Optimisation Combinatoire et Convexe.

Approximation results

Today

- Semidefinite relaxations
- Lagrangian relaxations for general QCQPs
- Randomization
- Bounds on suboptimality (MAXCUT)
- Exact relaxations, $\mathcal{S}$-lemma
- Concentration arguments
- Approximate $\mathcal{S}$-lemma
- Problems on graphs

Convex Optimization

Convex problem:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0, \ldots, f_m(x) \leq 0 \end{array}$$

$x \in \mathbb{R}^n$ is optimization variable; $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are **convex**:

$$f_i(\lambda x + (1-\lambda)y) \leq \lambda f_i(x) + (1-\lambda) f_i(y)$$

for all x, y, $0 \leq \lambda \leq 1$

- includes LS, LP, QP, and **many others**
- like LS, LP, and QP, convex problems are **fundamentally tractable**

Nonconvex Problems

Nonconvexity makes problems **essentially untractable**...

- Sometimes the result of bad problem formulation
- Natural limitation: fixed transaction costs, binary communications, ...

What can be done?... Use convex optimization results to

- Get exact solutions in **rare** cases.
- Find bounds on the optimal value, by **relaxation**.
- Get "good" feasible points via **randomization**.

Nonconvex Problems

- Focus first on a specific class of problems: general QCQPs
- Large range of applications...

A generic QCQP can be written

$$\begin{array}{ll} \mbox{minimize} & x^T P_0 x + q_0^T x + r_0 \\ \mbox{subject to} & x^T P_i x + q_i^T x + r_i \leq 0, \quad i=1,\ldots,m \end{array}$$

- If all P_i are p.s.d., this is a convex problem...
- We suppose at least one P_i is not p.s.d.

Example: Boolean Least Squares

Boolean least-squares problem:

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

- basic problem in digital communications
- could check all 2^n possible values of x . . .
- an NP-hard problem, and very hard in practice
- many heuristics for approximate solution

Example: Partitioning Problem

two-way partitioning problem described in §5.1.4 of the textbook

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

where $W \in \mathbf{S}^n$, with $W_{ii} = 0$.

- a feasible x corresponds to the partition

$$\{1, \ldots, n\} \;=\; \{i \mid x_i = -1\} \;\cup\; \{i \mid x_i = 1\}$$

- the matrix coefficient W_{ij} can be interpreted as the cost of having the elements i and j in the same partition.
- the objective is to find the partition with least total cost
- classic particular instance: MAXCUT $(W_{ij} \geq 0)$

Convex Relaxation

The original QCQP

$$\begin{array}{ll}\mbox{minimize} & x^TP_0x+q_0^Tx+r_0\\ \mbox{subject to} & x^TP_ix+q_i^Tx+r_i\leq 0,\quad i=1,\ldots,m\end{array}$$

can be bounded by, after writing $X=xx^T$,

$$\begin{array}{ll}\mbox{minimize} & \mathbf{Tr}(XP_0)+q_0^Tx+r_0\\ \mbox{subject to} & \mathbf{Tr}(XP_i)+q_i^Tx+r_i\leq 0,\quad i=1,\ldots,m\\ & X\succeq xx^T\\ & \mathbf{Rank}(X)=1\end{array}$$

the only nonconvex constraint is now $\mathbf{Rank}(X)=1$...

Convex Relaxation: Semidefinite Relaxation

- We can directly relax this last constraint, i.e. drop the nonconvex $\mathbf{Rank}(X) = 1$ to keep only $X \succeq xx^T$
- The resulting program gives a lower bound on the optimal value

$$\begin{array}{ll} \mbox{minimize} & \mathbf{Tr}(XP_0) + q_0^T x + r_0 \\ \mbox{subject to} & \mathbf{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\ & X \succeq xx^T \end{array}$$

Tricky. . . Can be improved?

Lagrangian Relaxation

From the original problem

$$\begin{array}{ll}\text{minimize} & x^T P_0 x + q_0^T x + r_0\\ \text{subject to} & x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m\end{array}$$

We can form the Lagrangian:

$$L(x,\lambda) = x^T \left(P_0 + \sum_{i=1}^m \lambda_i P_i\right) x + \left(q_0 + \sum_{i=1}^m \lambda_i q_i\right)^T x + r_0 + \sum_{i=1}^m \lambda_i r_i$$

in the variables $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}_+^m$...

Lagrangian Relaxation: Lagrangian

The dual can be computed explicitly as an (unconstrained) quadratic minimization problem, with:

$$\inf_{x\in\mathbb{R}} x^TPx + q^Tx + r = \begin{cases} r - \frac{1}{4}q^TP^\dagger q, & \text{if } P \succeq 0 \text{ and } q \in \mathcal{R}(P) \\ -\infty, & \text{otherwise} \end{cases}$$

we have:

$$\begin{aligned} \inf_x L(x,\lambda) = \quad & -\tfrac{1}{4}\left(q_0 + \textstyle\sum_{i=1}^m \lambda_i q_i\right)^T \left(P_0 + \textstyle\sum_{i=1}^m \lambda_i P_i\right)^\dagger \left(q_0 + \textstyle\sum_{i=1}^m \lambda_i q_i\right) \\ & + \textstyle\sum_{i=1}^m \lambda_i r_i + r_0 \end{aligned}$$

where we recognize a Schur complement...

Lagrangian Relaxation: Dual

The dual of the QCQP is then given by:

$$
\begin{array}{ll}
\mbox{maximize} & \gamma + \sum_{i=1}^m \lambda_i r_i + r_0 \\
\mbox{subject to} & \left[\begin{array}{cc} \left(P_0 + \sum_{i=1}^m \lambda_i P_i\right) & \left(q_0 + \sum_{i=1}^m \lambda_i q_i\right)/2 \\ \left(q_0 + \sum_{i=1}^m \lambda_i q_i\right)^T/2 & -\gamma \end{array}\right] \succeq 0 \\
& \lambda_i \geq 0, \quad i = 1, \ldots, m
\end{array}
$$

which is a semidefinite program in the variable $\lambda \in \mathbb{R}^m$ and can be solved efficiently.

Use semidefinite duality to compute the dual of this last program?

Lagrangian Relaxation: Bidual

Taking the dual again, we get an SDP is given by

$$\begin{array}{ll}\text{minimize} & \mathbf{Tr}(XP_0) + q_0^T x + r_0 \\ \text{subject to} & \mathbf{Tr}(XP_i) + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\ & \left[\begin{array}{cc} X & x^T \\ x & 1 \end{array}\right] \succeq 0 \end{array}$$

in the variables $X \in \mathbf{S}^n$ and $x \in \mathbb{R}^n$

- This is a convex relaxation of the original program
- We have recovered the semidefinite relaxation in an "automatic" way

Lagrangian Relaxation: Boolean LS

Using the previous technique, we can relax the original Boolean LS problem

$$\begin{array}{ll}\mbox{minimize} & \|Ax-b\|^2\\ \mbox{subject to} & x_i^2=1, \quad i=1,\ldots,n\end{array}$$

and relax it as an SDP

$$\begin{array}{ll}\mbox{minimize} & \mathbf{Tr}(AX)+2b^TAx+b^Tb\\ \mbox{subject to} & \left[\begin{array}{cc} X & x^T\\ x & 1\end{array}\right]\succeq 0\\ & X_{ii}=1, \quad i=1,\ldots,n\end{array}$$

this program then produces a lower bound on the optimal value of the original Boolean LS program

Lagrangian Relaxation: Partitioning

The partitioning problem defined above is

$$\begin{array}{ll}\text{minimize} & x^T W x\\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n\end{array}$$

the variable x disappears from the relaxation, which becomes

$$\begin{array}{ll}\text{minimize} & \mathbf{Tr}(WX)\\ \text{subject to} & X \succeq 0\\ & X_{ii} = 1, \quad i = 1, \ldots, n\end{array}$$

Feasible points?

- Lagrangian relaxations only provide lower bounds on the optimal value
- Can we compute good feasible points?
- Can we measure how suboptimal this lower bound is?

Randomization

The original QCQP

$$\begin{array}{ll}\text{minimize} & x^TP_0x+q_0^Tx+r_0\\ \text{subject to} & x^TP_ix+q_i^Tx+r_i\leq 0,\quad i=1,\ldots,m\end{array}$$

was relaxed into

$$\begin{array}{ll}\text{minimize} & \mathbf{Tr}(XP_0)+q_0^Tx+r_0\\ \text{subject to} & \mathbf{Tr}(XP_i)+q_i^Tx+r_i\leq 0,\quad i=1,\ldots,m\\ & \left[\begin{array}{cc} X & x^T\\ x & 1\end{array}\right]\succeq 0\end{array}$$

- The last (Schur complement) constraint is equivalent to $X-xx^T\succeq 0$
- Hence, if x and X are the solution to the relaxed program, then $X-xx^T$ is a covariance matrix...

Randomization

- Pick x as a Gaussian variable with $x \sim \mathcal{N}(x, X - xx^T)$
- x will solve the QCQP "on average" over this distribution

In other words, it will satisfy

$$\begin{array}{ll} \text{minimize} & \mathbf{E}[x^T P_0 x + q_0^T x + r_0] \\ \text{subject to} & \mathbf{E}[x^T P_i x + q_i^T x + r_i] \leq 0, \quad i = 1, \ldots, m \end{array}$$

a good feasible point can then be obtained by sampling enough x...

Linearization

Consider the constraint

$$x^T P x + q^T x + r \leq 0$$

we decompose the matrix P into its positive and negative parts

$$P = P_+ - P_-, \quad P_+,\ P_- \succeq 0$$

and original constraint becomes

$$x^T P_+ x + q_0^T x + r_0 \leq x^T P_- x$$

Linearization

Both sides of the inequality are now convex quadratic functions. We linearize the right hand side around an initial feasible point x_0 to obtain

$$x^T P_+ x + q_0^T x + r_0 \leq x^{(0)T} P_- x^{(0)} + 2x^{(0)T} P_-(x - x^{(0)})$$

- The right hand side is now an affine lower bound on the original function $x^T P_- x$ (see §3.1.3 in the book).
- The resulting constraint is convex and more conservative than the original one, hence the feasible set of the new problem will be a convex subset of the original feasible set
- We form a *convex restriction* of the problem

We can then solve the convex restriction to get a better feasible point $x^{(1)}$ and iterate...

Bounds on suboptimality

- In certain particular cases, it is possible to get a hard bound on the gap between the optimal value and the relaxation result
- A classical example is that of the MAXCUT bound

The MAXCUT problem is a particular case of the partitioning problem:

$$\begin{array}{ll} \text{maximize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

its Lagrangian relaxation is computed as:

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(WX) \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{array}$$

Bounds on suboptimality: MAXCUT

Let X be a solution to this program

- we look for a feasible point by sampling a normal distribution $\mathcal{N}(0, X)$
- we convert each sample point x to a feasible point by rounding it to the nearest value in $\{-1, 1\}$, i.e. taking

$$\hat{x} = \mathbf{sgn}(x)$$

crucially, when $\hat{x}$ is sampled using that procedure, the expected value of the objective $\mathbf{E}[\hat{x}^T W x]$ can be computed explicitly:

$$\mathbf{E}[\hat{x}^T W x] = \frac{2}{\pi} \sum_{i,j=1}^{n} W_{ij} \arcsin(X_{ij}) = \frac{2}{\pi} \mathbf{Tr}(W \arcsin(X))$$

Bounds on suboptimality: MAXCUT

- We are guaranteed to reach this expected value $2/\pi\,\mathbf{Tr}(W\arcsin(X))$ after sampling a few (feasible) points $\hat{x}$
- Hence we know that the optimal value OPT of the MAXCUT problem is between $2/\pi\,\mathbf{Tr}(W\arcsin(X))$ and $\mathbf{Tr}(WX)$

Furthermore, with $\arcsin(X) \succeq X$, we can simplify (and relax) the above expression to get:

$$\frac{2}{\pi}\mathbf{Tr}(WX) \leq OPT \leq \mathbf{Tr}(WX)$$

the procedure detailed above guarantees that we can find a feasible point at most $2/\pi$ suboptimal

Bounds on suboptimality: MAXCUT

Proposition

MAXCUT approximation. *Let OPT be the optimal value of the partitioning problem*

$$\begin{array}{ll} \textit{maximize} & x^T W x \\ \textit{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

where $W \succeq 0$, and let SDP be the optimal value of its Lagrangian relaxation

$$\begin{array}{ll} \textit{maximize} & \mathbf{Tr}(WX) \\ \textit{subject to} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{array}$$

then we have $\frac{2}{\pi}\mathbf{Tr}(WX) \leq OPT \leq \mathbf{Tr}(WX)$.

Proof. Suppose we sample $x \sim \mathcal{N}(0, X)$ then take $\hat{x} = \mathbf{sgn}(x)$. We get

$$\mathbf{E}[\hat{x}^T W x] = \frac{2}{\pi} \sum_{i,j=1}^{n} W_{ij} \arcsin(X_{ij}) = \frac{2}{\pi} \mathbf{Tr}(W \arcsin(X))$$

where $\arcsin(X)$ is taken elementwise, with

$$\arcsin(X)_{ij} = X_{ij} + \sum_{k=1}^{\infty} \frac{1 \cdot 3 \cdots (2k-1)}{2^k k!(2k+1)} X_{ij}^{2k+1}$$

which means

$$\arcsin(X) - X \succeq 0$$

because if we define the **elementwise matrix power** $[X]^k$ such that

$$[X]_{ij}^k = X_{ij}$$

then $[X]^k \succeq 0$ when $X \succeq 0$. This finally means that

$$\mathbf{E}[\hat{x}^T W x] = \frac{2}{\pi} \mathbf{Tr}(W \arcsin(X)) \geq \frac{2}{\pi} \mathbf{Tr}(WX) \quad \blacksquare$$

Numerical Example: Boolean LS

Boolean least-squares problem:

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|^2 \\ \text{subject to} & x_i^2 = 1, \quad i=1,\ldots,n \end{array}$$

with

$$\begin{aligned} \|Ax-b\|^2 &= x^TA^TAx - 2b^TAx + b^Tb \\ &= \mathbf{Tr}\, A^TAX - 2b^TA^Tx + b^Tb \end{aligned}$$

where $X = xx^T$, hence can express BLS as

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, A^TAX - 2b^TAx + b^Tb \\ \text{subject to} & X_{ii} = 1, \quad X \succeq xx^T, \quad \text{rank}(X) = 1 \end{array}$$

. . . still a very hard problem

SDP relaxation for BLS

Using Lagrangian relaxation, with

$$X \succeq xx^T \iff \left[\begin{array}{cc} X & x \\ x^T & 1 \end{array}\right] \succeq 0$$

we obtained the **SDP relaxation** (with variables X, x)

$$\begin{array}{ll} \mbox{minimize} & \mathbf{Tr}\, A^TAX - 2b^TA^Tx + b^Tb \\ \mbox{subject to} & X_{ii} = 1, \quad \left[\begin{array}{cc} X & x \\ x^T & 1 \end{array}\right] \succeq 0 \end{array}$$

- Optimal value of SDP gives **lower bound** for BLS
- If optimal matrix is rank one, we're done

Interpretation via randomization

- Can think of variables X, x in SDP relaxation as defining a normal distribution $z \sim \mathcal{N}(x, X - xx^T)$, with $\mathbf{E}\, z_i^2 = 1$
- SDP objective is $\mathbf{E}\, \|Az - b\|^2$

suggests randomized method for BLS:

- Find X^{opt}, x^{opt}, optimal for SDP relaxation
- Generate z from $\mathcal{N}(x^{\mathrm{opt}}, X^{\mathrm{opt}} - x^{\mathrm{opt}} x^{\mathrm{opt}T})$
- Take $x = \mathbf{sgn}(z)$ as approximate solution of BLS (can repeat many times and take best one)

Example

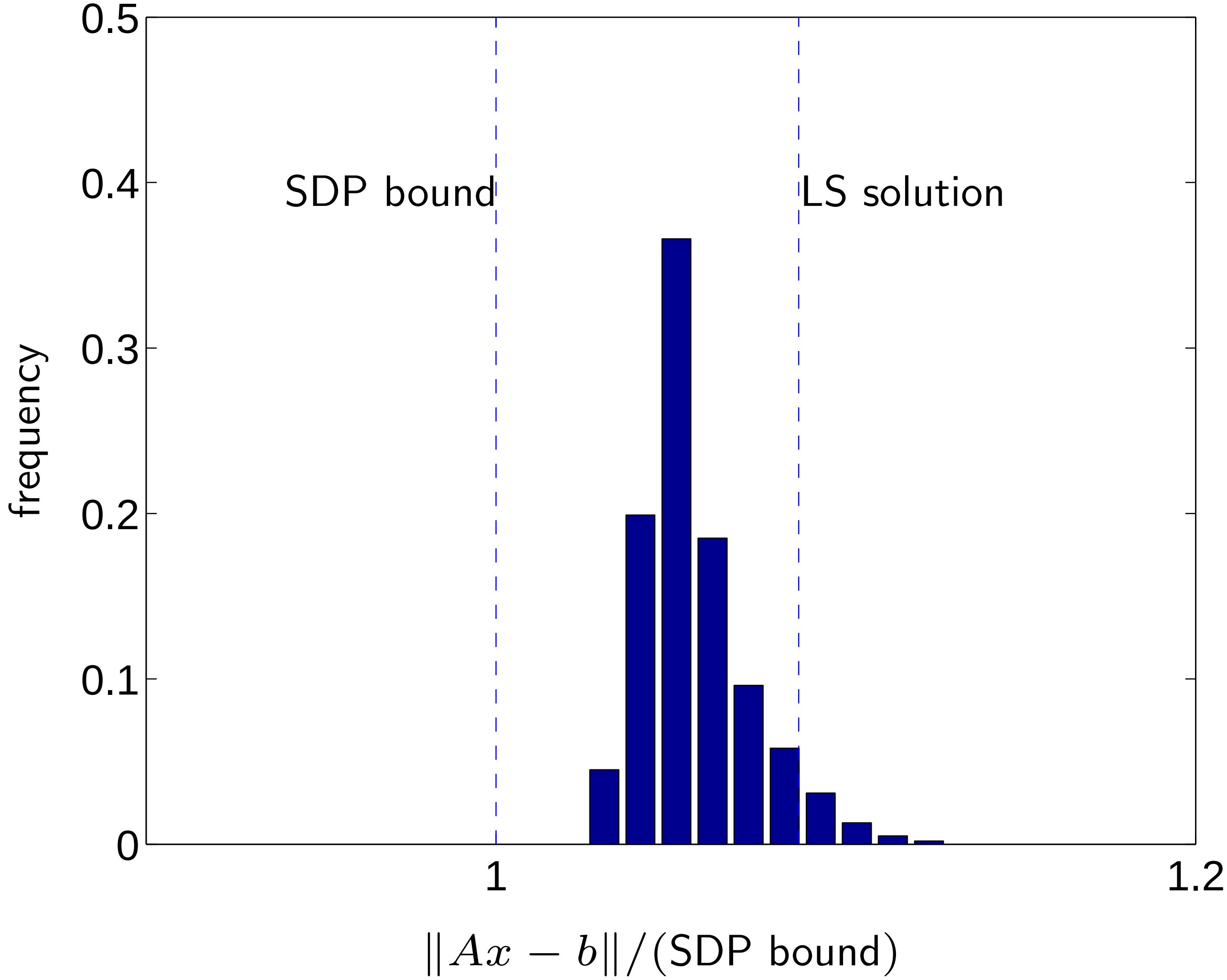
- (randomly chosen) parameters $A \in \mathbb{R}^{150\times100}$, $b \in \mathbb{R}^{150}$
- $x \in \mathbb{R}^{100}$, so feasible set has $2^{100} \approx 10^{30}$ points

LS approximate solution: minimize $\|Ax - b\|$ s.t. $\|x\|^2 = n$, then round yields objective 8.7% over SDP relaxation bound

randomized method: (using SDP optimal distribution)

- best of 20 samples: 3.1% over SDP bound
- best of 1000 samples: 2.6% over SDP bound

Example: Partitioning Problem



Example: Partitioning Problem

MAXCUT. Numerical example.

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

the Lagrange dual of this problem is given by the SDP:

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

Partitioning: Lagrangian relaxation

the dual of this SDP is another SDP

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, WX \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{array}$$

the solution X^{opt} gives a lower bound on the optimal value p^{opt} of the partitioning problem

- solve the previous SDP to find X^{opt} (and the bound p^{opt})
- let v denote an eigenvector of X^{opt} associated with its largest eigenvalue
- now let

$$\hat{x} = \mathbf{sgn}(v)$$

the vector $\hat{x}$ is our guess for a good partition

Partitioning: Randomization

Randomization.

- we generate independent samples $x^{(1)}, \ldots, x^{(K)}$ from a normal distribution with zero mean and covariance X^{opt}
- for each sample we consider the heuristic approximate solution

$$\hat{x}^{(k)} = \mathbf{sgn}(x^{(k)})$$

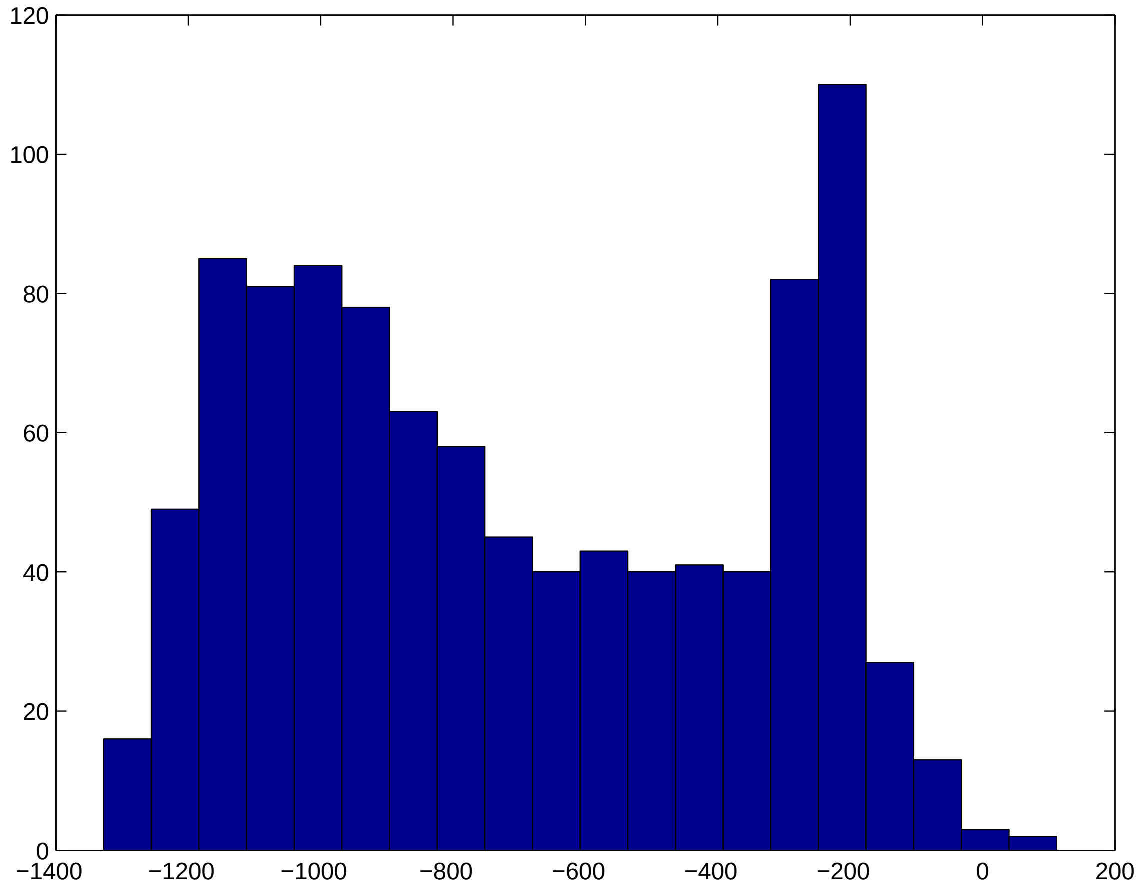
- we then take the one with lowest cost

On a randomly chosen problem:

- The optimal SDP lower bound p^{opt} is equal to -1641
- The simple $\mathbf{sgn}(x)$ heuristic gives a partition with total cost -1280

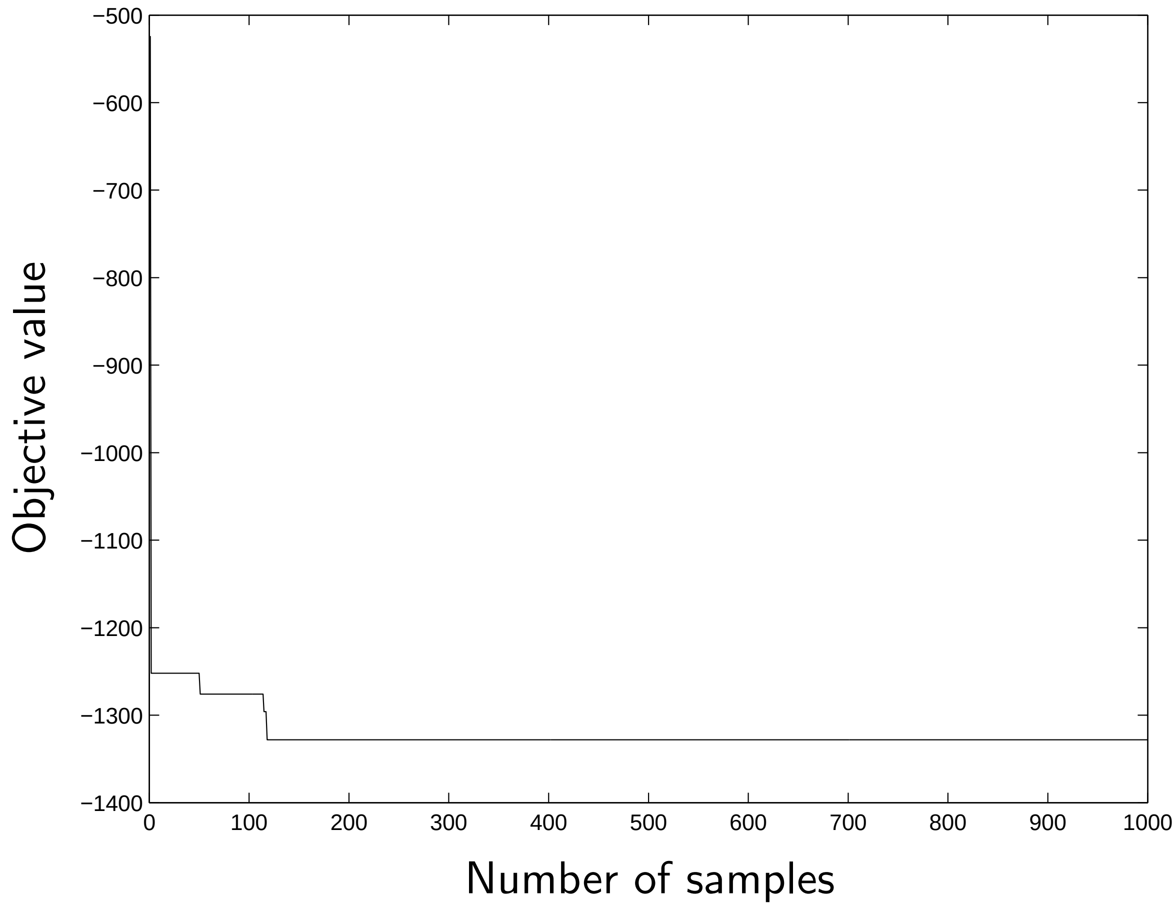
At this point, we can say that the optimal value is between -1641 and -1280

Partitioning: Numerical Example



Histogram of the objective obtained by the randomized heuristic, over 1000 samples: the minimum value reached here is -1328

Partitioning: Numerical Example



We know the optimum is between -1641 and -1328.

Greedy method

We can improve these results a little bit using the following simple **greedy heuristic**

- Suppose the matrix $Y = \hat{x}^T W \hat{x}$ has a column j whose sum $\sum_{i=1}^n y_{ij}$ is positive.
- Switching $\hat{x}_j$ to $-\hat{x}_j$ will decrease the objective by $2\sum_{i=1}^n y_{ij}$.
- if we pick the column y_{j_0} with largest sum, switch $\hat{x}_{j0}$ and repeat until all column sums are negative, we decrease the objective.

Applying this to the SDP heuristic gives an objective value of -1372, our best partition yet...

Hidden convexity, $\mathcal{S}$-lemma, . . .

$\mathcal{S}$-lemma

- In general, nonconvex quadratically constrained quadratic programming is hard.
- Yet, we have very efficient, very reliable algorithms to solve the following eigenvalue problem

$$\begin{array}{ll} \text{maximize} & x^T A x \\ \text{subject to} & x^T x = 1 \end{array}$$

  with complexity $O(n^2)$ (computing a sequence of matrix vector products).
- Why is this one easy?

$\mathcal{S}$-lemma. SDP relaxations of Nonconvex QPs with one quadratic constraint (two in some cases) are exact, hence these programs can be solved in polynomial time.

$\mathcal{S}$-lemma

Geometrically, the set $(x^T Ax, x^T Bx)$, where $x \in \mathbb{R}^n$, describes a **convex cone**. This has important consequences for semidefinite relaxations.

Proposition

Quadratic convexity. *Suppose $A, B \in \mathbf{S}_n$, then for all $X \succeq 0$, there exists $x \in \mathbb{R}^n$ with*

$$x^T Ax = \mathbf{Tr}(AX) \quad \text{and} \quad x^T Bx = \mathbf{Tr}(BX). \tag{1}$$

Proof. Suppose it is true for all $X \in \mathbf{S}_+^n$ with $2 \leq \mathbf{Rank}\, X \leq k$, i.e. there exists an x such that (1) holds. Let us show that it also holds if $\mathbf{Rank}\, X = k + 1$.

A matrix $X \in \mathbf{S}_+^n$ with $\mathbf{Rank}\, X = k + 1$ can be expressed as $X = yy^T + Z$ where $y \neq 0$ and $Z \in \mathbf{S}_+^n$ with $\mathbf{Rank}\, Z = k$. By assumption, there exists a z such that $\mathbf{Tr}(AZ) = z^T Az$, $\mathbf{Tr}(AZ) = z^T Bz$. Therefore

$$\mathbf{Tr}(AX) = \mathbf{Tr}(A(yy^T + zz^T)), \qquad \mathbf{Tr}(BX) = \mathbf{Tr}(B(yy^T + zz^T)).$$

$yy^T + zz^T$ has rank one or two, hence (1) by assumption.

It is therefore sufficient to prove the result if $\mathbf{Rank}\, X = 2$. If $\mathbf{Rank}\, X = 2$, we can factor X as $X = VV^T$ where $V \in \mathbb{R}^{n\times 2}$, with linearly independent columns v_1 and v_2.

Without loss of generality we can assume that V^TAV is diagonal. (If V^TAV is not diagonal we replace V with VP where $V^TAV = P\,\mathbf{diag}(\lambda)P^T$ is the eigenvalue decomposition of V^TAV.) We will write V^TAV and V^TBV as

$$V^TAV = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \qquad V^TBV = \begin{bmatrix} \sigma_1 & \gamma \\ \gamma & \sigma_2 \end{bmatrix},$$

and define

$$w = \begin{bmatrix} \mathbf{Tr}(AX) \\ \mathbf{Tr}(BX) \end{bmatrix} = \begin{bmatrix} \lambda_1 + \lambda_2 \\ \sigma_1 + \sigma_2 \end{bmatrix}.$$

We need to show that $w = (x^TAx, x^TBx)$ for some x. We distinguish two cases.

- First, assume $(0,\gamma)$ is a linear combination of the vectors (λ_1,σ_1) and (λ_2,σ_2):

$$0 = z_1\lambda_1 + z_2\lambda_2, \qquad \gamma = z_1\sigma_1 + z_2\sigma_2,$$

for some z_1, z_2. In this case we choose $x = \alpha v_1 + \beta v_2$, where α and β are determined by solving two quadratic equations in two variables

$$\alpha^2 + 2\alpha\beta z_1 = 1, \qquad \beta^2 + 2\alpha\beta z_2 = 1. \tag{2}$$

This will give the desired result, since

$$\begin{aligned}
&\left[\begin{array}{c} (\alpha v_1 + \beta v_2)^T A(\alpha v_1 + \beta v_2) \\ (\alpha v_1 + \beta v_2)^T B(\alpha v_1 + \beta v_2) \end{array}\right] \\
&= \alpha^2 \left[\begin{array}{c} \lambda_1 \\ \sigma_1 \end{array}\right] + 2\alpha\beta \left[\begin{array}{c} 0 \\ \gamma \end{array}\right] + \beta^2 \left[\begin{array}{c} \lambda_2 \\ \sigma_2 \end{array}\right] \\
&= (\alpha^2 + 2\alpha\beta z_1) \left[\begin{array}{c} \lambda_1 \\ \sigma_1 \end{array}\right] + (\beta^2 + 2\alpha\beta z_2) \left[\begin{array}{c} \lambda_2 \\ \sigma_2 \end{array}\right] \\
&= \left[\begin{array}{c} \lambda_1 + \lambda_2 \\ \sigma_1 + \sigma_2 \end{array}\right].
\end{aligned}$$

It remains to show that the equations (2) are solvable. To see this, we first note that α and β must be nonzero, so we can write the equations equivalently as

$$\alpha^2(1+2(\beta/\alpha)z_1) = 1, \qquad (\beta/\alpha)^2 + 2(\beta/\alpha)(z_2 - z_1) = 1.$$

The equation $t^2 + 2t(z_2 - z_1) = 1$ has a positive and a negative root. At least one of these roots (the root with the same sign as z_1) satisfies $1 + 2tz_1 > 0$, so we can choose

$$\alpha = \pm 1/\sqrt{1+2tz_1}, \qquad \beta = t\alpha.$$

This yields two solutions (α, β) that satisfy (2). (If both roots of $t^2 + 2t(z_2 - z_1) = 1$ satisfy $1 + 2tz_1 > 0$, we obtain four solutions.)

- Next, assume that $(0, \gamma)$ is not a linear combination of (λ_1, σ_1) and (λ_2, σ_2). In particular, this means that (λ_1, σ_1) and (λ_2, σ_2) are linearly dependent. Therefore their sum $w = (\lambda_1 + \lambda_2, \sigma_1 + \sigma_2)$ is a nonnegative multiple of (λ_1, σ_1), or (λ_2, σ_2), or both. If $w = \alpha^2(\lambda_1, \sigma_1)$ for some α, we can choose $x = \alpha v_1$. If $w = \beta^2(\lambda_2, \sigma_2)$ for some β, we can choose $x = \beta v_2$. ■

Strong duality.

- This shows directly that strong duality holds for

$$\begin{array}{ll} \text{maximize} & x^T A x \\ \text{subject to} & x^T B x \leq 0 \end{array}$$

  since the optimum value of this program is equal to that of its bidual, which is a semidefinite program in X.

- Some extensions of this result are possible, e.g. the inhomogeneous case

$$\begin{array}{ll} \text{maximize} & x^T A x + a^T x \\ \text{subject to} & x^T B x + b^T x + c = 0 \end{array}$$

  or the normalized case (known as Brickman's theorem)

$$\begin{array}{ll} \text{maximize} & x^T A x \\ \text{subject to} & x^T B x \leq 0 \\ & x^T x = 1 \end{array}$$

Concentration inequalities, approximate $\mathcal{S}$-lemma, etc. . .

Concentration inequalities

- We can extend randomization arguments to constrained problems.
- Concentration inequalities allow us to bound the probability that a constraint is feasible. Basically, if we match the constraints/objective on average, we can find w.h.p. a feasible point whose objective value is not too far off.

Theorem

Gaussian concentration. *Suppose $f(x) : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous with constant L with respect to the Euclidean norm, i.e.*

$$|f(y) - f(x)| \leq L\|x - y\|_2, \quad \textit{for all } x, y \in \mathbb{R}^n$$

then if g_i, $i = 1, \ldots, n$ are i.i.d. Gaussian variables with $g_i \sim \mathcal{N}(0,1)$, we have

$$\mathbf{Prob}\left[|M - f(g)| \geq Lt\right] \leq \exp(-t^2/2)$$

where $M = \mathbf{E}[f(g)]$ or its median.

Concentration inequalities

Similar concentration results also exist for binary random variables.

Theorem

Bernstein inequality. *Let* $u_i \in \{-1, 1\}$ *be i.i.d. random variables with* $\mathbf{E}[u_i] = 0$*, for any* $a \in \mathbb{R}^n$ *we have*

$$\textbf{Prob}\left[|a^T u| \geq t\|a\|_2\right] \leq \exp(-t^2/4)$$

Approximate $\mathcal{S}$-lemma

We can show the following result extending the $\mathcal{S}$-lemma to approximate the case with multiple quadratic constraints. (Inhomogeneous extensions are possible).

Theorem

Approximate $\mathcal{S}$-lemma. *Call OPT the optimal value of the following quadratic optimization problems*

$$\begin{array}{ll} \textit{maximize} & x^T A x \\ \textit{subject to} & x^T A_i x \leq c_i, \quad i = 1, \ldots, m \end{array}$$

in the variable $x \in \mathbb{R}^n$, where the matrix $A \in \mathbf{S}_n$ is arbitrary, $c_i > 0$, and $A_i \succeq 0$. Call SDP the optimal value of the semidefinite program (we assume strong duality holds and $SDP < \infty$)

$$\begin{array}{ll} \textit{maximize} & \mathbf{Tr}(AX) \\ \textit{subject to} & \mathbf{Tr}(A_i X), \quad i = 1, \ldots, m \end{array}$$

in the variable $X \in \mathbf{S}_n$. Then $OPT \leq SDP \leq 2 \ln \left(2 \sum_{i=1}^m \mathbf{Rank}(A_i)\right) OPT$.

Proof. We write X an optimal solution to SDP and $X^{1/2}AX^{1/2} = UDU^T$, the eigenvalue decomposition of $X^{1/2}AX^{1/2}$, with D diagonal and U orthogonal. We have, by construction

$$\mathbf{Tr}(D) = \mathbf{Tr}(UDU^T) = \mathbf{Tr}(X^{1/2}AX^{1/2}) = \mathbf{Tr}(AX) = SDP$$

We let $\xi_i \in \{-1, 1\}$ be i.i.d. random variables with $\mathbf{E}[\xi] = 0$. We define $\eta = X^{1/2}U\xi$, and write $D_i = U^T X^{1/2} A_i X^{1/2} U$, such that

$$\mathbf{Tr}(D_i) = \mathbf{Tr}(U^T X^{1/2} A_i X^{1/2} U) = \mathbf{Tr}(X^{1/2} A_i X^{1/2}) = \mathbf{Tr}(A_i X) \leq c_i$$

this means

$$\begin{array}{rcl} \eta^T A \eta & = & \xi^T U^T X^{1/2} A X^{1/2} U \xi \\ & = & \xi^T U^T U D U^T U \xi \\ & = & \xi^T D \xi = \mathbf{Tr}(D) = SDP, \end{array}$$

and similarly,

$$\begin{array}{rcl} \mathbf{E}[\eta^T A_i \eta] & = & \mathbf{E}[\xi^T U^T X^{1/2} A_i X^{1/2} U \xi] \\ & = & \mathbf{Tr}(U^T X^{1/2} A_i X^{1/2} U) = \mathbf{Tr}(A_i X) \leq c_i \end{array}$$

This shows that the vector η solves the SDP "on average". We now show how to construct vectors that satisfy approximately solve the QP with high probability. We can write

$$D_i = \sum_{j=1}^{k} d_j d_j^T, \quad k = \mathbf{Rank}(D_i) = \mathbf{Rank}(A_i),$$

Using the previous concentration inequality

$$\mathbf{Prob}[|d_j^T \xi| \geq \sqrt{t}\|d_j\|_2] \leq 2\exp(-t/2),$$

now, for each given ξ, if $\xi^T D_i \xi \geq t \sum_{j=1}^k \|d_j\|_2^2$ then for at least for one j, we have $|d_j^T \xi| \geq \sqrt{t}\|d_j\|_2$, hence

$$\begin{aligned}\mathbf{Prob}[\xi^T D_i \xi \geq t \sum_{j=1}^{k} \|d_j\|_2^2] &\leq \sum_{i=1}^{k} \mathbf{Prob}[|d_j^T \xi| \geq \sqrt{t}\|d_j\|_2] \\ &\leq 2\,\mathbf{Rank}(D_i)\exp(-t/2)\end{aligned}$$

Now, we have $\sum_{j=1}^k \|d_j\|_2^2 = \mathbf{Tr}(\sum_{j=1}^k d_j d_j^T) = \mathbf{Tr}(D_i) \leq c_i$. Hence we have showed

$$\mathbf{Prob}[\eta^T A_i \eta \geq t c_i] \leq 2\,\mathbf{Rank}(A_i)\exp(-t/2)$$

Let $\delta > 0$ and

$$\Theta = 2 \ln \left(\frac{\sum_{i=1}^m \mathbf{Rank}(A_i)}{1-\delta} \right),$$

using union bounds, with probability $\delta > 0$

$$\Theta^{-1/2} \eta$$

will be a feasible point of the QP, reaching an objective value of $\Theta^{-1} SDP$, hence

$$OPT \leq SDP \leq 2 \ln \left(2 \sum_{i=1}^m \mathbf{Rank}(A_i) \right) OPT \quad \blacksquare$$