) = \frac{\mu_D(d) - |S(d)|\eta}{1 - |E|\eta}, \qquad \widetilde{\mu}_S(s) = \frac{\mu_S(s) - |D(s)|\eta}{1 - |E|\eta}\,. \tag{3.6}$$

The pair (μ_D, μ_S) satisfies NCOND *iff the pair $(\widetilde{\mu}_D, \widetilde{\mu}_S)$ satisfies* NCOND *for η strictly positive and small enough.*

Using Lemmas 3.2.2 and 3.2.3, NCOND is satisfied iff MAXFLOW$(\mathcal{N}, \widetilde{\mu}_C, \widetilde{\mu}_S)$ returns 1 for η small enough. So the trick is to run MAXFLOW on the input $(\mathcal{N}, \widetilde{\mu}_C, \widetilde{\mu}_S)$ by considering η as a formal parameter made "as small as needed".

The precise meaning is the following. If $x_1, x_2, y_1, y_2 \in \mathbb{R}$, then: $(x_1 + y_1\eta) + (x_2 + y_2\eta) = (x_1 + x_2) + (y_1 + y_2)\eta$. Furthermore,

$$\begin{aligned} \big[x_1 + y_1\eta = x_2 + y_2\eta\big] \quad &\Longleftrightarrow \quad \big[x_1 = x_2,\ y_1 = y_2\big] \\ \big[x_1 + y_1\eta < x_2 + y_2\eta\big] \quad &\Longleftrightarrow \quad \Big[(x_1 < x_2) \ \text{ or } \ (x_1 = x_2,\ y_1 < y_2)\Big] \end{aligned} \tag{3.7}$$

So η is small enough not to reverse any strict inequality. When running MAXFLOW on $(\mathcal{N}, \widetilde{\mu}_D, \widetilde{\mu}_S)$, the algorithm deals with values of the type $(x + y\eta)$, and adds and compare them according to the above rules. Now observe that the algorithm stops in finite time, so it will have performed only a finite number of operations. Therefore, it would be possible, a posteriori, to assign to η a value which would be small enough to enforce (3.7).

The termination is obvious and the correctness follows from Lemmas 3.2.2 and 3.2.3.

Proposition 3.2.4. *Given a bipartite model $[(D, S, E), \mu]$, there exists an algorithm of time complexity $O((|D| + |S|)^3)$ to decide if* NCOND *is satisfied.*

3.3 Connectivity properties of the Markov chain

Define the following property for the transition graph of the Markov chain:

UTC : a unique (terminal) strictly connected component with all states leading to it.

Property UTC is necessary for stability as defined in Def. 3.2.1. But property UTC is not granted in bipartite matching models and counterexamples are given below (Examples 2 and 3). In fact, we will see that we are in an unusual situation: the necessary stability conditions NCOND turn out to be sufficient conditions for the property UTC (Theorem 3.3.2). Observe also that property UTC is weaker than irreducibility, and we will give an example of a model satisfying NCOND and UTC without being irreducible (Example 4).

3.3.1 Stable structures

To establish property UTC, we study a notion of independent interest: stable structures.

Definition 3.3.1. *A bipartite matching structure (D, S, E, F) is* stable *if there exists a probability measure μ satisfying (2.2) and whose marginals μ_D and μ_S satisfy* NCOND.

The justification for this terminology will appear in Section 3.6: we prove there that under the ML policy, any model satisfying NCOND is stable. So a structure is stable iff there exists an associated model which is stable.

First, there exist stable structures.

Example 1. *Consider $(D, S, E, D \times S)$, where (D, S, E) is the NN bipartite graph of Figure 3.1. Let*

$$\mu_D :\ \mu_D(1) = \mu_D(2) = 2/5,\ \mu_D(3) = 1/5, \qquad \mu_S :\ \mu_S(1') = \mu_S(2') = 2/5,\ \mu_S(3') = 1/5\,.$$

The product measure $\mu = \mu_D \times \mu_S$ has marginals μ_D and μ_S and we check that (μ_D, μ_S) satisfy NCOND. *Also it is easily proved that for any admissible greedy matching policy, the graph of the Markov chain is irreducible.*

On the other hand, there exist unstable structures. We illustrate this on two examples.

Example 2. *Consider the structure (D, S, E, F) where (D, S, E) is the NN graph of Figure 3.1, and where*

$$F = \big\{(1, 3'), (2, 2'), (3, 1')\big\}\,.$$

Consider any μ with $supp(\mu) = F$. We have $\mu_D(1) = \mu_S(3') = \mu(1, 3')$ which violates NCOND *for $V = \{3'\}$. We can also prove that the property* UTC *is not satisfied. Consider a state of the type (x, y) with $x = y = (0, 0, k)$, for some $k \geq 0$. Any one of the three possible arrivals leave the state unchanged. In particular, there is an infinite number of terminal components.*

Example 3. *Consider the bipartite matching structure defined in Figure 3.3. The graph (D, S, E) is represented on the left of the figure, while the graph (D, S, F) is represented on the right.*

Consider any μ with $supp(\mu) = F$. We have

$$\mu_S(\{1', 2'\}) = \mu(3, 1') + \mu(4, 2') \leq \mu_D(\{3, 4\})\,,$$

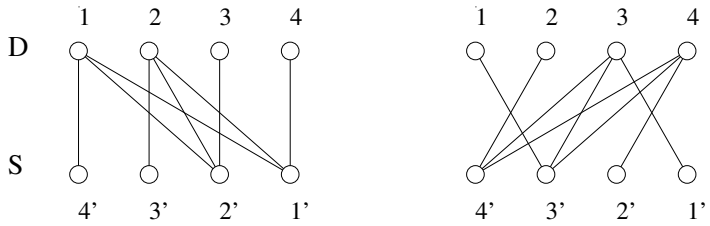


Figure 3.3: The matching graph (D, S, E) on the left, and the arrival graph (D, S, F) on the right.

which contradicts NCOND *for* $U = \{3, 4\}$. *We can also prove that the property* UTC *is not satisfied. Consider a state* (x, y) *with* $x_3 + x_4 = k > 0$. *Reducing the number of demand items of types 3/4 would require an arrival of type* $(1, 1')$ *or* $(1, 2')$ *or* $(2, 1')$ *or* $(2, 2')$. *But none of these pairs belong to* F. *Therefore it is impossible to reach a state* (x', y') *with* $x'_3 + x'_4 < x_3 + x_4$. *On the other hand an arrival of type* $(3, 3')$ *or* $(3, 4')$ *or* $(4, 3')$ *or* $(4, 4')$ *strictly increases the number of demand items of types 3/4. Hence all the states are transient, and there is no terminal strongly connected component.*

Stability of a structure is a decidable property. There exists a probability measure μ with the requested properties iff the following system of linear inequalities in variables $\mu(d, s),\ d \in D, s \in S$, have a solution:

$$\begin{cases} \sum_{(d,s)\in D\times S} \mu(d,s) = 1, & \\ \mu(d,s) > 0, & \forall (d,s) \in F, \\ \mu(d,s) = 0, & \forall (d,s) \in D \times S - F, \\ \mu_D(d) = \sum_{s\in S} \mu(d,s), & \forall d \in D, \\ \mu_S(s) = \sum_{d\in D} \mu(d,s), & \forall s \in S, \\ \text{NCOND}\ . & \end{cases} \tag{3.8}$$

However, the number of inequalities is exponential in $|D| + |S|$. We propose a criterion which is much simpler, both conceptually and algorithmically.

Consider a bipartite matching structure (D, S, E, F). Define $\widetilde{F} = \{(s, d) \mid (d, s) \in F\}$. Associate with the structure the **directed** graph $(D \cup S, E \cup \widetilde{F})$, in other words the nodes are $D \cup S$ and the arcs are

$$d \longrightarrow s, \quad \text{if } (d,s) \in E, \qquad s \longrightarrow d, \quad \text{if } (d,s) \in F\ .$$

We have represented in Figure 3.4 the directed graph associated with the structure of Example 3.

The graph of Figure 3.4 is not strongly connected: the four nodes on the right form a strongly connected component. Similarly, the directed graph associated with the structure of Example 2 is not strongly connected. On the other hand, the directed graph associated with the structure of Example 1 is strongly connected. This is not a coincidence.

Theorem 3.3.1. *Let* (D, S, E, F) *be a bipartite matching structure. The following two properties are equivalent:*

1. (D, S, E, F) *is a stable structure;*

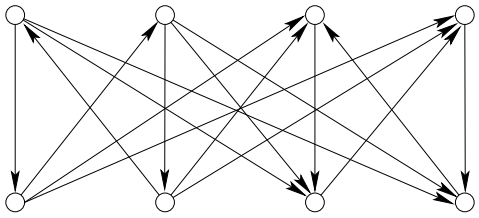


Figure 3.4: The directed graph associated with the structure of Figure 3.3.

2. $(D \cup S, E \cup \widetilde{F})$ *is strongly connected.*

In particular, one can decide if a structure is stable with an algorithm of time complexity $O(|D||S|)$ by testing the strong connectivity of $(D \cup S, E \cup \widetilde{F})$.

3.3.2 Property UTC

Theorem 3.3.2. *Consider a bipartite matching model* $[(D, S, E), \mu, \text{POL}]$. *Assume that the structure* (D, S, E, F) *is stable, equivalently that* $(D \cup S, E \cup \widetilde{F})$ *is strongly connected. Then the transition graph of the Markov chain of the bipartite matching model satisfies the property* UTC.

Example 4. *Consider a bipartite matching model associated with the structure* (D, S, E, F) *where* (D, S, E) *is the NN graph of Figure 3.1, and where*

$$F = \left\{(1, 1'), (2, 2'), (3, 3')\right\}.$$

The graph $(D \cup S, E \cup \widetilde{F})$ *is strongly connected. According to Theorem 3.3.1, the graph satisfies property* UTC. *But it is not irreducible. Indeed, it is impossible to reach the state* $((0, 1, 0); (0, 0, 1))$ *starting from the empty state.*

Below, we study the stability of bipartite matching models. Therefore, we always assume that the necessary conditions NCOND are satisfied. So we get the property UTC for the Markov chain as a consequence of Theorem 3.3.2.

3.4 Models that are stable for all admissible policies

Both the commutative and the non-commutative state space can be decomposed into facets, defined only by the non-zero classes.

Definition 3.4.1. *A* facet *is an ordered pair* (U, V) *such that:* $U \subset D, V \subset S$ *and* $U \times V \subset (D \times S - E)$. *The* zero-facet *is the facet* $(\emptyset, \emptyset)$, *we denote it shortly by* $\emptyset$.

For a facet $\mathcal{F} = (U, V)$, define:

$$D_\bullet(\mathcal{F}) = U, \qquad S_\bullet(\mathcal{F}) = V,$$
$$D_\circledcirc(\mathcal{F}) = D(V), \qquad S_\circledcirc(\mathcal{F}) = S(U),$$
$$D_\circ(\mathcal{F}) = D - (D_\bullet(\mathcal{F}) \cup D_\circledcirc(\mathcal{F})), \qquad S_\circ(\mathcal{F}) = S - (S_\bullet(\mathcal{F}) \cup S_\circledcirc(\mathcal{F})).$$

We alleviate the notations to $D_\bullet, S_\bullet, D_\circledcirc, \ldots$, when there is no possible confusion. The symbol $\bullet$ stands for the non-zero classes, the symbol $\circledcirc$ for the classes that are forced to be at zero (since they are matched with non-zero classes), and the symbol $\circ$ for the classes that happen to be at zero.

Graphical convention. A facet $\mathcal{F}$ can be represented graphically by coloring the nodes of the bipartite graph according to the above convention (see Figure 3.5 for an illustration):

- nodes in $D_\bullet(\mathcal{F})$ and $S_\bullet(\mathcal{F})$ are represented as filled circles;
- nodes in $D_\circledcirc(\mathcal{F})$ and $S_\circledcirc(\mathcal{F})$ are represented as double circles;
- nodes in $D_\circ(\mathcal{F})$ and $S_\circ(\mathcal{F})$ are represented as simple circles.

Definition 3.4.2. *A facet $\mathcal{F}$ is called saturated if $D_\circ(\mathcal{F}) = \emptyset$ or $S_\circ(\mathcal{F}) = \emptyset$.*

In Figure 3.5, the facet on the left is non-saturated, while the one on the right is saturated.

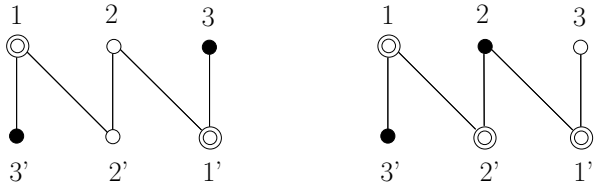


Figure 3.5: NN graph: facets $(\{3\}, \{3'\})$ and $(\{2\}, \{3'\})$.

Denote by $\mathfrak{F}$ the set of facets. Define the following conditions on μ:

$$\text{SCond}: \quad \mu_C(D_\circledcirc(\mathcal{F})) + \mu_S(S_\circledcirc(\mathcal{F})) \ > \ 1 - \mu(E \cap D_\circ(\mathcal{F}) \times S_\circ(\mathcal{F})), \quad \forall \mathcal{F} \in \mathfrak{F} - \{\emptyset\} \tag{3.9}$$

In particular, the subset of the inequalities (3.9) obtained by considering only the saturated facets gives precisely the inequalities NCond.

By application of the Lyapunov-Foster Theorem, see for instance [21, Section 5.1], for the linear Lyapunov function:

$$L(u, v) = |u|, \ (u, v) \in \mathcal{E},$$

counting the number of unmatched demand items, it follows that SCond are sufficient stability conditions.

Proposition 3.4.1. *A bipartite model with probability μ satisfying* SCond *is stable under any admissible matching policy.*

Corollary 3.4.2. *Consider a bipartite graph in which any non-zero facet is saturated. For any admissible matching policy, the stability region is maximal.*

The bipartite graph $(D = \{1, 2\}, S = \{1', 2'\}, D \times S - \{(2, 2')\})$ is such that any non-zero facet is saturated. Therefore, its stability region is maximal for any admissible policy. The same is true for the "almost complete graphs" $(D = \{1, \ldots, k\}, S = \{1', \ldots, k'\}, D \times S - \{(i, i'), \forall i\})$.

Example 5. *Consider the NN graph from Figure 3.1. The graph has only one non-zero facet that is non-saturated, facet $(\{3\}, \{3'\})$. For any admissible policy, the stability region is at least the polyhedron* SCond*, Proposition 3.4.1, which is defined by:*

$$\text{NCond}, \qquad \mu_D(1) + \mu_S(1') > 1 - \mu(2, 2') \, . \tag{3.10}$$

Assume now $\mu = \mu_D \times \mu_S$ *and* $\mu_D = \mu_S$. *Set* $x = \mu_D(1) = \mu_S(1')$ *and* $y = \mu_D(2) = \mu_S(2')$. *Then:*

$$\text{NCOND}: \begin{cases} x < 0.5 \\ 2x + y > 1 \end{cases} \qquad \text{SCOND}: \begin{cases} \text{NCOND} \\ 2x + y^2 > 1 \end{cases}$$

In Figure 3.6, the light (yellow) region corresponds to SCOND, *and the union of the light and*

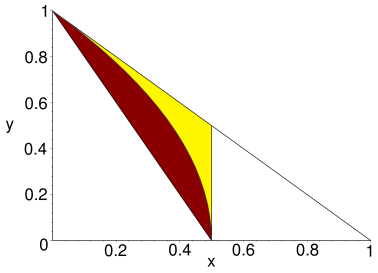


Figure 3.6: NCOND and SCOND for the NN-graph with $\mu = \mu_C \times \mu_S$ and $\mu_C = \mu_S$.

dark (red) regions corresponds to NCOND.

3.5 Priorities and MS are not always stable

Consider the NN bipartite graph of Figure 3.1 and Example 5. For this model, Proposition 3.4.1 does not allow to decide if the stability region is maximal (see Figure 3.6). In fact, we show below that for the PR and MS matching policies, the stability region is not maximal.

Proposition 3.5.1. *Consider the NN model with either the MS policy or the PR (priority) policy such that demand of class* d_1 *(resp. supply of class* s_1*) gives priority to supply of class* s_2 *and supply of class* s_1 *to demand of class* d_2 *(see Figure 3.7). For both policies, the stability*

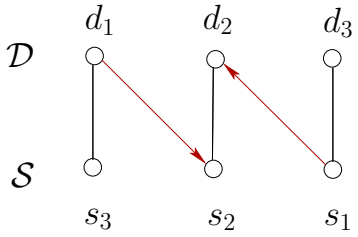


Figure 3.7: PR policy with non-maximal stability region.

region is not maximal.

The proof consist in finding a point in NCOND for which the Markov chain is not positive recurrent. For both policies, this is true for the following distribution: $\mu_D = (1/3, 2/5, 4/15)$, $\mu_S = \mu_D$, and $\mu = \mu_D \times \mu_S$.

3.6 ML is always stable

Theorem 3.6.1. *For any bipartite graph, the ML policy has a maximal stability region.*

The idea of the proof is as follows. Consider the quadratic Lyapunov function:

$$L(x,y) = \sum_{d\in D} x_d^2 + \sum_{s\in S} y_s^2, \quad (x,y) \in \mathcal{E}. \tag{3.11}$$

Observe that the ML policy minimizes the value of this Lyapunov function at each step. We introduce a facet-dependent randomized policy that depends on the arrival distribution μ. For this policy, we can prove that the quadratic Lyapunov function has a negative drift outside a finite region. By the Lyapunov-Foster's Theorem, see for instance [21, Section 5.1], the alternate matching policy is stable. Since the ML matching policy minimizes the value of the quadratic Lyapunov function, the ML policy is also stable. This auxiliary facet-dependent randomized policy is constructed using network flow arguments.

3.7 Discussion and related results for the non-bipartite case

Mairesse and Moyal [97] observed that the bipartite double cover (see for exemple [23]) of a GM model is in fact a special case of a EBM model. Given a graph $G = (\mathcal{V}, \mathcal{E})$, its bipartite double cover is the bipartite graph $2 \circ G = (2 \circ \mathcal{V}, 2 \circ \mathcal{E})$ defined as

$$2 \circ \mathcal{V} = \mathcal{V} \cup \{\tilde{u} \mid u \in \mathcal{V}\}, \quad 2 \circ \mathcal{E} = \{(u,\tilde{v}), (v,\tilde{u}) \mid (u,v) \in \mathcal{E}\}$$

where $\tilde{\mathcal{V}} = \{\tilde{u} \mid u \in \mathcal{V}\}$ is a disjoint copy of $\mathcal{V}$ (see Figure 3.8).

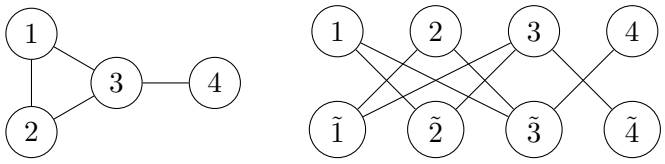


Figure 3.8: A graph and its double cover.

Based on this observation, they were able to extend many of the results presented in this chapter to the GM model. In particular, define the set of measures

$$\textsc{Ncond}(G) : \{\mu \in \mathcal{M}(\mathcal{V}) \,:\, \text{for any } \mathcal{I} \in \mathbb{I}(G), \mu(\mathcal{I}) < \mu(\mathcal{E}(\mathcal{I}))\}. \tag{3.12}$$

Theorem 1 in [97] states that this set is non-empty if and only if G is non-bipartite. Furthermore, Proposition 2 in [97] that for any admissible (i.e. markovian greedy) policy Φ the stability region of the GM model associated to (G, μ, Φ), is included in $\textsc{Ncond}(G)$. An admissible policy Φ is said to be *maximal* if these two sets coincide. Theorem 2 in [97] establishes the maximality of the matching policy 'Match the Longest' for any non-bipartite graph, however priority policies and the uniform-class policy (that matches a new arrival to an item of a class chosen uniformly among all non-empty compatibility classes) are not always maximal (respectively, Theorem 3 and Proposition 7 in [111]). Last, Theorem 2 in [97] states that any GM model on a graph G that is p-partite complete for $p \geq 3$ has a stability region that coincides with $\textsc{Ncond}(G)$ whatever the admissible matching policy; in other word *any* admissible policy is maximal in this case.

Extensions to hypergraphs were inversigated in [120].

Finally, stability of stochastic matching models was considered in the sense of positive recurrence for the Markov chain of the model. It may be also interesting to investigate the null-recurrent case, for example using ideas in [9].

Chapter 4

FCFM stochastic matching

The focus of this chapter is on FCFM policy, i.e. the policy that matches each incoming item with the compatible item that has been waiting the longest, if there is any. FCFM policy is perceived as fair, it maximizes the stability region and is tractable, due to its product form stationary distribution that will be discussed in this chapter. The intuitive interpretation for its maximum stability is that, if an item has been waiting longer than another, it is likely that this item is compatible with classes that are scarcer, so matching this item is a good heuristic to preserve stability.

We will present the product form results and the reversibility properties of GM and BM models. Although the results for the BM model were obtained first, we start by presenting in Section 4.1 the results for the GM model, for which the notation and the key ideas are simpler to introduce. Then in Section 4.2 we give product form results for the BM model. Based on the product form solution, in Section 4.3, we show a performance paradox for GM model, in the sense that adding edges to the compatibility graph can lead to larger mean queue lengths. We end the chapter by a discussion and related results.

This chapter is based on the following publications that contain more details and the proofs of the results presented in this chapter: [J7] (Section 4.1), [J11] (Section 4.2), [C10] (Section 4.3).

4.1 Product form and reversibility properties of GM model

Notation. For a word $w \in A^*$ of length $|w| = q$, we write $w = w_1 w_2 ... w_q$, i.e. w_i is the i-th letter of the word w. In the proofs below, we understand the word $w_1 ... w_k$ as $\emptyset$ whenever $k = 0$. Also, for any $w \in A^*$ and any $i \in [\![1, |w|]\!]$, we denote by $w_{[i]}$, the word of length $|w| - 1$ obtained from w by deleting its i-th letter.

4.1.1 FCFM general matching model

We consider a GM model from Section 2.1 with a connected matching graph $G = (\mathcal{V}, \mathcal{E})$ that will be considered fixed. We assume that the matching policy is 'First Come, First Matched' (FCFM), that is, the match of i is the oldest among all stored items of neighboring classes of i. For any $n \in \mathbb{N}$, let $V_n \in \mathcal{V}$ denote the class of the n-th incoming item. We assume that the sequence $(V_n)_{n\in\mathbb{N}}$ is iid, from the distribution μ on $\mathcal{V}$. Without loss of generality, we assume that μ has full support $\mathcal{V}$. According to the terminology in [97] and Section 2.1, we consider

the general matching (GM) model associated to (G, μ, FCFM).

Markov representation

Fix an integer $n_0 \geq 1$, and a realization $v_1, ..., v_{n_0}$ of $V_1, ..., V_{n_0}$. Define the word $\mathbf{v} \in \mathcal{V}^*$ by $\mathbf{v} := v_1 ... v_{n_0}$. Then, there exists a unique FCFM *matching* of the word $\mathbf{v}$, that is, a graph having set of nodes $\{v_1, ..., v_{n_0}\}$ and whose edges represent the matches performed in the system until time n_0, if the successive arrivals are given by $\mathbf{v}$ and the matching policy is FCFM. This matching is denoted by $M^{\text{FCFM}}(\mathbf{v})$. The state of the system is defined as the word $W^{\text{FCFM}}(\mathbf{v}) \in \mathcal{V}^*$, whose letters are the classes of the unmatched items at time n_0, i.e. the isolated vertices in the matching $M^{\text{FCFM}}(\mathbf{v})$, in their order of arrivals. The word $W^{\text{FCFM}}(\mathbf{v}) \in \mathcal{V}^*$ is called *queue detail* at time n_0. Any admissible queue detail belongs to

$$\mathbb{W} = \Big\{ w \in \mathcal{V}^* \ : \ \forall (i,j) \in \mathcal{E}, \ |w|_i |w|_j = 0 \Big\}. \tag{4.1}$$

Fix a (possibly random) word $Y \in \mathbb{W}$. Denote for all $n \geq 0$ by $W_n^{\{Y\}}$ the buffer content at time n (i.e. just before the arrival of item n) if the buffer content at time 0 was set to Y. Then the buffer-content sequence is stochastic recursive, since we clearly have that

$$\left\{ \begin{array}{ll} W_0^{\{Y\}} & = Y; \\ W_{n+1}^{\{Y\}} & = W_n^{\{Y\}} \odot_{\text{FCFM}} (V_n), \ n \in \mathbb{N}, \end{array} \right.$$

where for all $w \in \mathbb{W}$ and $v \in \mathcal{V}$,

$$w \odot_{\text{FCFM}} (v) = \left\{ \begin{array}{ll} wv & \text{if } \ |w|_{\mathcal{E}(v)} = 0; \\ w_{[\Phi(w,v)]} & \text{otherwise,} \end{array} \right.$$

where $\Phi(w, v) = \min\{k \in [\![1, |w|]\!], : \ w_k \in \mathcal{E}(v)\}$. Consequently, if we assume that the sequence $(V_n)_{n \in \mathbb{N}}$ is independent of Y, the queue detail $\left(W_n^{\{Y\}}\right)_{n \in \mathbb{N}}$ is a $\mathbb{W}$-valued $\mathcal{F}_n$-Markov chain, where $\mathcal{F}_0 = \sigma(Y)$ and $\mathcal{F}_n = \sigma\left(Y, V_0, ..., V_{n-1}\right)$ for all $n \geq 1$. The sequence $(W_n)_{n \in \mathbb{N}}$ is termed *natural chain* of the system.

For a connected graph G, the chain $(W_n)_{n \in \mathbb{N}}$ is clearly irreducible, as all states of $\mathbb{W}$ lead to $\emptyset$. In line with [97], we define the *stability region* of the GM model (G, μ, FCFM) by

$$\text{STAB}(G, \text{FCFM}) := \left\{ \mu \in \mathcal{M}\left(\mathcal{V}\right) \ : \ (W_n)_{n \in \mathbb{N}} \text{ is positive recurrent} \right\}, \tag{4.2}$$

which is clearly independent of the initial state Y in view of the above observation.

We show next that the policy FCFM has maximal stability region, and characterize the steady state of the system under the condition that $\mu \in \text{NCOND}(G)$.

4.1.2 Product form

The main result we established in [J7] is Theorem 4.1.1 that shows the maximality of the FCFM policy by constructing explicitly the stationary distribution of the natural chain on $\mathbb{W}$. This probability distribution has a remarkable product form.

Theorem 4.1.1. *Let $G = (\mathcal{V}, \mathcal{E})$ be a non-bipartite graph. Then the sets* $\text{STAB}(G, \text{FCFM})$ *and* $\text{NCOND}(G)$*, defined respectively by (4.2) and (3.12) coincide, in other words the GM model*

(G, μ, FCFM) is stable if and only if μ belongs to the set $\text{NCOND}(G)$*. In that case, the following is the only stationary probability of the natural chain $(W_n)_{n\in\mathbb{N}}$:*

$$\Pi_W(w) = \alpha \prod_{\ell=1}^{q} \frac{\mu(w_\ell)}{\mu\Big(\mathcal{E}\left(\{w_1, ..., w_\ell\}\right)\Big)}, \quad \textit{for any } w = w_1...w_q \in \mathbb{W}, \tag{4.3}$$

where

$$\alpha = \left\{ 1 + \sum_{\mathcal{I}\in\mathbb{I}(G)} \sum_{\sigma\in\mathfrak{S}_{|\mathcal{I}|}} \prod_{j=1}^{|\mathcal{I}|} \frac{\mu(i_{\sigma(j)})}{\mu\left(\mathcal{E}\left(\{i_{\sigma(1)}, ..., i_{\sigma(j)}\}\right)\right) - \mu\left(\{i_{\sigma(1)}, ..., i_{\sigma(j)}\}\right)} \right\}^{-1}. \tag{4.4}$$

Characteristics at equilibrium

We can easily deduce, from Theorem 4.1.1, closed form formulas for performance measures of the system in steady state. Denote by W_∞, the stationary queue detail of the system, that is, a random variable distributed following the stationary probability Π_W.

The average total number of items in storage at equilibrium is given by

$$\mathbf{E}\left[|W_\infty|\right] = \sum_{k\in\mathbb{N}} k\Pi_W\left(\{w \in \mathbb{W} \,:\, |w| = k\}\right).$$

According to Little's law, for any $i \in \mathcal{V}$, the waiting time before getting matched, for an item of class i entering the system in steady state, is given by

$$\frac{\mathbf{E}\left[|W_\infty|\right]}{\mu(i)} = \frac{1}{\mu(i)} \sum_{k\in\mathbb{N}} k\Pi_W\left(\{w \in \mathbb{W} \,:\, |w|_i = k\}\right).$$

In particular, the probability that a class i-item does have to wait before getting matched in a stationary system is given by

$$\mathbb{P}\left[|W_\infty|_{\mathcal{E}(i)} = 0\right] = \Pi_W\left(\{w \in \mathbb{W} \,:\, |w|_{\mathcal{E}(i)} = 0\}\right),$$

for Π_W in (4.3).

The proof of Theorem 4.1.1 is based on a subtle reversibility scheme that is related to the proof of reversibility for the BM model in [J11], presented in Section 4.2. However, GM model is not a particular case of BM model, so the proof presents many specificities with respect to [J11], and it turns out to be more elegant as we do not need to handle two populations of items. To outline the ideas of the proof, we introduce two auxiliary Markov representations of the system.

4.1.3 Auxiliary Markov representations

For $w = w_1w_2 \dots w_q \in \mathcal{V}^*$, we denote by $\check{w}$ the reversed version of w, i.e. $\check{w} = w_qw_{q-1}...w_2w_1$. Let $\overline{\mathcal{V}}$ be an independent copy of the set $\mathcal{V}$, i.e. a set of cardinality $|\mathcal{V}|$, disjoint of $\mathcal{V}$ and containing copies of the elements of $\mathcal{V}$. We denote by $\overline{a}$, the copy of any element a of $\mathcal{V}$, and also say that $\overline{a}$ is the *counterpart* of a, and vice-versa. For any $\overline{a} \in \overline{\mathcal{V}}$, let us denote $\overline{\overline{a}} = a$. Let $\mathbf{V} := \mathcal{V} \cup \overline{\mathcal{V}}$. For any word $\mathbf{w} \in \mathbf{V}^*$, denote by $\mathcal{V}(\mathbf{w})$ (respectively, $\overline{\mathcal{V}}(\mathbf{w})$) the set of letters of $\mathcal{V}$ (resp., $\overline{\mathcal{V}}$) that are present in $\mathbf{w}$:

$$\mathcal{V}(\mathbf{w}) = \Big\{a \in \mathcal{V} \,:\, |\mathbf{w}|_a > 0\Big\}; \qquad \overline{\mathcal{V}}(\mathbf{w}) = \Big\{\overline{a} \in \overline{\mathcal{V}} \,:\, |\mathbf{w}|_{\overline{a}} > 0\Big\}.$$

For any $\mathbf{w} \in \mathbf{V}^*$, the *restriction* of $\mathbf{w}$ to $\mathcal{V}$ (respectively, to $\overline{\mathcal{V}}$) is the word $\mathbf{w}|_{\mathcal{V}} \in \mathcal{V}^*$ (resp., $\mathbf{w}|_{\overline{\mathcal{V}}} \in \overline{\mathcal{V}}^*$) of size $|\mathbf{w}|_{\mathcal{V}}$ (resp. of size $|\mathbf{w}|_{\overline{\mathcal{V}}}$), obtained by keeping only the letters belonging to $\mathcal{V}$ (resp. to $\overline{\mathcal{V}}$) in $\mathbf{w}$, in the same order. The *dual* $\overline{\mathbf{w}}$ of the word $\mathbf{w} = \mathbf{w}_1 ... \mathbf{w}_q \in \mathbf{V}^*$ is the word obtained by exchanging the letters of $\mathbf{w}$ with their counterpart, i.e. $\overline{\mathbf{w}} = \overline{\mathbf{w}_1} ... \overline{\mathbf{w}_q}$.

Example 6. Take for instance $\mathbf{w} = a\, b\, \overline{a}\, c\, \overline{b}\, \overline{c}\, \overline{b}\, d\, a$. Then we obtain $\mathcal{V}(\mathbf{w}) = \{a, b, c, d\}$, $\overline{\mathcal{V}}(\mathbf{w}) = \left\{\overline{a}, \overline{b}, \overline{c}\right\}$, $\mathbf{w}|_{\mathcal{V}} = a\, b\, c\, d\, a$, $\mathbf{w}|_{\overline{\mathcal{V}}} = \overline{a}\, \overline{b}\, \overline{c}\, \overline{b}$, $\overline{\mathbf{w}} = \overline{a}\, \overline{b}\, a\, \overline{c}\, b\, c\, b\, \overline{d}\, \overline{a}$, $\tilde{\mathbf{w}} = a\, d\, \overline{b}\, \overline{c}\, \overline{b}\, c\, \overline{a}\, b\, a$.

Backwards detailed chain We define the $\mathbf{V}^*$-valued backwards detailed process $(B_n)_{n\in\mathbb{N}}$ as follows: $B_0 = \emptyset$ and for any $n \geq 1$,

- if $W_n = \emptyset$ (i.e. all the items arrived up to time n are matched at time n), then we set $B_n = \emptyset$;
- if not, we let $i(n) \leq n$ be the index of the oldest item in line. Then, the word B_n is of length $n - i(n) + 1$, and for any $\ell \in [\![1, n - i(n) + 1]\!]$, we set

$$B_n(\ell) = \begin{cases} V_{i(n)+\ell-1} & \text{if } V_{i(n)+\ell-1} \text{ has not been matched up to time } n; \\ \overline{V_k} & \text{if } V_{i(n)+\ell-1} \text{ is matched at or before time } n, \text{ with item } V_k \\ & \text{(where } k \leq n). \end{cases}$$

In other words, assuming that the initial system is empty, the word B_n gathers the class indexes of all unmatched items entered up to n, and the copies of the class indexes of the items matched after the arrival of the oldest unmatched item at n, at the place of the class index of the item they have been matched to. Observe that we necessarily have that $B_n(1) = V_{i(n)} \in \mathcal{V}$. Moreover, the word B_n necessarily contains all the letters of W_n. More precisely, we have

$$W_n^{\{\emptyset\}} = B_n|_{\mathcal{V}}, \;\; n \geq 0. \tag{4.5}$$

It is easily seen that $(B_n)_{n\in\mathbb{N}}$ also is a $\mathcal{F}_n$-Markov chain: for any $n \geq 0$, the value of B_{n+1} can be deduced from that of B_n and the class V_{n+1} of the item entered at time $n + 1$.

Forward detailed chain The $\mathbf{V}^*$-valued forward detailed process $(F_n)_{n\in\mathbb{N}}$ is defined as follows: $F_0 = \emptyset$ and for any $n \geq 1$,

- if $W_n = \emptyset$, then we also set $F_n = \emptyset$;
- if not, we let $\mathscr{U}_n$ be the set of items entered up to n and not yet matched at n (which is non empty since $W_n \neq \emptyset$), and set

$$j(n) = \sup \{m > n \,:\, V_m \text{ is matched with an element of } \mathscr{U}_n\}.$$

Notice that $j(n)$ is possibly infinite. Then, F_n is the word of $\mathbf{V}^*$ of size $j(n) - n$ (respectively of $\mathbf{V}^{\mathbb{N}}$ if $j(n) = +\infty$), such that for any $\ell \in [\![1, j(n) - n]\!]$ (resp., $\ell \in \mathbb{N}_+$),

$$F_n(\ell) = \begin{cases} V_{n+\ell} & \text{if } V_{n+\ell} \text{ is not matched with an item arrived up to } n; \\ \overline{V_k} & \text{if } V_{n+\ell} \text{ is matched with item } V_k, \text{ where } k \leq n. \end{cases}$$

In other words, assuming that the initial system is empty, the word F_n contains the copies of all the class indexes of the items entered up to time n and matched after n, together with the class indexes of all unmatched items entered before the last item matched with an item entered up to n, if any.

Recalling that $F_0 = \emptyset \in \mathbf{V}^*$, observe that $F_n \in V^*$ almost surely for all $n \in \mathbb{N}$. Indeed, for any $n \in \mathbb{N}$, and $\mathbf{w} \in \mathbf{V}^*$, on the event $\{F_n \in \mathbf{w}\}$, F_{n+1} being an infinite word would mean that the arriving item at time $n+1$ is never matched, an event with null probability given that G is connected. Observe that if $F_n \in \mathbf{V}^*$ is finite, then $F_n\left(j(n)-n\right) \in \overline{\mathcal{V}}$ since by definition, the item $V_{j(n)}$ is matched with some V_k for $k \leq n$, and therefore $F_n\left(j(n)-n\right) = \overline{V}_k$. It is also clear that $(F_n)_{n\in\mathbb{N}}$ is a $\mathcal{F}_n$-Markov chain, as for any $n \geq 0$, the value of F_{n+1} depends solely on F_n and the class index $V_{j(n)+1}$ of the item entered at time $n + j(n) + 1$.

Example 7. Consider the compatibility graph of Figure 4.1. An arrival scenario together with successive values of the natural chain, the backwards and the forwards chain are represented in Figure 4.2.

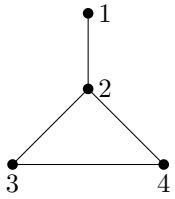


Figure 4.1: Compatibility graph of Example 7.

4.1.4 Dynamic reversibility

For both chains $(B_n)_{n\in\mathbb{N}}$ and $(F_n)_{n\in\mathbb{N}}$, a state $\mathbf{w} \in \mathbf{V}^*$ is said admissible if it can be reached by the chain under consideration under FCFM. We denote

$$\begin{aligned} \mathbb{B} &:= \left\{\mathbf{w} \in \mathbf{V}^* \,:\, \mathbf{w} \text{ is admissible for } (B_n)_{n\in\mathbb{N}}\right\}; \\ \mathbb{F} &:= \left\{\mathbf{w} \in \mathbf{V}^* \,:\, \mathbf{w} \text{ is admissible for } (F_n)_{n\in\mathbb{N}}\right\}. \end{aligned} \tag{4.6}$$

The two subsets $\mathbb{B}$ and $\mathbb{F}$ are isomorphic. More precisely,

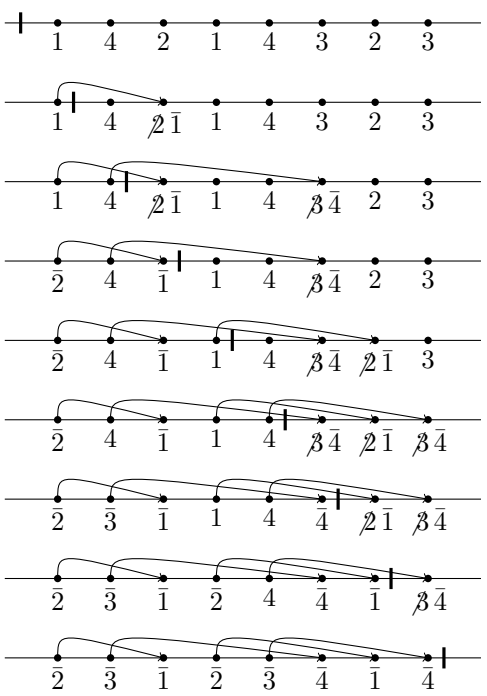
Lemma 4.1.2. *The mapping $\mathbf{w} \mapsto \overleftarrow{\mathbf{w}}$ is one-to-one from $\mathbb{B}$ into $\mathbb{F}$.*

The dynamics of $(B_n)_{n\in\mathbb{N}}$ and $(F_n)_{n\in\mathbb{N}}$ are related in the following sense

Lemma 4.1.3. *Let ν_B be the measure on $\mathbb{B}$ defined by (4.8). Then for any two admissible states $\mathbf{w}, \mathbf{w}'$ for $(B_n)_{n\in\mathbb{N}}$, the states $\overleftarrow{\mathbf{w}}$ and $\overleftarrow{\mathbf{w}'}$ are admissible for $(F_n)_{n\in\mathbb{N}}$ and we have that*

$$\nu_B(\mathbf{w})\mathbb{P}\left[B_{n+1} = \mathbf{w}' | B_n = \mathbf{w}\right] = \nu_B\left(\overleftarrow{\mathbf{w}'}\right)\mathbb{P}\left[F_{n+1} = \overleftarrow{\mathbf{w}} | F_n = \overleftarrow{\mathbf{w}'}\right]. \tag{4.7}$$

Proposition 4.1.4. *Suppose that μ belongs to NCOND(G) defined by (3.12). Then the Backwards detailed Markov chain $(B_n)_{n\in\mathbb{N}}$ is positive recurrent, and admits the following stationary*



$W_0 = \emptyset, \quad B_0 = \emptyset, \quad F_0 = \emptyset$

$W_1 = 1, \quad B_1 = 1, \quad F_1 = 4\bar{1}$

$W_2 = 14, \quad B_2 = 14, \quad F_2 = \bar{1}14\bar{4}$

$W_3 = 4, \quad B_3 = 4\bar{1}, \quad F_3 = 14\bar{4}$

$W_4 = 41, \quad B_4 = 4\bar{1}1, \quad F_4 = 4\bar{4}\bar{1}$

$W_5 = 411, \quad B_5 = 4\bar{1}11, \quad F_5 = \bar{4}\bar{1}\bar{4}$

$W_6 = 14, \quad B_6 = 14\bar{4}, \quad F_6 = \bar{1}\bar{4}$

$W_7 = 4, \quad B_7 = 4\bar{4}\bar{1}, \quad F_6 = \bar{4}$

$W_0 = \emptyset, \quad B_0 = \emptyset, \quad F_0 = \emptyset$

Figure 4.2: An arrival scenario on the graph of Figure 4.1, and the trajectories of the three Markov chains.

measure (unique up to a multiplicative constant) on $\mathbb{B}$*,*

$$\begin{cases} \nu_B(\emptyset) & = 1; \\ \nu_B(\mathbf{w}) & = \prod_{i=1}^{p} \mu(i)^{|\mathbf{w}|_i + |\overline{\mathbf{w}}|_i}, \ \mathbf{w} \in \mathbb{B} \setminus \{\emptyset\}. \end{cases} \tag{4.8}$$

Furthermore,

$$\nu_B(\mathbb{B}) = 1 + \sum_{\mathcal{I} \in \mathbb{I}(G)} \sum_{\sigma \in \mathfrak{S}_{|\mathcal{I}|}} \prod_{j=1}^{|\mathcal{I}|} \frac{\mu\left(i_{\sigma(j)}\right)}{\mu\left(\mathcal{E}\left(\{i_{\sigma(1)}, ..., i_{\sigma(j)}\}\right)\right) - \mu\left(\{i_{\sigma(1)}, ..., i_{\sigma(j)}\}\right)}, \tag{4.9}$$

so ν_B *is a finite measure on* $\mathbb{B}$*.*

Observe that the measure of a word $\mathbf{w}$ does not change whenever any of its letters a is exchanged with $\overline{a}$. In particular, we have $\nu_B(\overleftarrow{\overline{\mathbf{w}}}) = \nu_B(\mathbf{w})$ for any $\mathbf{w}$.

Furthermore, using the connection in Lemma 4.1.3 between the two processes, we prove that the $(F_n)_{n\in\mathbb{N}}$ is the reversed Markov chain of $(B_n)_{n\in\mathbb{N}}$, on a sample space where arrivals are reversed in time and exchanged with their match. In particular $(F_n)_{n\in\mathbb{N}}$ also admits ν_B as a stationary measure.

Sketch of proof of Theorem 4.1.1

We know from Proposition 2 of [97] that if μ is not an element of $\textsc{Ncond}(G)$, then the chain $(W_n)_{n\in\mathbb{N}}$ is transient or null recurrent. If we now assume that $\mu \in \textsc{Ncond}(G)$, then, first,

observe that in view of (4.9), the measure defined for all $\mathbf{w} \in \mathbb{B}$ by $\alpha\nu_B(\mathbf{w})$, for α in (4.4) and ν_B defined by (4.8), defines a probability measure on $\mathbb{B}$. Second, from Proposition 4.1.4, the auxiliary chain $(B_n)_{n\in\mathbb{N}}$ is positive recurrent, and admits ν_B as a stationary measure. So from (4.5), $(W_n)_{n\in\mathbb{N}}$ is also positive recurrent. As it is also irreducible on $\mathbb{W}$, it has a unique stationary probability. To check that the latter is given by Π_W defined by (4.3), it thus suffices to check that

$$\Pi_W(w) = \alpha \sum_{\mathbf{w}\in\mathbb{B}:\, \mathbf{w}|_{\mathcal{V}}=w} \nu_B(\mathbf{w}) \quad \text{for any } w \in \mathbb{W}.$$

4.2 Product form and reversibility properties of BM model

In this section, we summarize the contributions for the $[(\mathcal{D}, \mathcal{S}, \mathcal{E}), \mu, \text{FCFM}]$ bipartite matching (BM) model:

- We derive a Loynes' scheme, which enables to get to stationarity through sample path dynamics, and to prove the existence of a unique FCFM matching over $\mathbb{Z}$.
- We define a pathwise transformation in which we interchange the positions of the two items in a matched demand-supply pair, see Figure 4.4, and we prove the "dynamic reversibility" of the model under this transformation.
- We construct "primitive" Markov chains whose product form stationary distributions are obtained directly from the dynamic reversibility. Using these as building blocks, we derive product form stationary distributions for multiple "natural" Markov chains associated with the model, and we compute various non-trivial performance measures as a by-product.

Detailed results and the proofs can be found in [J11].

4.2.1 FCFM bipartite matching model

We use d_i, $i = 1, \ldots, I$ to denote the different types of demand, and we use d^m to denote the type of the m-th demand item arrived to the system. Similarly we use s_j, $j = 1, \ldots, J$ to denote the types of supply, and s^n denotes the type of the n-th supply item. In figures, we will simplify the notation by writing $d^m = i$ if $d^m = d_i$, and $s^n = j$ if $s^n = s_j$, and we will arrange the sequences in two lines, the top one containing the ordered demand, and the bottom one containing the ordered supply. We will sometimes swap items between the top and the bottom lines in a way to be explained then. In figures, we put an edge between d^m and s^n if they are matched, and we call this edge a *link* in the matching, characterized by the pair of times (m, n).

Assumptions: We assume a connected bipartite *matching* or *compatibility* graph $G = (\mathcal{D}, \mathcal{S}, \mathcal{E})$, with $\mathcal{D} = \{d_1, \ldots, d_I\}$, $\mathcal{S} = \{s_1, \ldots, s_J\}$, and $\mathcal{E} \subset \mathcal{D} \times \mathcal{S}$. Denote by α and β the marginals of μ, i.e. for all $d \in \mathcal{D}$, $\alpha(d) = \mu(d, \mathcal{S})$ and for all $s \in \mathcal{S}$, $\beta(s) = \mu(\mathcal{D}, s)$. In the bipartite matching (BM) model, it is assumed that $\forall d, s,\ \mu(d, s) = \alpha(d)\beta(s)$, i.e. we assume the independence between arriving demand and supply. We assume that $\alpha_d > 0$ for all $d \in \mathcal{D}$ and $\beta_s > 0$ for all $s \in \mathcal{S}$. We also assume that $\mu \in \text{NCOND}$ given by (3.4). We let $\mathcal{S}(c_i)$

be the set of supply types compatible with demand type d_i, and $\mathcal{D}(s_j)$ be the set of demand types compatible with supply type s_j. For $D \subset \mathcal{D}$ and $S \subset \mathcal{S}$, we define

$$\mathcal{S}(C) = \bigcup_{d_i \in D} \mathcal{S}(d_i), \qquad \mathcal{D}(S) = \bigcup_{s_j \in S} \mathcal{D}(s_j)\,.$$

We also define

$$\mathcal{U}(S) = \overline{\mathcal{D}(\overline{S})} = \mathcal{D} \setminus \mathcal{D}(\mathcal{S} \setminus S)$$

(where $\overline{\cdot}$ denotes the complement) which is the set of customer types that can only be served by the servers in S. We call these the *unique demand types* of S. Finally, we let Let

$$\alpha_D = \sum_{d_i \in D} \alpha_{d_i}, \qquad \beta_S = \sum_{s_j \in S} \beta_{s_j}\,.$$

For BM model, it is easy to see that NCOND is equivalent to conditions in the following lemma:

Lemma 4.2.1. *The following three conditions are equivalent (and equivalent to* NCOND*):*

$$\begin{aligned} &\forall D \subset \mathcal{D}, D \neq \emptyset, D \neq \mathcal{D}, \quad \alpha_D < \beta_{\mathcal{S}(D)} \\ &\forall S \subset \mathcal{S}, S \neq \emptyset, S \neq \mathcal{S}, \quad \beta_S < \alpha_{\mathcal{D}(S)} \\ &\forall S \subset \mathcal{S}, S \neq \emptyset, S \neq \mathcal{S}, \quad \beta_S > \alpha_{\mathcal{U}(S)}. \end{aligned} \tag{4.10}$$

These conditions are called *complete resource pooling* conditions in [3].

Consider two independent random sequences $(d^m)_{m\in T}$ and $(s^n)_{n\in T}$ which are chosen respectively i.i.d. from $\mathcal{D}$ according to $\alpha = \{\alpha_{d_1}, \ldots, \alpha_{d_I}\}$, and i.i.d. from $\mathcal{S}$ according to $\beta = \{\beta_{s_1}, \ldots, \beta_{s_J}\}$. The parameter set T can be finite, for instance $T = \{0, 1, 2, \ldots, N\}$, one sided infinite, $T = \mathbb{N}$, or two sided infinite, $T = \mathbb{Z}$.

We study first come first matched (FCFM) policy, as defined in [3]. Informally, this means that an item from S is matched to the earliest possible compatible item from D, and vice versa. More formally,

Definition 4.2.1. *Let $d = (d^m)_{m\in T_1}$ and $s = (s^n)_{n\in T_2}$ be some fixed ordered sequences of demand and supply. Here T_1 and T_2 are index sets which are either finite, one-sided infinite or bi-infinite.*

- *A* partial matching *of d and s is a set $A \subset T_1 \times T_2$ corresponding to demand-supply pairs satisfying:*

$$\begin{aligned} &(i) \quad (m,n) \in A \implies (d^m, s^n) \in \mathcal{E} \\ &(ii) \quad \forall m,\ \#\{n : (m,n) \in A\} \leq 1,\ \forall n,\ \#\{m : (m,n) \in A\} \leq 1\,. \end{aligned}$$

  For a partial matching A, we denote

$$A_d = \{m : \exists n,\, (m,n) \in A\}, \quad A_s = \{n : \exists m,\, (m,n) \in A\}\,.$$

- *A partial matching A is a* (complete) matching *if there are no unmatched compatible pairs left outside of A, that is:*

$$m \notin A_d,\ n \notin A_s \quad \implies \quad (d^m, s^n) \notin \mathcal{E}.$$

- *A matching A is FCFM if for every $(m,n) \in A$,*

  if $l < n$, $(d^m, s^l) \in \mathcal{E}$, then there exists $k < m$ such that $(k,l) \in A$ and
  if $k < m$, $(d^k, s^n) \in \mathcal{E}$, then there exists $l < n$ such that $(k,l) \in A$.

- *A matching is* perfect *if $T_1 = T_2$, and all demands and supplies are matched.*

Figure 4.3 illustrates FCFM matching of two sequences, for a given compatibility graph. Items are browsed from left to right. For instance, d^1 cannot be matched with s^1 nor s^2 since $d^1 = d_1, s^1 = s^2 = s_2$ and $(d_1, s_2) \notin \mathcal{E}$, but it will be matched with $s^3 = s_3$ which is the earliest possible match.

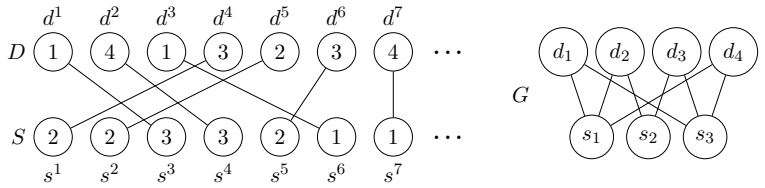


Figure 4.3: FCFM matching of two sequences (left), for the compatibility graph G (right)

Lemma 4.2.2 (Adan & Weiss, [3])**.** *For every finite index sets, there exists a complete FCFM matching, and it is unique.*

The sequences $(d^m)_{m\in\mathbb{N}}$ and $(s^n)_{n\in\mathbb{N}}$ are *matchable* if the number of type d_i demand and the number of type s_j supply is infinite for all $i = 1\ldots, I$ and $j = 1, \ldots, J$. The sequences $(d^m)_{m\in\mathbb{Z}}$ and $(s^n)_{n\in\mathbb{Z}}$ are *matchable* if the same also holds for times $-1, -2, \ldots$.

Theorem 4.2.3 (Adan & Weiss, [3])**.** *For any two matchable sequences $(d^m)_{m\in\mathbb{N}}$ and $(s^n)_{n\in\mathbb{N}}$, there exists a unique perfect FCFM matching.*

The matching can be obtained in a constructive way up to arbitrary length. We consider three methods of constructing a FCFM matching step by step, and define three Markov chains associated with them:

(i) Matching pair by pair: Proceeding from a complete FCFM matching of $(d^m, s^n)_{m,n\le N}$, we add the pair d^{N+1}, s^{N+1}, and match them FCFM to compatible previously unmatched supply and demand if possible, or to each other if possible, or else leave one or both unmatched. With each step we associate a state that consists of the ordered lists of the unmatched supply and demand. It is easy to see that the step by step evolution of the state defines a countable state discrete time irreducible and aperiodic Markov chain. We denote it by $O = (O_N)_{N\in\mathbb{N}}$, this is the 'natural' pair by pair FCFM Markov chain.

(ii) Matching supply by supply: Proceeding from the FCFM matching of all the supply items s^n, $n \le N$, we add the next supply s^{N+1}, and match it to the first compatible demand that has not yet been matched. With each step we associate a state that consists of the ordered list of skipped demands. This again defines a countable state discrete time irreducible and aperiodic Markov chain. We denote it by $Q^s = (Q^s_N)_{N\in\mathbb{N}}$, this is the 'natural' supply by supply FCFM Markov chain.

(iii) Matching demand by demand is analogous, resulting in a 'natural' demand by demand FCFM Markov chain $Q^d = (Q^d_N)_{N\in\mathbb{N}}$.

All three Markov chains have a common state if at time N a perfect matching of all previous demand and supply items is reached, in which case the state consists of empty lists and is denoted by $\emptyset$ or by 0. Because these Markov chains have a common state, they will be transient, null recurrent, or ergodic together.

Definition 4.2.2. *We will say that the FCFM bipartite matching is* ergodic *if the corresponding Markov chains are ergodic.*

Theorem 4.2.4 (Adan & Weiss [3])**.** *The FCFM bipartite matching is ergodic if and only if the conditions in Lemma 4.2.1 hold.*

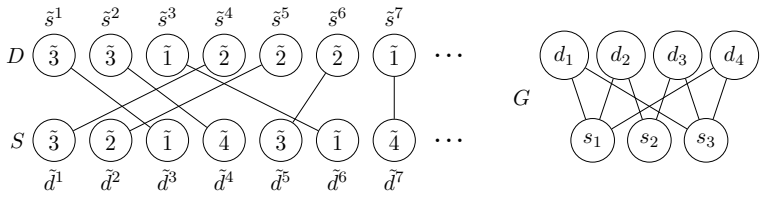


Figure 4.4: The matched sequences of Figure 4.3, after the exchange transformation

4.2.2 Loynes' construction of FCFM bipartite matching over $\mathbb{Z}$

In this section we show that if complete resource pooling holds, then for two independent bi-infinite sequences of i.i.d. demands and supplies, $(d^n, s^n)_{n\in\mathbb{Z}}$ there exists almost surely a unique FCFM matching. This matching coincides with the matchings obtained from the stationary versions of the various Markov chains described above. The matching is obtained using a Loynes' type scheme (see [94] for the original Loynes' construction).

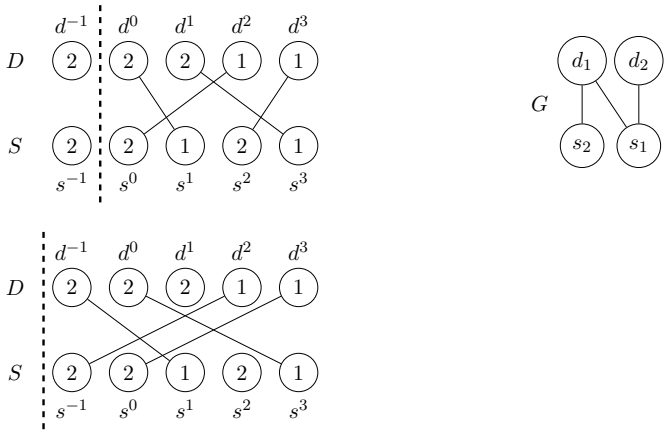


Figure 4.5: Backward step changes the matching

Consider the (unique by Theorem 4.2.3) matching of $(d^n, s^n)_{n>-K}$, and let $K \to \infty$. At first glance we notice that as we let K increase the matching changes. This is illustrated

in the example of Figure 4.5. In this example, if and $d^{-K-1} = d_2$ and $s^{-K-1} = s_2$, then starting empty at $-K-1$, the state at $-K$ cannot be empty, so the matching starting empty from $-K$ is different from the matching starting from $-K-1$. Nevertheless, in this simple example, classical results can be used to prove that the matching does converge as $K \to \infty$. The condition for that is that complete resource pooling should hold, which in this example happens when $\alpha_2 < \beta_1$.

We sketch the argument to provide some intuition for the general proof to come. Denote by $(O_n^{[-K]})_{n \geq -K}$, the pair by pair FCFM Markov chain associated with the sequence $(d^n, s^n)_{n \geq -K}$ (see the definition above). Denote by $X_n^{[-K]}$ the number of unmatched demands, all of them of type d_2, (or of unmatched supplies, all of them of type s_2) in $O_n^{[-K]}$. Observe that:

$$X_{n+1}^{[-K]} = \begin{cases} X_n^{[-K]} + 1 & \text{if} \quad (d^{n+1}, s^{n+1}) = (d_2, s_2) \\ \max(X_n^{[-K]} - 1, 0) & \text{if} \quad (d^{n+1}, s^{n+1}) = (d_1, s_1) \\ X_n^{[-K]} & \text{if} \quad (d^{n+1}, s^{n+1}) = (d_1, s_2) \text{ or } (d_2, s_1)\,. \end{cases}$$

So $(X_n^{[-K]})_{n \geq -K}$ can be interpreted as the queue-length process of a discrete-time version of an M/M/1 queue. Complete resource pooling then implies that the drift of this queue is negative. A straightforward adaptation of the original Loynes argument (designed for a continuous time G/G/1 queue with negative drift), yields the existence of a limiting process for $(X_n^{[-K]})_{n \geq -K}$ when $K \to +\infty$, see for instance Chapter 2.1 in [8]. Let us denote the limiting process by $(X_n^{\infty})_{n \in \mathbb{Z}}$. Again, the original Loynes argument shows that the set of indices $\{n \in \mathbb{Z},\ X_n^{\infty} = 0\}$ is a.s. infinite. These indices can be viewed as regeneration points for the processes $(X_n^{[-K]})_{n \geq -K}$. In particular, if $X_k^{\infty} = 0$ then $X_k^{[-K]} = 0$ for all $-K \leq k$. Now, observe that these regeneration points are also regeneration points for the processes $(O_n^{[-K]})_{n \geq -K}$, that is, $[X_k^{\infty} = 0] \implies [\forall K,\ -K \leq k,\ O_k^{[-K]} = \emptyset]$. Between two regeneration points, we have a finite complete matching of the sequences of demands and of supplies, and this matching will not change over time. Therefore, there exists a limiting bi-infinite matching which is obtained by simply concatenating the finite matchings between regeneration points.

In general, we prove the following result:

Theorem 4.2.5. *For two independent i.i.d. sequences $(d^n, s^n)_{n \in \mathbb{Z}}$, if complete resource pooling holds, there exists almost surely a unique FCFM matching over all of $\mathbb{Z}$, and it can be obtained by Loynes' scheme, of constructing a FCFM matching from $-K$ to ∞, and letting $K \to \infty$.*

The proof uses two pathwise results which do not depend on any probabilistic assumption: monotonicity and subadditivity.

Lemma 4.2.6 (Monotonicity)**.** *Consider $d^1, \ldots, d^M$ and $s^1, \ldots, s^N$, and complete FCFM matching between them. Assume there are K demands and L supplies left unmatched. Consider now an additional demand d^0, and the complete FCFM matching between $d^0, d^1, \ldots, d^M$ and $s^1, \ldots, s^N$. Then this matching will have no more than $K+1$ demands and L supplies unmatched.*

Lemma 4.2.7 (Subadditivity)**.** *Let $A' = (d^1, \ldots, d^m)$, $A'' = (d^{m+1}, \ldots, d^M)$ and $B' = (s^1, \ldots, s^n)$, $B'' = (s^{n+1}, \ldots, s^N)$ and let $A = (d^1, \ldots, d^M)$, $B = (s^1, \ldots, s^N)$. Consider*

the complete FCFM matching of A', B', of A'', B'', and of A, B and let K', K'', K be the number of unmatched demands and L', L'', L be the number of unmatched supplies in these three matchings. Then $K \leq K' + K''$ and $L \leq L' + L''$.

The third ingredient is the existence of a simple path (i.e. with no repeated nodes) in the compatibility graph G. This follows directly from the connectivity assumption for G.

Lemma 4.2.8. *Consider an incompatible pair c^0, s^0. Then there exists an h and a sequence $d^1, \ldots, d^h$, $s^1, \ldots, s^h$ with $h \leq \min\{I, J\} - 1$, where (s^i, d^i), $i = 1, \ldots, h$ are compatible, such that the FCFM matching of $d^0, d^1, \ldots, d^h, s^0, s^1, \ldots, s^h$ is perfect.*

Clearly, the probability of occurrence of such a sequence is strictly positive and lower bounded by: $\delta = \prod_{(d,s)\in\mathcal{E}} \alpha_d \beta_s$.

We assume from now on in this section that complete resource pooling holds. By Theorem 4.2.4, the pair by pair matching Markov chain $(O_N)_{N\in\mathbb{N}}$ is ergodic. Using the Kolmogorov extension theorem [113], we may define (in a non-constructive way) a stationary version $O^* = (O_N^*)_{N\in\mathbb{Z}}$ of the Markov chain. Define also $O^{[k]} = (O_N^{[k]})_{N\geq -k}$ the realization of the Markov chain that starts at $O_{-k}^{[k]} = \emptyset$.

The proof of Theorem 4.2.5 consists now of two steps:

Forward coupling. The proof uses Lemmas 4.2.6, 4.2.7, 4.2.8.

Proposition 4.2.9 (Forward coupling)**.** *The two processes $(O_n^*)_{n\in\mathbb{N}}$ and $(O_n^{[0]})_{n\in\mathbb{N}}$ will couple after a finite time τ, with $E(\tau) < \infty$.*

Note that once $O^{[0]}$ and O^* couple, they stay together forever.

Backward coupling. The second step is based on standard arguments to show backward coupling and convergence to a unique matching.

Proposition 4.2.10 (Backward coupling)**.** *Let O^* be the stationary pair by pair FCFM matching process, and let $O^{[-k]}$ be the process starting empty at time $-k$. Then $\lim_{k\to\infty} O_N^{[-k]} = O_N^*$ for all $N \in \mathbb{Z}$ almost surely.*

Each process $O_N^{[-k]}$ determines matches uniquely for all $N > -k$, so if we fix N, matches from N onwards are uniquely determined by $\lim_{k\to\infty} O_N^{[-k]}$. Hence $(O_N^*)_{N\in\mathbb{Z}}$ determines for every supply s^n and every demand d^n his match, uniquely, almost surely.

4.2.3 Exchange transformation and dynamic reversibility

The FCFM matching depends on the time direction in which it is constructed. The simple example in Figure 4.6 shows that FCFM is not preserved if the time direction is reversed.

Nevertheless, this model has a dynamic reversibility result. In this section we introduce the exchange transformation, in which we switch the positions of each matched pair of demand and supply. Figures 4.3 and 4.4 illustrate the exchange transformation. We show that the exchanged sequences are independent i.i.d. and that the matching is FCFM in reversed time.

Definition 4.2.3. *Consider a FCFM bipartite matching of sequences $(d^n, s^n)_{n\in T}$. The exchange transformation of the matched pair d^n, s^m, is the matched pair $\tilde{d}^m, \tilde{s}^n$ where $\tilde{s}^n = s^m$ and $\tilde{d}^m = d^n$.*

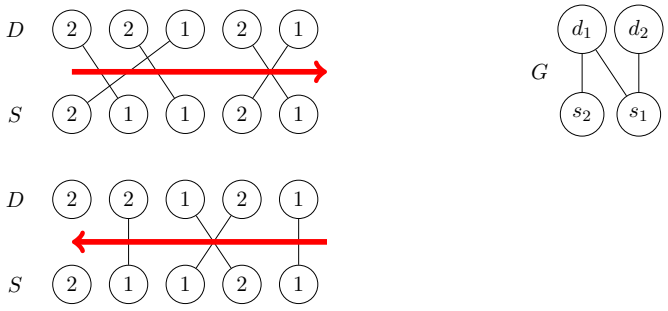


Figure 4.6: FCFM is not preserved when time is reversed

Lemma 4.2.11. *Let A be a perfect matching of $d^1, \dots, d^M$, and $s^1, \dots, s^M$. Let $\tilde{s}^1, \dots, \tilde{s}^M$, $\tilde{d}^1, \dots, \tilde{d}^M$ be the sequences obtained by the exchange transformation, retaining the same links of the matched pairs. The resulting matching of $\tilde{d}^1, \dots, \tilde{d}^M$, $\tilde{s}^1, \dots, \tilde{s}^M$ is the unique FCFM matching in reversed time.*

Lemma 4.2.12. *Consider the FCFM matching of two i.i.d. sequences, and let $\mathcal{O}_{(m+1,m+M)}$ be the block of demand and supply items for the times $[m+1, m+M]$. Then the conditional probability of observing values $\mathcal{O}_{(m+1,m+M)} = \big((d^{m+1}, \dots, d^{m+M}), (s^{m+1}, \dots, s^{m+M})\big)$ conditional on the event that the FCFM matching of these values is a perfect match is:*

$$\begin{aligned} &P\Big(\mathcal{O}_{m+1,m+M} = \big((d^{m+1}, \dots, d^{m+M}), (s^{m+1}, \dots, s^{m+M})\big) \\ &\qquad \Big| \, \mathcal{O}_{m+1,m+M} \text{ forms a perfect FCFM match}\Big) \\ &= \kappa_M \prod_{i=1}^{I} {\alpha_{d_i}}^{\#d_i} \prod_{j=1}^{J} {\beta_{s_j}}^{\#s_j} \end{aligned}$$

where κ_M is a constant that may depend on M, and $\#d_i$, $\#s_j$ count the number of type d_i demand and type s_j supply items in the block.

Corollary 4.2.13. *Let $\mathcal{O}_{m+1,m+M}$ be a FCFM perfectly matched block, and let $\overleftarrow{\mathcal{O}}_{m+1,m+M}$ be obtained from $\mathcal{O}_{m+1,m+M}$ by performing the exchange transformation and time reversal. Then $\overleftarrow{\mathcal{O}}^M$ is a FCFM perfectly matched block, and*

$$P(\overleftarrow{\mathcal{O}}_{m+1,m+M}) = P(\mathcal{O}_{m+1,m+M})$$

Theorem 4.2.14. *Consider a bipartite matching model under complete resource pooling (conditions in Lemma 4.2.1). Let $(d^n, s^n)_{n\in\mathbb{Z}}$ be the independent i.i.d. sequences of demand and supply items, with the unique FCFM matching between them. Then the exchanged sequences $(\tilde{d}^n, \tilde{s}^n)_{n\in\mathbb{Z}}$ are independent i.i.d. of the same law as $(d^n, s^n)_{n\in\mathbb{Z}}$. The unique FCFM matching in reverse time between them (using Loynes' construction in reversed time) consists of the same links as the matching between $(d^n, s^n)_{n\in\mathbb{Z}}$.*

We have found that for any two independent i.i.d. sequences of demand $D = (d^m)_{m\in\mathbb{Z}}$ and of supply items $S = (s^n)_{n\in\mathbb{Z}}$, under complete resource pooling, there is a unique FCFM matching almost surely. Furthermore, if we exchange every matched pair (d^n, s^m) of demand and supply and retain the matching, we obtain two permuted sequences, of matched and

exchanged demand $\tilde{D} = (\tilde{d}^n)_{n\in\mathbb{Z}}$, and of matched and exchanged supply $\tilde{S} = (\tilde{s}^m)_{m\in\mathbb{Z}}$. These new sequences are again independent and i.i.d., and the retained matching between them is FCFM in reversed time direction, and it is the unique FCFM matching of $\tilde{D}$, $\tilde{S}$ in reversed time.

4.2.4 Stationary distributions

We consider the Markovian evolution of the stationary FCFM matching on $\mathbb{Z}$, and derive stationary distributions of several Markov chains associated with it. The FCFM matching of D, S evolves moving step by step from the past up to position N, where we add matches and perform exchanges at each step from N to $N+1$. Four ways in which this can be done are illustrated in Figure 4.7 where light circles represent demand and dark circles represent supply (in original or exchanged positions). In each of these, if we reverse the time direction we get a Markovian construction of FCFM matches between $\tilde{D}$, $\tilde{S}$ that moves from $N+1$ to N and at each of these steps adds matches for elements $\tilde{d}^n, \tilde{s}^m$ and exchanges them back to d^m, s^n. We exploit this reversibility to derive the stationary distributions.

The outline of this subsection is as follows: we first describe in more details the four mechanisms and define a Markov chain associated with each. The states of these processes consist of the ordered lists of items in the region encircled by a dashed ellipse in each of the four panels in Figure 4.7. We call these the detailed Markov chains. Then we formulate Lemmas 4.2.15, 4.2.16 on time reversal, that associates each Markov chain in the forward time direction, with a corresponding Markov chain in the reversed time direction. We use time reversal to derive in Theorem 4.2.17 the stationary distributions of the detailed Markov chains. These are, up to a normalizing constant, simply the distributions of a finite sequence of multi-Bernoulli trials. Also, all these distributions possess the same normalizing constant.

Then we define a Markov chain with an augmented state description, and obtain its stationary distribution as a corollary to Theorem 4.2.17. The advantage of this augmented chain is that its state can be re-interpreted as the state of a queue with parallel servers which is overloaded, as described in [137, 3, 4]. Under this interpretation it is possible to sum over the detailed states and to obtain the stationary distribution of a host of other processes associated with FCFM matching. Furthermore, by summing over all the states we obtain the normalizing constant for the stationary distributions of Theorem 4.2.17. We conjecture that its calculation is ♯-P hard.

Finally, we again sum over states to obtain the stationary distribution of the 'natural' Markov chains. We illustrate this for the FCFM matching model of the "NN"-system.

Mechanisms for evolution of FCFM matching and detailed Markov chains

We consider four mechanisms for the Markovian evolution of the stationary FCFM matching, and define an associated Markov chain for each.

Supply by supply matching

At time N all supply items s^n, $n \le N$ have been matched and exchanged with the demand items to which they were matched, as illustrated in panel (i) of Figure 4.7. At this point the supply line has $\tilde{d}^n$, $n \le N$ demand items that matched and replaced by supply items s^n, $n \le N$, and supply items s^n, $n > N$ are still unmatched. On the demand line there is

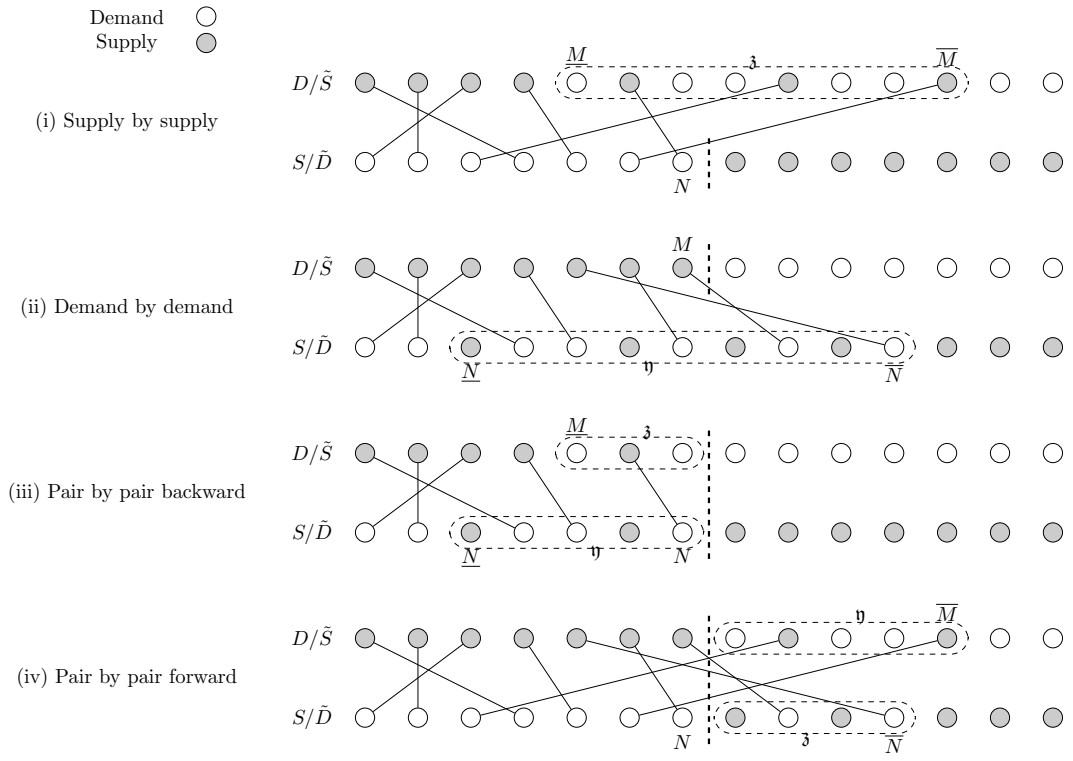


Figure 4.7: Four mechanisms of FCFM matching

a position $\underline{M}$ such that all the demand items d^m, $m < \underline{M}$ have been matched and replaced by $\tilde{s}^m$, and $d^{\underline{M}}$ is the first unmatched demand item, and there is a position $\overline{M}$ such that all demand items d^m, $m > \overline{M}$ have not yet been matched, and $d^{\overline{M}}$ is the last demand item that has been matched, so that now $\tilde{s}^{\overline{M}}$ is the matched and exchanged supply item in position $\overline{M}$. If the matching for $n \leq N$ is perfect then $\overline{M} = \underline{M} - 1 = N$, otherwise $L = \overline{M} - \underline{M} + 1 \geq 2$. We let $\mathfrak{z} = 0$ in the former case (sometimes we write $\mathfrak{z} = \emptyset$), and in the latter case we let $\mathfrak{z} = (z^1, \ldots, z^L)$ be the ordered sequence of unmatched demand and of matched and exchanged supply items so that $\mathfrak{z}^1 = d^{\underline{M}}$, $\mathfrak{z}^L = \tilde{s}^{\overline{M}}$ and z^l, $1 < l < L$ is either $d^{\underline{M}+l-1}$ if unmatched or $\tilde{s}^{\underline{M}+l-1}$ if matched and exchanged.

We define the *supply by supply FCFM detailed matching process* $Z^s = (Z^s_N)_{N\in\mathbb{Z}}$ with $Z^z_N = \mathfrak{z}$. It is a Markov chain where the transition from Z^s_N to Z^s_{N+1} depends on the current state $\mathfrak{z}$, and on the innovation variables which are the types of supply item s^{N+1} and of demand items d^m, $m > \overline{M}$.

Demand by demand matching

Similar to supply by supply matching, at time M all demand items d^m, $m \leq M$ have been matched and exchanged with supply items, as illustrated in panel (ii) of Figure 4.7. We define a *demand by demand FCFM detailed matching process*, $Z^d = (Z^d_M)_{M\in\mathbb{Z}}$ so that the state $Z^d_M = \mathfrak{y}$ is $\mathfrak{y} = 0$ for perfect match, and otherwise $\mathfrak{y} = (z^1, \ldots, z^L)$ where $z^1 = s^{\underline{N}}$ is the first unmatched supply item on the supply line, $z^L = \tilde{d}^{\overline{N}}$ is the last matched and exchanged demand item, and z^l, $1 < l < L$ is either $s^{\underline{N}+l-1}$ if unmatched or $\tilde{d}^{\underline{N}+l-1}$ if matched and exchanged. It is a Markov chain where the transition from Z^d_M to Z^d_{M+1} depends on the

current state $\mathfrak{y}$, and on the innovation variables which are the types of demand item d^{M+1} and of supply items s^n, $n > \overline{N}$.

Pair by pair backward matching

For pair by pair backward FCFM matching (illustrated in panel (iii) of Figure 4.7) we assume that all possible FCFM matches between s^n, d^m, $m, n \leq N$ have been made and exchanged, and in step $N+1$ we add the pair s^{N+1}, d^{N+1}, and if possible match and exchange each of them FCFM to previous unmatched items or to each other.

We define the pair by pair backwards detailed FCFM matching process $D = (D_N)_{N\in\mathbb{Z}}$ as $D_N = (\mathfrak{z}, \mathfrak{y})$, where $\mathfrak{z} = (z^1, \ldots, z^L)$ describes the demand line and $\mathfrak{y} = (y^1, \ldots, y^K)$ describes the supply line. Here z^1 is the first unmatched demand item, in position $N - L + 1$, and the remaining items of $\mathfrak{z}$ are either unmatched demand items or matched and exchanged supply items, y^1 is the first unmatched supply item, in position $N - K + 1$, and the remaining items of $\mathfrak{y}$ are either unmatched supply items or matched and exchanged demand items. The number of unmatched demand items in $\mathfrak{z}$ needs to be equal to the number of unmatched supply items in $\mathfrak{y}$. We may have $\mathfrak{z} = \mathfrak{y} = 0$ if there is a perfect match, otherwise both $L \geq 1$ and $K \geq 1$. This is a Markov chain, whose next state depends on the current state and the random innovation consists of the types of s^{N+1}, d^{N+1}.

Pair by pair forward matching

For pair by pair forward FCFM matching (illustrated in panel (iv) of Figure 4.7) we assume all demand items s^n, d^m, $m, n \leq N$ have been matched and exchanged. After step N we consider the pair in position $N+1$, which may contain either items which were matched and exchanged already, or items which are still unmatched, and then in step $N+1$ the items which are still unmatched after step N are matched and exchanged with each other or with items in positions $> N+1$.

We define the pair by pair forward FCFM matching process $E = (E_N)_{N\in T}$ as $E_N = (\mathfrak{y}, \mathfrak{z})$, where $\mathfrak{y} = (y^1, \ldots, y^K)$ lists items in positions $N+1, \ldots, N+K$ on the demand line, where y^K is the last matched and exchanged supply item $\tilde{s}^{N+K}$, and y^k, $1 \leq k < K$ is either an unmatched demand or a matched and exchanged supply item in position $N+k$, and where $(z^1, \ldots, z^L)$ lists items in positions $N+1, \ldots, N+L$ on the supply line, where z^L is the last matched and exchanged demand item $\tilde{d}^{N+L}$ and z^l, $1 \leq l < L$ is either an unmatched supply, or a matched and exchanged demand in position $N+k$. $E_N = \emptyset$ after a perfect match, otherwise $K, L \geq 1$. E_N is a Markov chain, whose next state depends on the current state, and the random innovation consists of the d^m, $m > N+L$ and s^n, $n > N+K$.

Time reversal of the detailed Markov chains

Examining panel (i) of Figure 4.7 we see that it illustrates FCFM matching and exchange of supply and demand lines D and S all the way from $-\infty$ up to position N, and at the same time it also illustrates matching and exchange of supply and demand lines $\tilde{D}$ and $\tilde{S}$ FCFM in reversed time, all the way from ∞ to $N+1$. Our main observation now is that if $Z_N^s = \mathfrak{z} = (z^1, \ldots, z^L)$, then $(z^L, \ldots, z^1)$ is exactly the state of the demand by demand FCFM matching in reversed time of the sequences $\tilde{D}$, $\tilde{S}$, when all demand items $\tilde{d}^n, n \geq N+1$ have been matched to some $\tilde{s}^m$, and exchanged back to a demand d^m and supply s^n. For $\mathfrak{z} = (z^1, \ldots, z^L)$ we denote $\overleftarrow{\mathfrak{z}} = (z^L \ldots, z^1)$. We denote the Markov chain of demand by

demand FCFM matching of $\tilde{D}$, $\tilde{S}$ in reversed time by $\overleftarrow{Z}$, so that $\overleftarrow{Z}^d_N$, is the state where all $\tilde{d}^n$, $n > N$ have been matched. We then state formally (see Figure 4.8):

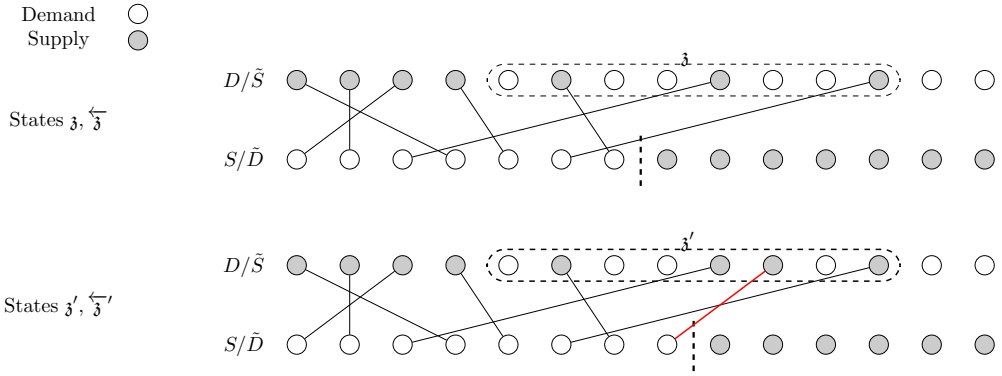


Figure 4.8: Single match and exchange of demand supply pair and its reversal

Lemma 4.2.15 (Time reversal)**.** *The Markov chain $\overleftarrow{Z}^d_N$ of demand by demand FCFM matching of $\tilde{D}$, $\tilde{S}$ in reversed time, is the time reversal of the Markov chain Z^s_N of supply by supply FCFM matching of D, S, in the sense that*

$$Z^s_N = \mathfrak{z},\ Z^s_{N+1} = \mathfrak{z}' \text{ if and only if } \overleftarrow{Z}^d_{N+1} = \overleftarrow{\mathfrak{z}'},\ \overleftarrow{Z}^d_N = \overleftarrow{\mathfrak{z}}. \tag{4.11}$$

This implies that the reversal of the transition $Z^s_N = \mathfrak{z} \to Z^s_{N+1} = \mathfrak{z}'$ is exactly the transition $\overleftarrow{Z}^d_{N+1} = \overleftarrow{\mathfrak{z}'} \to \overleftarrow{Z}^d_N = \overleftarrow{\mathfrak{z}}$. In other words, if the transition of $Z^s_N \to Z^s_{N+1}$ matches and exchanges s^n with d^m, then the transition of $\overleftarrow{Z}^d_{N+1} \to \overleftarrow{Z}^d_N$ matches and exchanges $\tilde{d}^n$ with $\tilde{s}^m$.

A similar observation on time reversal holds also for the pair by pair backward and forward detailed Markov chains. Examining panel (iii) of Figure 4.7 we again see that it illustrates FCFM matching and exchange of supply and demand lines D and S all the way from $-\infty$ up to position N, and at the same time it also illustrates matching and exchange of supply and demand lines $\tilde{D}$ and $\tilde{S}$ FCFM in reversed time, all the way from ∞ to $N+1$. Our main observation now is that if $Z^s_N = (\mathfrak{z}, \mathfrak{y}) = \big((z^1, \ldots, z^L), (y^1, \ldots, y^K)\big)$, then $\big((y^K, \ldots, y^1), (z^L, \ldots, z^1)\big) = (\overleftarrow{\mathfrak{y}}, \overleftarrow{\mathfrak{z}})$ is exactly the state of the pair by pair forward detailed FCFM matching in reversed time, of the sequences $\tilde{D}$, $\tilde{S}$, when all demand and supply items $\tilde{d}^n, \tilde{s}^m$, $m, n > N$ have been matched and exchanged back to a demand d^m and supply s^n. We denote the pair by pair forward detailed FCFM matching of $\tilde{D}$, $\tilde{S}$ in reversed time by $\overleftarrow{E}_N$. We then state formally (see Figure 4.9):

Lemma 4.2.16 (Time reversal)**.** *The Markov chain $\overleftarrow{E}_N$ of pair be pair forward FCFM matching of $\tilde{D}$, $\tilde{S}$ in reversed time, is the time reversal of the Markov chain D_N of pair by pair backward FCFM matching of D, S, in the sense that*

$$D_N = (\mathfrak{z}, \mathfrak{y}),\ D_{N+1} = (\mathfrak{z}', \mathfrak{y}') \text{ if and only if } \overleftarrow{E}_{N+1} = (\overleftarrow{\mathfrak{y}'}, \overleftarrow{\mathfrak{z}'}),\ \overleftarrow{E}_N = (\overleftarrow{\mathfrak{y}}, \overleftarrow{\mathfrak{z}}). \tag{4.12}$$

This implies that the reversal of the transition $D_N = \mathfrak{z} \to Z_{N+1} = \mathfrak{z}'$ is exactly the transition $\overleftarrow{E}_{N+1} = (\overleftarrow{\mathfrak{y}'}, \overleftarrow{\mathfrak{z}'}) \to \overleftarrow{E}_N = (\overleftarrow{\mathfrak{y}}, \overleftarrow{\mathfrak{z}})$. In other words, if the transition of $D_N \to D_{N+1}$ looks for matches for d^{N+1}, s^{N+1} and exchanges s^n with d^m, then the transition of $\overleftarrow{E}_{N+1} \to \overleftarrow{E}_N$ considers the elements in position N, and transforms them back.

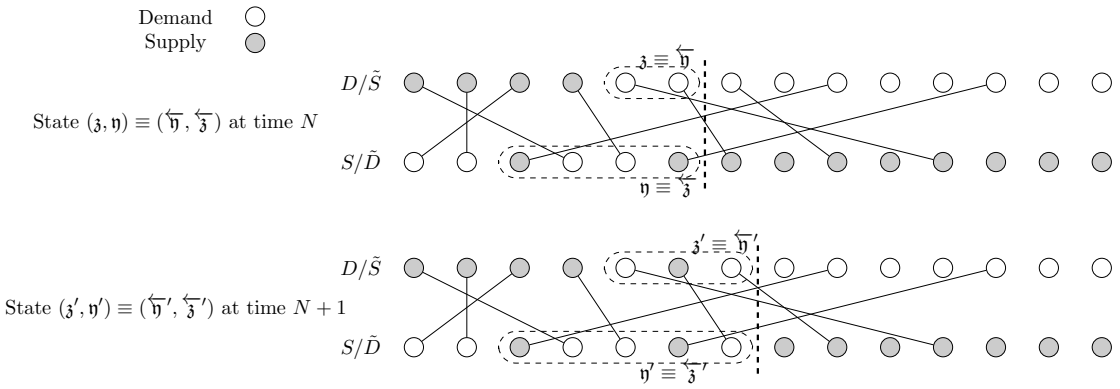


Figure 4.9: Adding pair backward and its reversal adding pair forward

Stationary distributions of the detailed Markov chains

We are now ready to derive the stationary distributions of the detailed Markov chains.

Theorem 4.2.17. *(i) The stationary distribution of Z^s_N and of Z^d_M is given, up to a normalizing constant, by*

$$\pi_{Z^s}(\mathfrak{z}) = \pi_{Z^d}(\overleftarrow{\mathfrak{z}}) = \prod_{i=1}^{I} \alpha_{d_i}{}^{\#d_i} \prod_{j=1}^{J} \beta_{s_j}{}^{\#s_j}, \tag{4.13}$$

where $\#d_i$ is the number of demand items of type d_i, and $\#s_j$ is the number of supply items of type s_j, as they appear in the state $\mathfrak{z}$.

(ii) The stationary distribution of D_N and of E_N is given, up to a normalizing constant, by

$$\pi_D(\mathfrak{z},\mathfrak{y}) = \pi_E(\overleftarrow{\mathfrak{y}}, \overleftarrow{\mathfrak{z}}) = \prod_{i=1}^{I} \alpha_{d_i}{}^{\#d_i} \prod_{j=1}^{J} \beta_{s_j}{}^{\#s_j}, \tag{4.14}$$

where $\#d_i$, $\#s_j$ count demand and supply items as they appear in the state $(\mathfrak{z},\mathfrak{y})$.

(iii) The normalizing constant is the same in all four distributions.

We prove (i) using the time reversal result and Kelly's Lemma ([83], Section 1.7): For a Markov chain X_t, if we can find non-negative $\pi(i)$ and $p_{j\to i}$ such that

$$\sum_i p_{j\to i} = 1 \text{ for all } j, \quad \text{and} \quad \pi(i)P(X_{t+1} = j \,|\, X_t = i) = \pi(j)p_{j\to i} \text{ for all } i,j \tag{4.15}$$

then π is the stationary distribution of X_t, and $p_{j\to i}$ are the transition rates of the reversed stationary process, $p_{j\to i} = P(X_t = i \,|\, X_{t+1} = j)$. The proof of (ii) is similar. To show (iii) we show that there is a 1-1 correspondence between states of Z^s_N and D_N.

Augmented state, marginals and the normalizing constant

The extreme simplicity of the stationary probabilities obtained in Theorem 4.2.17 is deceptive, since it does not indicate which states are possible, according to the compatibility graph and the FCFM matching policy. In particular, there seems to be no simple way of deciding what

are all the possible states of the four detailed Markov chains. As a result it is not at all obvious how to calculate the normalizing constant for the distributions (4.13), (4.14). To allow us to classify states in a convenient way, and thus to allow us to count them and to add up their stationary probabilities, we define an augmented detailed Markov chain. It also describes the supply by supply FCFM matching mechanism, but its states are augmentations of the states of Z^s, in that they are even more detailed.

Consider supply by supply FCFM matching when all supply items up to position N have been matched and exchanged. For the supply by supply FCFM detailed matching process we defined Z^s_N as the sequence of elements from position $\underline{M}$ of the first unmatched demand item to position $\overline{M}$ of the last matched and exchanged supply on the demand line. We now consider positions $\underline{N} < \underline{M}$ and $\overline{N} > \overline{M}$ such that the interval of positions $\underline{N}$ to $\overline{N}$ contains for each supply type at least one matched and exchanged supply item, and it contains for each demand type at least one unmatched demand, and the interval is minimal. Let $\mathfrak{z} = (z^1, \ldots, z^K)$ where $K = \overline{N} - \underline{N} + 1$, $z^1 = \tilde{s}^{\underline{N}}$, $z^K = d^{\overline{N}}$, and for $\underline{N} < l < \overline{N}$, z^l is either an unmatched demand or a matched and exchanged supply in position $\underline{N} + l - 1$. We consider the process ${}^oZ^s = ({}^oZ^s_N)_{N\in\mathbb{Z}}$ where ${}^oZ^s_N = \mathfrak{z}$. Note that ${}^oZ^s_N = \mathfrak{z}$ differs from Z^s_N by the addition of some supply items before $d^{\underline{M}}$ and some demand items after $\tilde{s}^{\overline{M}}$. We always have $K \geq I + J$.

We define also the *supply by supply FCFM augmented matching process* $\mathbb{Z} = (\mathbb{Z}_N)_{N\in T}$ with state $\mathfrak{z} = (z^1, \ldots, z^L)$ with $L = \overline{M} - \underline{N} + 1 \geq J$, which includes elements from positions $\underline{N}$ to $\overline{M}$ on the demand line, starting with $z^1 = \tilde{s}^{\underline{N}}$ and ending with $z^L = \tilde{s}^{\overline{M}}$

Corollary 4.2.18. *The stationary distributions of ${}^oZ^s$ and of $\mathbb{Z}$ are given by*

$$\pi_{{}^oZ^s}(\mathfrak{z}) = B \prod_{i=1}^{I} \alpha_{c_i}{}^{\#c_i} \prod_{j=1}^{J} \beta_{s_j}{}^{\#s_j}, \tag{4.16}$$

$$\pi_{\mathbb{Z}}(\mathfrak{z}) = B \prod_{i=1}^{I} \alpha_{c_i}{}^{\#c_i} \prod_{j=1}^{J} \beta_{s_j}{}^{\#s_j}, \tag{4.17}$$

where $\#d_i$ is the number of demand items of type d_i in $\mathfrak{z}$, and $\#s_j$ is the number of supply items of type s_j in $\mathfrak{z}$.

The motivation for considering the augmented process $\mathbb{Z}$ is that each state $\mathbb{Z}_N = \mathfrak{z}$ can be written in a different form, and in that form we can actually enumerate all the possible states. This enables us to obtain stationary distributions of various marginal processes, and finally to derive an explicit expression for the normalizing constant B. We now rewrite the state $\mathfrak{z} = z^1, \ldots, z^L$ as follows: Let S_J be the type of supply $z^L = \tilde{s}^{\overline{M}}$. Define recursively, for $1 \leq j < J$, S_j as the type of the last supply in the sequence $z^1, \ldots, z^L$ which is different from $S_{j+1}, \ldots, S_J$. Then $R = (S_1, \ldots, S_J)$ is a permutation of the supply types $s_1, \ldots, s_J$. Let $\mathfrak{w}_1, \ldots, \mathfrak{w}_{J-1}$ be the subsequences of demand and supply types between the locations of $S_1, \ldots, S_J$ in $\mathfrak{z}$. We will then write the state as $\mathfrak{z} = (S_1, \mathfrak{w}_1, \ldots, \mathfrak{w}_{J-1}, S_J)$. The idea of presenting the state in this form stems from [138, 137] and was used in [3, 4].

The main feature of $\mathfrak{z} = (S_1, \mathfrak{w}_1, \ldots, \mathfrak{w}_{J-1}, S_J)$ is that all the demand items in $\mathfrak{w}_\ell$ are of types in $\mathcal{U}(S_1, \ldots, S_\ell)$ and all the supply items in $\mathfrak{w}_\ell$ are of types in $\{S_{\ell+1}, \ldots, S_J\}$. Of course

we can write the stationary distribution of states of $\mathbb{Z}$, given in (4.17), also as:

$$\pi_{\mathbb{Z}}(S_1, \mathfrak{w}_1, \ldots, \mathfrak{w}_{J-1}, S_J) = B \prod_{j=1}^{J} \beta_{s_j} \prod_{\ell=1}^{J-1} \left(\prod_{d_i \in \mathsf{U}\{S_1,\ldots,S_\ell\}} \alpha_{d_i}^{\#(d_i,\mathfrak{w}_\ell)} \prod_{s_j \in \{S_{\ell+1},\ldots,S_J\}} \beta_{s_j}^{\#(s_j,\mathfrak{w}_\ell)} \right), \tag{4.18}$$

where $\#(d_i, \mathfrak{w}_\ell)$, $\#(s_j, \mathfrak{w}_\ell)$ count the number of type d_i demand and of type s_j supply in $\mathfrak{w}_\ell$. We will use the notation $B^s = B \prod_{j=1}^{J} \beta_{s_j}$.

We now consider the process $R_N = (S_1, \ldots, S_N)$ which is the permutation of supply types after the Nth match. It is derived by aggregating the states of the detailed augmented Markov chain $\mathbb{Z}$.

Theorem 4.2.19. *The stationary distributions of R is given by:*

$$\pi_R(S_1, \ldots, S_J) = B^s \prod_{\ell=1}^{J-1} \left(\beta_{\{S_1,\ldots,S_\ell\}} - \alpha_{\mathsf{U}\{S_1,\ldots,S_\ell\}} \right)^{-1}. \tag{4.19}$$

We are now ready to obtain the normalizing constant B (see [3]).

Theorem 4.2.20. *The normalizing constant B is given by:*

$$B = \left(\prod_{j=1}^{J} \beta_{s_j} \times \sum_{(S_1,\ldots,S_J)\in\mathcal{P}_J} \prod_{\ell=1}^{J-1} \left(\beta_{\{S_1,\ldots,S_\ell\}} - \alpha_{\mathcal{U}(\{S_1,\ldots,S_\ell\})} \right)^{-1} \right)^{-1} \tag{4.20}$$

$$= \left(\prod_{i=1}^{I} \alpha_{d_i} \times \sum_{(D_1,\ldots,D_I)\in\mathcal{P}_I} \prod_{\ell=1}^{I-1} \left(\beta_{\mathcal{S}(\{D_1,\ldots,D_\ell\})} - \alpha_{\{D_1,\ldots,D_\ell\}} \right)^{-1} \right)^{-1}. \tag{4.21}$$

where the summation is over all permutations of $s_1, \ldots, s_J$ in the first expression, and over all permutations of $d_1, \ldots, d_I$ in the second expression.

By observing when B is finite we obtain:

Corollary 4.2.21. *A necessary and sufficient condition for ergodicity of all the FCFM matching Markov chains is complete resource pooling (4.10).*

Corollary 4.2.22. *The conditional distributions of the numbers of unmatched demand and of matched and exchanged supply, given the permutation is a product of geometric probabilities:*

$$P(X_N = (S_1, n_1, \ldots, n_{J-1}, S_J) \,|\, S_1, \ldots, S_J) = \prod_{\ell=1}^{J-1} \left(\frac{\alpha_{\mathcal{U}\{S_1,\ldots,S_\ell\}}}{\beta_{\{S_1,\ldots,S_\ell\}}} \right)^{n_\ell} \left(1 - \frac{\alpha_{\mathcal{U}\{S_1,\ldots,S_\ell\}}}{\beta_{\{S_1,\ldots,S_\ell\}}} \right),$$

$$P(Y_N = (S_1, m_1, \ldots, m_{J-1}, S_J) \,|\, S_1, \ldots, S_J) = \prod_{\ell=1}^{J-1} \left(\frac{\beta_{\{S_{\ell+1},\ldots,S_J\}}}{\alpha_{\mathcal{D}\{S_{\ell+1},\ldots,S_J\}}} \right)^{m_\ell} \left(1 - \frac{\beta_{\{S_{\ell+1},\ldots,S_J\}}}{\alpha_{\mathcal{D}\{S_{\ell+1},\ldots,S_J\}}} \right).$$

Stationary distribution of the 'natural' Markov chains

We now consider the 'natural' Markov chains O of pair by pair and Q^s, Q^d of supply by supply and demand by demand FCFM matching. The state consists of the ordered list of unmatched demand and/or supply.

Theorem 4.2.23. *The stationary distributions for $Q^s = (Q^s_N)_{N\in\mathbb{Z}}$, for $Q^d = (Q^d_N)_{N\in\mathbb{Z}}$ and for $O = (O_N)_{N\in\mathbb{Z}}$ are given by*

$$\pi_{Q^s}(d^1,\dots,d^L) = B\big(1-\beta_{\mathcal{S}(\{d^1,\dots,d^L\})}\big)\prod_{\ell=1}^{L}\frac{\alpha_{d^\ell}}{\beta_{\mathcal{S}(\{d^1,\dots,d^\ell\})}}, \tag{4.22}$$

$$\pi_{Q^d}(s^1,\dots,s^K) = B\big(1-\beta_{\mathcal{D}(\{s^1,\dots,s^K\})}\big)\prod_{k=1}^{K}\frac{\beta_{s^k}}{\alpha_{\mathcal{D}(\{s^1,\dots,s^k\})}}. \tag{4.23}$$

$$\pi_O(d^1,\dots,d^L,s^1,\dots,s^L) = B\prod_{\ell=1}^{L}\frac{\alpha_{d^\ell}}{\beta_{\mathcal{S}(\{d^1,\dots,d^\ell\})}}\frac{\beta_{s^\ell}}{\alpha_{\mathcal{D}(\{s^1,\dots,s^\ell\})}}. \tag{4.24}$$

An example: the "NN" system

We consider the "NN" system, which is illustrated in Figure 4.10. This system was studied in [31],

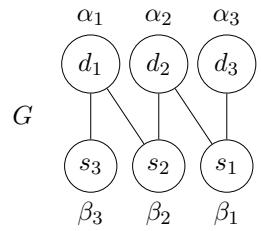


Figure 4.10: Compatibility graph and probabilities of the "NN" system

where ergodicity under complete resource pooling was demonstrated but stationary probabilities could not at that time be obtained.

We can easily calculate stationary probabilities of the pair by pair, supply by supply and demand by demand 'natural' FCFM processes, using formulae (4.22)–(4.24), (4.20). The conditions for stability, i.e. for complete resource pooling, are:

$$\beta_1 > \alpha_3, \qquad \alpha_1 > \beta_3, \qquad \alpha_1+\beta_1 < 1.$$

Some examples are (we write α_i, β_j for $\alpha_{d_i}, \beta_{s_j}$):

$$P(Q^s_N = d_1,d_1,d_1,d_1) = B\beta_1\left(\frac{\alpha_1}{\beta_2+\beta_3}\right)^4,$$

$$P(Q^s_N = d_3,d_3,d_3,d_3,d_3) = B(1-\beta_1)\left(\frac{\alpha_3}{\beta_1}\right)^5,$$

$$P(Q^c_N = s_3,s_3,s_3,s_2,s_3,s_2) = B\alpha_3\left(\frac{\beta_3}{\alpha_1}\right)^3\left(\frac{\beta_2}{\alpha_1+\alpha_2}\right)^2\frac{\beta_3}{\alpha_1+\alpha_2},$$

$$P\big(O_N = (d_3,d_3,d_2,d_3,d_2),(s_3,s_3,s_3,s_3,s_3)\big) = B\left(\frac{\alpha_3}{\beta_1}\right)^2\left(\frac{\alpha_2}{\beta_1+\beta_2}\right)^2\left(\frac{\alpha_3}{\beta_1+\beta_2}\right)\left(\frac{\beta_3}{\alpha_1}\right)^5.$$

The value of the normalizing constant is:

$$B = \frac{(\alpha_1-\beta_3)(\beta_1-\alpha_3)(1-\alpha_1-\beta_1)}{\alpha_1\alpha_2\beta_1\beta_2}.$$

4.2.5 Calculation of performance measures

Matching rates

Assume ergodicity (complete resource pooling) holds. The *matching rate* r_{d_i,s_j} is the a.s. limit of the fraction of matches of demand of type d_i with supply of type s_j, in the complete FCFM matching of $(s^n, d^m)_{0\le m,n\le N}$, as $N\to\infty$. An expression for r_{d_i,s_j} was derived in [3]. We include this expression here for completeness and also because of its close similarity to the derivation of the distribution of *link lengths* L_{s_j}, L_{d_i,s_j}.

Both r_{d_i,s_j} and the distribution of L_{d_i,s_j} are obtained by considering the state of the process ${}^oZ^s$ which is $\mathfrak{s}=(S_1,\mathfrak{w}_1,S_2,\mathfrak{w}_2,\ldots,S_J,\mathfrak{w}_J)$ (or equivalently, of the process $\mathbb{Z}$ with the addition of $\mathfrak{w}_J$, where $\mathfrak{w}_J$ is the sequence of $d^{\overline{M}+1},\ldots,d^{\overline{N}}$). The final expressions include summation over all the permutations $S_1,\ldots,S_J\in\mathcal{P}_J$ of the supply types $s_1,\ldots,s_J$.

For convenience we use the following notations relative to each permutation $S_1,\ldots,S_J$:

$$\alpha_{(k)}=\alpha_{\mathcal{U}\{S_1,\ldots,S_k\}},\qquad \beta_{(k)}=\beta_{\{S_1,\ldots,S_k\}}=\beta_{S_1}+\cdots+\beta_{S_k}.$$

Note that if $\mathcal{U}\{S_1,\ldots,S_k\}=\emptyset$ then $\alpha_{(k)}=0$. Further,

$$\phi_k=\frac{\alpha_{\mathcal{U}\{S_1,\ldots,S_k\}\cap\{d_i\}}}{\alpha_{\mathcal{U}\{S_1,\ldots,S_k\}}},\qquad \psi_k=\frac{\alpha_{\mathcal{U}\{S_1,\ldots,S_k\}\cap(\mathcal{D}(s_j)\setminus\{d_i\})}}{\alpha_{\mathcal{U}\{S_1,\ldots,S_k\}}},\qquad \chi_k=1-\phi_k-\psi_k,$$

where ϕ_k,ψ_k,χ_k express the conditional probability that $s^{N+1}=s_j$ and $d^m\in\mathfrak{w}_k$ form an (s_j,d_i) match, or an (s_j,d_k), $d_k\neq d_i$ match, or no match at all, respectively. By convention $0/0=0$.

The expression for the matching rate is:

$$\begin{aligned} r_{d_i,s_j} &= \beta_{s_j}\sum_{(S_1,\ldots,S_J)\in\mathcal{P}_J}\pi_R(S_1,\ldots,S_J)\\ &\quad\left(\sum_{k=1}^{J-1}\phi_k\frac{\alpha_{(k)}}{\beta_{(k)}-\alpha_{(k)}\chi_k}\prod_{l=1}^{k-1}\frac{\beta_{(l)}-\alpha_{(l)}}{\beta_{(l)}-\alpha_{(l)}\chi_l}+\frac{\phi_J}{\phi_J+\psi_J}\prod_{l=1}^{J-1}\frac{\beta_{(l)}-\alpha_{(l)}}{\beta_{(l)}-\alpha_{(l)}\chi_l}\right). \end{aligned}\tag{4.25}$$

Link lengths

Assume ergodicity (equivalently, complete resource pooling) holds, and consider a stationary FCFM matching over $\mathbb{Z}$. If s^n is matched to d^m we let $L(s^n,d^m)=m-n$ denote the link length. We define the random variable L_{s_j} to have the stationary distribution of link lengths of supply of type s_j. We define the random variable L_{s_j,d_i} to have the stationary distribution of link lengths of matches between supply of type s_j and demand of type d_i. We derive the distributions of L_{s_j} and of L_{s_j,d_i}. They are mixtures of convolutions of some positively signed and some negatively signed geometric random variables. We summarize the results in terms of generating functions.

Theorem 4.2.24. *The generating functions of the distributions of L_{s_j}, L_{d_i,s_j} are:*

$$\begin{aligned} E(\mathsf{Z}^{L_{s_j}}) &= \sum_{(S_1,\ldots,S_J)\in\mathcal{P}_J}\pi_R(S_1,\ldots,S_J)\sum_{\ell=1}^{J}\frac{\alpha_{(\ell)}(\phi_\ell+\psi_\ell)}{\beta_{(\ell)}-\alpha_{(\ell)}\chi_\ell}\prod_{k=1}^{\ell-1}\frac{\beta_{(k)}-\alpha_{(k)}}{\beta_{(k)}-\alpha_{(k)}\chi_k}\\ &\quad\times\prod_{k=1}^{\ell}\Big(\frac{\beta_{(k)}-\alpha_{(k)}\chi_k}{\beta_{(k)}-\alpha_{(k)}\chi_k\mathsf{Z}}\Big)\times\prod_{k=\ell}^{J}\Big(\frac{\beta_{(k)}-\alpha_{(k)}}{1-\alpha_{(k)}-(1-\beta_{(k)})\mathsf{Z}^{-1}}\Big)\times\frac{1}{\mathsf{Z}^{J-\ell}} \end{aligned}$$

$$E(\mathsf{Z}^{L_{s_j,c_i}}) = \sum_{(S_1,\ldots,S_J)\in\mathcal{P}_J} \pi_R(S_1,\ldots,S_J) \sum_{\ell=1}^{J} \frac{\alpha_{(\ell)}\phi_\ell}{\beta_{(\ell)}-\alpha_{(\ell)}(\psi_\ell+\chi_\ell)} \prod_{k=1}^{\ell-1} \frac{\beta_{(k)}-\alpha_{(k)}}{\beta_{(k)}-\alpha_{(k)}(\psi_k+\chi_k)}$$
$$\times \prod_{k=1}^{\ell} \Big(\frac{\beta_{(k)}-\alpha_{(k)}(\psi_k+\chi_k)}{\beta_{(k)}-\alpha_{(k)}\psi_k-\alpha_{(k)}\chi_k\mathsf{Z}}\Big) \times \prod_{k=\ell}^{J} \Big(\frac{\beta_{(k)}-\alpha_{(k)}}{1-\alpha_{(k)}-(1-\beta_{(k)})\mathsf{Z}^{-1}}\Big) \times \frac{1}{\mathsf{Z}^{J-\ell}}$$

4.3 Performance paradox in FCFM systems

This section addresses the performance paradox in the FCFM general matching model. We analyze the impact of adding edges to the matching graph. This can be seen as increasing the matching flexibility, and intuitively, one might expect that this improves the overall system performance.

For example, consider a carpooling system where compatibilities between classes represent the geographic proximity of the users. Consider now the case where one class of users declares to be willing to walk or drive longer to be eligible to be matched with a geographical further location. This situation corresponds to a new matching system with a matching graph with an additional edge. Another example is asking a family registering for social housing to list less requirements in order to be compatible with more housing units. We show that this can lead to overall longer average queue lengths, and therefore also longer average waiting times.

This performance paradox can be seen as a reminiscent of the Braess paradox. This connection will be further discussed in Section 4.4. More closely related to our present work, performance issues due to flexibility were studied in skill based routing models such as queueing systems with redundant requests in [63]. They demonstrate that adding redundancy to a class improves its mean response time but can hurt the mean response time of other classes. Skill based routing models are used in many applications, such as for instance call centers [61], and have several connections to matching models, that will be discussed in discussed in Section 4.4. However, these connections between redundancy service model and stochastic matching models rely on specific assumptions, that we do not suppose in our work, such as bipartite compatibility graph and server by server Markov representation, or ignoring some items if they are not immediately matched upon arrival.

Adding an edge to the compatibility graph leads to a bigger set of possible matchings. Therefore, it is clear that, when we add an edge to the compatibility graph, the performance of the system will always improve under an optimal policy. However, this is less obvious for other matching policies. We study the conditions under which the mean number of unmatched items under FCFM policy decreases when we add an edge to the compatibility graph. The FCFM assumption allows us to use the product-form result from Theorem in Theorem 4.1.1.

We start in 4.3.1 by computing a closed-form expression for the mean total number of items present in the system under the stationary distribution given in Theorem 4.1.1. Specifically, we show that it can be written as a finite sum over all independent sets. This is then used in 4.3.2 to derive sufficient conditions for the existence or the non-existence of a performance paradox in matching models under an asymptotic assumption similar to the heavy-traffic assumptions in queueing systems [86]. We assume that the sum of the arrival rates of an independent set, which we call saturated, grows to its capacity (the sum of the arrival rates of its neighbors), while the other independent sets stay strictly within their capacity. We prove that a performance paradox exists when the saturated independent set has both nodes of the added edge as neighbors. We also prove that a performance paradox does not exist when the

saturated independent set contains at least one of the nodes of the added edge. The intuition behind is that the performance paradox occurs when the added edge in the compatibility graph disrupts the draining of the saturated set.

All the proofs of this section can be found in [C10]. Discussion on the extensions to other matching policies is provided in Section 4.4.

4.3.1 Expected total number of unmatched items

We consider a general matching model (G, μ, FCFM) with matching graph $G = (\mathcal{V}, \mathcal{E})$ that is connected, and $\mu \in \text{NCOND}(G)$ given in (2.1).

We denote by $\mathbf{E}[Q] = \mathbf{E}[|W_\infty|]$ the mean total number of items present in the system under the stationary distribution π given in Theorem 4.1.1. We start by computing a closed-form expression for $\mathbf{E}[Q]$. This result will be used in the proof of the existence of a performance paradox in the next subsection.

For an independent set $\mathcal{I} \in \mathbb{I}$, consider an ordered version of $\mathcal{I}$, noted $\mathcal{I}^o = (i_1, \cdots, i_{|\mathcal{I}|})$. We note σ a permutation of its elements, i.e $\mathcal{I}^{\sigma(o)} = (i_{\sigma(1)}, \cdots, i_{\sigma(|\mathcal{I}|)})$ and $\mathfrak{S}_{|\mathcal{I}|}$ the set of all permutations of $[\![1, |\mathcal{I}|]\!]$.

We define

$$T_{\mathcal{I}^o} = \prod_{k=1}^{|\mathcal{I}|} \frac{\mu_{i_k}}{|\mu_{\mathcal{E}(\{i_1,\cdots,i_k\})}| - |\mu_{\{i_1,\cdots,i_k\}}|}$$

and $T_{\mathcal{I}} = \sum_{\sigma \in \mathfrak{S}_{|\mathcal{I}|}} T_{\mathcal{I}^{\sigma(o)}}$. We also define

$$E_{\mathcal{I}^o} = \sum_{l=1}^{|\mathcal{I}|} \frac{|\mu_{\mathcal{E}(\{i_1,\cdots,i_l\})}|}{|\mu_{\mathcal{E}(\{i_1,\cdots,i_l\})}| - |\mu_{\{i_1,\cdots,i_l\}}|} \times \prod_{k=1}^{|\mathcal{I}|} \frac{\mu_{i_k}}{|\mu_{\mathcal{E}(\{i_1,\cdots,i_k\})}| - |\mu_{\{i_1,\cdots,i_k\}}|}$$

and $E_{\mathcal{I}} = \sum_{\sigma \in \mathfrak{S}_{|\mathcal{I}|}} E_{\mathcal{I}^{\sigma(o)}}$.

The normalizing constant of the stationary distribution in Theorem 4.1.1 can be written as the sum of terms over the independent sets. In the following result, we show that the expected total number of remaining items, defined as an infinite sum over all possible words, can be also written a finite sum over all independent sets.

Proposition 4.3.1. *Let $\mathbf{E}[Q]$ be the expected value of Q under the stationary distribution π. Then*

$$\mathbf{E}[Q] = \left(1 + \sum_{\mathcal{I} \in \mathbb{I}} T_{\mathcal{I}}\right)^{-1} \left(\sum_{\mathcal{I} \in \mathbb{I}} E_{\mathcal{I}}\right).$$

4.3.2 Performance paradox

Consider another compatibility graph $\widetilde{\mathcal{G}} = (V, \widetilde{\mathcal{E}})$, obtained from $G = (\mathcal{V}, \mathcal{E})$ by adding the edge (i^*, j^*), i.e $\widetilde{\mathcal{E}} = \mathcal{E} \cup \{(i^*, j^*)\}$. We denote by $\widetilde{W}$ the Markov chain defined by $\widetilde{\mathcal{G}}$ on $\widetilde{\mathbb{W}}$ and by $\mathbb{E}[\widetilde{Q}]$ the mean total number of items present in the system with the added edge. Since $\mathcal{E}(\mathcal{I}) \subseteq \widetilde{\mathcal{E}}(\mathcal{I})$ for all $\mathcal{I} \in \mathbb{I}$ and we assume $\mu \in \text{NCOND}(G)$, $\widetilde{W}$ is also positive recurrent.

We say that there exists a performance paradox if there exists an edge (i^*, j^*) such that

$$\mathbf{E}\left[\widetilde{Q}\right] > \mathbf{E}[Q],$$

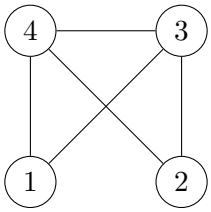


Figure 4.11: Example of a compatibility graph.

i.e. the mean number of items can be increased by adding an edge to the compatibility graph.

For example, we will compare the performance between a matching model with the compatibility graph of Fig. 4.11 and a matching model with a complete compatibility graph of four nodes (i.e we add an edge between node 1 and node 2 in Fig. 4.11).

Let $\widetilde{\mathbb{I}}$ be the set of independent sets of $\widetilde{\mathcal{G}}$. Since $\widetilde{\mathcal{E}}$ only differ from $\mathcal{E}$ in the added edge (i^*, j^*), $\widetilde{\mathbb{I}}$ can be obtained from $\mathbb{I}$ by removing all the independent sets that contain both i^* and j^*.

We prove sufficient conditions for the existence or non-existence of a performance paradox in matching models under a heavy-traffic assumption. This asymptotic assumption is similar to the assumption (A1) used in [C45], and that will be presented in Section 5.3 to prove an approximate optimality result for a carefully designed matching policy.

For any $\mathcal{I} \in \mathbb{I}$, denote by $|W_t|_{\mathcal{I}} = \sum_{i\in\mathcal{I}} |W_t|_i$, $t \geq 0$ and

$$\Delta_{\mathcal{I}} = |\mu_{\mathcal{E}(\mathcal{I})}| - |\mu_{\mathcal{I}}|.$$

Under FCFM policy, for any $t \geq 0$, we have

$$\mathbf{E}\left[|W_{t+1}|_{\mathcal{I}} - |W_t|_{\mathcal{I}}\right] \geq -\Delta_{\mathcal{I}},$$

with the equality that is achieved for example when $|W_t|_i > 0,\ \forall i \in \mathcal{I}$ and $|W_t|_i = 0,\ \forall i \in \mathcal{E}(\mathcal{E}(\mathcal{I}))\backslash\mathcal{I}$. A set $\mathcal{I} \in \arg\min_{\mathcal{I}\in\mathbb{I}} \Delta_{\mathcal{I}}$ will be called a *bottleneck set* in the sense that it has the smallest maximal draining speed. Let

$$\bar{\delta} = \min_{\mathcal{I}\in\mathbb{I}} \Delta_{\mathcal{I}} = \min_{\mathcal{I}\in\mathbb{I}} (|\mu_{\mathcal{E}(\mathcal{I})}| - |\mu_{\mathcal{I}}|).$$

We have $\bar{\delta} > 0$ as we assume that $\mu \in \textsc{Ncond}(G)$. We select a bottleneck set $\hat{\mathcal{I}} \in \arg\min_{\mathcal{I}\in\mathbb{I}} \Delta_{\mathcal{I}}$ with the highest cardinality, i.e.

$$|\hat{\mathcal{I}}| = \max_{\mathcal{I}\in\mathbb{I} \text{ s.t. } \Delta_{\mathcal{I}}=\bar{\delta}} |\mathcal{I}|.$$

We are interested in how the performance of the system will evolve by adding an edge when $\Delta_{\hat{\mathcal{I}}}$ tends towards 0. First, we define a parameterized family of item class distributions:

$$\mu_i^{\delta} = \begin{cases} \mu_i + \frac{\bar{\delta}}{2}\frac{\mu_i}{|\mu_{\hat{\mathcal{I}}}|} - \frac{\delta}{2}\frac{\mu_i}{|\mu_{\hat{\mathcal{I}}}|} & \text{if } i \in \hat{\mathcal{I}} \\ \mu_i - \frac{\bar{\delta}}{2}\frac{\mu_i}{|\mu_{\mathcal{E}(\hat{\mathcal{I}})}|} + \frac{\delta}{2}\frac{\mu_i}{|\mu_{\mathcal{E}(\hat{\mathcal{I}})}|} & \text{if } i \in \mathcal{E}(\hat{\mathcal{I}}) \\ \mu_i & \text{otherwise} \end{cases}$$

for all $0 < \delta \leq \bar{\delta}$. It is clear that $\mu^{\bar{\delta}} = \mu$. By definition of μ^{δ}, when δ tends to 0, then $|\mu^{\delta}_{\mathcal{E}(\hat{\mathcal{I}})}| - |\mu^{\delta}_{\hat{\mathcal{I}}}| = \delta$ tends to 0. Then μ^{δ} is a distribution with full support, and $\mu^{\delta} \in \textsc{Ncond}(G)$.

Lemma 4.3.2. *Let $0 < \delta \leq \bar{\delta}$. We have $\mu_i^\delta > 0$ for all $i \in V$ and $\sum_{i \in V} \mu_i^\delta = 1$. Furthermore, $\mu^\delta \in \text{NCOND}(G)$.*

We have constructed a continuous path of distributions μ^δ, for $0 < \delta \leq \bar{\delta}$, so we can consider $\mathbf{E}[Q]$ as a function of δ and take its limit when δ tends to 0. We can rewrite μ^δ as a linear combination of δ, i.e $\mu_i^\delta = a_i + b_i\delta$ with

$$a_i = \begin{cases} \mu_i + \frac{\bar{\delta}}{2}\frac{\mu_i}{|\mu_{\hat{\mathcal{I}}}|} & \text{if } i \in \hat{\mathcal{I}} \\ \mu_i - \frac{\bar{\delta}}{2}\frac{\mu_i}{|\mu_{\mathcal{E}(\hat{\mathcal{I}})}|} & \text{if } i \in \mathcal{E}(\hat{\mathcal{I}}) \\ \mu_i & \text{otherwise} \end{cases}$$

and

$$b_i = \begin{cases} -\frac{\mu_i}{2|\mu_{\hat{\mathcal{I}}}|} & \text{if } i \in \hat{\mathcal{I}} \\ \frac{\mu_i}{2|\mu_{\mathcal{E}(\hat{\mathcal{I}})}|} & \text{if } i \in \mathcal{E}(\hat{\mathcal{I}}) \\ 0 & \text{otherwise} \end{cases}.$$

Definition 4.3.1. *An independent set $\mathcal{I}$ is called saturated if $\Delta_{\mathcal{I}}^\delta = |\mu_{\mathcal{E}(\mathcal{I})}^\delta| - |\mu_{\mathcal{I}}^\delta|$ tends to 0 when δ tends to 0, i.e. if $|a_{\mathcal{E}(\mathcal{I})}| - |a_{\mathcal{I}}| = 0$.*

Proposition 4.3.3. *The vector $a = (a_1, \cdots, a_n)$ is stochastic, $a \in \text{NCOND}(G)$, for all $\mathcal{I} \in \mathbb{I} \setminus \{\hat{\mathcal{I}}\}$ and $|a_{\hat{\mathcal{I}}}| = |a_{\mathcal{E}(\hat{\mathcal{I}})}|$, i.e. $\hat{\mathcal{I}}$ is the only saturated independent set for our parametrized family μ^δ.*

We now present the main result about the performance paradox.

Theorem 4.3.4. *If $\hat{\mathcal{I}}$ has both i^* and j^* as neighbors, then there exists a performance paradox for δ sufficiently small. If $\hat{\mathcal{I}}$ contains i^* or j^* and $\mathcal{E}(\hat{\mathcal{I}}) \subsetneq \widetilde{\mathcal{E}}(\hat{\mathcal{I}})$, then there does not exist a performance paradox for δ sufficiently small.*

Example 8. *Consider a matching model with the compatibility graph of Fig. 4.11 and let us define μ as $\mu_1 = \mu_2 = 0.22$, $\mu_3 = 0.45$ and $\mu_4 = 0.11$. The independent set is $\hat{\mathcal{I}} = 3$ is the only bottleneck set that achieves the smallest maximal draining speed $\bar{\delta} = \Delta_{\{3\}} = |\mu_{\{1,2,4\}}| - |\mu_3| = 0.1$. Define a new collection of conditional probability distributions μ^δ based on the bottleneck set $\hat{\mathcal{I}} = \{3\}$, for all $0 < \delta \leq 0.1$, i.e $\mu_1^\delta = \mu_2^\delta = 0.2 + \frac{\delta}{5}$, $\mu_3^\delta = 0.5 - \frac{\delta}{2}$ and $\mu_4^\delta = 0.1 + \frac{\delta}{10}$. Thus $\Delta_{\hat{\mathcal{I}}}^\delta = |\mu_{\mathcal{E}(\hat{\mathcal{I}})}^\delta| - |\mu_{\hat{\mathcal{I}}}^\delta| = \delta$ tends to 0 when δ tends to 0 and $\hat{\mathcal{I}}$ is the only saturated independent set. Comparing this matching model with the matching model with the complete compatibility graph leads to a paradox for δ sufficiently small. Indeed, the added edge $(i^*, j^*) = (1, 2)$ has both nodes as neighbors of $\hat{\mathcal{I}} = \{3\}$. In this example, $\mathbf{E}\left[\widetilde{Q}\right] > \mathbf{E}[Q]$ if and only if $0 < \delta < 0.0818369$.*

The intuition behind the results of Theorem 4.3.4 is that the performance paradox occurs when the added edge in the compatibility graph disrupts the draining of a bottleneck. Indeed, as δ decreases, the maximal draining speed of the saturated independent set $\hat{\mathcal{I}}$ becomes very small compared to the other independent sets. Thus, the set of classes $\hat{\mathcal{I}}$ is a critical part of the system and every item of those classes should be matched anytime a compatible item arrives. However, adding an edge in the compatibility graph between two neighbors of $\hat{\mathcal{I}}$ means that sometimes the FCFM policy will match items of those two neighbors together

instead of having them available when items of $\hat{\mathcal{I}}$ arrive. In that case, it is quite intuitive that the load would increase on this already critical part of the system which would lead to degrading performances. Assume now that the added edge in the compatibility graph is between a node within $\hat{\mathcal{I}}$ and a node that was not a neighbor of $\hat{\mathcal{I}}$. Then, items of the new neighbor of $\hat{\mathcal{I}}$ can now be sometimes matched with the latter. Thus, reducing the load on the critical part of the system and improving performance.

4.4 Discussion and related results

Both BM and GM models are somewhat related to queuing systems with known product forms. For BM, this was first observed by Adan and Weiss [3]. They show that the supply by supply Markov chain description is related to a queuing system with multi-type customers and multi-type servers from [137], that also has a product form distribution, under some special conditions. Using this connection, Adan and Weiss in [3] were able to prove that FCFM BM has maximal stability region and they established a product form stationary distribution result for this Markov chain description. However, the product form stationary distribution was derived by partial balance, similar to [137]. Reversibility and Loynes construction for FCFM BM was first established in [J11], together with product form results for other markovian descriptions and the closed form expressions for link lengths in the stationary perfect matching. This stationary distribution was then used to derive expressions for the matching rates.

Connections between FCFS parallel server systems and matching models were further studied in [2]. They established very strong connections between FCFS-ALIS parallel queuing model from [4], redundancy service model from [63] and a parallel FCFS matching queue in which arriving customers join a queue of waiting customers and arriving servers are matched to a first compatible waiting customer, or lost otherwise. They show that the continuous-time Markov chains that describe all three service models share the same stationary distribution, which leads the way to comparing their performance measures. In particular, the redundancy service model and the matching queue are equivalent in the sense that they share the same continuous-time Markov chain. By introducing a new discrete FCFS infinite directed matching model, similar to the one in [3], but in which servers are matched only to customers arrived before them, and embedding all three previous models into this new one, they obtained a version of Burke's Theorem for the redundancy service and for the matching queue systems. An overview of connections between product forms for FCFS queuing models with server-job compatibilities is provided in [62].

In [3], the authors show that the FCFS infinite bipartite matching model has the same state description and stationary distribution as the "First Come First Served-Assign the Longest Idle Server" (FCFS-ALIS) parallel queueing model conditioned on all the servers being busy, which implies heavy-traffic assumptions on the arrival rates.

For GM models, the connection with order independent loss queues was first established by Comte in [48]. Comte shows that a FCFM GM model is a loss variant of order-independent queues introduced in [13]. This allowed to derive closed-form expressions for several performance metrics, like the waiting probability or the mean matching time, that can be computed recursively over the family of independent sets of items. Some of these results have been then translated to the FCFM BM model in [49].

So far, FCFM BM and FCFM GM are the only instances of FCFM EBM that we know

to have a product form. The reversibility proof for FCFM BM heavily uses the independence assumption between the supply and demand classes, so investigating other instances of EBM will require a new approach.

An extension to a variant of FCFM GM in which the items of a given class can be matched together, i.e. to the case of a compatibility graph with self-loops, have been simultaneously proposed in [11] and [C6], using different proof techniques.

Loynes type constructions have been investigated further in [W4] for GM and in [W7] for EBM models, under general matching policies. In [W7], we proposed an explicit construction of the stationary state of EBM model. We use a Loynes-type backwards scheme, allowing to show the existence and uniqueness of a bi-infinite perfect matching under various conditions, for a large class of matching policies and of bipartite matching structures. The key algebraic element of our construction is the sub-additivity of a suitable stochastic recursive representation of the model, satisfied under most usual matching policies. By doing so, we also derive stability conditions for the system under general stationary ergodic assumptions, subsuming the classical markovian settings. The extension to GM is studied in [W4]. We prove that most common matching policies (including FCFM, priorities and random) satisfy a particular sub-additive property, which we exploit to show in many cases, the coupling-from-the-past to the steady state, using a backwards scheme *à la* Loynes. We then use these results to explicitly construct perfect bi-infinite matchings, and to build a perfect simulation algorithm in the case where the buffer of the system is finite.

Extensions of performance paradox to greedy matching policies are studied further in [W2]. What is intriguing is that the same paradox occurs even if we consider the whole family of greedy policies, i.e. the policies that always match a new arrival to a waiting item if there are any compatible items waiting in the system. An example is given in [W2]. In stochastic matching model, greedy matching policies can be interpreted as selfish behavior of new arrivals, so this performance paradox is to some extent similar to a Braess paradox observed in transportation networks [20]. Braess paradox states that, when the agents can take self-interested decisions, the travelling times of the agents can increase if we add a new road. The idea behind this phenomenon is that the extension of the network might cause a redistribution of the traffic that increases the congestion and, as a result, the delay of agents. More precisely, the Braess paradox shows that the travel time in the Nash equilibrium (the set of strategies such as no agent has incentive to deviate unilaterally) can increase if we add a shortcut in the network. This result reflects that the selfish behavior of agents in a network might lead to a situation whose performance is sub-optimal. The existence of a Braess paradox has been explored in several contexts related to queueing networks [10, 33, 46, 47, 79].

In the next chapter, we will consider optimal control problems in stochastic matching systems. The existence of performance paradox for any greedy policies shows that it is necessary to extend the class of admissible policies to include non-greedy behavior.

Chapter 5

Optimal control in bipartite stochastic matching

The choice of matching decisions can be cast as an optimal control problem for a dynamic matching model. The goal is to obtain a better understanding of the structure of optimal policies. The focus in this chapter is on the infinite-horizon average-cost optimal control problem for a linear cost function c on buffer levels.

We start by giving a complete characterization of an optimal policy in the `N` case in Section 5.2. We show that there exists an optimal policy that gives priority to the pendant edges of the matching graph, and that is of threshold type for the diagonal edge. We also fully characterize the optimal threshold. Under some assumptions in the costs of the nodes, this threshold-based structure of the optimal matching policy extends to quasi-complete graphs (i.e. complete graphs minus one edge). In an arbitrary acyclic graph, we show that, when the cost of the pendant nodes is larger or equal than the cost of its neighbors, the optimal policy always gives priority to the pendant edges.

In Section 5.3, a new class of policies is introduced: the h-MaxWeight with threshold. The policy is the solution to a linear program that minimizes the drift of the function h, subject to non-idling constraints. Performance analysis is based on a one-dimensional relaxation of the stochastic control problem, which is found to be a special case of an inventory model, as treated in the classical theory of Clark and Scarf [45]. Consequently, the optimal policy for the relaxation admits a closed-form expression based on a threshold rule. These observations inform the choice of function h, and the choice of threshold. For a parameterized family of models in which the network load approaches capacity, this policy is shown to be approximately optimal, *with bounded regret*, even though the average cost grows without bound.

This chapter is based on on the following publications that contain more details and the proofs of the results presented in this chapter: [J5, C45].

5.1 MDP model

The bipartite matching model considered in this chapter is a variant of bipartite stochastic matching model. It consists of a queueing network model with two sets of buffers, distinguished by their role as providing supply or demand of resources. Let $\ell_{\mathcal{S}}$ denote the number of supply classes, $\ell_{\mathcal{D}}$ the number of demand classes, and define the following index sets:

$\mathcal{D}$: Indices of demand classes. $\mathcal{S}$: Indices of supply classes.

$\mathcal{E}$: Matching pairs, $\mathcal{E} \subset \mathcal{D} \times \mathcal{S}$, $\mathcal{A}$: Arrival pairs, $\mathcal{A} \subset \mathcal{D} \times \mathcal{S}$.

The bipartite graph $(\mathcal{D} \cup \mathcal{S}, \mathcal{E})$ is called the *matching graph*. The graph is assumed connected.

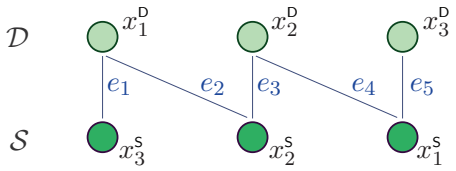


Figure 5.1: The NN network.

The NN-network shown in Fig. 5.1 is an example in which $\ell_{\mathcal{D}} = \ell_{\mathcal{S}} = 3$, and the set $\mathcal{E}$ denotes the edges (e_i) shown. Each of the three integers $\{x_i^{\mathcal{D}} : i = 1, 2, 3\}$ corresponds to units of demand of a particular class, and $\{x_i^{\mathcal{S}} : i = 1, 2, 3\}$ correspond to units of the three different types of supply.

A discrete-time Markov Decision Process (MDP) model is introduced next to capture temporal dynamics. The main departure from traditional queueing networks is that there are no servers. Instead of "service", activities in this model correspond to instantaneous matching a particular unit of supply with a unit of demand.

The vector of buffer levels for the dynamic matching model is denoted $X(t)$. It takes values in $\mathbb{Z}_+^\ell$, where $\ell = \ell_{\mathcal{D}} + \ell_{\mathcal{S}}$. The following notation is used to emphasize the different roles for supply or demand buffers:

$$X(t) = (X_1^{\mathcal{D}}(t), \dots, X_{\ell_{\mathcal{D}}}^{\mathcal{D}}(t), X_1^{\mathcal{S}}(t), \dots, X_{\ell_{\mathcal{S}}}^{\mathcal{S}}(t))^T \tag{5.1}$$

It is often convenient to drop the super-scripts. In this case, for $i \in \mathcal{D} := \{1, \dots, \ell_{\mathcal{D}}\}$, the integer $X_i(t)$ denotes the number of units of demand of class i, and for $j \in \mathcal{S} := \{\ell_{\mathcal{D}} + 1, \dots, \ell_{\mathcal{D}} + \ell_{\mathcal{S}}\}$, the integer $X_j(t)$ denotes the units of supply of class j.

An i.i.d. arrival process is denoted $\boldsymbol{A}$. As in Chapter 3, we assume that a single pair arrive at each time slot – one of demand and one of supply. That is, for each t, $A(t)$ takes values in the set

$$\mathsf{A}_\diamond = \{\mathsf{1}^i + \mathsf{1}^j : (i, j) \in \mathcal{A}\}, \tag{5.2}$$

where $\mathsf{1}^i$ denotes a column vector with ith component equal to 1 and zero elsewhere.

Let $\xi^0 = (1, \dots, 1, -1, \dots, -1)^T$, the vector with $\ell_{\mathcal{D}}$ entries of $+1$, followed by $\ell_{\mathcal{S}}$ entries of -1. The queue length vector is subject to the following balance constraint:

$$\xi^0 \cdot X(t) = 0 \tag{5.3}$$

An input process $\boldsymbol{U}$ represents the sequence of matching activities, and the queue dynamics are

$$X(t+1) = X(t) - U(t) + A(t+1), \quad t \ge 0\,. \tag{5.4}$$

Input constraints are captured by the input space:

$$\mathsf{U}_\diamond = \Big\{ u = \sum_{e \in \mathcal{E}} n_e u^e : \; n_e \in \mathbb{Z}_+ \Big\} \tag{5.5}$$

The vectors $\{u^e\}$ are an enumeration of all single matches across edges of the matching graph: that is, $u^e = \mathsf{1}^i + \mathsf{1}^j$ for $e = (i, j) \in \mathcal{E}$. There are also implicit constraints on $U(t)$, since the components of $X(t)$ are constrained to non-negative integer values.

The input is further constrained by $U(t) \in \mathsf{U}_\diamond(x)$ when $X(t) = x$, where

$$\mathsf{U}_\diamond(x) = \big\{u \in \mathsf{U}_\diamond : x - u \geq 0\big\}\,, \qquad x \in \mathsf{X}_\diamond \tag{5.6}$$

Based on (5.2) and (5.15) we have

$$\xi^0 \cdot U(t) = 0 \quad \textit{and} \quad \xi^0 \cdot A(t) = 0\,, \qquad \textit{a.s.} \tag{5.7}$$

Consequently, the constraint (5.3) holds automatically under (5.15) and (5.2), provided it holds at time $t = 0$. That is, for each t, $X(t)$ takes values in the set

$$\mathsf{X}_\diamond = \{x \in \mathbb{Z}_+^\ell : \xi^0 \cdot x = 0\}. \tag{5.8}$$

The existence of an optimal policy for this MDP model requires stabilizability of the network.

We assume a linear cost function on buffer levels and aim to minimize the average cost given by

$$\eta = \limsup_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} \mathsf{E}\big[c(X(t))\big]. \tag{5.9}$$

Stabilizability. Let $\mathcal{S}(i)$ denote the set of supply classes that can be matched with a class i demand, and let $\mathcal{D}(j)$ denote the set of demand classes that can be matched with a class j supply. This definition and the extension to subsets $D \subset \mathcal{D}$ and $S \subset \mathcal{S}$ are formalized as follows:

$$\mathcal{S}(i) = \{j \in \mathcal{S} : (i, j) \in \mathcal{E}\}\,, \quad \mathcal{D}(j) = \{i \in \mathcal{D} : (i, j) \in \mathcal{E}\}$$

$$\mathcal{S}(D) = \bigcup_{i\in D} \mathcal{S}(i)\,, \qquad \mathcal{D}(S) = \bigcup_{j\in S} \mathcal{D}(j)$$

For any vector $x \in \mathbb{R}_+^\ell$ denote, $|x_D| = \sum_{i\in D} x_i$, $|x_S| = \sum_{j\in S} x_j$.

The MDP model is said stabilizable if there exists a policy for which the controlled MDP model is positive Harris recurrent. The necessary and sufficient condition for stabilizability of the MDP model is given as follows, based on the mean arrival rate vector $\alpha = \mathsf{E}[A(t)]$:
`NCond`: For all non-empty subsets $D \subsetneq \mathcal{D}$ and $S \subsetneq \mathcal{S}$,

$$|\alpha_D| < |\alpha_{\mathcal{S}(D)}| \quad \textit{and} \quad |\alpha_S| < |\alpha_{\mathcal{D}(S)}| \tag{5.10}$$

The proof can be found in [J22] and it was presented in Chapter 3. The sufficiency part is constructive: it is shown that the following *Match the Longest* (ML) policy is stabilizing under `NCond`:

$$\phi(x) = \arg\max\{u \cdot \nabla h\,(x) : u \in \mathsf{U}_\diamond(x)\}, \qquad x \in \mathsf{X}_\diamond, \tag{5.11}$$

with $h(x) = \|x\|^2$, the usual ℓ_2-norm.

Under `NCond` we are also assured of the existence of an optimal policy for the MDP model. The proof of Proposition 5.1.1 follows from Theorem 9.0.2 of [109].

Proposition 5.1.1. *If* `NCond` *holds, then the optimal average cost* η^* *exists as a deterministic constant, independent of initial conditions.*

In stability and performance analysis it is sometimes convenient to consider the process $Q(t) = X(t) - A(t)$, $t \geq 1$, which evolves according to a recursion similar to (5.4):

$$Q(t+1) = Q(t) - U(t) + A(t)\,, \qquad t \geq 0 \tag{5.12}$$

If $\boldsymbol{U}$ is defined using a stationary policy $U(t) = \phi(X(t))$, it then follows that $\boldsymbol{Q}$ is a Markov chain evolving on $\mathsf{X}_\diamond$.

Furthermore, in a part of the analysis it is useful to allow $U(t)$ to depend on both $Q(t)$ and $A(t)$ rather than a function of the sum $X(t)$. To apply MDP theory it is therefore necessary to defined a second MDP model in which the state is the pair process $Y(t) = (Q(t), A(t))$. It is assumed that the input process $\boldsymbol{U}$ is non-anticipative (a function of present and past values of $\boldsymbol{Y}$). A stationary (state feedback) policy is of the form $U(t) = \psi(Y(t))$, for some function $\psi\colon \mathsf{X}_\diamond \times \mathsf{A}_\diamond \to \mathsf{U}_\diamond$. We allow for randomized stationary policies in our analysis.

Let η_Y^* denote the optimal average cost for the MDP with the larger state process $\boldsymbol{Y}$. Given the greater information, it is immediate that $\eta_Y^* \leq \eta^*$, with η^* defined in Prop. 5.1.1. In fact, the two are identical.

Proposition 5.1.2. *If* `NCond` *holds, then the optimal average cost exists as a deterministic constant, independent of initial conditions, for either MDP* $\boldsymbol{X}$ *or* $\boldsymbol{Y}$*. Moreover, the average costs are equal:* $\eta_Y^* = \eta^*$. □

5.2 Optimality results for specific graphs

For simplicity, in this section we assume independence between demand and supply arrivals. The demand item arrives to the queue d_i with probability α_i and the supply item arrives to the queue s_j with probability β_j, i.e:

$$\forall (i,j) \in \mathcal{A} \quad \mathbb{P}(A(n) = e_{(i,j)}) = \alpha_i \beta_j > 0$$

with $\sum_{i=1}^{n_\mathcal{D}} \alpha_i = 1$, $\sum_{j=1}^{n_\mathcal{S}} \beta_j = 1$ and where $\mathcal{A} = \mathcal{D} \times \mathcal{S}$, $e_{(i,j)} = e_{d_i} + e_{s_j}$ and $e_k \in \mathbb{N}^{n_\mathcal{D}+n_\mathcal{S}}$ is the vector of all zeros except in the k-th coordinate where it is equal to one, $k \in \mathcal{D} \cup \mathcal{S}$. We assume that the α_i and β_j are chosen such that the arrival distribution satisfies the necessary and sufficient conditions for stabilizability of the MDP model: `Ncond` given by (5.10), i.e $\forall D \subsetneq \mathcal{D}$, $\forall S \subsetneq \mathcal{S}$:

$$\sum_{d_i \in D} \alpha_i < \sum_{s_j \in \mathcal{S}(D)} \beta_j \text{ and } \sum_{s_j \in S} \beta_j < \sum_{d_i \in \mathcal{D}(S)} \alpha_i. \tag{5.13}$$

As $A(t)$ are i.i.d., to ease the notation from now on, we denote by A a random variable with the same distribution as $A(1)$. For a given function v, $Y(t) = (q, a)$, $x = q + a$, $u \in \mathsf{U}_\diamond(x)$, we define:

$$\begin{aligned} T_u v(q,a) &= c(q,a) + \mathbb{E}[v(q+a-u, A)] = c(x) + \mathbb{E}[v(x-u, A)] \\ Tv(q,a) &= c(q,a) + \min_{u \in U_x} \mathbb{E}[v(q+a-u, A)] = c(x) + \min_{u \in U_x} \mathbb{E}[v(x-u, A)] \end{aligned}$$

A solution of the average cost problem can be obtained as a solution of the Bellman fixed point equation $\eta_Y^* + v = Tv$.

We say that a value function v or a decision rule u is structured if it satisfies a special property, such as being increasing, decreasing or convex. Throughout this section, by increasing we mean nondecreasing and we will use strictly increasing for increasing. A policy is called structured when it only uses structured decision rules.

We use property preservation framework for the dynamic programming operator T. First, we identify a set of structured value functions V^σ and a set of structured deterministic Markovian decision rules D^σ such that if the value function belongs to V^σ an optimal decision rule belongs to D^σ. Then, we show that the properties of V^σ are preserved by the Dynamic Programming operator and that they hold in the limit. This framework is described in [118]. In particular, we use in our analysis [118, Theorem 6.11.3], and its variant [76, Theorem 1], for the discounted cost case and [118, Theorem 8.11.1] for the average cost case.

Assumption 1. The cost function c is a nonnegative function with the same structured properties as the value function, i.e $c \in V^\sigma$.

All bipartite and connected graphs of at most 4 nodes, except the N-shaped graph that will be studied in Section 5.2.1, are complete, i.e $\mathcal{E} = \mathcal{D} \times \mathcal{S}$. Without loss of generality, we can focus on the matching graphs such that there are more demand nodes than supply nodes, i.e $n_\mathcal{D} \geq n_\mathcal{S}$ (see Figure 5.2).

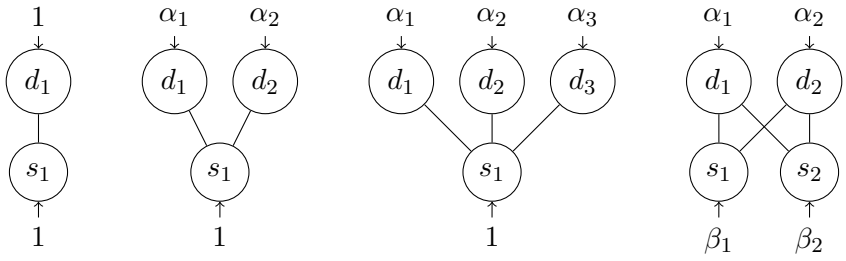


Figure 5.2: Bipartite and connected matching graphs of at most 4 nodes (except the N-shaped matching graph).

For these matching graphs, the stabilizability conditions are $0 < \alpha_i < 1$ for $i \in \{1, 2, 3\}$ and $0 < \beta_j < 1$ for $j \in \{1, 2\}$. The optimal policy for complete graphs is very intuitive. We can empty the system from any state and because we only consider costs on the number of items at each node and not rewards on which edge we match, it does not matter how do we empty the system. However, leaving some items in the system will only increase the total cost in the long run. Thus, in complete graphs it is optimal to match all compatible items (i.e. any greedy policy is optimal).

5.2.1 N-model

The first model where difficulties arise is the N-shaped graph in Figure 5.3). For this graph, we have $\mathcal{D} = \{d_1, d_2\}$, $\mathcal{S} = \{s_1, s_2\}$ and $\mathcal{E} = \{(1, 1), (1, 2), (2, 2)\}$. Let us also define $(2, 1)$ as the imaginary edge between d_2 and s_1 (imaginary because $(2, 1) \notin \mathcal{E}$) that we introduce to ease the notations. To ensure stability, we assume that $\alpha > \beta$.

We will show that the optimal policy for this case has a specific structure. For this purpose, we first present the properties of the value function. Then, we show how these properties characterize the optimal decision rule and how they are preserved by the dynamic programming operator. Finally, we state the main result in Theorem 5.2.3.

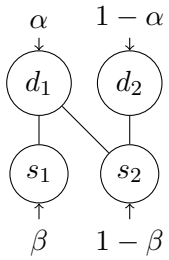


Figure 5.3: The N-shaped matching graph.

Value Function Properties

Definition 5.2.1 (Increasing property)**.** *Let $(i,j) \in \mathcal{E}$. We say that a function v is increasing in (i,j) or $v \in \mathcal{I}_{(i,j)}$ if*

$$\forall a \in \mathcal{A}, \forall q \in \mathcal{Q}, \quad v(q + e_{(i,j)}, a) \geq v(q,a).$$

We also note $\mathcal{I} = \cup_{(i,j)\in\mathcal{E}} \mathcal{I}_{(i,j)}$.

We will use the increasing property in $(1,1)$ and $(2,2)$. We also define an increasing property in the imaginary edge $(2,1)$ using the same definition as in Definition 5.2.1 (even if $(2,1) \notin \mathcal{E}$) that we will note $\mathcal{I}_{(2,1)}$.

Remark 2. *The increasing property in $(2,1)$ can be interpreted as the fact that we prefer to match $(1,1)$ and $(2,2)$ rather than to match $(1,2)$. Indeed, $v(q + e_{(1,1)} + e_{(2,2)} - e_{(1,2)}, a) = v(q + e_{(2,1)}, a) \geq v(q,a)$.*

We also define the convexity in $(1,2)$ and $(2,1)$ as follows:

Definition 5.2.2 (Convexity property)**.** *A function v is convex in $(1,2)$ or $v \in \mathcal{C}_{(1,2)}$ if $v(q + e_{(1,2)}, a) - v(q,a)$ is increasing in $(1,2)$, i.e., $\forall a \in \mathcal{A}$, $\forall q \in \mathcal{Q}$ such that $q_{d_1} \geq q_{s_1}$, we have*

$$v(q + 2e_{(1,2)}, a) - v(q + e_{(1,2)}, a) \geq v(q + e_{(1,2)}, a) - v(q,a).$$

Likewise, v is convex in $(2,1)$ or $v \in \mathcal{C}_{(2,1)}$ if $v(q + e_{(2,1)}, a) - v(q,a)$ is increasing in $(2,1)$, i.e., $\forall a \in \mathcal{A}$, $\forall q \in \mathcal{Q}$ such that $q_{s_1} \geq q_{d_1}$, we have

$$v(q + 2e_{(2,1)}, a) - v(q + e_{(2,1)}, a) \geq v(q + e_{(2,1)}, a) - v(q,a).$$

Definition 5.2.3 (Boundary property)**.** *A function $v \in \mathcal{B}$ if*

$$\forall a \in \mathcal{A}, \quad v(0,a) - v(e_{(2,1)}, a) \leq v(e_{(1,2)}, a) - v(0,a).$$

As we will show in Proposition 5.2.2, the properties $\mathcal{I}_{(1,1)}$, $\mathcal{I}_{(2,2)}$, $\mathcal{I}_{(2,1)}$ and $\mathcal{C}_{(1,2)}$ characterize the optimal decision rule. On the other hand, $\mathcal{C}_{(2,1)}$ and $\mathcal{B}$ are required to show that $\mathcal{C}_{(1,2)}$ is preserved by the operator T.

We will consider the following set of structured value functions

$$V^\sigma = \mathcal{I}_{(1,1)} \cap \mathcal{I}_{(2,2)} \cap \mathcal{I}_{(2,1)} \cap \mathcal{C}_{(1,2)} \cap \mathcal{C}_{(2,1)} \cap \mathcal{B}. \tag{5.14}$$

We show that the properties of the value function are preserved by the dynamic programming operator. In other words, we show that if $v \in V^\sigma$, then $Tv \in V^\sigma$.

Optimal decision rule

In this section, we show that, for any $v \in V^\sigma$, there is a control of threshold-type in $(1,2)$ with priority to $(1,1)$ and $(2,2)$ that minimizes $T_u v$.

Definition 5.2.4 (Threshold-type decision rule)**.** *A decision rule u_x is said to be of threshold type in $(1,2)$ with priority to $(1,1)$ and $(2,2)$ if:*

1. *it matches all of $(1,1)$ and $(2,2)$.*
2. *it matches $(1,2)$ only if the remaining items (in d_1 and s_2) are above a specific threshold, denoted by t (with $t \in \mathbb{N} \cup \infty$).*

i.e, $u_x = \min(x_{d_1}, x_{s_1}) e_{(1,1)} + \min(x_{d_2}, x_{s_2}) e_{(2,2)} + k_t(x) e_{(1,2)}$ *where*

$$k_t(x) = \begin{cases} 0 & if\ x_{d_1} - x_{s_1} \leq t \\ x_{d_1} - x_{s_1} - t & otherwise \end{cases}.$$

We define D^σ as the set of decision rules that are of threshold type in $(1,2)$ with priority to $(1,1)$ and $(2,2)$ for any $t \in \mathbb{N} \cup \infty$.

If $t = \infty$, the decision rule will never match $(1,2)$. Otherwise, the decision rule will match $(1,2)$ until the remaining items in d_1 and s_2 are below or equal to the threshold t.

In the next proposition, we establish that there exists an optimal decision rule with priority to $(1,1)$ and $(2,2)$.

Proposition 5.2.1. *Let $v \in \mathcal{I}_{(1,1)} \cap \mathcal{I}_{(2,2)} \cap \mathcal{I}_{(2,1)}$, let $0 \leq \theta \leq 1$. For any $q \in \mathcal{Q}$ and $a \in \mathcal{A}$, $x = q + a$, there exists $u^* \in U_x$ such that $u^* \in \arg\min_{u \in U_x} T_u v(q,a)$, $u^*_{(1,1)} = \min(x_{d_1}, x_{s_1})$ and $u^*_{(2,2)} = \min(x_{d_2}, x_{s_2})$.*

From this result, it follows that there exists an optimal decision rule that matches all possible $(1,1)$ and $(2,2)$. Our goal now is to find the optimal number of matchings in $(1,2)$. We introduce first some notation:

Definition 5.2.5. *Let $x \in \mathcal{Q}$. We define:*

$$K_x = \begin{cases} \{0\} & if\ x_{d_1} \leq x_{s_1} \\ \{0, \cdots, \min(x_{d_1} - x_{s_1}, x_{s_2} - x_{d_2})\} & otherwise \end{cases}$$

the set of possible matching in $(1,2)$ after having matched all possible $(1,1)$ and $(2,2)$.

Remark 3. *The state of the system after having matched all possible $(1,1)$ and $(2,2)$ is of the form $(0,l,l,0)$ if $x_{d_1} \leq x_{s_1}$ and of the form $(l,0,0,l)$ otherwise (because of the definition of $\mathcal{Q}$ and U_x).*

Finally, we prove that a decision rule of threshold type in $(1,2)$ with priority to $(1,1)$ and $(2,2)$ is optimal. This is done by choosing the right t for different cases such that $k_t(x)$ is the optimal number of matchings in $(1,2)$ for a given x.

Proposition 5.2.2. *Let $v \in \mathcal{I}_{(1,1)} \cap \mathcal{I}_{(2,2)} \cap \mathcal{I}_{(2,1)} \cap \mathcal{C}_{(1,2)}$. For any $q \in \mathcal{Q}$ and for any $a \in \mathcal{A}$, $x = q + a$, there exists $u^* \in D^\sigma$ (see Definition 5.2.4) such that $u^* \in \arg\min_{u \in U_x} T_u v(q,a)$.*

Structure of the optimal policy

The following theorem shows that there exists an optimal stationary Markovian matching policy which is formed of a sequence of decision rules that belong to D^σ (with a fixed threshold).

Theorem 5.2.3. *An optimal control for the average cost problem is of threshold type in* $(1,2)$ *with priority to* $(1,1)$ *and* $(2,2)$.

Computing the optimal threshold

We consider the matching policy of threshold type in $(1,2)$ with priority to $(1,1)$ and $(2,2)$ in the average cost case.

Proposition 5.2.4. *Let* $\rho = \frac{\beta(1-\alpha)}{\alpha(1-\beta)} \in (0,1)$, $R = \frac{c_{s_1}+c_{d_2}}{c_{d_1}+c_{s_2}}$ *and* $\Pi^{T_{(1,2)}}$ *be the set of matching policy of threshold type in* $(1,2)$ *with priority to* $(1,1)$ *and* $(2,2)$. *Assume that the cost function is a linear function. The optimal threshold* t^*, *which minimizes the average cost on* $\Pi^{T_{(1,2)}}$, *is*

$$t^* = \begin{cases} \lceil k \rceil & if\ f(\lceil k \rceil) \leq f(\lfloor k \rfloor) \\ \lfloor k \rfloor & otherwise \end{cases}$$

where $k = \frac{\log \frac{\rho-1}{(R+1)\log \rho}}{\log \rho} - 1$ *and* $f(x) = (c_{d_1}+c_{s_2})x + (c_{d_1}+c_{d_2}+c_{s_1}+c_{s_2})\frac{\rho^{x+1}}{1-\rho} - (c_{d_1}+c_{s_2})\frac{\rho}{1-\rho} + ((c_{d_1}+c_{s_1})\alpha\beta + (c_{d_2}+c_{s_2})(1-\alpha)(1-\beta) + (c_{d_2}+c_{s_1})(1-\alpha)\beta + (c_{d_1}+c_{s_2})\alpha(1-\beta))$.

Quasi-complete graphs

Further results for specific graphs are obtained in [J5]. We show that, under some assumptions on the costs of the nodes, this threshold-based structure of the optimal matching policy extends to quasi-complete graphs (i.e. complete graphs minus one edge).

5.2.2 Acyclic graphs

In Section 5.2.1, we fully characterized the optimal matching control of the N-shaped matching graph, which is an acyclic graph. In this section, we study the optimal matching control of an arbitrary acyclic matching graph. We show that, under certain assumptions on the costs of the nodes, the optimal matching policy consists of prioritizing the matching of the pendant edges.

For an acyclic matching graph, we say that (i,j) is an pendant edge if the unique adjacent node of d_i is s_j or if the unique adjacent node of s_j is d_i. We denote by $\mathcal{E}^*$ the set of pendant edges. We say that an edge (i_1,j_1) belongs to the neighborhood of an edge (i_2,j_2) if $i_1 = i_2$ or $j_1 = j_2$. We denote by $N((i,j))$ the neighborhood of an edge (i,j).

We assume that the neighborhood of all the edges is not empty, i.e., the matching graph is connected, and that in the neighborhood of a pendant edge there are no pendant edges. An example is provided in Figure 5.4. The set of pendant edges is $\mathcal{E}^* = \{(1,1),(3,3),(6,5)\}$ and the neighborhood of $(1,1)$ is $N((1,1)) = \{(1,2)\}$, whereas that of $(3,3)$ is $N((3,3)) = \{(2,3),(4,3)\}$.

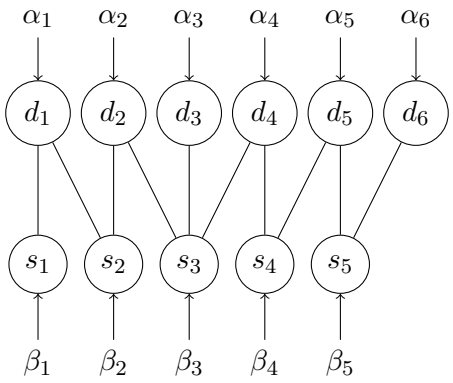


Figure 5.4: An example of an acyclic matching graph.

Optimal decision rule

Definition 5.2.6 (Prioritizing the pendant edges)**.** *We say that a matching policy prioritizes the pendant edges if it matches all the items in the pendant edges. This means that, for all* $(i,j) \in \mathcal{E}^*$, $u_{(i,j)} = \min(x_{d_i}, x_{s_j})$.

We consider that D^σ is the decision rule that prioritizes the pendant edges.

Proposition 5.2.5. *Let* $a \in \mathcal{A}$, $q \in \mathcal{Q}$, $x = q + a$, $v \in V^\sigma$. *There exists* $u^* \in U_x$ *such that* $u^* \in \arg\min_{u \in U_x} T_u v(q,a)$ *and* $u^*_{(i,j)} = \min(x_{d_i}, x_{s_j})$ *for all* $(i,j) \in \mathcal{E}^*$.

Structure of the optimal policy

Theorem 5.2.6. *The optimal control for the average cost problem prioritizes the pendant edges.*

For the above results, we have assumed that the pendant edges do not have other pendant edges in their neighborhoods. We now explain how the results of this section also hold when in the neighborhood of an pendant edge there are other pendant edges. An example the matching models we now study consists of the matching graph of Figure 5.4 with an additional demand node d_7 and an edge $(7,5)$.

If the cost of the pendant edges that are neighbors is the same, these edges can be merged and seen as a single edge whose arrival probability is equal to the sum of their arrival probabilities. Otherwise, using similar arguments it can be shown that the optimal policy prioritizes the most expensive pendant edges.

5.3 Approximate optimality with bounded regret

We will restrict here additionally the input space to be finite, and use:

$$\mathsf{U}_\diamond = \Big\{ u = \sum_{e \in \mathcal{E}} n_e u^e : \ n_e \in \mathbb{Z}_+, \ \ |n| \le \overline{n}_u \Big\} \tag{5.15}$$

where $|n| = \sum n_e$, and $\overline{n}_u \ge 1$ is a fixed integer. The integer $\overline{n}_u$ must be chosen sufficiently large to ensure stabilizability of the network. This constraint on the input is imposed only

to simplify Taylor series approximations used to obtain performance bounds. It is assumed henceforth that $\overline{n}_u \geq 4$, to ensure the feasibility of the randomized policies used in the proof of Theorem 5.3.6.

5.3.1 h-MaxWeight policies

For a differentiable function $h\colon \mathbb{R}_+^\ell \to \mathbb{R}_+$, the is exactly of the form (5.48):

$$\phi(x) = \arg\max\{u \cdot \nabla h\,(x) : u \in \mathsf{U}_\diamond(x)\}, \qquad x \in \mathsf{X}_\diamond \tag{5.16}$$

Here we present initial motivation, and conditions for stability.

Recall that the main goal of this paper is to approximation the solution to the average cost optimal control problem, in which $c\colon \mathbb{R}_+^\ell \to \mathbb{R}_+$ is linear: $c(x) = c \cdot x$ for some vector c with strictly positive entries. That is, we search for a stationary policy ϕ for which the average cost is approximately equal to the optimal value η^* defined in Prop. 5.1.1.

The approximately optimal policy introduced in Section 5.3.4 is a variation of h-MW, with a particular choice of the function h; it is denoted $\tilde{c}$, and defined as follows. For a fixed constant $\beta > 0$, let $\tilde{x}$ denote the function of x with entries,

$$\tilde{x}_k = x_k + \beta(e^{-x_k/\beta} - 1)\,, \quad x_k \in \mathbb{R}_+ \tag{5.17}$$

The right hand side vanishes at $x_k = 0$, as does its first derivative. The constant β is chosen so that its derivative with respect to x_k is small for x_k within some neighborhood of the origin. We then define

$$\tilde{c}(x) := c(\tilde{x})\,, \quad x \in \mathsf{X}_\diamond \tag{5.18}$$

The motivation for the perturbation is that the resulting function satisfies the following zero-marginal cost property: with $h = \tilde{c}$,

$$\frac{\partial}{\partial x_k} h(x) = 0 \qquad \text{if } x_k = 0 \tag{5.19}$$

This eliminates boundary effects in Lyapunov function stability analysis. It also indirectly provides safety-stocks when using the $\tilde{c}$-MW policy (an example is provided for a scheduling model in Fig. 3 of [105]).

The $\tilde{c}$-MW policy is similar to the original MW policy if β is large, since the function $\tilde{c}$ is approximately quadratic. This is seen through the second order Taylor series expansion

$$e^{-x_k/\beta} = 1 - x_k/\beta + \tfrac{1}{2}(x_k/\beta)^2 + O\Big((x_k/\beta)^3 e^{-x_k/\beta}\Big)\,, \quad \beta \sim \infty \tag{5.20}$$

so that for large β,

$$\beta\,\tilde{c}(x) \approx h_c(x) := \frac{1}{2}\sum_{k=1}^{\ell} c_k x_k^2 \tag{5.21}$$

A stationary policy $U(t) = \phi(X(t))$ is said to be stabilizing if the resulting Markov chain $\boldsymbol{X}$ is positive recurrent. It is shown in [24] From Chapter 3 it follows that the h_c-MW policy is stabilizing under `NCond`. The $\tilde{c}$-MW policy is known to be stabilizing for a class of scheduling models, provided $\beta > 0$ is chosen sufficiently large (see [105, Section 2.2.1]). This result is extended to the matching model in Section 5.3.1.

Why we need perturbation

When the function h is linear, $h(x) = c \cdot x$, then the h-MW policy reduces to the priority policy:

$$\phi_c(x) = \arg\max\{c \cdot u : u \in \mathsf{U}_\diamond(x)\} \tag{5.22}$$

where $\mathsf{U}_\diamond(x)$ is the permissible input set defined in (5.6).

Consider for example the NN model shown in Fig. 5.1, with cost function

$$c(x) = (x_1^{\mathcal{D}} + x_3^{\mathcal{S}}) + 2(x_2^{\mathcal{D}} + x_2^{\mathcal{S}}) + 4(x_3^{\mathcal{D}} + x_1^{\mathcal{S}})$$

The resulting c-MW policy gives priority to "vertical edges". This is precisely the example discussed in Chapter 3 for which an arrival process can be constructed such that `NCond` holds, yet the resulting Markov chain $\boldsymbol{X}$ is transient. It is likely that the $\tilde{c}$-MW policy will also give rise to a transient Markov chain for $\beta > 0$ sufficiently small.

Stability of the $\tilde{c}$-MW policy

A sufficient condition for stability requires the following lower bound on the parameter β:

$$\beta > \frac{1}{\gamma}\frac{c_{\max}}{c_{\min}}\big[2\|\alpha\|^2 + m_A^2 + m_U^2\big] \tag{5.23}$$

where $c_{\min} = \min_i c_i$, $c_{\max} = \max_i c_i$, $m_A^2 = \mathsf{E}[\|A(t)\|^2]$, $m_U^2 = \max\{\|u\|^2 : u \in \mathsf{U}_\diamond\}$, and $\gamma = \frac{1}{|\mathcal{E}|}\min_{k\in\mathcal{D}\cup\mathcal{S}} \alpha_k$.

Proposition 5.3.1. *Suppose that* `NCond` *holds, and that β satisfies the bound* (5.23)*. Then the $\tilde{c}$-MW policy is stabilizing, and the controlled Markov chain $\boldsymbol{X}$ is geometrically ergodic.*

5.3.2 Workload

The monograph [109] presents a general theory of workload based on a fluid model, which is taken of the general form

$$\tfrac{d}{dt}x(t) = B\zeta(t) + \alpha$$

where α is the arrival rate, $\zeta(t)$ is a vector of "activity rates", and B is a matrix of suitable dimension. The fluid model is intended to describe the average behavior of (5.4) over a long time horizon, which leads to the following definitions for the matching model: the state process is constrained exactly as in the MDP model: $x(t) \in \mathsf{X} = \{x \in \mathbb{R}_+^\ell : \xi^0 \cdot x = 0\}$. We take B the matrix consistent with (5.15):

$$B = -[u^1 \mid \cdots \mid u^{|\mathcal{E}|}]$$

That is, the columns of $-B$ are equal to the single matches $\{u^e\}$ defined below (5.15). The input is constrained to the non-negative orthant: $\zeta(t) \in \mathbb{R}_+^{|\mathcal{E}|}$ (the upper bound $\overline{n}_u$ is relaxed).

A geometric description is given as follows. Denote by V_0 the velocity space for the arrival free model:

$$V_0 = \{v = B\zeta : \zeta \in \mathbb{R}_+^{|\mathcal{E}|}\} \tag{5.24}$$

and $V = \{v + \alpha : v \in V_0\}$. The fluid model is then described by the differential inclusion, $\frac{d}{dt}x(t) \in V$.

It is evident that the set V_0 is a cone with vertex equal to the origin. The Weyl-Minkowski Theorem asserts that it is represented as the intersection of half spaces: there are vectors $\{\xi^i : 1 \le i \le \ell_w\} \subset \mathbb{R}^\ell$ such that

$$V_0 = \{v : \xi^i \cdot v \ge 0 : 1 \le i \le \ell_w\} \tag{5.25}$$

This geometry is illustrated in Fig. 5.5.

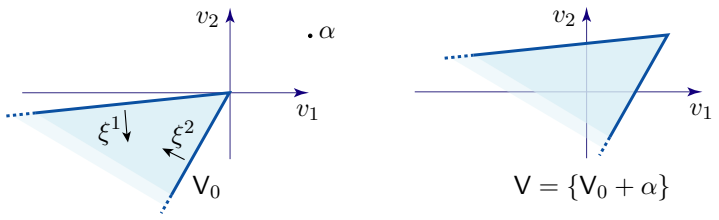


Figure 5.5: Velocity sets for a matching model: V_0 is a cone, and V is an affine translation.

The minimal draining time is defined as follows: for $x \in \mathsf{X}$,

$$T^*(x) = \inf\{t \ge 0 : x + tv = 0, \quad v \in V\}$$

The following result follows exactly as in the proof of eq. (6.13) of [109, Prop. 6.1.3].

Proposition 5.3.2. *Suppose that* `NCond` *holds. Then* $T^*(x) < \infty$ *for each* $x \in \mathsf{X}$*, and has the following representation: If* $\xi^i \cdot x \le 0$ *for each* i*, then* $T^*(x) = 0$*. Otherwise,*

$$T^*(x) = \max_i \frac{x \cdot \xi^i}{|\alpha \cdot \xi^i|} \tag{5.26}$$

□

The next step is to identify the workload vectors. For any set $D \subsetneq \mathcal{D}$ we let $\xi^D \in \mathbb{R}^\ell$ denote the vector whose components are 1 for $i \in D$, -1 for $i \in \mathcal{S}(D)$, and zero elsewhere.

Proposition 5.3.3. *The representation* (5.25) *holds with*

$$\{\xi^i\} = \{\xi^D : D \subset \mathcal{D}\} \cup \{-1^j : j \in \mathcal{D} \cup \mathcal{S}\}$$

We could introduce symmetric notation for $S \subset \mathcal{S}$, but this is unnecessary: for each $S \subsetneq \mathcal{S}$, denote $\widehat{S} = \cup\{S' \ : \ \mathcal{D}(S') = \mathcal{D}(S)\}$. By definition of $\widehat{S}$ and the connectivity of the graph, $\mathcal{S}(\mathcal{D}(S)^c) = \widehat{S}^c$. Thus, with $\bar{D} = \mathcal{D}(S)^c$,

$$\xi^S \le \xi^{\widehat{S}} = \xi^{\bar{D}} - \xi^0.$$

Moreover, by (5.24) we have $v \cdot \xi^0 = 0$, and also $v_i \le 0$ for each i for any $v \in V_0$, giving

$$v \cdot \xi^S \ge v \cdot \xi^{\bar{D}} \ge 0\,, \qquad v \in V_0$$

Consequently, it is sufficient to consider only demand in a representation of V_0.

Based on this geometry and on Prop. 5.3.2, we see that the vectors $\{\xi^D\}$ play a role similar to workload vectors in standard queueing models. Moreover, condition `NCond` can be equivalently expressed,

$$\delta(D) := -\xi^D \cdot \alpha > 0, \qquad \text{for all } D \subsetneq \mathcal{D}$$

which means that $-\alpha$ is in the interior of V_0: a typical example is shown in Fig. 5.5. A heavily loaded system is one in which $\delta(D)$ is very close to zero for one or more sets $D \subsetneq \mathcal{D}$.

We can now define a workload process. For a particular set $D \subsetneq \mathcal{D}$ we take $W(t) = \xi^D \cdot X(t)$, and $\delta = \delta(D)$.

Proposition 5.3.4. *The workload process evolves according to the recursion,*

$$W(t+1) = W(t) - \delta + I(t) + N(t+1), \qquad t \ge 0, \tag{5.27}$$

in which $\delta > 0$, $N(t+1) = \delta + \xi^D \cdot A(t+1)$, and $I(t) = -\xi^D \cdot U(t)$.

(i) *$I(t)$ takes values in the non-negative integers $\mathbb{Z}_+$, and is zero if and only if there is no matching between $\mathcal{S}(D)$ and D^c at time t.*

(ii) *The sequence $\boldsymbol{N}$ is zero-mean, i.i.d., and takes values in $\{-1+\delta, \delta, 1+\delta\}$.*

(iii) *The first and second order statistics are represented as follows:*

$$\begin{aligned} \delta &= p_N^- - p_N^+ \\ \sigma_N^2 &= p_N^- + p_N^+ - \delta^2 \end{aligned} \tag{5.28}$$

in which $p_N^+ = \mathsf{P}\{\xi^T A(t) = 1\}$ and $p_N^- = \mathsf{P}\{\xi^T A(t) = -1\}$.

Given a convex cost function $c\colon \mathbb{R}_+^\ell \to \mathbb{R}_+$, the *effective cost* is defined as the solution to the convex program,

$$\overline{c}(w) := \min\{c(x) : x \in \mathbb{R}_+^\ell,\ \xi^D \cdot x = w\,,\ \xi^0 \cdot x = 0\}, \qquad w \in \mathbb{R} \tag{5.29}$$

An optimizer is called an *effective state*. We denote by $\mathcal{X}^*$ a continuous function, and satisfying

$$\mathcal{X}^*(w) \in \arg\min_{x \in \mathbb{R}_+^\ell} \{c(x) : \xi^D \cdot x = w\,,\ \xi^0 \cdot x = 0\}, \qquad w \in \mathbb{R} \tag{5.30}$$

It is assumed throughout most of the paper that $c\colon \mathbb{R}_+^\ell \to \mathbb{R}_+$ is a linear function of the state, $c(x) = \sum c_i x_i$, with $c_i > 0$ for each i. It follows that $\overline{c}$ is piecewise linear.

Lemma 5.3.5. *For a linear cost function c, the solution to the linear program* (5.29) *results in a piecewise linear function of w:*

$$\overline{c}(w) = \max(\overline{c}_+ w, -\overline{c}_- w) \tag{5.31}$$

where $\overline{c}_+ = \overline{c}_+^D + \overline{c}_+^S$ and $\overline{c}_- = \overline{c}_-^D + \overline{c}_-^S$, with

$$\begin{aligned} \overline{c}_+^D &= \min\{c_i : i \in D\}\,, \quad \overline{c}_+^S &&= \min\{c_j : j \in S^c\} \\ \overline{c}_-^S &= \min\{c_j : j \in S\}\,, \quad \overline{c}_-^D &&= \min\{c_i : i \in D^c\} \end{aligned} \tag{5.32}$$

An optimizer x^* for (5.29) *exists in which exactly two entries are non-zero. The form depends on the sign of w:*

$$\begin{aligned} w \geq 0: \quad & x_i^* = w \quad \textit{for some } i \in D \textit{ satisfying } c_i = \overline{c}_+^D, \\ & x_j^* = w \quad \textit{for some } j \in S^c \textit{ satisfying } c_j = \overline{c}_+^S. \\ w < 0: \quad & x_i^* = |w| \quad \textit{for some } i \in D^c \textit{ satisfying } c_i = \overline{c}_-^D, \\ & x_j^* = |w| \quad \textit{for some } j \in S \textit{ satisfying } c_j = \overline{c}_-^S. \end{aligned} \tag{5.33}$$

□

We henceforth assume that $x^* = \mathcal{X}^*(w)$ is of the form (5.33) for a fixed pair of indices i, j.

The workload relaxation is an MDP model, described as a one-dimensional controlled random walk, defined on the same probability space, and evolves as,

$$\widehat{W}(t+1) = \widehat{W}(t) - \delta + \hat{I}(t) + N(t+1), \qquad t \geq 0, \tag{5.34}$$

in which $\widehat{W}(0) \in \mathbb{R}$ is given, and $\boldsymbol{N}$ is an i.i.d. sequence in $\mathbb{R}$ with zero mean. The process $\hat{I}(t)$ is interpreted as the input for this MDP model: it is adapted to $\widehat{\boldsymbol{W}}$ and takes on non-negative values. The controlled random walk (5.34) with cost function $\overline{c}$ is thus a relaxation of the original MDP model.

The controlled random walk (5.34) is considered in [109, Sec. 7.4 and Sec. 9.7]: In [109, Theorem 9.7.2] it is shown that an optimal policy is determined by a threshold policy of the following form: There is a scalar $\tau^\bullet > 0$ so that

$$\hat{I}(t) = \max\{\widehat{W}(t) - \tau^\bullet, 0\} \tag{5.35}$$

Under this policy, the stochastic process $\{\Phi(t) = \widehat{W}(t) - N(t) + \delta\}$ is a reflected random walk on $[-\tau^\bullet, \infty)$. Equation (7.37) of [109] defines the diffusion heuristic, intended to approximate this threshold based on a reflected-Brownian motion model, giving

$$\tau^* = \tfrac{1}{2}\frac{\sigma_N^2}{\delta} \log\Big(1 + \frac{\overline{c}_+}{\overline{c}_-}\Big) \tag{5.36}$$

where δ and σ_N^2 are given in (5.28). The approximation $|\tau^\bullet - \tau^*| = O(1)$, independent of δ, is established.

5.3.3 *h*-MaxWeight with threshold

The structure of the policy for the relaxation is the inspiration for the following refinement of the h-MaxWeight policy in (5.48).

For a differentiable function $h\colon \mathbb{R}^\ell \to \mathbb{R}_+$, and a threshold $\tau \geq 0$, the h-MWT (h-MaxWeight with threshold) policy is obtained as the solution to the constrained non-linear program,

$$\begin{aligned} \phi(x) = \arg\max \quad & u \cdot \nabla h\,(x) \\ \text{subject to} \quad & u \in \mathsf{U}_\diamond(x) \quad \textit{and} \quad \xi^D \cdot u \geq \min(w + \tau, 0) \end{aligned} \tag{5.37}$$

Based on the definition of workload and idleness (5.27), the constraint $\xi^D \cdot u \geq \min(w+\tau, 0)$ is equivalently expressed

$$I(t) \leq \max(-W(t) - \tau, 0), \qquad \textit{when } X(t) = x \textit{ and } U(t) = u.$$

This constraint is motivated by the definition of a threshold policy (5.35) for the workload relaxation.

We take $\tau = \tau^*$ in our main results, and the function h (see (5.46)) is also designed using inspiration from the workload relaxation.

The choice of $h = \tilde{c}$, as defined in (5.18), will also be considered in numerical experiments in Section 8.3.

5.3.4 Asymptotic optimality

To evaluate performance we consider an asymptotic setting: Assume that we have a family of arrival processes $\{A^\delta(t)\}$ parameterized by $\delta \in [0, \bar{\delta}_\bullet]$, where $\bar{\delta}_\bullet \in (0,1)$. Each is assumed to satisfy (5.2). The following additional assumptions are imposed throughout:

(A1) For one set $D \subsetneq \mathcal{D}$ we have $\xi^D \cdot \alpha^\delta = -\delta$, where α^δ denotes the mean of $A^\delta(t)$.

Moreover, there is a fixed constant $\underline{\delta} > 0$ such that $\xi^{D'} \cdot \alpha^\delta \leq -\underline{\delta}$ for any $D' \subsetneq \mathcal{D}$, $D' \neq D$, and $\delta \in [0, \bar{\delta}_\bullet]$.

(A2) The distributions are continuous at $\delta = 0$, with linear rate: For some constant b,

$$\mathsf{E}[\|A^\delta(t) - A^0(t)\|] \leq b\delta. \tag{5.38}$$

(A3) The sets $\mathcal{E}$ and $\mathcal{A}$ do not depend upon δ, and the graph associated with $\mathcal{E}$ is connected. Moreover, there exists $i_0 \in \mathcal{S}(D)$, $j_0 \in D^c$, and $p_I > 0$ such that

$$\mathsf{P}\{A^\delta_{i_0}(t) \geq 1 \textit{ and } A^\delta_{j_0}(t) \geq 1\} \geq p_I, \qquad 0 \leq \delta \leq \bar{\delta}_\bullet. \tag{5.39}$$

We suppress the dependency of $\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{U}$ on δ when there is no risk of confusion. We also let $\xi = \xi^D$, so that $\delta = -\xi \cdot \alpha$.

We are now prepared to state the main result this section, establishing asymptotic optimality of a family of h-MWT policies. We let η^* denote the optimal average cost for the MDP model, $\widehat{\eta}^*$ the optimal average cost for (5.34), and the following is shown to approximate each of these values:

$$\widehat{\eta}^{**} = \tau^* \overline{c}_- = \tfrac{1}{2} \frac{\sigma^2_\Delta}{\delta} \, \overline{c}_- \log\Big(1 + \frac{\overline{c}_+}{\overline{c}_-}\Big) \tag{5.40}$$

Theorem 5.3.6 (Asymptotic Optimality With Bounded Regret)**.** *Suppose that Assumptions (A1)–(A3) hold. For each $\delta \in (0, \bar{\delta}_\bullet]$, there is a function h such that the h-MWT policy using the threshold τ^* has finite average cost η, satisfying the following bounds,*

$$\widehat{\eta}^* \leq \eta^* \leq \eta \leq \widehat{\eta}^* + O(1)$$

where the constant $O(1)$ does not depend upon δ. Moreover, the average cost for the relaxation is approximated by the value in (6.11)*:*

$$\widehat{\eta}^* = \widehat{\eta}^{**} + O(1)$$

Construction of the h-MWT policy. The ACOE for the MDP model is

$$\min_{u\in\mathsf{U}_\diamond} \mathsf{E}[c(X(t)) + h^*(X(t+1)) \mid X(t)=x\,,\ U(t)=u] = h^*(x) + \eta^*, \qquad x\in\mathsf{X}_\diamond, \tag{5.41}$$

in which η^* is the optimal average cost, and h^* is the relative value function. The approximation is taken as the sum of two terms:

$$h(x) = \hat{h}(\xi\cdot x) + h_c(x)$$

Similar to [105], the function h_c is introduced to penalize deviations between $c(x)$ and $\overline{c}(\xi\cdot x)$.

The first term is a function of workload: it solves for $w\ge -\tau^*$ the differential equation,

$$-\delta\hat{h}'(w) + \tfrac{1}{2}\sigma_\Delta^2\hat{h}''(w) = -\overline{c}(w) + \widehat{\eta}^{**}, \tag{5.42}$$

where the threshold τ^* is defined in (5.36), and the optimal average cost $\widehat{\eta}^{**}$ is given in (6.11). There is a solution that is convex and increasing on $[-\tau^*,\infty)$, with $\hat{h}'(-\tau^*)=\hat{h}''(-\tau^*)=0$:

$$\hat{h}(w) = \begin{cases} A_+w^2 + B_+w & w\ge 0 \\ A_-w^2 + B_-w + C_- + D_-e^{\Theta w} & -\tau^*\le w\le 0 \end{cases} \tag{5.43}$$

where $\Theta^{-1} = \sigma_\Delta^2/(2\delta)$, $A_+ = \overline{c}_+/(2\delta)$, $B_+ = \delta^{-1}\big(\sigma_\Delta^2 A_+ - \widehat{\eta}^{**}\big)$, $A_- = -\overline{c}_-/(\Theta\sigma_\Delta^2)$, $B_- = 2\Theta^{-1}A_- - \delta^{-1}\widehat{\eta}^{**}$, $D_- = \Theta^{-1}\big(B_+ - B_-\big)$. The constant C_- is obtained by imposing continuity at zero.

The domain is extended to obtain a convex, C^2 function on all of $\mathbb{R}$. We fix a parameter $\delta_+\in(0,p_I)$, where $p_I>0$ is used in Assumption (A3). The sum $\delta+\delta_+$ is interpreted as the desired idleness rate when $w<-\tau^*$. Fix another constant $\theta>0$, and for $w<-\tau^*$ define

$$\hat{h}(w) = \hat{h}(-\tau^*) + \frac{\overline{c}_-}{\delta_+}\Big[\tfrac{1}{2}(w+\tau^*)^2 + \frac{1}{\theta}(w+\tau^*) + \frac{1}{\theta^2}\Big(1-\exp(\theta(w+\tau^*))\Big)\Big] \tag{5.44}$$

where $\hat{h}(-\tau^*)$ is given in (5.43).

We now turn to the construction of h_c. For this we might take a constant times $[c(x)-\overline{c}(\xi\cdot x)]^2$. This fails because of positive drift on the boundary of $\mathsf{X}_\diamond$.

Let $\tilde{x}$ denote the function of x with entries, $\tilde{x}_k = x_k + \beta(e^{-x_k/\beta}-1)$, where $\beta>0$ is a constant. The right hand side vanishes at the origin, as does its first derivative. The constant β is chosen so that its derivative with respect to x_k is small for x_k in some interval $[0,\overline{q}]$.

A similar transformation for workload is used,

$$\widetilde{w} = \operatorname{sign}(w)\big[|w| + \beta(e^{-|w|/\beta}-1)\big] \tag{5.45}$$

If $w=\xi\cdot x$ then the definition does not change, but $\widetilde{w}$ is of course a function of x; the perturbation ensures that $\overline{c}(\widetilde{w})$ is C^1 as a function of x.

The second term is defined to be $h_c(x) = [c(x)-\overline{c}(w)]^2$; the function h and its gradient are thus

$$h(x) = \hat{h}(w) + h_c(x) = \hat{h}(w) + \kappa[c(\tilde{x}) - \overline{c}(\widetilde{w})]^2 \tag{5.46}$$

$$\nabla h\,(x) = \hat{h}'(w)\xi + \nabla h_c(x)\,, \qquad x\in\mathbb{R}^\ell_+,\ w=\xi\cdot x. \tag{5.47}$$

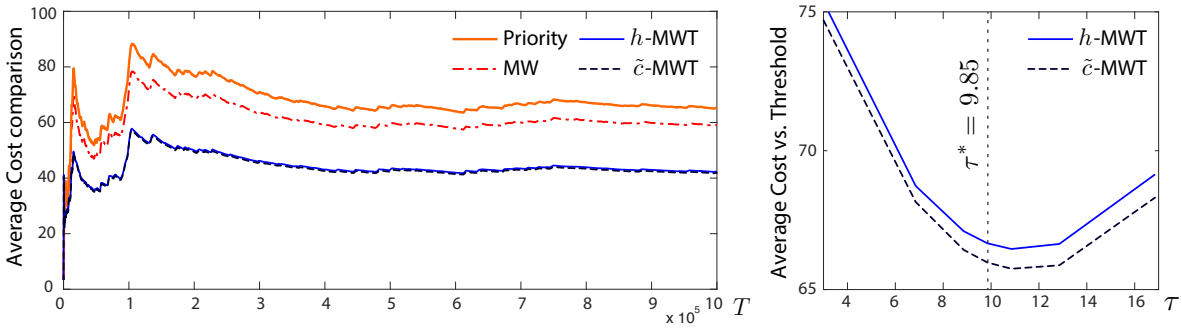


Figure 5.6: The plots on the left hand side compare the average cost for $T \leq 10^6$, obtained using h-MWT, $\tilde{c}$-MWT, MaxWeight (given by (5.48) with $h(x) = \sum c_i x_i^2$), and a priority policy. Shown on the right is the average cost at $T = 10^5$ for the two MWT policies in which the threshold τ was taken as a parameter.

5.3.5 Numerical experiments

Numerical experiments were performed for the NN-network from Figure 5.1, for various cost functions and arrival statistics. Results from one experiment are summarized here. The parameters were chosen so that the set that defines maximal workload was $D = \{3^{\mathcal{D}}\}$. The value $\delta = -\xi^{\mathcal{T}}\alpha$ was smaller than $-\xi^{D'} \cdot \alpha$ for any other set $D' \subsetneq \mathcal{D}$. The cost was taken to be $c(x) = x_1^{\mathcal{D}} + 2x_2^{\mathcal{D}} + 3x_3^{\mathcal{D}} + 3x_1^{\mathcal{S}} + 2x_2^{\mathcal{S}} + x_3^{\mathcal{S}}$. Numerical values were chosen so that $\delta = 0.01$.

Four policies were considered: two versions of the h-MWT policy (5.37), one with h given in (5.46) and the other with $h(x) = c(\tilde{x})^2$; the static policy giving priority to vertical matches (edges e_1, e_3 and e_5 in Fig. 5.1); the cost-weighted MaxWeight policy:

$$\phi(x) = \arg\max\{u \cdot \nabla h\,(x) : u \in \mathsf{U}_\diamond(x)\}, \qquad x \in \mathsf{X}_\diamond, \tag{5.48}$$

with $h(x) = \sum c_i x_i^2$.

A comparison of the average cost $T^{-1}\sum_{t=1}^{T} c(Q(t))$ is shown on the left hand side of Fig. 5.6. The average cost under either of the MWT policies (using the $\tau = \tau^*$) performed the best – about 30% lower than the cost-weighted MaxWeight policy.

The figure on the right hand side shows the average cost for $T = 10^5$ obtained for the two MWT policies for a range of threshold values τ. It is clear that the best value of τ obtained through simulation is very close to the value τ^* predicted by the RBM model.

It is conjectured that $c(\tilde{x})$-MWT is also asymptotically optimal.

5.4 Discussion and related results

In this chapter we considered bipartite matching graphs with linear costs in the buffer size. We model this problem as a Markov decision problem for the average cost problem.

In Section 5.2, we fully characterize the optimal policy for the N-shaped matching graph. We show that there exists an optimal policy that is of threshold type for the diagonal edge with priority to the pendant edges of the matching graph. We also fully characterize the optimal threshold value. Under some assumptions in the costs of the nodes, this threshold-based structure of the optimal matching policy extends to quasi-complete graphs (i.e. complete graphs minus one edge). In an arbitrary acyclic graph, we show that, when the cost of the

pendant nodes is larger or equal than the cost of its neighbors, the optimal policy always gives priority to the pendant edges. In [J5] we also provide similar optimality results for the discounted cost problem, and we discuss the W-shaped matching model. We conjecture that the optimal matching policy for the W-model is also of threshold type with priority to the extreme edges when the cost of the extreme nodes is higher that the one of other nodes.

In Section 5.3, we have shown how relaxation techniques can lead to insight for the construction of good policies with low complexity. The key argument is a correspondence with models in inventory theory.

Many of the results in [109] on workload relaxations are based on *stabilizability of the arrival-free model*. That is, it is assumed that the network without arrivals can be stabilized using some policy. This assumption *fails* for matching models. Consider the case of organ donation (e.g. [1]): if there is a patient waiting for a kidney, and no donors arrive, then the patient will wait for eternity. Nevertheless, there is a natural formulation of workload for these models, described in Section 5.3.2. Each component of the multi-dimensional workload process can take on positive and negative values, much like what is found in inventory models. It is found that optimal policies will have structure similar to what is found in inventory theory, such as the classical work of Clark and Scarf [45]. In particular, based on a one-dimensional relaxation, an approximating model is obtained that can be identified as an inventory model of a special form, so that an optimal policy for the relaxation is obtained via a one-dimensional threshold policy. We introduced a class of policies that take into account the structure found in the optimal policy for a workload relaxation, leading to the h-MaxWeight with threshold policy, or h-MWT. This policy is greedy subject to a non-idling constraint whenever the workload is above a pre-determined threshold. The non-idling constraint is relaxed when the workload is below this threshold. It is demonstrated that this simple policy is approximately optimal, in the sense of Theorem 5.3.6, so that the regret is bounded as a function of network load.

These conclusions imply that optimal policies do not follow the conventions of Section 3. Optimal policies may perform no matches at certain time instances, even though matches are possible.

Although the theoretical results are based on a heavy-traffic setting, we expect that this structure will play some role even when the assumptions of this work are violated.

The conclusion of Theorem 5.3.6 and the analysis used are different from the prior work [105] or [70, 82]. The paper [105] is most similar, in which the main result is a logarithmic bound on regret in heavy-traffic. This is a weaker conclusion compared to our result, but the main difference is an entirely different mode of analysis. The proof of optimality in [105, 70, 82] relies on the simplicity of the optimal solution, which is approximately path-wise minimal in heavy-traffic. The proof of logarithmic regret amounts to quantifying this approximation. The proof of bounded regret in our work is based on "lifting" the dynamic programming equations from a diffusion approximation to the original stochastic network. This diffusion approximation is entirely different than encountered in this prior work.

The prior work [66] considers a general class of matching models, with performance analysis based on an asymptotic heavy-traffic setting. The conclusions are very different because of the different assumptions imposed on the network model: when specialized to connected bipartite graphs, their Assumption 1 implies that bipartite graph reduces to a star network.

The matching graphs considered in [66] allow more general topologies, including certain hypergraphs. Assumption 1 is useful in their analysis because it is then possible to establish a form of path-wise optimality for the workload model under a particular policy (as in [105,

70, 82]). This is not possible for the models considered in the present paper because the workload process takes on positive and negative values: an average-cost optimal policy has a threshold form, which is inconsistent with path-wise optimality (see Section 5.5 of [109] for further discussion). A related average reward maximization problem has been considered in [112]. An extension of the greedy primal-dual algorithm was developed and proved to be asymptotically optimal for the long-term average matching reward. However their results do no longer hold if there is a cost function on the queue sizes.

Part II

Control of distributed power demand

Chapter 6

Balancing the power grid with flexible loads

The power system transformation brings new challenges and opportunities due to changes and uncertainties in electricity consumption and generation. In power networks, it is necessary that the electricity production is equal to the demand at all times. In addition to ensuring sufficient electricity production, there is also a need for flexible resources that can quickly adapt their production / consumption to compensate for demand forecasting errors and ensure real-time balancing between production and demand. This service is an example of the system services (also called ancillary services) essential to the proper functioning of power grids.

Matching electricity supply and demand used to be relatively straightforward, with large and controllable power plants on the one hand, and demand that was relatively easy to predict on the other. Slowly ramping cheaper generators were committed in advance to follow the predicted demand. The real time balancing was done by ramping up or down the most responsive power plants, such as gas turbines or hydro, when available. They were operating at lower capacity, leaving the possibility to ramp up or down their generation. This was providing balancing reserves used to correct the forecasting errors and follow the demand in real time. In recent years, there has been a significant increase in participation of intermittent renewable generation. Balancing service from traditional power plants is becoming very expensive due to the need to compensate for the missed opportunity cost the power plants are facing while operating at a lower set-point to be able to ramp up and down more aggressively than in the past. Providing new flexibility resources is crucial to integrate renewable energies into the power grid. At the same time, the rapid development of "smart technologies" (e.g. Linky meter and the connected appliances) has opened new possibilities for innovation on the demand side, as well as new control solutions on the grid level.

There is an enormous flexibility potential in the power consumption of the majority of electric loads (e.g., thermal loads such as water heaters, air-conditioners and refrigerators; electric vehicles, etc.). Their power consumption can be shifted in time to some extent without any significant impact to the consumer needs. This flexibility can be exploited to create "virtual batteries". The best example of this is the heating, ventilation, and air conditioning (HVAC) system of a building: There is no perceptible change to the indoor climate if the airflow rate is increased by 10% for 20 minutes, and decreased by 10% for the next 20 minutes. Power consumption deviations follow the airflow deviations closely, but indoor temperature will be essentially constant.

The major issue lies in the distributed nature of this flexibility resource: piloting the flexible demand in real time requires a design of (simple) incentives for millions of devices. Moreover, many residential devices are on-off (e.g. water-heater or air-conditioner). In order to provide valuable balancing service, the aggregate must provide a predictable response. The future power grids will contain millions of smart components, which completely prohibits centralized decision making using standard stochastic optimization techniques, such as stochastic dynamic programming and Markov decision processes (MDP), as they do not scale well with the number of different components in the system (both the state space and the control space of the model grow exponentially with the number of components).

This part of the manuscript provides an overview of our probabilistic distributed control approach for balancing the power grid using flexible loads. The proposed approach relies on new smart technologies allowing for automatic control of devices. The objective is to control a great amount of devices to provide services to the system (load shaping or ancillary services) while: i) maintaining the quality of service for the users; ii) minimizing communications between controllable devices and the central controller. The proposed approach combines the techniques from the theory of controlled Markov processes, mean-field theory, and linear control theory.

We start in Chapter 6 with an overview of our probabilistic distributed control approach for an online tracking problem: The objective is to control the average consumption of a population of N devices to track the reference signal (R_t), which is progressively revealed by the grid at discrete time steps $t = 1, ..., T$ (online reference tracking problem). Through load-level and grid-level control design, high-quality ancillary service for the grid is obtained without impacting quality of service delivered to the consumer. This approach to grid regulation is called demand dispatch: loads are providing service continuously and automatically, without consumer interference.

One theoretical contribution will be presented more in detail in Chapter 7. It was motivated by the need to extend the initial distributed control approach to include the randomness that cannot be controlled (e.g. hot water usage or the weather conditions) [C41]. This lead to a new ODE method for solving a parametrized family of Markov Decision Processes [J10]. Besides power applications, this new technique also has its potential applications in machine learning and robotics [C33].

Demand dispatch has many advantages:

- *minimal communication:* a unique control signal is sent from the central entity to the loads, without the communication from the loads to the centralized entity;
- local control design enables strict guarantees for the quality of service for the users;
- randomized control limits the synchronization of the response of the loads.

However, in this online reference tracking formulation of the problem, the target consumption is revealed in real time (there is no anticipation of the target by probabilistic forecasts). The fact of not allowing any anticipation of the target consumption does not make it possible to fully integrate the constraints of the different devices in terms of energy consumed over a given period. To overcome this limitation, we have proposed an offline reference tracking approach that takes into account a deterministic forecast of the target consumption over a period of anticipation (e.g. day ahead) and solves directly the tracking problem at the population level, formalized as a Kullback-Leibler-Quadratic (KLQ) optimal control problem in discrete

[C20, C8, J1], or continuous time [C22]. This new Kullback-Leibler-Quadratic (KLQ) control approach can be seen as a special case of a finite horizon stochastic optimal control problem with the objective function that is composed of two terms: quadratic tracking error cost and a relative entropy control cost that penalizes the deviation from the nominal behavior of the load. An overview for the discrete time case is provided in Chapter 8.

This part is based on the following publications: [B2] (Chapter 6), [C33, J10] (Chapter 7), and [C8, J1] (Chapter 8).

6.1 Introduction

Inexpensive energy from the wind and the sun comes with unwanted volatility, such as ramps with the setting sun or a gust of wind. Controllable generators manage supply-demand balance of power today, but this is becoming increasingly costly with increasing penetration of renewable energy. It has been argued since the 1980s that consumers should be put in the loop: "demand response" will help to create needed supply-demand balance. However, consumers use power for a reason, and expect that the quality of service (QoS) they receive will lie within reasonable bounds. For example, the temperature in a building or refrigerator must lie within strict bounds. Moreover, the behavior of some consumers is unpredictable, while the grid operator requires predictable controllable resources to maintain reliability. The goal of this chapter is to describe distributed control solutions for *demand dispatch* that will create *virtual energy storage* from flexible loads. By design, the grid-level services from flexible loads will be as controllable and predictable as a generator or fleet of batteries. Strict bounds on QoS will be maintained in all cases. The potential economic impact of these new resources is enormous. California spends billions of dollars on batteries that provide only a small fraction of the balancing services that could be obtained using demand dispatch. The potential impact on society is enormous: a sustainable energy future is possible with the right mix of infrastructure and control systems.

Supply-demand balance in a power grid. As more wind and solar energy come online, the system operators who run the power grid are faced with a problem: how do they compensate for the variable nature of renewable energy resources? The control systems diagram in Fig. 6.1 provides a simple view of how the grid is operated today, in which wind and solar are viewed as sources of disturbances. In North America, the **GRID** block is in fact a subset of the grid called a *balancing region*. The block denoted **Compensation** represents a *balancing authority* (BA). The grid-level measurements obtained by the BA are summarized as a scalar function of time called the area control error (ACE). It is a linear combination of two error signals: the deviation of local grid frequency from the nominal 60 Hz, and the tie-line error — defined as the mismatch between scheduled and actual flow of power out of the balancing region. Command signals are broadcast to resources such as controllable generators so that the ACE signal is kept within desired bounds.

The compensator G_c is designed by the BA in a particular region. For example, PJM (an RTO in the Eastern U.S.) creates their RegA and RegD signals by passing the ACE signal first through a PI compensator, and then through a bandpass filter. In this case the compensator G_c in Fig. 6.1 is taken to be a PI controller, and the bandpass filters are embedded in the block denoted **Actuation**. The decomposition "$H = H_a + H_b + \cdots$" represents many resources acting in parallel to provide actuation, which may include controllable generation

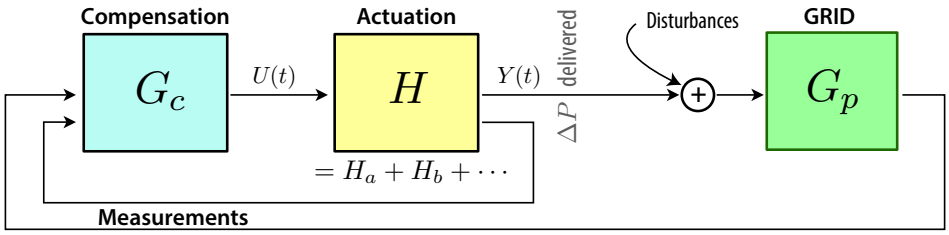


Figure 6.1: Power Grid Control Loop.

and batteries.

It is anticipated that the basic architecture illustrated in Fig. 6.1 will remain in place for many decades to come. The grid will become more adaptable to persistent disturbances or crisis through a combination of control techniques and hardware.

The term *ancillary service* refers to resources required to maintain supply-demand balance in the grid, but do not necessarily supply energy. While controllable generation is the most common source of most ancillary services today, other technologies such as flywheels and batteries are increasingly popular.

Virtual energy storage. Batteries may be a clean source of ancillary service, but currently they are still an expensive solution. In addition to the large space required for large systems, batteries have finite life time, and waste energy as they are charged and discharged to service the grid [55].

Distributed control architectures are described in this chapter to create *virtual energy storage* (VES) based on the inherent flexibility in power consumption from a majority of loads. The ancillary services that can be obtained include regulation (such as automatic generation control, or AGC), smooth peaks in load, address ramps from wind or solar generation, and help to recover gracefully from contingencies such as generation faults. It is believed that VES is a low-cost complement to batteries and power plants, and may in the future provide the majority of required ancillary services.

The term *Demand Dispatch* is a convenient alternative to *Demand Response*; the latter is defined by policy makers and regulatory bodies (such as FERC) as load-shedding in exchange for some monetary reward. *Load shedding is not the goal of the technology surveyed here.* In applications to both regulation and ramping services, the distributed control algorithms are designed so that power consumption is increased and decreased over time, while keeping the total energy deviation over time at zero — *just like charging and discharging of a battery.*

The control architectures described in this chapter are based on a series of papers on distributed control [C50, J16, J14, J13].

The proposed frequency decomposition of VES resources was first introduced in [68, 67] in the context of commercial buildings, and generalized in [C44]. The key novel contribution in all of this work is the focus on "intelligence at the load", based on local control loops. There are many benefits:

(i) Communication infrastructure requirements are reduced, which hopefully leads to both improved security, and higher consumer confidence regarding privacy.

(ii) A simple control problem at the BA, since the single-input / single-output system is highly

controllable.

(iii) Strict bounds on quality of service (QoS) to the consumer are guaranteed.

This chapter does not consider market issues. It is assumed that consumer engagement will be achieved through contractual agreements and periodic credits, such as automatic water-heater control by EDF, or those offered by Florida Power and Light in their OnCall® program.

The remainder of the chapter is organized as follows. Section 6.2 contains a high-level description of the control architecture, with details on distributed control contained in Section 6.3. Section 6.4 provides an application to control of TCLs. Related results are discussed in Section 6.5.

6.2 Distributed Control Architecture

The grid operator requires resources to balance the grid at all times. A significant proportion of the needed resources can come in the form of virtual storage from flexible loads. Reliable grid services can be obtained from loads, but this requires a well-designed control architecture.

A particular hierarchical control architecture is proposed. One realization is illustrated by the feedback structure shown in Fig. 6.1, in which the actuation block is composed of many resources acting in parallel, including generation, batteries, and virtual energy storage.

Assumptions regarding this control structure include

(i) *Local control:* This will be based in part on randomized decision rules, that provide necessary degrees of freedom in shaping aggregate dynamics. Randomization also helps to prevent synchronization of the response from loads.

(ii) *Information flow from loads:* Two-way information exchange between the BA and individual loads is *not* a component of this architecture. In [J16] it is assumed that the BA measures aggregate power consumption from the loads under its authority. Alternatively, each load broadcasts its power state several times per day, and aggregate power consumption is estimated at the BA [C43].

In [C40, C42] it is argued that it is possible to create a reliable control system in which direct information flow from loads to BA is entirely absent. This requires more complex control at each load, and hence is beyond the scope of this chapter.

(iii) *Information flow from the BA:* A single regulation signal is broadcast to each collection of loads of the same class, as illustrated in Fig. 6.2. This signal is designed based on grid level measurements, and a model of the aggregate behavior of the loads in each class.

The value of "local intelligence" at each load is vital for the envisioned architecture. Feedback loops at each load are used to ensure that QoS constraints are met, and also so that the aggregate of loads will appear to the grid operator as a reliable resource – much like a battery system, or a controllable generator.

Consumer choice will be an input to any VES system, and a monetary reward may be part of the arrangement. A contract for services can be established so that the consumer is rewarded for participation, without exposing him or her to the complexity and uncertainty of the grid. In this way the BA or aggregator can design the system so that highly reliable grid services are obtained, while respecting the QoS constraints of the consumer.

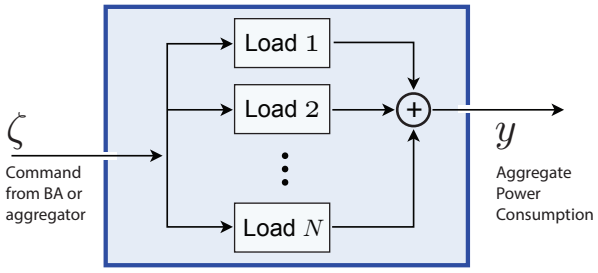


Figure 6.2: *Control architecture*: a common command signal is transmitted to each load in a particular class. The resulting input-output system from ζ to power consumption y is regarded as virtual energy storage.

In the future it is possible that some loads will be grid-friendly by design; the consumer will never know that their refrigerator is helping to regulate the grid.

Local control The lowest level of control in the proposed architecture is at an individual load, such as a water heater, refrigerator, agricultural water pump, or air-conditioner. The load is equipped with sensors. For example, the microprocessor in a water heater receives measurements of water temperature at one or more locations in the unit. It is also assumed that the load receives measurements from the grid. This could be purely local, such as the grid frequency measured locally [101, 100]. We will consider here the setting where each load receives a signal from the BA, in which the theory is best developed.

The local control loop is designed to meet these potentially conflicting goals: 1. Ensure that the load is providing the desired services to the consumer, respecting strict bounds on QoS, and 2. Ensure that the *aggregate* of loads responds to a signal from the BA in a manner that is both predictable and beneficial to the grid.

One obvious challenge: the degrees of freedom are extremely limited for a typical load of interest. For example, a residential water heater or refrigerator can be in only one of a small number of power states. Contained in Section 6.3 are several design techniques for local control that result in smooth aggregate behavior. This is possible without the introduction of complex scheduling rules, or the solution of real-time optimization problems at the BA.

Macro control This high-level control layer may be a part of the traditional BA, or through a load aggregator. The balancing challenges are of many different categories, on many different time-scales:

(i) Automatic Generation Control (AGC): time scales of seconds to 20 minutes.

(ii) Balancing reserves. In the Bonneville Power Authority, the balancing reserves include both AGC and balancing on timescales of many hours.[1]

(iii) Contingencies (e.g., a generator outage).

(iv) Peak shaving.

[1]Balancing on a slower time-scale is achieved through real time markets in some other regions of the U.S., and in every region under the jurisdiction of an RTO.

(v) Smoothing ramps from solar or wind generation.

In this chapter it is assumed that these high-level control problems are addressed as they are today: the BA receives measurements of the grid, and based on this information sends out signals to each resource in its domain. In many cases control loops are based on standard PI (proportional-integral) control design.

The difference here is that some resources are virtual, such as a collection of water heaters. A large collection of batteries distributed across the region might be regarded as a single resource – in this case, local control loops will be installed in each battery system so that the aggregate behaves as a single massive battery.

6.3 Mean-Field Control Design

Standard approaches for solving a stochastic control problem include stochastic dynamic programming and Markov decision processes (MDP) [117]. The future power grids will contain millions of smart components, which prohibits centralized decision making using these techniques as they do not scale well with the number of different components in the system (both the state space and the control space of the model grow exponentially with the number of components). The extension of MDP models to the case of optimization problems involving many agents that are making decisions based on partial knowledge of the system is called DEC-POMDP (decentralized partially observable MDP). These problems are NEXP-hard for the finite horizon optimization case [14], and undecidable in the infinite horizon case [96].

In physics and probability theory, mean field theory (MFT) approximates the behavior of a large number of small individual components which interact with each other. The effect of all the other individuals on any given individual is approximated by a single averaged effect, thus reducing a many-body problem to a one-body problem. The mean-field ideas first appeared in physics in the work of Pierre Curie and Pierre Weiss to describe phase transitions [78, 139]. Approaches inspired by these ideas have seen applications in epidemic models [19], computer network performance and game theory [91, 74]. In power systems, they were first used to model the aggregate dynamic of the collection of water-heaters in [98], and more recently in [87, 135, 130]. However, the global objective optimization under mean-field interactions remains very challenging and an exact analysis is possible only under restrictive assumptions on the local dynamics and the cost structure. There is still a significant gap between the theoretical assumptions and the applications, and the results may be sensitive to the modeling errors.

The approach in [38, 39, 26] combines the mean-field theory with classical feedback control. The main idea consists in defining a parametrized family of randomized local decision rules that lead to an aggregate behavior with desirable control properties (e.g. passivity for the linearized aggregate input-output system).

This section provides an overview of key concepts and results of this approach.

6.3.1 Mean-field model

A nominal Markovian model for an individual load is created based on its typical operating behavior. This is described as a Markov chain with transition matrix denoted P_0, with state space $\mathsf{X} = \{x^1, \dots, x^d\}$. For example, a water chiller turns on or off depending upon the

temperature of the water. In this case, a state value x^i encodes water temperature as well as the power state (on or off).

A family of transition matrices $\{P_\zeta : \zeta \in \mathbb{R}\}$ is then constructed to define local decision making. Each load evolves as a controlled Markov chain on X, with common input $\boldsymbol{\zeta} = (\zeta_0, \zeta_1, \dots)$. It is assumed that the scalar signal $\boldsymbol{\zeta}$ is broadcast to each load. If a load is in state x at time t, and the value ζ_t is broadcast, then the load transitions to the state x' with probability $P_{\zeta_t}(x, x')$. Letting X_t^i denote the state of the ith load at time t, and assuming N loads, the empirical pmf (probability mass function) is defined as the average,

$$\mu_t^N(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\{X_t^i = x\}, \qquad x \in \mathsf{X}.$$

The mean-field model is the deterministic system defined by the evolution equations,

$$\mu_{t+1} = \mu_t P_{\zeta_t}, \quad t \geq 0, \tag{6.1}$$

in which μ_t is a row vector of dimension d. Under general conditions on the model and on μ_0 it can be shown that μ_t^N is approximated by μ_t.

In [38, 39, 107] it is assumed that average power consumption is obtained through measurements or state estimation: Let $\mathcal{U}(x)$ denote power consumption when the load is in state x, where $\mathcal{U}\colon \mathsf{X} \to \mathbb{R}_+$. The average power consumption is denoted

$$y_t^N = \frac{1}{N}\sum_{i=1}^{N} \mathcal{U}(X_t^i)\,,$$

which is approximated using the mean-field model:

$$y_t = \sum_x \mu_t(x)\mathcal{U}(x), \quad t \geq 0. \tag{6.2}$$

The next subsection describes the linearized dynamics. Sections 6.3.2–6.3.3 provide an overview of design techniques for the parametrized transition family $\{P_\zeta : \zeta \in \mathbb{R}\}$, to ensure that the linearized input-output model has desirable properties for control at the grid level.

Linearized mean-field model The mean-field model (6.1) is a state space model that is linear in the state μ_t, and nonlinear in the input ζ_t. The observation equation (6.2) is also linear as a function of the state. Assumptions imposed in [38, 39, 107] imply that the input is a continuous function of these values. The design of the feedback law $\zeta_t = \phi_t(y_0, \dots, y_t)$ is based on a linearization of this state space model.

The linearized input-output model requires additional notation. The derivative of the transition matrix is also a $d \times d$ matrix, denoted

$$\mathcal{E}_\zeta = \frac{d}{d\zeta} P_\zeta \tag{6.3}$$

Denote $\widetilde{\mathcal{U}}_\zeta = \mathcal{U} - \bar{u}_\zeta$, with $\bar{u}_\zeta = \pi_\zeta(\mathcal{U})$.

The invariant pmf π_ζ for P_ζ is regarded as the equilibrium state for the mean-field model (6.1), with respect to the constant input value $\zeta_t \equiv \zeta$. The linearization about this equilibrium is described in Prop. 6.3.1. The proof can be found in [107, Prop. 2.4].

Proposition 6.3.1. *Consider a family of transition matrices $\{P_\zeta : \zeta \in \mathbb{R}\}$ that are continuously differentiable in ζ. Assume also that P_ζ is irreducible and aperiodic for each ζ. The unique invariant pmf π_ζ is an equilibrium for* (6.1) *when $\boldsymbol{\zeta}$ takes on this constant value. The input-output model with state evolution* (6.1)*, input $\boldsymbol{\zeta}$, and output* (6.2) *admits a linearization about this equilibrium. It is described as a d-dimensional state space model with transfer function,*

$$G_\zeta(z) = C[Iz - A]^{-1}B \tag{6.4}$$

in which $A = P_\zeta^T$, $C_i = \widetilde{\mathcal{U}}_\zeta(x^i)$ for each i, and

$$B_i = \sum_x \pi_\zeta(x)\mathcal{E}_\zeta(x, x^i), \qquad 1 \le i \le d \tag{6.5}$$

6.3.2 Local control design

It is assumed throughout this chapter that the family of transition matrices used for distributed control is of the form,

$$P_\zeta(x, x') := P_0(x, x') \exp\big(h_\zeta(x, x') - \Lambda_{h_\zeta}(x)\big) \tag{6.6}$$

in which h_ζ is continuously differentiable in ζ, and Λ_{h_ζ} is the normalizing constant

$$\Lambda_{h_\zeta}(x) := \log\Big(\sum_{x'} P_0(x, x') \exp\big(h_\zeta(x, x')\big)\Big) \tag{6.7}$$

Each P_ζ is irreducible and aperiodic under the assumption that this is true for P_0.

Myopic design and the exponential family. A simple choice is the *myopic design*. This is obtained by setting $h_\zeta(x, x') = \zeta\mathcal{U}(x')$,

$$P_\zeta^{myop}(x, x') := P_0(x, x') \exp\big(\zeta\mathcal{U}(x') - \Lambda_\zeta(x)\big) \tag{6.8}$$

with the normalizing constant $\Lambda_\zeta(x) := \log\Big(\sum_{x'} P_0(x, x') \exp\big(\zeta\mathcal{U}(x')\big)\Big)$. This corresponds to a tilted probability transition matrix, favoring the transitions to states with lower power consumption when $\zeta < 0$, and to states with higher power consumption when $\zeta > 0$.

Advantages of this design include ease of implementation, and the straightforward generalization to the continuous state space case. This generalization will be illustrated in Section 6.4.

It is possible to consider any other family of functions, linear with respect to ζ, leading to an exponential family for $\{P_\zeta : \zeta \in \mathbb{R}\}$,

$$h_\zeta(x, x') = \zeta H_0(x, x'). \tag{6.9}$$

The choice of H_0 will typically correspond to the linearization of a more advanced design around the value $\zeta = 0$ (or some other fixed value of ζ). One example is given in Section 6.3.3.

Individual Perspective Design. Consider a finite-time-horizon optimization problem: For a given terminal time T, let p_0 denote the pmf on strings of length T:

$$p_0(x_1, \dots, x_T) = \prod_{i=0}^{T-1} P_0(x_i, x_{i+1}) \,,$$

where $x_0 \in \mathsf{X}$ is assumed to be given. The scalar $\zeta \in \mathbb{R}$ is interpreted as a weighting parameter in the following definition of total welfare. For any pmf p, this is defined as the weighted difference,

$$\mathcal{W}_T(p) = \zeta \mathsf{E}_p\Big[\sum_{t=1}^{T} \mathcal{U}(X_t)\Big] - D(p\|p_0) \tag{6.10}$$

where the expectation is with respect to p, and D denotes relative entropy:

$$D(p\|p_0) := \sum_{x_1,\dots,x_T} \log\Big(\frac{p(x_1, \dots, x_T)}{p_0(x_1, \dots, x_T)}\Big) p(x_1, \dots, x_T)$$

It is easy to check that the myopic design is an optimizer for the horizon $T = 1$,

$$P_\zeta^{myop}(x_0, \cdot) \in \arg\max_p \mathcal{W}_1(p).$$

The infinite-horizon mean welfare is denoted,

$$\eta_\zeta^* = \lim_{T\to\infty} \frac{1}{T} \mathcal{W}_T(p_T^*) \tag{6.11}$$

The two terms in the welfare function (6.10) represent the two conflicting goals: To provide service to the grid and to reduce deviation of the load's behavior from the nominal. If the controlled probability p is chosen to be different from p_0, it potentially reduces the QoS to the consumer, which is modeled by the term "$-D(p\|p_0)$."

Recall that $\mathcal{U}(X_t)$ is equal to the power consumption of the load at time t. If the grid operator desires lower power demand than the nominal value, this goal is modeled through the first term in (6.10) whenever the parameter ζ is negative.

A solution to the infinite horizon problem is given by a time-homogenous Markov chain whose transition matrix is obtained following the solution of an eigenvector problem, based on the $d \times d$ matrix,

$$\widehat{P}(x, x') = \exp(\zeta \mathcal{U}(x)) P_0(x, x') \,, \qquad x, x' \in \mathsf{X} \,. \tag{6.12}$$

Let $\lambda > 0$ denote the Perron-Frobenious eigenvalue, and v the eigenvector with non-negative entries satisfying,

$$\widehat{P} v = \lambda v \tag{6.13}$$

The proof of Prop. 6.3.2 is contained in [107, Section II], following [132].

Proposition 6.3.2. *If P_0 is irreducible, an optimizing p^* that achieves* (6.11) *is defined by a time-homogeneous Markov chain whose transition probability is defined by,*

$$\check{P}_\zeta(x, x') = \frac{1}{\lambda} \frac{1}{v(x)} \widehat{P}(x, x') v(x') \,, \qquad x, x' \in \mathsf{X} \,. \tag{6.14}$$

6.3.3 Uncontrolled dynamics

In many cases it is not possible to apply the IPD solution in the form (6.14) because a portion of the stochastic dynamics are not directly controllable. Consider a load model in which the full state space is the Cartesian product $\mathsf{X} = \mathsf{X}^{\mathsf{u}} \times \mathsf{X}^{\mathsf{n}}$, where X^{u} are components of the state that can be directly manipulated through control.

In [27, 26], the following conditional-independence structure is assumed: for each state $x = (x_u, x_n)$, and each $\zeta \in \mathbb{R}$,

$$\begin{aligned} \check{P}_\zeta(x, x') &= R_\zeta(x, x'_u) Q_0(x, x'_n)\,, \\ R_\zeta(x, x'_u) &= R_0(x, x') \exp\bigl(h_\zeta(x, x'_u) - \Lambda_{h_\zeta}(x)\bigr) \end{aligned} \tag{6.15}$$

where $\sum_{x'_u} R_\zeta(x, x'_u) = \sum_{x'_n} Q_0(x, x'_n) = 1$ for each x and ζ. The matrix Q_0 is out of our control – this models load dynamics and exogenous disturbances. For example, it may be used to model the impact of the weather on the climate of a building. The matrices $\{R_\zeta\}$ are a product of design.

It is reasonable to assume that $\mathcal{U}$ is a function only of X^{u}; the power state is directly controllable. In this case the myopic design (6.8) is unchanged, $h_\zeta(x, x'_u) = \zeta \mathcal{U}(x'_u)$.

The formulation of the IPD optimization problem is unchanged, but its solution is not in the form (6.14). A computational ODE approach is introduced in [27, 26] and will be presented more in detail in Chapter 7: for a vector field $\mathcal{V}$ whose domain and range are functions on $\mathsf{X} \times \mathsf{X}^{\mathsf{u}}$,

$$\frac{d}{d\zeta} h_\zeta = \mathcal{V}(h_\zeta)\,, \quad \zeta \in \mathbb{R}\,, \qquad h_0 \equiv 1\,.$$

Besides its computational value, this approach provides a useful alternative to the myopic design. The function $H_0 = \mathcal{V}(h_0)$ can be used in the exponential family design (6.9). It is shown in [27] that this function is a solution to Poisson's equation for the nominal model: $P_0 H_0 = H_0 - \widetilde{\mathcal{U}}_0$.

Motivation for the IPD design or its exponential family approximation is in part empirical. In nearly every numerical experiment conducted to-date, it is found that the resulting input-output mean field model appears nearly linear over a large range of ζ, and also minimum phase. Moreover, in nearly all cases the linearization (6.4) is *passive* when the delay is removed. That is, the transfer function $zC[Iz - A]^{-1}B$ is strictly positive real.

Passivity can be established mathematically for a restricted class of models [25], or using a different ODE called the system perspective design (SPD) [26].

6.3.4 Quality of service and opt-out

In analysis of QoS it is convenient to consider a steady-state setting: the state process for each load is assumed to be a stationary process on the two-sided time interval. It is also useful to consider a functional form for QoS – the following conventions were introduced in [38].

Several QoS metrics may be considered simultaneously, but each are assumed to be of the following form. Assumed given is a function $\ell\colon \mathsf{X} \to \mathbb{R}$, defined so that $L_t^i := \ell(X_t^i)$ describes a "snap-shot" indication of QoS for the ith load at time t. The function ℓ may represent the temperature of a TCL, cycling of an on/off load, or power consumption as a function of $x \in \mathsf{X}$.

Second is a stable transfer function denoted $H_{\mathcal{L}}$. The QoS of the ith load at time t is defined by passing $\boldsymbol{L}^i$ through the transfer function $H_{\mathcal{L}}$. Two classes of transfer functions $H_{\mathcal{L}}$ are considered:

(i) Summation over a finite time horizon T_f:

$$\mathcal{L}_t^i = \sum_{k=0}^{T_f} \ell(X_{t-k}^i). \tag{6.16}$$

(ii) Discounted sum, with discount factor $\beta \in [0, 1)$:

$$\mathcal{L}_t^i = \sum_{k=0}^{\infty} \beta^k \ell(X_{t-k}^i)\,. \tag{6.17}$$

When β is close to unity, or T_f is very large, then these QoS metrics can be approximated by Gaussian random variable by appealing to the Central Limit Theorem [38]. A Gaussian distribution indicates that QoS for some individuals in the population will sometimes take on unacceptable values.

QoS can be constrained by imposing an additional layer of control at each load. A simple mechanism is *opt-out control*.

The opt-out mechanism is based on pre-defined upper and lower limits, denoted b_+ and b_-. A load ignores a command from grid operator if it will result in $\mathcal{L}_{t+1}^i \notin [b_-, b_+]$, and takes an alternative action so that $\mathcal{L}_{t+1}^i \in [b_-, b_+]$. This ensures that the QoS metric of each load remains within the predefined interval for all time.

Numerical examples are presented in [38] for both residential pools and TCLs.

6.4 Example: Thermostatically Controlled Loads

This special case is dominant in much of the literature on demand dispatch. Examples of thermostatically controlled loads (TCLs) include refrigerators, water heaters and air-conditioning. Each of these loads is already equipped with primitive "local intelligence" based on a *deadband* (or *hysteresis interval*): there is a sensor that measures the temperature of the unit, and turns the power on when the measured value reaches one end of this deadband.

The state process for a TCL at time t will be of the form

$$X(t) = (X_u(t), X_n(t)) = (m(t), \Theta(t))\,, \tag{6.18}$$

in which $m(t) \in \{0, 1\}$ denotes the power mode (the value "1" indicating the unit is on), and $\Theta(t)$ the inside temperature of the load. Exogenous disturbances that directly influence $\boldsymbol{\Theta}$ include ambient temperature, and usage: the inside temperature of a refrigerator is impacted by an open door, and the temperature of water in a water heater is influenced by the rate of flow of water out of the unit.

The remainder of this section is restricted to a residential water heater (WH). This will simplify discussion, and extensions to other TCLs are often straightforward.

Nominal model. The standard ODE model of a water heater is the first-order linear system:

$$\tfrac{d}{dt}\Theta(t) = -\lambda[\Theta(t) - \Theta^a(t)] + \gamma m(t) - \alpha[\Theta(t) - \Theta^{in}(t)]f(t)\,, \tag{6.19}$$

for constants $(\lambda, \gamma, \alpha)$, in which $\Theta(t)$ is the temperature of the water in the tank, $\Theta^a(t)$ is ambient temperature, $\Theta^{in}(t)$ is temperature of the cold water entering the tank (degrees Fahrenheit), $f(t)$ is flow rate of hot water from the WH (gallons/s), and $m(t)$ is the power mode of the WH ("on" indicated by $m(t) = 1$). The corresponding power consumed by a WH when $m(t) = 1$ is denoted P_{on}.

The upper and lower temperature limits that define the deadband are denoted Θ_-, Θ_+, respectively. A standard residential water heater in the U.S. has the following typical behavior: At the moment that $\Theta(t)$ reaches the lower limit Θ_-, the unit turns on, and remains on until the time t_+ at which $\Theta(t_+) = \Theta_+$. The unit then turns off and begins to cool. It may take 6 hours to once again reach the lower limit, while the time to heat the water is much shorter.

The nominal model used for local control design is based on an approximation of this typical behavior, in which with some probability the unit turns on before $\Theta(t)$ reaches Θ_-, and the unit may also turn off before reaching the maximum temperature Θ_+. The definition of the nominal model is based on the specification of two cumulative distribution functions (CDFs) for the temperature at which the load turns on or turns off, denoted $F^{\oplus}$ and $F^{\ominus}$. Random variables with these CDFs are denoted $\widetilde{\Theta}^{\oplus}$ and $\widetilde{\Theta}^{\ominus}$, so that

$$F^{\oplus}(\theta) = \mathsf{P}\{\widetilde{\Theta}^{\oplus} \le \theta\}, \quad F^{\ominus}(\theta) = \mathsf{P}\{\widetilde{\Theta}^{\ominus} \le \theta\}\,, \quad \theta \in \mathbb{R}\,.$$

It is always assumed that $\widetilde{\Theta}^{\oplus}$ and $\widetilde{\Theta}^{\ominus}$ take values in the interval $[\Theta_-, \Theta_+]$, which implies that $F^{\oplus}(\theta) = F^{\ominus}(\theta) = 1$ for $\theta \ge \Theta_+$ and $F^{\oplus}(\theta) = F^{\ominus}(\theta) = 0$ for $\theta < \Theta_-$.

A particular design for $F^{\ominus}$ is obtained on fixing three parameters $\theta_0^{\ominus} \in [\Theta_-, \Theta_+]$, and constants $\varrho \in (0, 1)$ and $\kappa > 1$:

$$F^{\ominus}(\theta) = (1 - \varrho)\frac{[\theta - \theta_0^{\ominus}]_+^{\kappa}}{[\Theta_+ - \theta_0^{\ominus}]^{\kappa}}\,, \qquad \theta \in [\Theta_-, \Theta_+)\,,$$

where $[x]_+ := \max(0, x)$ for $x \in \mathbb{R}$. In a symmetric model, the other CDF is defined by the transformation,

$$F^{\oplus}(\theta) = 1 - \lim_{\theta' \downarrow \theta} F^{\ominus}(\Theta_+ + \Theta_- - \theta')$$

Fig. 6.3 illustrates a particular special case of the symmetric model.

It is assumed that the local control operates in discrete-time. By choice of time-units, without loss of generality it is assumed that the sampling interval is 1 unit. At time instance k, if the water heater is on (i.e., $m(k) = 1$), then it turns off at time $k + 1$ with probability,

$$p^{\ominus}(k+1) = \frac{[F^{\ominus}(\Theta(k+1)) - F^{\ominus}(\Theta(k))]_+}{1 - F^{\ominus}(\Theta(k))}$$

If $\Theta(k+1) \le \Theta(k)$, then this probability is zero. Similarly, if the load is off, then it turns on with probability

$$p^{\oplus}(k+1) = \frac{[F^{\oplus}(\Theta(k)) - F^{\oplus}(\Theta(k+1))]_+}{F^{\oplus}(\Theta(k))}$$

The nominal behavior of the power mode can be expressed

$$\begin{aligned}\mathsf{P}\{m(k)=1 \mid \theta(k-1),\theta(k),m(k-1)=0\} &= p^{\oplus}(k)\\ \mathsf{P}\{m(k)=0 \mid \theta(k-1),\theta(k),m(k-1)=1\} &= p^{\ominus}(k)\end{aligned} \tag{6.20}$$

The IPD and SPD designs were obtained in [26] based on a similar nominal model for a residential refrigerator.

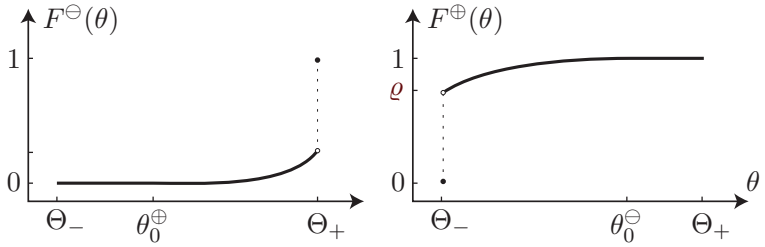


Figure 6.3: Nominal model for a water heater

The myopic design (7.11) is obtained through an exponential tilting:

$$p_\zeta^{\oplus}(k) := \frac{p^{\oplus}(k)e^{\zeta}}{p^{\oplus}(k)e^{\zeta}+1-p^{\oplus}(k)}\,, \qquad p_\zeta^{\ominus}(k) := \frac{p^{\ominus}(k)}{p^{\ominus}(k)+(1-p^{\ominus}(k))e^{\zeta}}$$

If $p^{\oplus}(k) > 0$, then the probability $p_\zeta^{\oplus}(k)$ is strictly increasing in ζ, approaching 1 as $\zeta \to \infty$; it approaches 0 as $\zeta \to -\infty$, provided $p^{\oplus}(k) < 1$.

System identification. Power, temperature, and usage data from residential water heaters was obtained through our partners at ORNL.[2] The constants $(\lambda, \gamma, \alpha)$ were estimated using least squares. The parameter values listed in Table 6.1 reflect the range of values observed in actual data.

A testbed was created to simulate a collection of $N = 100,000$ water heaters with usage. Each evolves according to the ODE (6.19), but parameters were different for each of the N loads: parameters were chosen via uniform sampling of the values in Table 6.1. A simulation model for usage at each load was created, based on sampling from historical usage of actual water heaters.

The mean-field model is a nonlinear input-output system with input $\boldsymbol{\zeta}$ and output equal to power deviation, $\boldsymbol{y}$. An approximate linear model was obtained through least squares, in which the input $\boldsymbol{\zeta}$ was taken to be the swept-sine: $\zeta(t) = 1.5\sin(10^{-7}t^2)$ for $0 \le t \le 432\times10^5$ sec. (5 days). Fig. 6.4 shows results from the estimation experiment for two different model orders. The Bode plots shown represent the approximate model in continuous time. The 5th order model predicts that the gain of the linearization vanishes as the frequency tends to zero (DC). This is a physical reality for this example.

The linearization is minimum phase and stable. Its gain is approximately constant in the frequency range $[5\times10^{-4}, 10^{-2}]$ rad/s. It is expected that a collection of water heaters can accurately track signals in this frequency range.

[2]Water heater data provided by Ecotope, Inc., with funding from the Northwest Energy Efficiency Alliance (NEEA) and the Bonneville Power Administration (BPA).

Temp. Ranges	ODE Pars.	Loc. Control
$\Theta_+ \in [118, 122]$ F	$\lambda \in [8, 12.5] \times 10^{-6}$	$T_s = 15$ sec
$\Theta_- \in [108, 112]$ F	$\gamma \in [2.6, 2.8] \times 10^{-2}$	$\kappa = 4$
$\Theta^a \in [68, 72]$ F	$\alpha \in [6.5, 6.7] \times 10^{-2}$	$\varrho = 0.8$
$\Theta^{in} \in [68, 72]$ F	$P_{\text{on}} = 4.5$ kW	$\theta_0 = \Theta_-$

Table 6.1: Parameters for nominal model for water heaters.

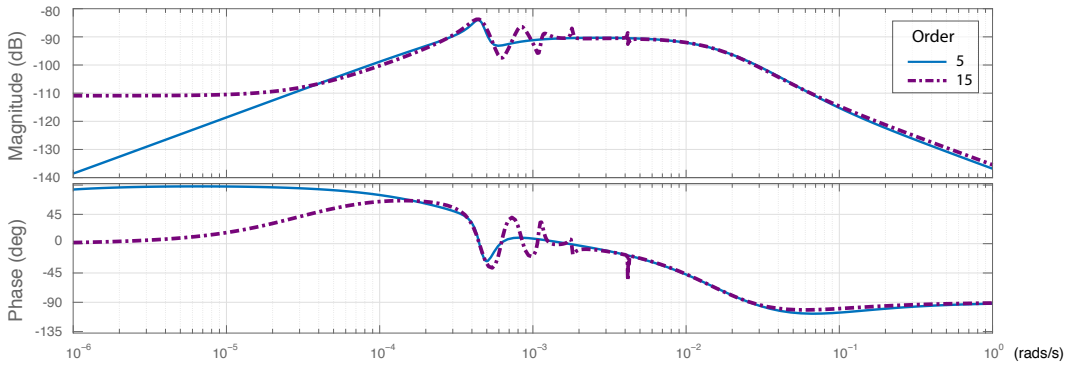


Figure 6.4: Least square estimates of the transfer function for water heaters.

Tracking. Design at the macro level is most easily performed for a model in continuous time. A PI controller $G_c(s) = K_P + K_I/s$ was designed based on the linearized mean-field model. The values $K_P = 10^5$ and $K_I = 500$ result in a crossover frequency $\omega_c = 0.03$ rad/s (corresponding to a time period of approximately 3.5 minutes), with a 75° phase margin.

The balancing reserves signal from the Bonneville Power Administration (BPA) was used in the tracking experiments described in this section. A typical windy day, February 19, 2016, was chosen for the experiments described here. The signal was filtered using a second-order Butterworth high pass filter with a cut-off frequency at 8×10^{-4} rad/s (corresponding to a sine wave with period of approximately 2 hours).

Fig. 6.5 shows results from several numerical experiments. The three rows are differentiated by the regulation signal: In the first row $\boldsymbol{r} \equiv 0$, in the second the absolute value of the regulation signal takes a maximum value of about 8 MW, and in the final row the prior regulation signal was multiplied by 4. Exact tracking is not feasible over the entire period for the largest regulation signal (results shown in the bottom left plot), but the performance remains nearly perfect over time periods for which $|r_t|$ does not exceed about 90% of the nominal power consumption.

The second column shows evolution of temperature and the power mode for a typical load in the three cases. The seed for the random number generator was identical in each of the three experiments. It is amazing to see that the evolution of temperature and power mode is hardly impacted by local control.

These loads are equally valuable for contingency and ramping services. Fig. 6.6 shows recent results that illustrate the potential. In these experiments the water flow was set to zero; in this case, the nominal power consumption for 100,000 loads is approximately 8 MW. Each plot is a particular saw-tooth wave, scaled to reach the maximum lower limit of −8 MW.

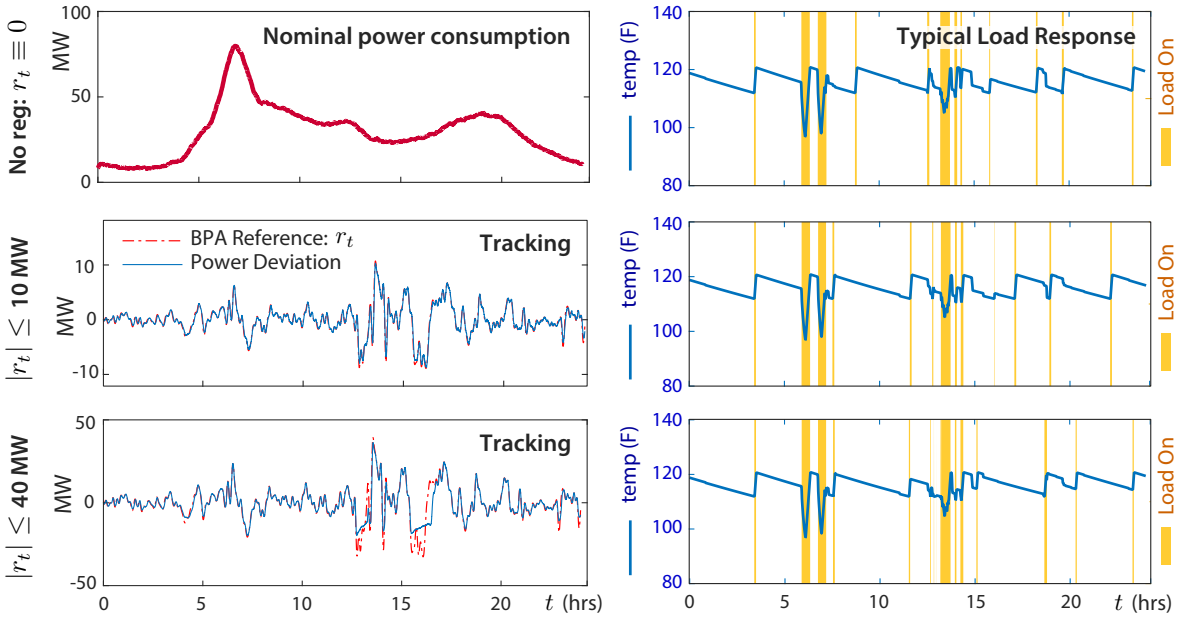


Figure 6.5: Tracking results with 100,000 water heaters, and the behavior of a single water heater in three cases, distinguished by the reference signal $\boldsymbol{r}$. The morning peak in nominal power consumption is consistent with typical water usage included in the simulation experiments.

6.5 Discussion and related results

With appropriate filtering and local control, loads can provide excellent grid services without two-way communication. While there is some cost to install hardware on appliances that can receive a signal from a balancing authority, in the long run this will be far less costly than batteries.

The numerical results presented in this chapter, in particular the tracking results illustrated in Figures 6.5, 6.6, show that VES working in conjunction with traditional resources can provide balancing services, ramping services and contingency reserves simultaneously.

Future research questions include:

(i) The application of reinforcement learning may be valuable for learning the local control law.

(ii) Further research is required to better estimate capacity in terms of both energy and power.

(iii) The impact of usage is not entirely understood. Numerical results presented in Section 6.4 suggest that this is not an obstacle in the case of water heaters. Air-conditioning is a greater challenge because variations in load are much greater.

(iv) A question posed in [100]: Does the load need to receive a signal from the BA? It is possible that some VES resources can provide valuable services using only local measurements. Frequency (as well as voltage) measurements can be obtained inexpensively at loads, and these measurements are similar to those used by the BA to construct analogs of our "$\boldsymbol{\zeta}$"

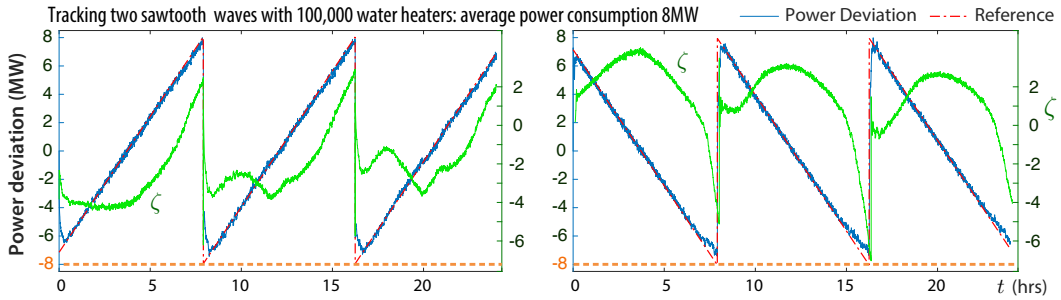


Figure 6.6: Tracking a pair of saw tooth waves with 10^5 water heaters.

today. The advantage of distributed control is reduced cost due to reduced communication between a BA and loads.

Some history and further reading. In the early eighties, Schweppe wrote a series of influential articles on the value of new architectures for the grid [124, 123], with an emphasis on demand response based on either automation or prices. Tools for analysis were lacking at the time, but many researchers came to fill the void. An influential example is the paper [98], that introduced ideas from statistical mechanics to model a large population of thermostatically controlled loads (TCLs).

There is substantial literature on indirect load control, where customers are encouraged to shift their electricity usage in response to real-time prices. Dynamic prices can introduce uncertain dynamics, such as cyclical price fluctuations and increased sensitivity to exogenous disturbances, and present a risk to system stability [121, 32, 43].

Randomization is an essential element of the distributed control architecture described in this chapter. Its value has been widely recognized in academia as well as industry [126].

On the academic side, Matheiu's dissertation [102] and references [103, 104] were highly influential, motivating in part the research surveyed in this chapter and others [44, 88, 135, 142]. The control model in [102] is based on the mean-field setting of [98], with the introduction of a control signal from a central authority: at each time slot, a BA or aggregator broadcasts probability values $\{p_\tau^\oplus, p_\tau^\ominus : \tau \in \mathbb{R}\}$ where $p_\tau^\oplus$ ($p_\tau^\ominus$) denotes the probability of turning the device on (off) when the temperature of the device is τ. The temperatures are binned to obtain a finite state-space aggregate model. This model is bilinear and partially observed, where the state x is the histogram of load temperature and power consumption. The bilinear control system is transformed to a linear model by defining products of probability and state as an input. The resulting linear state space model has the same state, but the vector-valued input is now defined as products of the form $u_k = p_\tau^m x_j$ for some $\tau(k)$, $j(k)$, and $m(k) \in \{\oplus, \ominus\}$. Feedback control design is performed based on LQR. However, it is still necessary to recover the probability vector $\{p_\tau^m\}$. In this prior work, this is defined as the ratio of components of the input $u(t)$, and components of the *estimate* of the state at time t (see e.g. eq. (11) of [104]). It is assumed that estimates are computed by the BA based on measurements of aggregate power consumption. A current challenge with this approach is the creation of sufficiently accurate state estimates for an inherently infinite-dimensional system. Challenges to state estimation are discussed in [39], where it is shown that the linearized mean field model may not be observable. Robustness of this approach to bilinear control systems is another an

important area for future research.

The approach to distributed control surveyed in Sections 6.2 and 6.3 involves an entirely different approach to local control at each load. One example is the *Individual Perspective Design* (IPD) described in Section 6.3.2. This can be regarded as an application of the MDP technique of Todorov [132], but only in one special case: the construction of [132] can be applied only if there is no exogenous stochastic disturbance in the load model. Contained in Section 6.3.2 are techniques to extend this design to a broader class of load models. These ideas were first applied to demand-dispatch in [108], and have seen many extensions since. For more history the reader is referred to [107, 42], in addition to the papers surveyed in Section 6.3.2. While beyond the scope of this article, it is important to note that Todorov's 'linearly solvable' MDP model [132] is similar to prior work such as [81], and the form of the solution could have been anticipated from well-known results in the theory of large-deviations for Markov chains [27]. It is pointed out in [133] that this approach has a long history in the context of controlled stochastic differential equations [57].

Chapter 7

ODE method for Markov decision processes

This chapter concerns computation of optimal policies in which the one-step reward function contains a cost term that models Kullback-Leibler divergence with respect to nominal dynamics. This technique was introduced by Todorov in [132], where it was shown under general conditions that the solution to the average-reward optimality equations reduce to a simple eigenvector problem. Since then many authors have sought to apply this technique to control problems and models of bounded rationality in economics.

A crucial assumption is that the input process is essentially unconstrained. For example, if the nominal dynamics include randomness from nature (e.g., the impact of wind on a moving vehicle), then the optimal control solution does not respect the exogenous nature of this disturbance.

We introduce a technique to solve a more general class of action-constrained MDPs. The main idea is to solve an entire parameterized family of MDPs, in which the parameter is a scalar weighting the one-step reward function.

This chapter is based on the publications [J10, C33], that contain more details and the proofs of the results summarized in this chapter.

7.1 Introduction

Consider a Markov Decision Process (MDP) with finite state space X, general action space U, and one-step reward function $w\colon \mathsf{X} \times \mathsf{U} \to \mathbb{R}$. Two standard optimal control criteria are *finite-horizon*:

$$\mathcal{W}_T^*(x) = \max \sum_{t=0}^{T} \mathsf{E}[w(X(t), U(t)) \mid X(0) = x] \tag{7.1}$$

where $T \geq 0$ is fixed, and *average reward*:

$$\eta^*(x) = \max \Big\{ \liminf_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathsf{E}[w(X(t), U(t)) \mid X(0) = x] \Big\} . \tag{7.2}$$

where $\boldsymbol{X} = \{X(t) : t \geq 0\}$, $\boldsymbol{U} = \{U(t) : t \geq 0\}$ denote the state and input sequences.

In either case, the maximum is over all admissible input sequences; it is obtained as deterministic state feedback under general conditions. In the average-reward framework the

optimal policy is typically stationary: $U(t) = \phi^*(X(t))$ for a mapping $\phi^* \colon \mathsf{X} \to \mathsf{U}$, and $\eta^*(x)$ does not depend upon the initial condition x (see [119, 15]).

A special class of MDP models was introduced by [132], for which either optimal control problem has an attractive solution. The reward function is assumed to be the sum of two terms:

$$w(x, \mu) = \mathcal{U}(x) - D(\mu \| P_0(x, \cdot)).$$

The first term is a function $\mathcal{U} \colon \mathsf{X} \to \mathbb{R}$ that is completely unstructured. The second is a "control cost", defined using Kullback–Leibler (K-L) divergence (also known as *relative entropy*). The control cost is based on deviation from nominal (control-free) behavior; modeled by a nominal transition matrix P_0:

$$D(\mu \| P_0(x, \cdot)) := \sum_{x'} \mu(x') \log\Big(\frac{\mu(x')}{P_0(x, x')}\Big).$$

It is shown that the solution with respect to the average reward criterion is obtained as the solution to the following eigenvector problem: let (λ, v) denote the Perron-Frobenius eigenvalue-eigenvector pair for the positive matrix with entries $\widehat{P}(x, x') = \exp(\mathcal{U}(x)) P_0(x, x')$, $x, x' \in \mathsf{X}$. The eigenvector property $\widehat{P} v = \lambda v$ implies that the "twisted" matrix

$$\check{P}(x, x') = \frac{1}{\lambda} \frac{v(x')}{v(x)} \widehat{P}(x, x'), \quad x, x' \in \mathsf{X}. \tag{7.3}$$

is a transition matrix on X. This transition matrix defines the dynamics of the model under optimal control. A similar model was introduced in the earlier work of [81], but without the complete solution reviewed here.

Since the publication of [132] there has been significant theoretical advancement, with proposed applications to economics [64], distributed control [107], and neuroscience [53].

It is appealing to imagine that rational economic agents are solving an eigenvector problem to maximize their utility. However, a careful look at the controlled dynamics (7.3) suggests a limitation of this MDP formulation: *how can this transformation respect exogenous disturbances from nature?* An essential assumption in this prior work is that for each x, and any pmf μ, it is possible to choose the action so that $P(x, x') = \mu(x')$. This is equivalent to the assumption that the action space U consists of all probability mass functions on X, and the controlled transition matrix is entirely determined by the input as follows:

$$\mathsf{P}\{X(t+1) = x' \mid X(t) = x, U(t) = \mu\} = \mu(x'), \qquad x, x' \in \mathsf{X},\ \mu \in \mathsf{U}. \tag{7.4}$$

This modeling assumption presents a significant limitation, as pointed out in [133]: "*It prevents us from modeling systems subject to disturbances outside the actuation space*". The optimal solution cannot take the form (7.3) when this additional randomness is included in the model, since this would mean our control action would modify the weather.

Contributions The main contribution is broadening the K-L cost framework to include constraints on the pmf μ appearing in (7.4). The new approach to computation is based on the solution of an entire family of MDP problems, parameterized by a scalar ζ appearing as a weighting factor in the one-step reward function. Letting X_t denote the state, and R_t denote the randomized policy at time t, this one-step reward is of the form

$$w(X_t, R_t) = \zeta \mathcal{U}(X_t) - c_{\text{KL}}(X_t, R_t) \tag{7.5}$$

in which c_{KL} denotes relative entropy with respect to nominal dynamics (see (7.14)).

The main results are contained in Theorems 7.3.1 and 7.4.2, with parallel results for the total- and average-reward control problems. In each case, it is shown that *the solution to an entire family of MDPs can be obtained through the solution of a single ordinary differential equation* (ODE).

The ODE solution is most elegant in the average-reward setting. For each ζ, the solution to the average-reward optimization problem is based on a relative value function $h^*_\zeta \colon \mathsf{X} \to \mathbb{R}$. For the MDP with d states, each function is viewed as a vector in $\mathbb{R}^d$ with entries $\{h^*_\zeta(x^i) : 1 \le i \le d\}$. A vector field $\mathcal{V} \colon \mathbb{R}^d \to \mathbb{R}^d$ is constructed so that these functions solve the ODE

$$\tfrac{d}{d\zeta} h^*_\zeta = \mathcal{V}(h^*_\zeta)\,, \qquad \text{with boundary condition } h^*_0 \equiv 0.$$

One step in the construction of $\mathcal{V}$ is differentiating each side of the dynamic programming equations; a starting point of the 50 year old sensitivity theory of [122], and more recent [128]. More closely related is the sensitivity theory surrounding Perron-Frobenius eigenvectors that appears in the theory of large deviations [89, Prop. 4.9]. The goals of this prior work are different, and we are not aware of comparable algorithms that simultaneously solve the family of control problems.

The optimal control formulation is far more general than in the aforementioned work [132, 64, 107], as it allows for inclusion of exogenous randomness in the MDP model. The dynamic programming equations become significantly more complex in this generality, so that in particular, the Perron-Frobenious computational approach used in prior work is no longer applicable.

In addition to its value as a computational tool, there is a significant benefit to solve the entire collection of optimal control problems for a range of the parameter ζ. For example, this provides a means to understand the tradeoff between state cost and control effort. Simultaneous computation of the optimal policies is also an essential ingredient of the distributed control architecture introduced in [107] and that was presented in Section 6.3.3.

The remainder of the chapter is organized as follows. Section 7.2 describes the new Kullback–Leibler cost criterion. Section 7.3 provides an overview of numerical techniques for the MDP solutions for finite time horizon and Section 7.4 for average cost criterion. Conclusions and future discussion are contained in Section 7.5.

7.2 MDPs with Kullback–Leibler cost

7.2.1 MDP model

The dynamics of the MDP are assumed of the form (7.4), where the action space consists of a convex subset of probability mass functions (pmf) on X. An explanation of the one-step reward (7.5) will be provided after a few preliminaries.

A transition matrix P_0 is given that describes nominal (control-free) behavior. It is assumed to be *irreducible and aperiodic*. It follows that P_0 admits a unique invariant pmf, denoted π_0. For any other transition matrix, with unique invariant pmf π, the *Donsker-Varadhan rate function* is denoted,

$$K(P\|P_0) = \sum_{x,x'} \pi(x)P(x,x') \log\Big(\frac{P(x,x')}{P_0(x,x')}\Big) \tag{7.6}$$

under the usual convention that "$0\log(0) = 0$". It is called a "rate function" because it defines the relative entropy rate between two stationary Markov chains, see [51].

As in [132, 64, 107], the rate function is used here to model the cost of deviation from the nominal transition matrix P_0. The two control objectives surveyed in the introduction will be specialized as follows, based on the utility function $\mathcal{U}\colon \mathsf{X} \to \mathbb{R}$ and a scaling parameter $\zeta \geq 0$. For the finite-horizon optimal control problem,

$$\mathcal{W}_T^*(x,\zeta) = \max \sum_{t=0}^{T} \mathsf{E}_x[\zeta \mathcal{U}(X(t)) - c_{\mathrm{KL}}(X(t), P_t)]\,, \tag{7.7}$$

where the expectation is conditional on $X(0) = x$, and

$$c_{\mathrm{KL}}(x,P) = D(P(x,\,\cdot\,)\|P_0(x,\,\cdot\,)) := \sum_{x'} P(x,x') \log\Big(\frac{P(x,x')}{P_0(x,x')}\Big)$$

for any $x \in \mathsf{X}$ and transition matrix P.

The average reward optimization problem is analogous:

$$\eta^*(\zeta) = \max\Big(\liminf_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathsf{E}_x\left[\zeta \mathcal{U}(X(t)) - c_{\mathrm{KL}}(X(t), P_t)\right]\Big)\,. \tag{7.8}$$

In each case, the maximum is over all transition matrices $\{P_t\}$. The average reward optimization problem can be cast as the solution to the convex optimization problem,

$$\eta_\zeta^* = \max_{\pi,P} \big\{\zeta\pi(\mathcal{U}) - K(P\|P_0) : \pi P = \pi\big\} \tag{7.9}$$

where the maximum is over all transition matrices.

In this context, the one-step reward appearing in (7.1, 7.2) is a function of pairs (x, P):

$$w(x,P) := \zeta \mathcal{U}(x) - c_{\mathrm{KL}}(x,P) \tag{7.10}$$

for any $x \in \mathsf{X}$ and transition matrix P. There is practical value to considering a parameterized family of reward functions. For one, it is useful to understand the sensitivity of the control solution to the relative weight given to utility and the penalty on control action. This is well understood in classical linear control theory – consider for example the celebrated symmetric root locus in linear optimal control [59].

Nature & nurture Exogenous randomness from nature imposes additional constraints in the optimal control problem (7.7) or (7.8).

It is assumed that the state space is the cartesian product of two finite sets: $\mathsf{X} = \mathsf{X}^{\mathsf{u}} \times \mathsf{X}^{\mathsf{n}}$, and the state is similarly expressed $X(t) = (X_u(t), X_n(t))$. At a given time t it is assumed that $X_n(t+1)$ is conditionally independent of the input at time t, given the value of $X(t)$. This is formalized by the following conditional-independence assumption:

$$P(x,x') = R(x,x'_u)Q_0(x,x'_n), \quad x = (x_u, x_n) \in \mathsf{X},\ x'_u \in \mathsf{X}^{\mathsf{u}},\ x'_n \in \mathsf{X}^{\mathsf{n}} \tag{7.11}$$

The matrix R defines the randomized decision rule for $X_u(t+1)$ given $X(t)$. The matrix Q_0 is fixed and models the distribution of $X_n(t+1)$ given $X(t) = x$, and each are subject to the pmf constraint: $\sum_{x'_u} R(x,x'_u) = \sum_{x'_n} Q_0(x,x'_n) = 1$ for each x.

Subject to the constraint (7.11), the two optimal control problems (7.8, 7.10) are transformed to the final forms considered here:

$$\mathcal{W}_T^*(x,\zeta) = \max \sum_{t=0}^{T} \mathsf{E}_x[w(X(t),R(t))] \tag{7.12}$$

$$\eta^*(\zeta) = \max\Big\{\liminf_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} \mathsf{E}_x[w(X(t),R(t))]\Big\} \tag{7.13}$$

where in each case the maximum is over sequences of randomized decision rules $\{R(0),\dots,R(T)\}$,

$$\begin{aligned} w(x,R) &:= \zeta\mathcal{U}(x) - c_{\mathrm{KL}}(x,R) \\ \textit{and}\quad c_{\mathrm{KL}}(x,R) &:= \sum_{x'} P(x,x')\log\Big(\frac{P(x,x')}{P_0(x,x')}\Big) = \sum_{x'_u} R(x,x'_u)\log\Big(\frac{R(x,x'_u)}{R_0(x,x'_u)}\Big) \end{aligned} \tag{7.14}$$

7.2.2 Notation

For any transition matrix P, an invariant pmf is interpreted as a row vector, so that invariance can be expressed $\pi P = \pi$. Any function $f\colon \mathsf{X}\to\mathbb{R}$ is interpreted as a d-dimensional column vector, and we use the standard notation $Pf\,(x) = \sum_{x'} P(x,x')f(x')$, $x\in\mathsf{X}$. The *fundamental matrix* is the inverse,

$$Z = [I - P + 1\otimes\pi]^{-1} \tag{7.15}$$

where $1\otimes\pi$ is a matrix in which each row is identical, and equal to π. If P is irreducible and aperiodic, then it can be expressed as the power series $Z = \sum_{n=0}^{\infty}[P - 1\otimes\pi]^n$, with $[P-1\otimes\pi]^0 := I$ (the $d\times d$ identity matrix), and $[P-1\otimes\pi]^n = P^n - 1\otimes\pi$ for $n\ge 1$.

Any function $g\colon \mathsf{X}\times\mathsf{X}\to\mathbb{R}$ is regarded as an unnormalized log-likelihood ratio: Denote for $x,x'\in\mathsf{X}$,

$$P_g(x,x') := P_0(x,x')\exp\big(g(x'\mid x) - \Lambda_g(x)\big), \tag{7.16}$$

in which $g(x'\mid x)$ is the value of g at $(x,x')\in\mathsf{X}\times\mathsf{X}$, and $\Lambda_g(x)$ is the normalization constant,

$$\Lambda_g(x) := \log\Big(\sum_{x'} P_0(x,x')\exp\big(g(x'\mid x)\big)\Big) \tag{7.17}$$

The rate function can be expressed in terms of its invariant pmf π_g, the bivariate pmf $\Pi_g(x,x') = \pi_g(x)P_g(x,x')$, and the log moment generating function (7.17):

$$\begin{aligned} K(P_g\|P_0) &= \sum_{x,x'} \Pi_g(x,x')\big[g(x'\mid x) - \Lambda_g(x)\big] \\ &= \sum_{x,x'} \Pi_g(x,x')g(x'\mid x) - \sum_x \pi_g(x)\Lambda_g(x) \end{aligned} \tag{7.18}$$

The unusual notation is introduced because $g(x'\mid x)$ will take the form of a conditional expectation in all of the results that follow: given any function $h\colon \mathsf{X}\to\mathbb{R}$ we denote

$$h(x'_u\mid x) = \sum_{x'_n} Q_0(x,x'_n)h(x'_u,x'_n)\,. \tag{7.19}$$

In this case the transformation only transforms the dynamics of $\boldsymbol{X}_u$:

$$P_h(x,x') = R_h(x,x'_u)Q_0(x,x'_n)\,,\quad R_h(x,x'_u) := R_0(x,x'_u)\exp\big(h(x'_u\mid x) - \Lambda_g(x)\big)\,.$$

7.3 ODE for finite time horizon

Here an ODE is constructed to compute the value functions $\{\mathcal{W}^*_\tau(x,\zeta) : 1 \le \tau \le T\,, \zeta \ge 0\}$. To aide exposition it is helpful to first look at the general problem: Assume that the state space X is finite, the action space U is *general*, and let $\{P_u(x,x')\}$ denote the controlled transition matrix. The one-step reward on state-action pairs is of the form $w(x,u) = \zeta\mathcal{U}(x) - c(x,u)$, where $c\colon \mathsf{X}\times\mathsf{U} \to \mathbb{R}_+$. Assume that $c(x,u) \equiv 0$ for a unique value $u = u_0$.

For each $1 \le \tau \le T$ denote, as in (7.1),

$$\mathcal{W}^*_\tau(x,\zeta) = \max \sum_{t=0}^{\tau} \mathsf{E}_x[w(X(t),U(t))] \tag{7.20}$$

where the maximum is over all admissible inputs $\{U(t) = \phi_t(X(0),\dots,X(t))\}$. Each value function can be regarded as the maximum over functions $\{\phi_t\}$ (subject to measurability conditions and hard constraints on the input). It is assumed that the maximum (7.20) is finite for each (x,ζ).

The dynamic programming equation (principle of optimality) holds: for $\tau \ge 1$,

$$\mathcal{W}^*_\tau(x,\zeta) = \max_u \Big\{\zeta\mathcal{U}(x) - c(x,u) + \sum_{x'} P_u(x,x')\mathcal{W}^*_{\tau-1}(x')\Big\} \tag{7.21}$$

Assume that a maximizer $\phi^*_{\tau-1,\zeta}(x)$ exits for each τ,ζ, and x.

A crucial observation is that for each x, the value function appearing in (7.20) is the maximum of functions that are affine in ζ. It follows that $\mathcal{W}^*_\tau(x,\zeta)$ is convex as a function of ζ, and hence absolutely continuous. Consequently, the right derivative $H^*_\tau(x,\zeta) := \frac{d^+}{d\zeta}\mathcal{W}^*_\tau(x,\zeta)$ exists everywhere. A recursive equation follows from (7.21):

$$H^*_\tau(x,\zeta) = \mathcal{U}(x) + \sum_{x'} \check{P}_{\tau-1,\zeta}(x,x')H^*_{\tau-1}(x',\zeta) \tag{7.22}$$

where $\check{P}_{\tau-1,\zeta}(x,x') = P_{u^*}(x,x')$ with $u^* = \phi^*_{\tau-1,\zeta}(x)$.

In matrix notation this becomes $H^*_\tau = \check{Z}_{\tau-1,\zeta}\mathcal{U}$, where $\check{Z}_{0,\zeta} = I$, and for any $1 \le \tau \le T$,

$$\check{Z}_{\tau-1,\zeta} = I + \check{P}_{\tau-1,\zeta} + \check{P}_{\tau-1,\zeta}\check{P}_{\tau-2,\zeta} + \check{P}_{\tau-1,\zeta}\check{P}_{\tau-2,\zeta}\cdots\check{P}_{0,\zeta} \tag{7.23}$$

This is similar to a truncation of the power series representation of the fundamental matrix (7.15).

Denote $\mathcal{W}^*_\zeta(x) = \{\mathcal{W}^*_k(x,\zeta) : 0 \le k \le T\}$, regarded as a vector in $\mathbb{R}^{|\mathsf{X}|\times(T+1)}$, parameterized by the non-negative constant ζ. The following result follows from the preceding arguments:

Theorem 7.3.1. *The family of functions $\{\mathcal{W}^*_\zeta\}$ solves the ODE $\frac{d^+}{d\zeta}\mathcal{W}^*_\zeta = \mathcal{V}(\mathcal{W}^*_\zeta)$, $\zeta \ge 0$, with boundary condition $\mathcal{W}^*_0 = 0$. The vector field can be described in block-form as follows, with $T+1$ blocks:*

$$\tfrac{d^+}{d\zeta}\mathcal{W}^*_k(\,\cdot\,,\zeta) = \mathcal{V}_k(\mathcal{W}^*_\zeta)\,, \quad 0 \le k \le T\,.$$

The identity $\mathcal{V}_0(\mathcal{W}) = \mathcal{U}$ holds for any $\mathcal{W}$. For $k \geq 1$, the right hand side depends on its argument only through the associated policy: for any sequence of functions $\mathcal{W} = (\mathcal{W}_0, \dots, \mathcal{W}_T)$,

$$
\begin{aligned}
\mathcal{V}_k(\mathcal{W}) &= Z_{k-1}\mathcal{U} \\
\textit{where} \qquad Z_{k-1} &= I + P_{k-1} + P_{k-1}P_{k-2} + P_{k-1}P_{k-2}\cdots P_0 \\
P_i(x,x') &= P_{\phi_i(x)}(x,x')\,, \quad \textit{all } i,\ x,\ x', \\
\phi_i(x) &= \arg\max_u \Big\{ -c(x,u) + \sum_{x'} P_u(x,x')\mathcal{W}_i(x') \Big\}\,, \qquad 1 \leq i,k \leq T.
\end{aligned}
$$

□

The theorem provides valuable computational tools for models of moderate cardinality and moderate time-horizon. Two questions remain:

(i) What is ϕ_i for the problem under study ?

(ii) Can a tractable ODE be constructed in infinite-horizon optimal control problems?

The answer to the second question is the focus of Section 7.4. The answer to (i) is contained in the following. For any function $\mathcal{W}\colon \mathsf{X} \to \mathbb{R}$, denote

$$
R_{\mathcal{W}}(x,\,\cdot\,) = \arg\max_R \Big\{ w(x,R) + \sum_{x'} P(x,x')\mathcal{W}(x') \Big\}\,, \quad x \in \mathsf{X}\,,
$$

subject to the constraint that P depends on R via (7.11), and with w defined in (7.14).

Proposition 7.3.2. *For any function $\mathcal{W}$ the maximizer $R_{\mathcal{W}}$ is unique and can be expressed*

$$
R_{\mathcal{W}}(x,x'_u) = R_0(x,x'_u)\exp\big(\mathcal{W}(x'_u \mid x) - \Lambda(x)\big)
$$

where $\mathcal{W}(x'_u \mid x) = \sum_{x'_n} Q_0(x,x'_n)\mathcal{W}(x'_u,x'_n)$ for each $x \in \mathsf{X}$, $x'_u \in \mathsf{X}^{\mathsf{u}}$, and $\Lambda(x)$ is a normalizing constant, defined so that $R_{\mathcal{W}}(x,\,\cdot\,)$ is a pmf for each x.

It follows from the proposition that the vector field is smooth in a neighborhood of the optimal solution $\{\mathcal{W}^*_\zeta : \zeta \geq 0\}$. These results are central to the average-reward case considered next.

7.4 Average reward formulation

We consider now the case of average reward (7.13), subject to the structural constraint (7.11). The associated average reward optimization equation (AROE) is expressed as follows:

$$
\max_R \Big\{ w(x,R) + \sum_{x'} P(x,x')h^*_\zeta(x') \Big\} = h^*_\zeta(x) + \eta^*(\zeta) \tag{7.24}
$$

In which $\eta^*(\zeta)$ is the optimal average reward, and h^*_ζ is the *relative value function*. The maximizer defines a transition matrix:

$$
\check{P}_\zeta = \arg\max_P \big\{ \zeta\pi(\mathcal{U}) - K(P\|P_0) : \pi P = \pi \big\} \tag{7.25}
$$

Recall that the relative value function is not unique, since a new solution is obtained by adding a non-zero constant; the normalization $h^*_\zeta(x^\circ) = 0$ is imposed, where $x^\circ \in \mathsf{X}$ is a fixed state.

The proof of Theorem 7.4.1 (i) is a consequence of Prop. 7.3.2. The second result is obtained on combining Lemmas B.2–B.4 of [27].

Theorem 7.4.1. *There exist optimizers $\{\check\pi_\zeta, \check P_\zeta : \zeta \in \mathbb{R}\}$, and solutions to the AROE $\{h^*_\zeta, \eta^*(\zeta) : \zeta \in \mathbb{R}\}$ with the following properties:*

(i) *The optimizer $\check P_\zeta$ can be obtained from the relative value function h^*_ζ as follows:*

$$\check P_\zeta(x, x') := P_0(x, x') \exp\big(h_\zeta(x'_u \mid x) - \Lambda_{h_\zeta}(x)\big) \tag{7.26}$$

where for $x \in \mathsf{X}$, $x'_u \in \mathsf{X}^{\mathsf{u}}$,

$$h_\zeta(x'_u \mid x) = \sum_{x'_n} Q_0(x, x'_n) h^*_\zeta(x'_u, x'_n), \tag{7.27}$$

and $\Lambda_{h_\zeta}(x)$ is the normalizing constant (7.17) with $h = h_\zeta$.

(ii) *$\{\check\pi_\zeta, \check P_\zeta, h^*_\zeta, \eta^*(\zeta) : \zeta \in \mathbb{R}\}$ are continuously differentiable in the parameter ζ.* □

Representations for the derivatives in Theorem 7.4.1 (ii), in particular the derivative of $\Lambda_{h^*_\zeta}$ with respect to ζ, lead to a representation for the ODE used to compute the transition matrices $\{\check P_\zeta\}$.

It is convenient to generalize the problem slightly here: let $\{h^\circ_\zeta : \zeta \in \mathbb{R}\}$ denote a family of functions on X, continuously differentiable in the parameter ζ. They are not necessarily relative value functions, but we maintain the structure established in Theorem 7.4.1 for the family of transition matrices. Denote,

$$h_\zeta(x'_u \mid x) = \sum_{x'_n} Q_0(x, x'_n) h^\circ_\zeta(x'_u, x'_n), \quad x \in \mathsf{X},\ x'_u \in \mathsf{X}^{\mathsf{u}} \tag{7.28}$$

and then define as in (7.16),

$$P_\zeta(x, x') := P_0(x, x') \exp\big(h_\zeta(x'_u \mid x) - \Lambda_{h_\zeta}(x)\big) \tag{7.29}$$

The function $\Lambda_{h_\zeta} : \mathsf{X} \to \mathbb{R}$ is a normalizing constant, exactly as in (7.17):

$$\Lambda_{h^\circ_\zeta}(x) := \log\Big(\sum_{x'} P_0(x, x') \exp\big(h_\zeta(x'_u \mid x)\big)\Big)$$

We begin with a general method to construct a family of functions $\{h^\circ_\zeta : \zeta \in \mathbb{R}\}$ based on an ODE. The ODE is expressed,

$$\tfrac{d}{d\zeta} h^\circ_\zeta = \mathcal{V}(h^\circ_\zeta)\,, \qquad \zeta \in \mathbb{R}, \tag{7.30}$$

with boundary condition $h^\circ_0 \equiv 0$. A particular instance of the method will result in $h^\circ_\zeta = h^*_\zeta$ for each ζ. Assumed given is a mapping $\mathcal{H}^\circ$ from transition matrices to functions on X. Following this, the vector field $\mathcal{V}$ is obtained through the following two steps: For a function $h\colon \mathsf{X} \to \mathbb{R}$,

(i) Define a new transition matrix via (7.16),

$$P_h(x,x') := P_0(x,x')\exp\big(h(x'_u \mid x) - \Lambda_h(x)\big), \quad x,x' \in \mathsf{X}, \tag{7.31}$$

in which $h(x'_u \mid x) = \sum_{x'_n} Q_0(x,x'_n)h(x'_u,x'_n)$, and $\Lambda_h(x)$ is a normalizing constant.

(ii) Compute $H^\circ = \mathcal{H}^\circ(P_h)$, and define $\mathcal{V}(h) = H^\circ$. It is assumed that the functional $\mathcal{H}^\circ$ is constructed so that $H^\circ(x^\circ) = 0$ for any h.

We now specify the functional $\mathcal{H}^\circ$, whose domain consists of transition matrices that are irreducible and aperiodic. For any transition matrix P in this domain, the fundamental matrix Z is obtained using (7.15), and then $H^\circ = \mathcal{H}^\circ(P)$ is defined as

$$H^\circ(x) = \sum_{x'} [Z(x,x') - Z(x^\circ,x')]\mathcal{U}(x'), \qquad x \in \mathsf{X} \tag{7.32}$$

The function H° is a solution to Poisson's equation,

$$PH^\circ = H^\circ - \mathcal{U} + \bar{u}\,, \qquad \text{where } \bar{u} := \pi(\mathcal{U}) := \sum_x \pi(x)\mathcal{U}(x). \tag{7.33}$$

Theorem 7.4.2. *Consider the ODE* (7.30) *with boundary condition* $h_0^\circ \equiv 0$*, and with* $H^\circ = \mathcal{H}^\circ(P)$ *defined using* (7.32) *for each transition matrix* P *that is irreducible and aperiodic. The solution to this ODE exists, and the resulting functions* $\{h_\zeta^\circ : \zeta \in \mathbb{R}\}$ *coincide with the relative value functions* $\{h_\zeta^* : \zeta \in \mathbb{R}\}$*. Consequently,* $\check{P}_\zeta = P_{h_\zeta}$ *for each* ζ*.*

7.5 Discussion and related results

The ODE approach for solving MDPs has simple structure for the class of models considered in this chapter. An interesting possible extension are approaches to approximate dynamic programming as has been successful in the unconstrained model [133].

It is likely that the ODE has special structure for other classes of MDPs, such as the "rational inattention" framework of [127, 125]. The computational efficiency of this approach will depend in part on numerical properties of the ODE, such as its sensitivity for complex models. Applications to distributed control were the original motivation for this work, with particular attention to "demand dispatch" [41], presented in Chapter 6.

Chapter 8

Kullback-Leibler-Quadratic optimal control

This chapter presents approaches to mean-field control, motivated by distributed control of multi-agent systems. Control solutions are based on a convex optimization problem, whose domain is a convex set of probability mass functions (pmfs). The main contributions follow:

Kullback-Leibler-Quadratic (KLQ) optimal control is a special case, in which the objective function is composed of a control cost in the form of Kullback-Leibler divergence between a candidate pmf and the nominal, plus a quadratic cost on the sequence of marginals. Transform techniques are introduced to reduce complexity of the KLQ solution, motivated by the need to consider time horizons that are much longer than the inter-sampling times required for reliable control.

Infinite-horizon KLQ leads to a state feedback control solution with attractive properties. It can be expressed as either state feedback, in which the state is the sequence of marginal pmfs, or an open loop solution is obtained that is more easily computed.

A main application is to distributed control of residential loads, with objective to provide grid services, similar to utility-scale battery storage. The results show that KLQ optimal control enables the aggregate power consumption of a collection of flexible loads to track a time-varying reference signal, while simultaneously ensuring each individual load satisfies its own quality of service constraints.

This chapter is based on the publications [C8, J1], that contain more details and the proofs of the results summarized in this chapter.

8.1 Introduction

The goal of this work is to obtain control solutions for mean-field models. The optimization problems considered are generalizations of standard Markov Decision Process (MDP) objectives, in both finite-horizon and average-cost settings.

8.1.1 Mean field control

The mean-field control problem is an approach to distributed control of a collection of $\mathcal{N}$ homogeneous "agents", with $\mathcal{N} \gg 1$, modeled as discrete-time stochastic systems, with state processes at time k denoted $\{X_k^i : 1 \le i \le \mathcal{N}\}$. To avoid a long detour on notation it is

assumed that the common state space X is finite. For a single value k and time horizon $K \geq 1$, the empirical distributions are denoted

$$p^{\mathcal{N}}(\vec{x}) = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \mathbb{I}\{(X_0^i, \ldots, X_K^i) = \vec{x}\} \qquad \vec{x} \in \mathsf{X}^{K+1} \tag{8.1a}$$

$$\nu_k^{\mathcal{N}}(x) = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \mathbb{I}\{X_k^i = x\}\,, \qquad x \in \mathsf{X}\,, \tag{8.1b}$$

where $\vec{x} = (x_0, \ldots, x_K)$ denotes an arbitrary element of X^{K+1}. The set of pmfs on X^{K+1} is denoted by $\mathcal{S}(\mathsf{X}^{K+1})$ for $K \geq 1$, and $\mathcal{S}(\mathsf{X})$ for $K = 0$.

The integer $\mathcal{N}$ is regarded as a parameter in mean-field theory, and assumptions imply that there is convergence as $\mathcal{N} \to \infty$,

$$\lim_{\mathcal{N}\to\infty} p^{\mathcal{N}}(\vec{x}) = p(\vec{x})\,, \qquad \lim_{\mathcal{N}\to\infty} \nu_k^{\mathcal{N}}(x_k) = \nu_k(x)\,,$$

where $\nu_k \in \mathcal{S}(\mathsf{X})$ is the kth marginal of $p \in \mathcal{S}(\mathsf{X}^{K+1})$ for $0 \leq k \leq K$.

This limit is achieved by assuming homogeneity of the statistics of each agent: for each i the state evolution is consistent with p:

$$\mathsf{P}\{X_{k+1}^i = x_{k+1} \mid (X_0^i, \ldots, X_k^i) = \vec{x}_0^k\} = p(x_{k+1} \mid \vec{x}_0^k) \tag{8.2}$$

where the conditional pmfs are obtained from Bayes rule.

We propose a design of p to balance two objectives, based on a reference signal $\{r_k\}$, and function $\mathcal{Y}\colon \mathsf{X} \to \mathbb{R}$:

(i) $\nu_k \sim \nu_k^0$, where $\{\nu_k^0\}$ models *nominal behavior*.

(ii) $\langle \nu_k, \mathcal{Y} \rangle := \sum_{x\in\mathsf{X}} \nu_k(x)\mathcal{Y}(x) \approx r_k$.

The agents considered in Section 8.3 represent a population of residential water heaters, and $\mathcal{Y}\colon \mathsf{X} \to \mathbb{R}_+$ is chosen so that $\langle \nu_k^{\mathcal{N}}, \mathcal{Y} \rangle$ is the average power consumption over the population of loads.

Two approaches to design are developed.

Feedforward control: A sequence $\{\mathcal{C}_k : 1 \leq k \leq K\}$ of real-valued cost functions on the marginals is specified, and p^* is obtained as the solution to

$$J^\star(\nu_0^0) = \min_p \sum_{k=1}^{K} \mathcal{C}_k(\nu_k) \tag{8.3}$$

where the minimum is over all pmfs with first marginal ν_0^0. The two goals motivate the following objective function,

$$\mathcal{C}_k(\nu) = \mathcal{D}(\nu, \nu_k^0) + \frac{\kappa}{2}\big[\langle \nu, \mathcal{Y} \rangle - r_k\big]^2\,, \qquad \nu \in \mathcal{S}(\mathsf{X})\,, \tag{8.4}$$

in which $\kappa > 0$ is a penalty parameter, and $\mathcal{D}$ penalizes deviation from nominal behavior. The finite-horizon optimal control problem is thus

$$J^\star(\nu_0^0) = \min_p \sum_{k=1}^{K} \left[\mathcal{D}(\nu_k, \nu_k^0) + \frac{\kappa}{2}\big[\langle \nu_k, \mathcal{Y} \rangle - r_k\big]^2\right] \tag{8.5}$$

This finite horizon optimal control problem can be a component of a model predictive control (MPC) strategy, with time horizons for computation updates dictated by performance requirements and model accuracy.

Feedback control: If the nominal model is Markovian, then the evolution of the marginals follow the dynamics of a controlled nonlinear state space model,

$$\nu_{k+1} = f_k(\nu_k, \phi_k)\,, \qquad k \ge 0\,, \quad \nu_0^0 \text{ given} \tag{8.6}$$

where $\{\phi_k\}$ is the input sequence, evolving on an abstract set Φ. A feedback policy takes the form $\phi_k = \mathcal{K}_k(\nu_k)$.

Design choices for $\mathcal{K}_k$ are proposed in [J1] based on an infinite-horizon solution of (8.5). Justification requires further assumptions, including time-homogenous dynamics for (8.6), which holds if the nominal model is a time-homogeneous Markov chain.

We survey in this chapter only the results on feedforward control, and refer the reader to [J1] for feedback control design.

8.1.2 MDPs and mean-field control

The Markovian assumption for the nominal model is based on the standard controlled Markov chain model used in MDPs.

The model considered here is specified by a state space denoted S, input space U, and we denote $\mathsf{X} := \mathsf{S} \times \mathsf{U}$ (assumed finite). The joint state-input process is denoted $\boldsymbol{X} = \{X_k = (S_k, U_k) : k \ge 0\}$. In finite-horizon optimal control the model includes a sequence of controlled transition matrices $\{T_k : k \ge 0\}$ and cost functions $\{c_k : k \ge 0\}$, with $c_k \colon \mathsf{X} \to \mathbb{R}$ for each k.

The dynamics of $\boldsymbol{X} = (\boldsymbol{S}, \boldsymbol{U}) = \{S_k, U_k : k \ge 0\}$ are determined by the transition matrices as follows. It is assumed that $\boldsymbol{X}$ is adapted to a filtration $\{\mathcal{F}_k : k \ge 0\}$ (so that X_k is $\mathcal{F}_k$-measurable for each k), and

$$\mathsf{P}\{S_{k+1} = s' \mid \mathcal{F}_k;\ S_k = s\,,\ U_k = u\} = T_k(x, s')\,, \qquad x = (s,u) \in \mathsf{X}\,, s' \in \mathsf{S} \tag{8.7}$$

The set of functions from S to the simplex $\mathcal{S}(\mathsf{U})$ is denoted Φ, and we let ϕ denote a generic element of Φ. The decision rule defining the input sequence is assumed to be Markovian:

$$\mathsf{P}\{U_k = u \mid \mathcal{F}_{k-1};\ S_k = s\} = \phi_k(u \mid s)\,, \qquad x = (s,u) \in \mathsf{X} \tag{8.8}$$

with $\phi_k \in \Phi$ for each k.

The finite-horizon optimal control problem of MDP theory is a special case of (8.3), in which $\mathcal{C}_k$ linear for each k; in this case $\mathcal{C}_k(\nu_k) = \langle \nu_k, c_k \rangle = \sum_{x \in \mathsf{X}} \nu_k(x) c_k(x)$ for each k, and the sum on the right hand side of (8.3) may be expressed

$$\sum_{k=1}^{K} \langle \nu_k, c_k \rangle = \sum_{k=1}^{K} \mathsf{E}[c_k(X_k)]\,, \qquad X_k \sim \nu_k\,,$$

where $\boldsymbol{X}$ evolves according to the controlled Markovian dynamics. This interpretation is the first step in the linear programming (LP) approach to MDPs introduced by Manne [18, 99]. The second step is to recognize that the dynamics can be expressed as a sequence of linear constraints on the marginals,

$$\sum_{u'} \nu_k(s', u') = \sum_{s,u} \nu_{k-1}(s,u) T_{k-1}(x, s')\,, \quad s' \in \mathsf{S}\,,\ 1 \le k \le K\,, \quad \nu_0^0 \text{ given.} \tag{8.9}$$

Another special case of (8.3) is variance-penalized optimal control, for which $\mathcal{C}_k(\nu_k) = \langle \nu_k, c\rangle + \kappa\big[\langle \nu_k, c^2\rangle - \langle \nu_k, c\rangle^2\big]$, with $\kappa > 0$ a penalty parameter. The solution to the optimization problem (8.3) can be expressed using a randomized state feedback policy of the form (8.8) [6, 119, 106].

8.1.3 Kullback-Leibler-Quadratic control

In this approach to feedforward control we choose a Markovian model of the form (8.7,8.8) to define nominal behavior: for a collection $\{\phi_k^0\} \subset \Phi$,

$$p^0(\vec{x}) = \nu_0^0(x_0)P_0^0(x_0, x_1)P_1^0(x_1, x_2)\cdots P_{K-1}^0(x_{K-1}, x_K) \tag{8.10a}$$

$$P_k^0(x, x') = T_k(x, s')\phi_{k+1}^0(u' \mid s')\,, \qquad x, x' \in \mathsf{X} \tag{8.10b}$$

Any other $\{\phi_k\} \subset \Phi$ defines a Markov chain $\boldsymbol{X}$ with transition matrices,

$$P_k(x, x') := \mathsf{P}\{X_{k+1} = x' \mid X_k = x\} = T_k(x, s')\phi_{k+1}(u' \mid s')\,. \tag{8.11}$$

The marginals evolve according to linear dynamics, similar to (8.9):

$$\nu_k = \nu_{k-1}P_{k-1}\,, \qquad 1 \le k \le K \tag{8.12}$$

in which ν_k is interpreted as an n-dimensional row vector, with $n = |\mathsf{X}|$.

We obtain a convex program by optimizing over $\{\nu_k\}$, similar to the LP approach of [99]. Scalar variables $\{\gamma_k\}$ are introduced to simplify the objective, in anticipation of a Lagrangian decomposition:

$$J^\star(\nu_0^0) := \min_{\nu,\gamma} \Big[\sum_{k=1}^K \mathcal{D}(\nu_k, \nu_k^0) + \frac{\kappa}{2}\sum_{k=1}^K \gamma_k^2\Big] \tag{8.13a}$$

$$\text{s.t.}\;\; \gamma_k = \langle\, \nu_k, \mathcal{Y}\,\rangle - r_k\,, \tag{8.13b}$$

$$\sum_{u'} \nu_k(s', u') = \sum_{s,u} \nu_{k-1}(s, u)T_{k-1}(x, s')\,, \qquad 1 \le k \le K \tag{8.13c}$$

The *relative entropy rate* is adopted as the cost of deviation:

$$\mathcal{D}(\nu_k, \nu_k^0) := \sum_{s,u} \nu_k(s, u)\log\Big(\frac{\phi_k(u \mid s)}{\phi_k^0(u \mid s)}\Big) \tag{8.14}$$

The terminology is justified through the following steps. First, we have seen that any randomized policy gives rise to a pmf $p \in \mathcal{S}(\mathsf{X}^{K+1})$ that is Markovian:

$$p(\vec{x}) = \nu_0^0(x_0)P_0(x_0, x_1)P_1(x_1, x_2)\cdots P_{K-1}(x_{K-1}, x_K)$$

The *relative entropy* (Kullback-Leibler divergence) is the mean log-likelihood:

$$D(p\|p^0) = \sum L(\vec{x})\, p(\vec{x}) \tag{8.15}$$

where $L = \log(p/p^0)$ is an extended-real-valued function on X^{K+1}. The expression for P_k in (8.11) and the analogous formula for P_k^0 using ϕ_{k+1}^0 gives

$$L(\vec{x}) = \log\Big(\frac{p(\vec{x})}{p^0(\vec{x})}\Big) = \sum_{k=0}^{K-1} \log\Big(\frac{P_k(x_k, x_{k+1})}{P_k^0(x_k, x_{k+1})}\Big) \;\; = \sum_{k=1}^{K} \log\Big(\frac{\phi_k(u_k \mid s_k)}{\phi_k^0(u_k \mid s_k)}\Big) \tag{8.16}$$

Consequently, $D(p\|p^0) = \sum_{k=1}^K \mathcal{D}(\nu_k, \nu_k^0)$.

Proposition 8.1.1. *With $\mathcal{D}$ chosen as the relative entropy rate* (8.14)*, the optimization problem* (8.13) *is convex in $\{\nu_k, \gamma_k : 1 \le k \le K\}$. Furthermore, the linear constraints in* (8.13c) *are equivalent to* (8.12). □

8.1.4 Main contributions

KLQ optimal control. Consideration of the dual of the convex optimization problem (8.13) leads to many insights. The main conclusions summarized here are a special case of Theorem 8.2.1:

Theorem 8.1.2. [KLQ solution]. *Consider the convex program* (8.13)*. An optimizer $\{\phi_k^\star : 1 \le k \le K\}$ exists, is unique, and is of the form:*

$$\phi_k^\star(u \mid s) = \phi_k^0(u \mid s) \exp\Big(\sum_{s'} T_k(x, s') g_{k+1}^\star(s') + \lambda_k^\star \mathcal{Y}(s, u) - g_k^\star(s)\Big)\,, \tag{8.17a}$$

$$\textit{where}\quad g_k^\star(s) = \log\Big(\sum_u \phi_k^0(u \mid s) \exp\Big(\sum_{s'} T_k(x, s') g_{k+1}^\star(s') + \lambda_k^\star \mathcal{Y}(s, u)\Big)\Big) \tag{8.17b}$$

and $\{\lambda_k^\star : 1 \le k \le K\}$, $\{g_k^\star(s) : 1 \le k \le K\}$ are the Lagrange multipliers for the constraints (8.13b) *and* (8.13c)*, respectively, and $g_{K+1} \equiv 0$.*

Proposition 8.2.2 motivates a two-step approach in which $\lambda^\star$ is obtained as the solution to a convex program that maximizes the dual function $\varphi^\star$, and then $g^\star$ are computed through the nonlinear recursion (8.17b). Hence the larger computational challenge is computing $\lambda^\star$. Expressions for the derivatives of $\varphi^\star$ involve means and variances of $\mathcal{Y}(X_k)$, which invites the application of Monte-Carlo techniques when the state space is large or even uncountable—see Section 8.2.3.

Application to Demand Dispatch. The original motivation for the research surveyed here is application to distributed control of power systems. The term *Demand Dispatch* was introduced in the conceptual article [22] to describe the possibility of distributed intelligence in electric loads, designed so that the population would help provide supply-demand balance in the power grid.

The numerical results surveyed in Section 8.3 illustrate the application of KLQ to control a large population of residential loads. As expected, tracking error can be made arbitrarily small with large $\kappa > 0$, provided the reference signal is feasible.

It is found in numerical experiments that the histograms defining the state of the mean-field model rapidly "forget" their initial conditions. For example, Figure 8.1 shows the evolution of the histograms over time from six different degenerate initial conditions; within a few hours, they become nearly identical. If this phenomenon holds under general conditions, then it has important implications for control design. Further discussion is contained in Section 8.4.

Organization The remainder of this chapter is organized as follows: Section 8.2 describes a relaxation technique motivated by the desire to reduce computational complexity, along with a full analysis of the convex program (8.13) and its dual. Results from numerical experiments are collected together in Section 8.3. Conclusions and directions for future research are contained in Section 8.4.

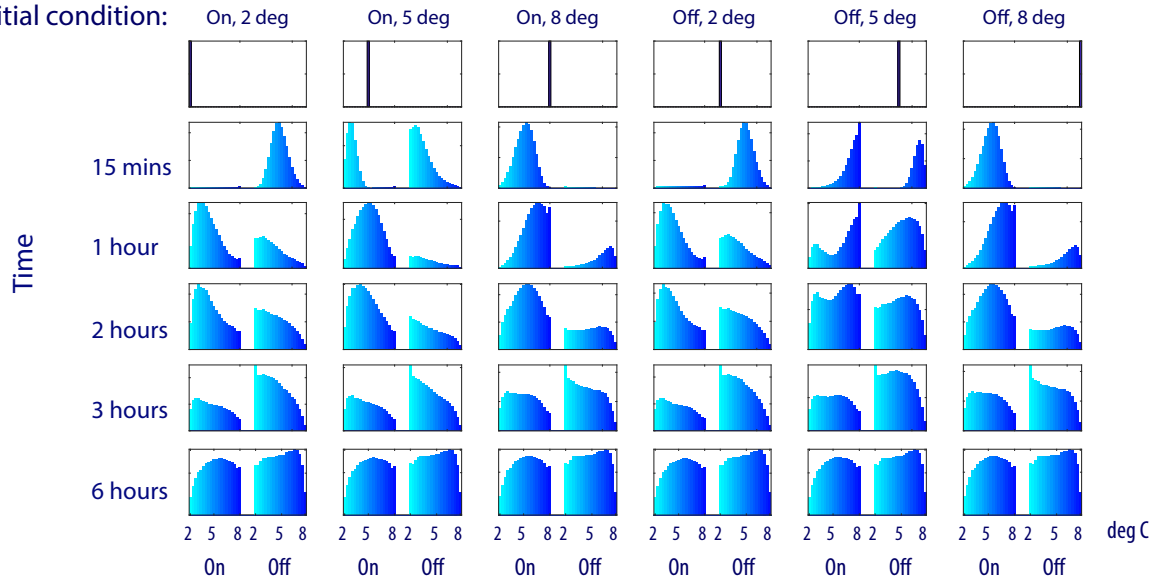


Figure 8.1: Evolution of the marginals $\{\nu_k^\star\}$ of individual agents with $\kappa = 150$, from six different initial conditions. The histograms nearly coincide after about three hours.

8.2 Kullback-Leibler-Quadratic Optimal Control

8.2.1 Subspace relaxation

A relaxation of the convex program (8.13) is described here. Motivation is most clear from consideration of distributed control of a collection of residential water heaters. These loads are valuable as sources of virtual energy storage since they in fact are energy storage devices (in the form of heat rather than electricity), and are also highly flexible. Flexibility comes in part from their extremely non-symmetric behavior: a typical unit may be on for just five minutes, and off continuously for more than six hours. The inter-sampling time at the load should be far less than five minutes to obtain a reliable model for control.

On the other hand, it is valuable for the time horizon to be on the order of several hours. For example, peak-shaving is more effective when water heaters have advance warning to pre-heat the water tanks. To obtain a useful control solution will thus require a very large value of K in (8.13). To reduce complexity, an approach is proposed here based on lossy compression of $\{r_k\}$ using transform techniques.

The transformations are based on a collection of functions $\{w_n : 1 \le n \le N\}$, with $w_n \colon \{0, 1, \dots, K\} \to \mathbb{R}$ for each n, and $N \ll K$. The transformed signal is the N-dimensional vector $\hat{r}$ with $\hat{r}_n = \sum_k w_n(k) r_k$ for each n, and the transformed function on X^{K+1} is denoted

$$\widehat{\mathcal{Y}}_n(\vec{x}) = \sum_{k=1}^{K} w_n(k) \mathcal{Y}(x_k)\,, \qquad 1 \le n \le N$$

The goal is to achieve the approximation $\langle p, \widehat{\mathcal{Y}}_n \rangle \approx \hat{r}_n$ for each n, while maintaining $p \approx p^0$. For example, a Fourier series can be used, with frequency $\omega > 0$, and N is necessarily odd:

$$\{w_n(k) : 1 \le n \le N\} = \{1, \sin(\omega m k), \cos(\omega m k) : 1 \le m \le (N-1)/2\}$$

The *degenerate* family is defined by

$$w_n(k) = \mathbb{I}\{n = k\}\,, \qquad 1 \le n, k \le K \tag{8.18}$$

so that $N = K$ in this case.

The optimal control problem with *subspace relaxation* is defined as the optimal control problem

$$J^\star(\nu_0^0) := \min_{\nu,\gamma} \sum_{k=1}^K \mathcal{D}(\nu_k, \nu_k^0) + \frac{\kappa}{2}\sum_{n=1}^N \gamma_n^2 \tag{8.19a}$$

$$\text{s.t.} \quad \gamma_n = \langle\, p, \widehat{\mathcal{Y}}_n \,\rangle - \hat{r}_n\,, \quad 1 \le n \le N \tag{8.19b}$$

$$\sum_{u'} \nu_k(s', u') = \sum_{s,u} \nu_{k-1}(s,u) T_{k-1}(x, s')\,, \qquad 1 \le k \le K\,,\ s' \in \mathsf{S} \tag{8.19c}$$

This reduces to (8.13) in the degenerate case (8.18).

The theory that follows is based in part on a relaxation of the dynamical constraints (8.19c), through the introduction of a Lagrange multiplier for each k. This is precisely the first step in the construction of the Hamiltonian in the Minimum Principle approach to optimal control [95].

8.2.2 Duality

Structure for the solution of (8.19) will be obtained by consideration of a dual, in which $\lambda \in \mathbb{R}^N$ and $g \in \mathbb{R}^{K \times J^\star}$ denote the vectors of Lagrange multipliers for the first and second set of constraints, respectively. The matrix g is interpreted as a sequence of functions $g_k \colon \mathsf{S} \to \mathbb{R}$ that are entirely analogous to the co-state variables in the Minimum Principle (the Lagrange multipliers for the dynamical constraints) [95].

The Lagrangian is denoted

$$\begin{aligned}\mathcal{L}(\nu, \gamma, \lambda, g) = &\sum_{k=1}^K \mathcal{D}(\nu_k, \nu_k^0) + \frac{\kappa}{2}\sum_{n=1}^N \gamma_n^2 + \sum_{n=1}^N \lambda_n \Big(\gamma_n + \sum_{k=1}^K w_n(k)\big[r_k - \langle \nu_k, \mathcal{Y}\rangle\big]\Big) \\ &+ \sum_{k=1}^K \sum_{s'} \Big(\sum_{u'} \nu_k(s', u') - \sum_{s,u} \nu_{k-1}(s,u) T_{k-1}(x, s')\Big) g_k(s')\end{aligned} \tag{8.20}$$

and the dual function is defined to be its minimum:

$$\varphi^\star(\lambda, g) := \min_{\nu,\gamma} \mathcal{L}(\nu, \gamma, \lambda, g)$$

The *dual* of the optimization problem (8.19) is defined as the maximum of the dual function $\varphi^\star$ over λ and g (see [95] for a complete and accessible treatment of this theory). We will see that there is no duality gap, so that for a quadruple $(\nu^\star, \gamma^\star, \lambda^\star, g^\star)$,

$$J^\star(\nu_0^0) = \mathcal{L}(\nu^\star, \gamma^\star, \lambda^\star, g^\star) = \varphi^\star(\lambda^\star, g^\star)\,.$$

In the following subsections a representation of the dual function is obtained that is suitable for optimization, which results in a valuable representation for the optimal policy. Properties of the dual function are contained in Theorem 8.2.1 and Proposition 8.2.2 that follow.

The statement of these results requires additional notation: define a function $\mathcal{T}_k^\lambda \colon \mathbb{R}^{|\mathsf{S}|} \to \mathbb{R}^{|\mathsf{S}|}$, for $f \colon \mathsf{S} \to \mathbb{R}$ and $\lambda \in \mathbb{R}^N$, via

$$\begin{aligned} \mathcal{T}_k^\lambda(f;s) &= \log\Big(\sum_u \phi_k^0(u \mid s) \exp\big(\sum_{s'} T_k(x,s') f(s') + \check{\lambda}_k \mathcal{Y}(s,u)\big)\Big)\,, \quad s \in \mathsf{S}\,, \\ \textit{where} \quad \check{\lambda}_k &= \sum_{n=1}^N \lambda_n w_n(k) \end{aligned} \tag{8.21}$$

The maximum of the dual function over g is denoted

$$\varphi^\star(\lambda) := \max_g \phi^\star(\lambda, g) = \varphi^\star(\lambda, g^\lambda)$$

where g^λ is a maximizer, $g^\lambda \in \arg\max_g \phi^\star(\lambda, g)$. It is shown in Proposition 8.2.2 that the vector valued function g^λ satisfies the recursion

$$g_k^\lambda = \mathcal{T}_k^\lambda(g_{k+1}^\lambda)\,, \quad 1 \le k \le K\,, \quad \text{where} \quad g_{K+1}^\lambda \equiv 0\,. \tag{8.22}$$

This forms part of the proof of Theorem 8.2.1, with complete details postponed to **??**.

Theorem 8.2.1. *There exists a maximizer $\{\lambda_n^\star, g_k^\star : 1 \le n \le N, 1 \le k \le K\}$ for $\varphi^\star$, and there is no duality gap:*

$$\varphi^\star(\lambda^\star, g^\star) = J^\star(\nu_0^0)$$

The optimal policy is obtained from $\{g_k^\star\}$ via:

$$\begin{aligned} \phi_k^\star(u \mid s) &= \phi_k^0(u \mid s) \exp\big(\sum_{s'} T_k(x,s') g_{k+1}^\star(s') + \check{\lambda}_k^\star \mathcal{Y}(s,u) - g_k^\star(s)\big) \\ \textit{where } g_k^\star(s) &= \mathcal{T}_k^\lambda(g_{k+1}^\star; s) \textit{ for } 1 \le k \le K, \textit{ and } g_{K+1}^\star \equiv 0\,, \end{aligned} \tag{8.23}$$

and $\{\check{\lambda}_k^\star\}$ are obtained from $\{\lambda_n^\star\}$ via (8.21). □

Denote for each k,

$$G_k^\lambda(x) = \sum_s T_{k-1}(x,s) g_k^\lambda(s) \tag{8.24}$$

Proposition 8.2.2. *The following hold for the dual of (8.19): for each $\lambda \in \mathbb{R}^N$,*

(i) *A maximizer g^λ is given by (8.22)*

(ii) *The maximum of the dual function over g is the concave function*

$$\varphi^\star(\lambda) = \lambda^T \hat{r} - \frac{1}{2\kappa}\|\lambda\|^2 - \langle\, \nu_0^0, G_1^\lambda \,\rangle \tag{8.25}$$

(iii) *The function (8.25) is continuously differentiable, and*

$$\frac{\partial}{\partial \lambda_n} \varphi^\star(\lambda) = \hat{r}_n - \frac{1}{\kappa}\lambda_n - \sum_{k=1}^K w_n(k) \langle\, \nu_k^\lambda, \mathcal{Y} \,\rangle\,, \qquad 1 \le n \le N \tag{8.26}$$

where $\{\nu_k^\lambda\}$ is the sequence of marginals obtained from the randomized policy defined in (8.23), substituting $\{g_k^\star\}$ by $\{g_k^\lambda\}$ defined in (i). □

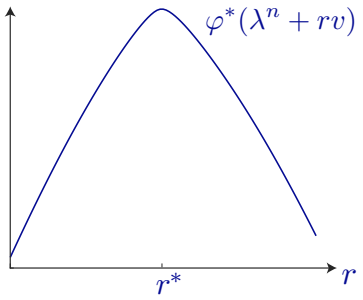


Figure 8.2: Dual function along a line-segment

To conclude this section, we provide representations of the log-likelihood ratio, $L(\vec{x})$, relative entropy $D(p^\lambda \| p^0)$, and primal objective function for the pmf $p^\lambda \in \mathcal{S}(\mathsf{X}^{K+1})$ obtained from the randomized policy defined in (8.23), substituting $\{g_k^\star\}$ by $\{g_k^\lambda\}$ defined in Proposition 8.2.2, part (i).

Corollary 8.2.3. *The following hold for all* $\{\check{\lambda}_k, g_k^\lambda : 1 \le k \le K\}$*:*

(i) *The log-likelihood ratio can be expressed:*

$$L(\vec{x}) = \sum_{k=1}^{K} \{\Delta_k(x_{k-1}, s_k) + \check{\lambda}_k \mathcal{Y}(x_k)\} - G_1^\lambda(x_0) \tag{8.27}$$

where for each k (recalling $x_k = (s_k, u_k)$)*,*

$$\Delta_k(x_{k-1}, s_k) = G_k^\lambda(x_{k-1}) - g_k^\lambda(s_k) \tag{8.28}$$

(ii) *The relative entropy is given by*

$$D(p^\lambda \| p^0) = \sum_{k=1}^{K} \check{\lambda}_k \langle\, \nu_k^\lambda, \mathcal{Y} \,\rangle - \langle\, \nu_0^0, G_1^\lambda \,\rangle \tag{8.29}$$

(iii) *The value of the primal is given by*

$$J(p^\lambda, \nu_0^0) := D(p^\lambda \| p^0) + \frac{\kappa}{2} \sum_{n=1}^{N} \gamma_n^2 \tag{8.30a}$$

$$= -\langle\, \nu_0^0, G_1^\lambda \,\rangle + \sum_{k=1}^{K} \check{\lambda}_k \langle\, \nu_k^\lambda, \mathcal{Y} \,\rangle + \frac{\kappa}{2} \sum_{n=1}^{N} \gamma_n^2 \tag{8.30b}$$

with $\gamma_n = \langle\, p^\lambda, \widehat{\mathcal{Y}}_n \,\rangle - \hat{r}_n$. □

The stochastic process $\{\Delta_k(X_{k-1}, S_k)\}$ is a martingale difference sequence; it vanishes when nature is deterministic, reducing to the solution obtained in [34].

8.2.3 Algorithms

Given the simple form of the derivative (8.26), it is tempting to apply gradient ascent to obtain $\lambda^\star$. The difficulty with standard first-order methods is illustrated in Figure 8.2. This is a plot of a typical example in which $\lambda^n \in \mathbb{R}^N$ is given, $v = \nabla\varphi^\star(\lambda^n)$, and the plot shows $\varphi^\star(\lambda^n + rv)$ for a range of positive r. We have found in examples that using gradient ascent on this cone-shaped curve may be slow to converge, likely due to a large "overshoot" when applying standard first-order methods.

In the numerical results that follow we opt for proximal gradient methods [114].

Monte Carlo methods. The gradient of the dual function may be expressed in terms of the first-order statistics of the random variables $\{\widehat{\mathcal{Y}}_n(\vec{X}) : 1 \le n \le N\}$ when $\vec{X} \sim p^\lambda$:

$$\begin{aligned}\mathsf{E}[\widehat{\mathcal{Y}}_n(\vec{X})] &= \sum_{\vec{x}} p^\lambda(\vec{x}) \sum_{k=1}^K w_n(k)\mathcal{Y}(x_k) \\ &= \sum_{k=1}^K w_n(k) \sum_{x_k} \sum_{x_i, i\neq k} p^\lambda(\vec{x})\mathcal{Y}(x_k) = \sum_{k=1}^K w_n(k)\langle\, \nu_k^\lambda, \mathcal{Y}\,\rangle\end{aligned} \tag{8.31}$$

Lemma 8.2.4 follows from (8.26) combined with (8.31):

Lemma 8.2.4. *For any $\lambda \in \mathbb{R}^N$ and $1 \le n \le N$,*

$$\frac{\partial}{\partial\lambda_n}\varphi^\star(\lambda) = \hat{r}_n - \frac{1}{\kappa}\lambda_n - \mathsf{E}[\widehat{\mathcal{Y}}_n(\vec{X})]\,, \quad \textit{in which } \vec{X} \sim p^\lambda. \tag{8.32}$$

□

See [28] for more on Monte Carlo methods and KLQ.

8.3 Applications to Demand Dispatch

An application of the control framework described in the previous sections is *Demand Dispatch*: an evolving science for automatically controlling flexible loads to help maintain supply-demand balance in the power grid. The goal of demand dispatch (DD) is to modify the behavior of flexible loads such that the aggregate power consumption tracks a reference signal that is broadcast by a balancing authority (BA).

In the numerical examples here we focus entirely on the mean-field model. We know from prior work that evolution of the empirical distributions does closely track this idealization: for reasonably large $\mathcal{N}$, following the notation (8.1b), the approximation $\nu_k^{\mathcal{N}} \approx \nu_k$ holds and the covariance of the error grows slowly with k (error is reduced with feedback [37, 39]). Although the control architecture in this prior work is very different, it should not surprise the reader that the law of large numbers and associated central limit theorem hold in this setting.

Although these techniques can be applied to any flexible load, the experiments in this section demonstrate distributed control of a population of residential water heaters or refrigerators. An MDP model is constructed in which the state is the standard used to capture hysteresis control, $S_k = (\theta_k, U_{k-1})$, in which $\theta_k \in \mathbb{R}$ is the temperature, and $U_k \in \{0, 1\}$ denotes power mode for each k. Remember the physical system operates in continuous time, and k represents the kth sampling time. This means that U_{k-1} represents the power mode during the sampling interval ending at the kth sampling time.

8.3.1 Designing the nominal model

Construction of the nominal model with transition matrices $\{P_k^0\}$ of the form (8.10b) requires specification of dynamics of nature and the nominal policy. In the case of water heaters, the sequence of transition matrices $\{T_k\}$ for nature were based on input-output data obtained from Oak Ridge National Laboratories [41]. For refrigerators, T was taken independent of k, constructed based on simulations of the standard linear TCL model:

$$\theta_{k+1} = \theta_k + \alpha[\theta^a - \theta_k] - \beta U_k + D_{k+1}\,, \tag{8.33}$$

in which $\alpha, \beta > 0$, θ^a denotes the (time-invariant) ambient air temperature, and the disturbance process $\boldsymbol{D}$ captures modeling error and usage.

In all cases the nominal policy was chosen time-homogeneous: $\phi_k^0 \equiv \phi^0$ is a fixed randomized policy, designed to approximate deterministic hysteresis control. We describe the construction for water heaters, following [107, 41].

The upper and lower temperature limits that define a deadband are denoted Θ_-, Θ_+, respectively. A standard residential water heater switches deterministically when it reaches the limits, but ϕ^0 is constructed so that the power mode will switch stochastically, often before reaching one of the two limits. The randomized decision rule is represented by two CDFs: $F^{\oplus}$ is the CDF for the temperature at which power turns on, and $F^{\ominus}$ is the CDF for the temperature at which power turns off.

A particular design for these CDFs, already discussed in Section 6.4 and shown in Figure 6.3, is obtained on fixing parameters $\theta_0^{\oplus}, \theta_0^{\ominus} \in [\Theta_-, \Theta_+]$, $\varsigma \in (0,1)$ and $\eta > 1$:

$$\begin{aligned} F^{\ominus}(\theta) &= \varsigma(\theta - \theta_0^{\ominus})_+^{\eta}, \\ F^{\oplus}(\theta) &= 1 - \varsigma(\theta_0^{\oplus} - \theta)_+^{\eta}\,, \qquad \theta \in [\theta_-, \theta_+)\,. \end{aligned}$$

Without loss of generality it is assumed that the sampling interval is 1 unit. At time instance k, the decision rule is:

(i) If $U_k = 0$ then $U_{k+1} = 1$ with probability $\phi_k^0(1|s) = \dfrac{[F^{\oplus}(\theta_{k-1}) - F^{\oplus}(\theta_k)]_+}{F^{\oplus}(\theta_{k-1})}$.

(ii) If $U_k = 1$ then $U_{k+1} = 0$ with probability $\phi_k^0(0|s) = \dfrac{[F^{\ominus}(\theta_k) - F^{\ominus}(\theta_{k-1})]_+}{1 - F^{\ominus}(\theta_{k-1})}$.

8.3.2 Tracking

In practical applications the aggregate power is of interest, which is approximated by $\varrho \mathcal{N} y_k$ at time k, where ϱ is the rated power of a single load. Hence the total population size $\mathcal{N}$ must be taken into account in any tracking problem. In plots that follow, we choose to focus on the "normalized" response, defined as follows: $y_k^{\text{ref}} = r_k/\varrho\,, \hat{y}_k^{\text{ref}} = \hat{r}_k/\varrho\,, y_k = \langle \nu_k, \mathcal{Y} \rangle / \varrho$. In this context, y_k can be interpreted as the probability of a load being on.

The two sets of plots in Figure 8.3 are distinguished by the reference signal. In each case the reference signal is a square wave. In (a) the signal is feasible, and in (b) it violates the energy limits of the collection of water heaters [69]. In Figure 8.3 (a) it is seen that tracking is nearly perfect for sufficiently large κ. Tracking of the larger reference signal would require temperature deviations to exceed the deadband of the water heater. Instead, we observe in Figure 8.3 (b) a graceful truncation of the reference signal.

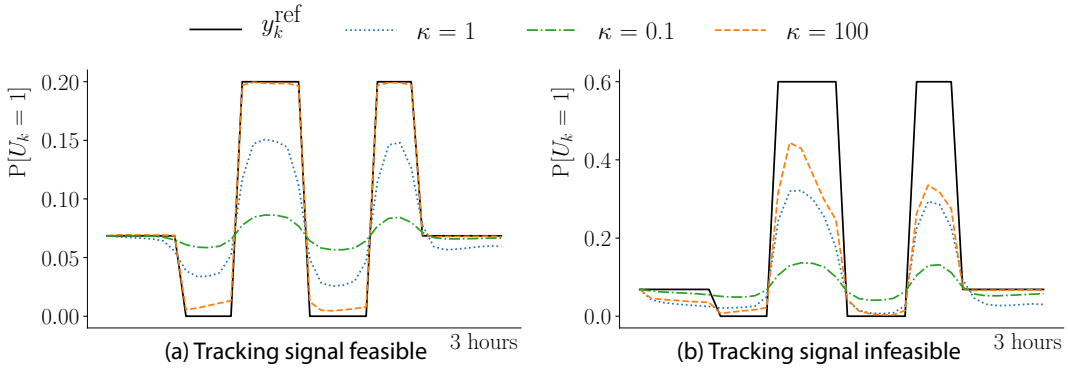


Figure 8.3: Evolution of $\mathcal{N}\langle \nu_k, \mathcal{Y} \rangle$: (a) reference signal is feasible; (b) reference signal is infeasible.

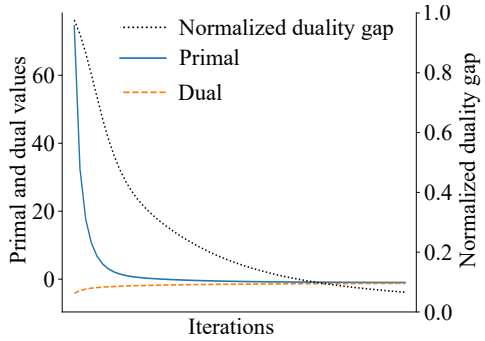


Figure 8.4: The normalized duality gap approaches zero in each experiment.

Figure 8.1 displays the results of a tracking experiment comparing six different initial conditions. Observe how their marginal distributions become nearly identical within a few hours. Recall that this control problem requires knowledge of the initial distribution ν_0. These results suggest that an accurate estimate of the global marginal distribution can be readily available at each load.

8.4 Discussion and related results

Literature review

Mean field control. The optimization problem (8.3) is inspired by mean-field game theory [91, 73, 74, 29, 65, 140] (see [35, 36, 30, 129] for recent surveys).

Mean-field control differs from mean-field game theory only because of greater control at the microscopic layer: we do not assume that an individual in the population is free to optimize based on its local objective function, so we avoid the fragility of Nash equilibria. This description is similar to *ensemble control* in physics (see [92] for history), and many in the power systems area opt for this term rather than mean-field control (see [42, 41] and their references).

Demand Dispatch. The goal of Demand Dispatch is to modify the behavior of loads so that their aggregate power consumption tracks a reference signal $\{r_k\}$ that is synthesized by

a *balancing authority* (BA). Randomized control techniques have been proposed in [104, 130, 107, 5, 42, 12] based on various control architectures.

The following control strategy is common to the approaches described in [107, 41]. It is assumed that a family of transition matrices $\{P_\zeta : \zeta \in \mathbb{R}\}$ is available at each load. A sequence $\{\zeta_0, \zeta_1, \dots\}$ is broadcast from the BA, based on measurements of the grid, and at time k the ith load transitions according to this law:

$$\mathsf{P}\{X^i_{k+1} = x' \mid X^i_k = x,\ \zeta_k = \zeta\} = P_\zeta(x, x')$$

IPD. The paper [107] re-interprets the control solution of [132] as a technique to create the family $\{P_\zeta\}$ through the solution to the nonlinear program:

$$P_\zeta := \arg\max \left\{\zeta \langle \pi, \mathcal{Y} \rangle - \mathcal{R}(P\|P^0)\right\}, \qquad \zeta \in \mathbb{R}, \tag{8.34}$$

where $\mathcal{R}$ denotes the rate function of Donsker and Varadhan [51, 90],

$$\mathcal{R}(P\|P^0) := \sum_{x,x'} \pi(x)P(x,x') \log\Big(\frac{P(x,x')}{P^0(x,x')}\Big) \tag{8.35}$$

in which π is the invariant pmf for P. The maximum in (8.34) is over all (π, P) subject to the invariance constraint $\pi P = \pi$ [107, 26]. The convex program (8.34) is called the *Individual Perspective Design* (IPD) in [26].

The finite-horizon version of (8.34) is also considered in [107, 26], similar to the KLQ formulation:

$$p^\zeta := \arg\max_p \Big\{\zeta \mathsf{E}_p\Big[\sum_{k=1}^K \mathcal{Y}(x_k)\Big] - D(p\|p^0)\Big\}. \tag{8.36}$$

Provided the entries of $T_k(x,s)$ take on only binary values, the finite-horizon IPD solution is obtained as a tilting of the nominal model:

$$p^\zeta(\vec{x}) = p^0(\vec{x}) \exp\Big(\zeta \sum_{k=1}^K \mathcal{Y}(x_k) - \Lambda(\zeta)\Big), \quad \textit{with } \Lambda(\zeta) \textit{ a normalizing constant.} \tag{8.37}$$

Further research

This chapter provides a complete theory for KLQ, without the restriction to deterministic dynamics imposed in [34, 42]. Plans for future research include:

(i) Monte-carlo approaches for both feedforward and feedback control designs.

(ii) Investigate alternative transform techniques.

KLQ and optimal transport. Extensions of the KLQ objective will likely provide useful relaxations of the classical *optimal transport* problem, in which the goal is to steer p^0 to a given target pmf $p^\star$ [136, 115, 40]. Rather than match the target pmf, we might match M generalized moments, minimizing $D(p\|p^0)$ subject to $\langle p, \mathcal{G}_i\rangle = \langle p^\star, \mathcal{G}_i\rangle$ for each i, with $\mathcal{G}_i \colon \mathsf{X}^{K+1} \to \mathbb{R}$.

A special case is the tracking problem,

$$\min_p \big\{ D(p\|p^0) \quad \text{subject to } \mathsf{E}_p\big[\mathcal{Y}(X_k)\big] = r_k\,, \quad 1 \le k \le K \big\} \tag{8.38}$$

This optimization problem is proposed in [42, Section 5], along with the explicit solution

$$p^\star(\vec{x}) = p^0(\vec{x}) \exp\Big(\sum_{k=1}^{K} \beta_k \mathcal{Y}(x_k) - \Lambda(\beta)\Big) \tag{8.39}$$

in which $\beta \in \mathbb{R}^K$ are Lagrange multipliers corresponding to the K constraints, and $\Lambda(\beta)$ a normalizing constant.

Preliminary results are summarized in [50].

Information architectures. On the application side, the choice of information architecture is an interesting topic for future research. Here are three possibilities:

(i) *Smart BA*: The BA uses the reference signal $\{r_k\}$ and its estimate of ν_0^0 to compute $\lambda^\star$ and broadcast it to the loads.

(ii) *Smart Load*: The BA broadcasts $\{r_k\}$ to the loads. Each load computes $\lambda^\star$ based on its internal model and $\nu_0^0 = \delta_{x_0}$, with $x_0 \in \mathsf{X}$ its current state.

(iii) *Genius Load*: The BA broadcasts $\{r_k\}$ to the loads. Each load computes $\lambda^\star$ based on its internal model and its estimate of ν_0^0.

Each approach has its strengths and weaknesses. Approaches (i) and (iii) require knowledge of the initial marginal pmf of the population, ν_0^0. If a perfect estimate is assumed, then the total cost in cases (i) and (iii) is equal to $J^\star(\nu_0^0)$. But, how can a load estimate the marginal pmf of the population? Recall the coupling shown in Figure 8.1: the histograms are nearly identical after about three hours, regardless of the initial condition. If enough time has passed since the latest MPC iteration, the pmfs $\{\nu_k\}$ computed locally can be used to approximate the marginal pmf of the population (perhaps smoothed using the techniques of [37, 39]).

In contrast, the total cost for case (ii) is the sum, $\sum_{i=1}^{d} \nu_0^0(x^i) J^\star(\delta_{x^i})$ since each load optimizes according to its own initial state, x^i. Even when the aggregate can easily track $\{r_k\}$, the cost $J^\star(\delta_{x^i})$ may be very large for individuals that are at odds with the reference signal. For example, an increase in power consumption could be requested while a water heater is near its upper temperature limit and must turn off. So, it is possible that approach (ii) will impose greater stress on the loads as compared to the other two options, or will lead to reduced capacity.

Bibliography

[1] United Network for Organ Sharing. Online, `https://www.unos.org/docs/Living_Donation_KidneyPaired.pdf`.

[2] I. Adan, I. Kleiner, R. Righter, and G. Weiss. FCFS parallel service systems and matching models. *Performance Evaluation*, 127-128:253 – 272, 2018.

[3] I. Adan and G. Weiss. Exact FCFS matching rates for two infinite multitype sequences. *Oper. Res.*, 60(2):475–489, 2012.

[4] I. Adan and G. Weiss. A skill based parallel service system under FCFS-ALIS—steady state, overloads, and abandonments. *Stoch. Syst.*, 4(1):250–299, 2014.

[5] M. Almassalkhi, J. Frolik, and P. Hines. Packetized energy management: asynchronous and anonymous coordination of thermostatically controlled loads. In *Proc. of the American Control Conf.*, pages 1431–1437. IEEE, 2017.

[6] E. Altman. *Constrained Markov decision processes*. Stochastic Modeling. Chapman & Hall/CRC, Boca Raton, FL, 1999.

[7] F. Baccelli and P. Brémaud. *Elements of Queueing Theory (2nd ed)*. Springer, 2002.

[8] F. Baccelli and P. Brémaud. *Elements of queueing theory*, volume 26 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition, 2003. Palm martingale calculus and stochastic recurrences, Stochastic Modelling and Applied Probability.

[9] F. Baccelli, M.-O. Haji-Mirsadeghi, and S. Khaniha. Coupling from the Past for the Null Recurrent Markov Chain. 31 Pages, 8 figures, Feb. 2022.

[10] N. G. Bean, F. P. Kelly, and P. G. Taylor. Braess's paradox in a loss network. *Journal of Applied Probability*, 34(1):155–159, 1997.

[11] J. Begeot, I. Marcovici, P. Moyal, and Y. Rahme. A general stochastic matching model on multigraphs, 2020.

[12] E. Benenati, M. Colombino, and E. Dall'Anese. A tractable formulation for multi-period linearized optimal power flow in presence of thermostatically controlled loads. In *IEEE Conference on Decision and Control*, pages 4189–4194. IEEE, 2019.

[13] S. A. Berezner and A. E. Krzesinski. Order independent loss queues. *Queueing Syst. Theory Appl.*, 23:331–335, 1996.

[14] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein. The complexity of decentralized control of Markov decision processes. *Math. Oper. Res.*, 27(4):819–840, 2002.

[15] D. Bertsekas and S. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, Belmont, MA, 1996.

[16] S. Boersma, B. Doekemeijer, M. Vali, J. Meyers, and J.-W. van Wingerden. A control-oriented dynamic wind farm model: Wfsim. *Wind Energy Science*, 3(1):75–95, 2018.

[17] V. S. Borkar. On minimum cost per unit time control of Markov chains. *SIAM J. Control Optim.*, 22(6):965–978, 1984.

[18] V. S. Borkar. Convex analytic methods in Markov decision processes. In *Handbook of Markov decision processes*, volume 40 of *Internat. Ser. Oper. Res. Management Sci.*, pages 347–375. Kluwer Acad. Publ., Boston, MA, 2002.

[19] J. Y. L. Boudec, D. McDonald, and J. Mundinger. A generic mean field convergence result for systems of interacting objects. In *Fourth International Conference on the Quantitative Evaluation of Systems (QEST 2007)*, pages 3–18, Sept 2007.

[20] D. Braess. Über ein paradoxon aus der verkehrsplanung. *Unternehmensforschung*, 12(1):258–268, 1968.

[21] P. Brémaud. *Markov chains*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues.

[22] A. Brooks, E. Lu, D. Reicher, C. Spirakis, and B. Weihl. Demand dispatch. *IEEE Power and Energy Magazine*, 8(3):20–29, May 2010.

[23] R. A. Brualdi, F. Harary, and Z. Miller. Bigraphs versus digraphs via matrices. *J. Graph Theory*, 4:51–73, 1980.

[24] A. Bušić, V. Gupta, and J. Mairesse. Stability of the bipartite matching model. *Adv. Appl. Probab.*, 45(2):351–378, 2013.

[25] A. Bušić and S. Meyn. Passive dynamics in mean field control. In *Proc. of the IEEE Conf. on Dec. and Control*, pages 2716–2721, Dec 2014.

[26] A. Bušić and S. Meyn. Distributed randomized control for demand dispatch. In *Proc. of the IEEE Conf. on Dec. and Control*, pages 6964–6971, Dec 2016.

[27] A. Bušić and S. Meyn. Ordinary Differential Equation Methods for Markov Decision Processes and Application to Kullback–Leibler Control Cost. *SIAM J. Control Optim.*, 56(1):343–366, 2018.

[28] A. Bušić, S. Meyn, and N. Cammardella. Learning optimal policies in mean field models with Kullback-Leibler regularization. *IEEE Conference on Decision and Control*, 2023 (submitted).

[29] P. E. Caines. Mean field games. In J. Baillieul and T. Samad, editors, *Encyclopedia of Systems and Control*, pages 706–712. Springer London, London, 2015.

[30] P. E. Caines. Mean field games. In J. Baillieul and T. Samad, editors, *Encyclopedia of Systems and Control*, pages 1197–1202. Springer London, London, 2021.

[31] R. Caldentey, E. H. Kaplan, and G. Weiss. FCFS infinite bipartite matching of servers and customers. *Adv. in Appl. Probab.*, 41(3):695–730, 2009.

[32] D. Callaway and I. Hiskens. Achieving controllability of electric loads. *Proceedings of the IEEE*, 99(1):184 –199, January 2011.

[33] B. Calvert, W. Solomon, and I. Ziedins. Braess's paradox in a queueing network with state-dependent routing. *Journal of Applied Probability*, 34(1):134–154, 1997.

[34] N. Cammardella, A. Bušić, Y. Ji, and S. Meyn. Kullback-Leibler-Quadratic optimal control of flexible power demand. In *Proc. of the IEEE Conf. on Dec. and Control*, pages 4195–4201, Dec. 2019.

[35] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I: Mean Field FBSDEs, Control, and Games*. Probability Theory and Stochastic Modelling. Springer Intl. Publishing, 2018.

[36] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications II: Mean Field Games with Common Noise and Master Equations*. Probability Theory and Stochastic Modelling. Springer Intl. Publishing, 2018.

[37] Y. Chen. *Markovian demand dispatch design for virtual energy storage to support renewable energy integration*. PhD thesis, University of Florida, Gainesville, FL, USA, 2016.

[38] Y. Chen, A. Bušić, and S. Meyn. Estimation and control of quality of service in demand dispatch. *IEEE Transactions on Smart Grid*, 9(5):5348–5356, Sept 2018.

[39] Y. Chen, A. Bušić, and S. Meyn. State estimation for the individual and the population in mean field control with application to demand dispatch. *IEEE Transactions on Automatic Control*, 62(3):1138–1149, March 2017.

[40] Y. Chen, T. T. Georgiou, and M. Pavon. Optimal transport in systems and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 4:89–113, 2020.

[41] Y. Chen, M. U. Hashmi, J. Mathias, A. Bušić, and S. Meyn. Distributed control design for balancing the grid using flexible loads. In S. Meyn, T. Samad, I. Hiskens, and J. Stoustrup, editors, *Energy Markets and Responsive Grids: Modeling, Control, and Optimization*, pages 383–411. Springer, New York, NY, 2018.

[42] M. Chertkov and V. Y. Chernyak. Ensemble control of cycling energy loads: Markov Decision Approach. In *IMA volume on the control of energy markets and grids*. Springer, 2017.

[43] I.-K. Cho and S. P. Meyn. Efficiency and marginal cost pricing in dynamic competitive markets with friction. *Theoretical Economics*, 5(2):215–239, 2010.

[44] K. Christakou, D.-C. Tomozei, J.-Y. Le Boudec, and M. Paolone. GECN: primary voltage control for active distribution networks via real-time demand-response. *IEEE Trans. on Smart Grid*, 5(2):622–631, March 2014.

[45] A. J. Clark and H. E. Scarf. Optimal policies for a multi-echelon inventory problem. *Management Sci.*, 6:465–490, 1960.

[46] J. E. Cohen and C. Jeffries. Congestion resulting from increased capacity in single-server queueing networks. *IEEE/ACM transactions on networking*, 5(2):305–310, 1997.

[47] J. E. Cohen and F. P. Kelly. A paradox of congestion in a queuing network. *Journal of Applied Probability*, 27(3):730–734, 1990.

[48] C. Comte. Stochastic non-bipartite matching models and order-independent loss queues. *Stochastic Models*, 38(1):1–36, 2022.

[49] C. Comte and J. L. Dorsman. Performance evaluation of stochastic bipartite matching models. In P. Ballarini, H. Castel, I. Dimitriou, M. Iacono, T. Phung-Duc, and J. Walraevens, editors, *Performance Engineering and Stochastic Modeling - 17th European Workshop, EPEW 2021, and 26th International Conference, ASMTA 2021, Virtual Event, December 9-10 and December 13-14, 2021, Proceedings*, volume 13104 of *Lecture Notes in Computer Science*, pages 425–440. Springer, 2021.

[50] T. L. Corre and S. M. Ana Bušić and. Feature projection for optimal transport. *IEEE Conference on Decision and Control (submitted) and arXiv:2208.01958*, 2023.

[51] A. Dembo and O. Zeitouni. *Large Deviations Techniques And Applications*. Springer-Verlag, New York, second edition, 1998.

[52] A. M. Devraj and S. P. Meyn. Zap Q-learning. In *Proc. of the Intl. Conference on Neural Information Processing Systems*, pages 2232–2241, 2017.

[53] K. Doya. How can we learn efficiently to act optimally and flexibly? *Proceedings of the National Academy of Sciences*, 106(28):11429–11430, 2009.

[54] J. Edmonds and R. M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM*, 19(2):248?264, 1972.

[55] P. Fairley. Energy storage: Power revolution. *Nature*, 526:S102–S104, October 29 2015.

[56] P. Fleming, J. Annoni, J. J. Shah, L. Wang, S. Ananthan, Z. Zhang, K. Hutchings, P. Wang, W. Chen, and L. Chen. Field test of wake steering at an offshore wind farm. *Wind Energy Science*, 2(1):229–239, 2017.

[57] W. H. Fleming and S. K. Mitter. Optimal control and nonlinear filtering for nondegenerate diffusion processes. *Stochastics*, 8(1):63–77, 1982.

[58] L. R. Ford, Jr. and D. R. Fulkerson. *Flows in networks*. Princeton University Press, Princeton, N.J., 1962.

[59] G. F. Franklin, M. L. Workman, and D. Powell. *Digital Control of Dynamic Systems*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3rd edition, 1997.

[60] P. Gács. Reliable cellular automata with self-organization. *J. Statist. Phys.*, 103(1-2):45–267, 2001.

[61] N. Gans, G. Koole, and A. Mandelbaum. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, 5(2):79–141, 2003.

[62] K. Gardner and R. Righter. Product forms for FCFS queueing models with arbitrary server-job compatibilities: an overview. *Queueing Syst. Theory Appl.*, 96(1-2):3–51, 2020.

[63] K. Gardner, S. Zbarsky, S. Doroudi, M. Harchol-Balter, E. Hyytiä, and A. Scheller-Wolf. Queueing with redundant requests: exact analysis. *Queueing Systems*, 83(3-4):227–259, 2016.

[64] P. Guan, M. Raginsky, and R. Willett. Online Markov decision processes with Kullback-Leibler control cost. *IEEE Trans. Automat. Control*, 59(6):1423–1438, June 2014.

[65] O. Guéant, J.-M. Lasry, and P.-L. Lions. *Mean Field Games and Applications*, pages 205–266. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[66] I. Gurvich and A. Ward. On the dynamic control of matching queues. *Stoch. Syst.*, 4(2):479–523, 2014.

[67] H. Hao, Y. Lin, A. Kowli, P. Barooah, and S. Meyn. Ancillary service to the grid through control of fans in commercial building HVAC systems. *IEEE Trans. on Smart Grid*, 5(4):2066–2074, July 2014.

[68] H. Hao, T. Middelkoop, P. Barooah, and S. Meyn. How demand response from commercial buildings will provide the regulation needs of the grid. In *50th Allerton Conference on Communication, Control, and Computing*, pages 1908–1913, 2012.

[69] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent. Aggregate flexibility of thermostatically controlled loads. *IEEE Trans. on Power Systems*, 30(1):189–198, Jan 2015.

[70] J. M. Harrison and R. J. Williams. Brownian models of open queueing networks with homogeneous customer populations. *Stochastics*, 22(2):77–115, 1987.

[71] G. V. Houtum, W. Zijm, I. Adan, and J. Wessels. Bounds for performance characteristics: a systematic approach via cost structures. *Communications in Statistics. Stochastic Models*, 14(1-2):205–224, 1998.

[72] M. F. Howland, S. K. Lele, and J. O. Dabiri. Wind farm power optimization through wake steering. *Proceedings of the National Academy of Sciences*, 116(29):14495–14500, 2019.

[73] M. Huang, P. E. Caines, and R. P. Malhame. Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ε-Nash equilibria. *IEEE Trans. Automat. Control*, 52(9):1560–1571, 2007.

[74] M. Huang, R. P. Malhame, and P. E. Caines. Large population stochastic dynamic games: closed-loop Mckean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information and Systems*, 6(3):221–251, 2006.

[75] M. Huber. Perfect sampling using bounding chains. *The Annals of Applied Probability*, 14(2):734–753, 2004.

[76] E. Hyon and A. Jean-Marie. Scheduling services in a queuing system with impatience and setup costs. *The Computer Journal*, 55(5):553–563, 2012.

[77] J. M. Jonkman, J. Annoni, G. Hayman, B. Jonkman, and A. Purkayastha. *Development of FAST.Farm: A New Multi-Physics Engineering Tool for Wind-Farm Design and Analysis*. 2017.

[78] L. P. Kadanoff. More is the same; phase transitions and mean field theories. *J. Stat. Phys.*, 137(5-6):777–797, 2009.

[79] H. Kameda. How harmful the paradox can be in the braess/cohen-kelly-jeffries networks. In *Proceedings. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 1, pages 437–445. IEEE, 2002.

[80] E. H. Kaplan. *Managing the demand for public housing*. PhD thesis, Massachusetts Institute of Technology, 1984.

[81] M. Kárný. Towards fully probabilistic control design. *Automatica*, 32(12):1719 –1722, 1996.

[82] F. Kelly and C. Laws. Dynamic routing in open queueing networks: Brownian models, cut constraints and resource pooling. *Queueing Syst. Theory Appl.*, 13:47–86, 1993.

[83] F. P. Kelly. *Reversibility and stochastic networks*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester, 1979.

[84] W. S. Kendall and J. Moller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32(3):844–865, 2000.

[85] A. C. Kheirabadi and R. Nagamune. A quantitative review of wind farm control with the objective of wind farm power maximization. *Journal of Wind Engineering and Industrial Aerodynamics*, 192:45–73, 2019.

[86] J. F. C. Kingman. On queues in heavy traffic. *Journal of the Royal Statistical Society. Series B (Methodological).*, 24(2):383–392, 1962.

[87] A. Kizilkale and R. Malhame. Mean field based control of power system dispersed energy storage devices for peak load relief. In *Proc. of the IEEE Conf. on Dec. and Control*, pages 4971–4976, 2013.

[88] A. Kizilkale and R. Malhame. A class of collective target tracking problems in energy systems: Cooperative versus non-cooperative mean field control solutions. In *Proc. of the IEEE Conf. on Dec. and Control*, pages 3493–3498, 2014.

[89] I. Kontoyiannis and S. P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.*, 13:304–362, 2003.

[90] I. Kontoyiannis and S. P. Meyn. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.*, 10(3):61–123 (electronic), 2005.

[91] J. M. Lasry and P. L. Lions. Mean field games. *Japan. J. Math.*, 2:229–260, 2007.

[92] J.-S. Li. Ensemble control of finite-dimensional time-varying linear systems. *Trans. on Automatic Control*, 56(2):345–357, 2010.

[93] R. M. Loynes. The stability of a queue with non-independent inter-arrival and service times. *Mathematical Proceedings of the Cambridge Philosophical Society*, 58(3):497–520, 1962.

[94] R. M. Loynes. The stability of a queue with non-independent interarrival and service times. *Proc. Cambridge Philos. Soc.*, 58:497–520, 1962.

[95] D. G. Luenberger. *Optimization by vector space methods*. John Wiley & Sons Inc., New York, 1969. Reprinted 1997.

[96] O. Madani, S. Hanks, and A. Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artificial Intelligence*, 147(1-2):5–34, 2003.

[97] J. Mairesse and P. Moyal. Stability of the stochastic matching model. *J. Appl. Probab.*, 53(4):1064–1077, 2016.

[98] R. Malhamé and C.-Y. Chong. Electric load model synthesis by diffusion approximation of a high-order hybrid-state stochastic system. *IEEE Trans. Automat. Control*, 30(9):854 – 860, Sep 1985.

[99] A. S. Manne. Linear programming and sequential decisions. *Management Sci.*, 6(3):259–267, 1960.

[100] J. Mathias, A. Bušić, and S. Meyn. Demand dispatch with heterogeneous intelligent loads. In *Proc. 50th Annual Hawaii International Conference on System Sciences (HICSS), and arXiv 1610.00813*, 2017.

[101] J. Mathias, R. Kaddah, A. Bušić, and S. Meyn. Smart fridge / dumb grid? Demand Dispatch for the power grid of 2020. In *Proc. 49th Annual Hawaii International Conference on System Sciences (HICSS)*, pages 2498–2507, Jan 2016.

[102] J. Mathieu. *Modeling, Analysis, and Control of Demand Response Resources*. PhD thesis, University of California at Berkeley, 2012.

[103] J. Mathieu and D. Callaway. State estimation and control of heterogeneous thermostatically controlled loads for load following. In *45th International Conference on System Sciences*, pages 2002–2011, Hawaii, 2012. IEEE.

[104] J. Mathieu, S. Koch, and D. Callaway. State estimation and control of electric loads to manage real-time energy imbalance. *IEEE Trans. Power Systems*, 28(1):430–440, 2013.

[105] S. Meyn. Stability and asymptotic optimality of generalized MaxWeight policies. *SIAM J. Control Optim.*, 47(6):3259–3294, 2009.

[106] S. Meyn. *Control Systems and Reinforcement Learning*. Cambridge University Press, Cambridge, 2022.

[107] S. Meyn, P. Barooah, A. Bušić, Y. Chen, and J. Ehren. Ancillary service to the grid using intelligent deferrable loads. *IEEE Trans. Automat. Control*, 60(11):2847–2862, Nov 2015.

[108] S. Meyn, P. Barooah, A. Bušić, and J. Ehren. Ancillary service to the grid from deferrable loads: The case for intelligent pool pumps in Florida. In *Proc. of the IEEE Conf. on Dec. and Control*, pages 6946–6953, Dec 2013.

[109] S. P. Meyn. *Control Techniques for Complex Networks*. Cambridge University Press, 2007. Pre-publication edition available online.

[110] P. Moyal, A. Bušić, and J. Mairesse. A product form for the general stochastic matching model. *J. Appl. Probab.*, 58(2):449–468, 2021.

[111] P. Moyal and O. Perry. On the instability of matching queues. *Ann. Appl. Probab.*, 27(6):3385–3434, 2017.

[112] M. Nazari and A. L. Stolyar. Reward maximization in general dynamic matching systems. *Queueing Syst.*, 91(1-2):143–170, 2019.

[113] B. Oksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.

[114] N. Parikh and S. Boyd. *Proximal Algorithms*. Foundations and Trends in Optimization. Now Publishers, 2013.

[115] G. Peyré and M. Cuturi. Computational optimal transport. *arXiv:1803.00567*, 2020.

[116] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1-2):223–252, 1996.

[117] M. L. Puterman. *Markov Decision Processes*. Wiley, New York, 1994.

[118] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. Wiley series in probability and statistics. Wiley-Interscience, 2005.

[119] M. L. Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[120] Y. Rahme and P. Moyal. A stochastic matching model on hypergraphs. *Advances in Applied Probability*, 53(4):951–980, 2021.

[121] M. Roozbehani, M. A. Dahleh, and S. K. Mitter. Volatility of power grids under real-time pricing. *IEEE Transactions on Power Systems*, 27(4):1926–1940, 2012.

[122] P. J. Schweitzer. Perturbation theory and finite Markov chains. *J. Appl. Prob.*, 5:401–403, 1968.

[123] F. Schweppe, R. Tabors, J. Kirtley, H. Outhred, F. Pickel, and A. Cox. Homeostatic utility control. *IEEE Trans. on Power Apparatus and Systems*, PAS-99(3):1151 –1163, May 1980.

[124] F. C. Schweppe. Power systems '2000': Hierarchical control strategies. *IEEE Spectrum*, pages 42–47, July 1978.

[125] E. Shafieepoorfard, M. Raginsky, and S. P. Meyn. Rationally inattentive control of Markov processes. *SIAM J. Control Optim.*, 54(2):987–1016, 2016.

[126] J. Sharp. Electrical load disconnect device with electronic control, Dec. 2012. US Patent 8,328,110.

[127] C. A. Sims. Rational inattention: Beyond the linear-quadratic case. *The American economic review*, pages 158–163, 2006.

[128] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 12, 1999.

[129] A. Taghvaei and P. G. Mehta. A survey of feedback particle filter and related controlled interacting particle systems. *arXiv preprint arXiv:2301.00935*, 2023.

[130] S. H. Tindemans, V. Trovato, and G. Strbac. Decentralized control of thermostatic loads for flexible demand response. *IEEE Transactions on Control Systems Technology*, 23(5):1685–1700, Sept 2015.

[131] T.M.Liggett. *Interacting particle systems. Classics in Mathematics*. Springer-Verlag, Berlin, 2005. reprint of the 1985 original.

[132] E. Todorov. Linearly-solvable Markov decision problems. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Proc. Advances in Neural Information Processing Systems*, pages 1369–1376, Cambridge, MA, 2007.

[133] E. Todorov. Efficient computation of optimal actions. *Proceedings of the National Academy of Sciences*, 106(28):11478–11483, 2009.

[134] A. Toom, N. Vasilyev, O. Stavskaya, L. Mityushin, G. Kurdyumov, and S. Pirogov. Discrete local markov systems. In R. Dobrushin, V. Kryukov, and A. Toom, editors, *Stochastic cellular systems: ergodicity, memory, morphogenesis*. Manchester University Press, 1990.

[135] L. C. Totu. *Large Scale Demand Response of Thermostatic Loads*. PhD thesis, Faculty of Engineering and Science, Aalborg University, 2015.

[136] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[137] J. Visschers, I. Adan, and G. Weiss. A product form solution to a system with multi-type jobs and multi-type servers. *Queueing Syst.*, 70(3):269–298, 2012.

[138] J. W. C. H. Visschers. *Random walks with geometric jumps*. Eindhoven University of Technology, Eindhoven, 2000. Dissertation, Technische Universiteit Eindhoven, Eindhoven, 2000.

[139] P. Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *J. Phys. Theor. Appl.*, 6(1):661–690, 1907.

[140] H. Yin, P. Mehta, S. Meyn, and U. Shanbhag. Synchronization of coupled oscillators is a game. *IEEE Trans. on Automatic Control*, 57(4):920–935, 2012.

[141] Y. Q. Zhao and D. Liu. The censored markov chain and the best augmentation. *Journal of Applied Probability*, 33(3):623–629, 1996.

[142] C. Ziras, E. Vrettos, and G. Andersson. Primary frequency control with refrigerators under startup dynamics and lockout constraints. In *IEEE Power & Energy Society General Meeting*, 2015.

RÉSUMÉ

Ce manuscrit résume une partie de mes travaux de recherche. L'accent est mis sur la décision et le contrôle dans les réseaux avec une demande et une offre stochastiques qui doivent être équilibrées par une entité centrale, ou en utilisant une architecture de contrôle distribuée. Deux contextes différents sont considérés : les systèmes d'appariement stochastique et l'équilibrage en temps réel de la demande et de l'offre stochastiques dans les réseaux électriques. L'objectif est similaire, mais les modèles diffèrent considérablement et utilisent des techniques issues de divers domaines, notamment les systèmes de files d'attente et les réseaux de flot de la théorie des graphes pour l'appariement stochastique, ainsi que les approximations du champ moyen et la théorie du contrôle pour l'équilibrage de l'offre et de la demande en temps réel. Les deux utilisent des processus markoviens de décision.

Le manuscrit est organisé en deux parties : la première partie traite les systèmes d'appariement stochastique et la deuxième, l'équilibrage de la demande et de l'offre stochastiques dans les réseaux électriques. Les deux parties peuvent être lues indépendamment l'une de l'autre.

ABSTRACT

This manuscript summarizes one part of my research. The focus is on the decision and control in networks with stochastic demand and supply that have to be balanced by a central entity, or using a distributed control design. Two different settings are considered: stochastic matching systems and real-time balancing of stochastic demand and supply in power grids. The objective is similar, yet the models differ considerably and use techniques from various fields including queueing systems and network flows from graph theory for stochastic matching, and mean-field approximations and control theory for real-time demand-supply balancing. Both settings use Markov decision processes.

The manuscript is organized in two parts: Part I covers stochastic matching systems and Part II balancing of stochastic demand and supply in power grids. The two parts can be read independently of each other.