An Approximate Shapley-Folkman Theorem.

Alexandre d'Aspremont,

CNRS & D.I., École Normale Supérieure.

With Thomas Kerdreux, Igor Colin (ENS).

Jobs

Postdoc position (1.5 years) in **ML / Optimization**.

At Ecole Normale Supérieure in Paris, competitive salary, travel funds.

Introduction

Minimizing finite sums.

$$\begin{array}{ll}\mbox{minimize} & \sum_{i=1}^n f_i(x_i)\\ \mbox{subject to} & x \in \mathcal{C}\end{array}$$

- Ubiquitous in statistics, machine learning.
- Better computational complexity (SAGA, SVRG, MISO, etc.).
- **Today.** More robust to nonconvexity issues.

Introduction

Minimizing finite sums.

- **Penalized regression.** Given a penalty function $g(x)$ such as ℓ_1, ℓ_0, SCAD, solve

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n z_i^2 + \lambda \sum_{i=1}^p g(x_i) \\ \text{subject to} & z = Ax - b \end{array}$$

- **Empirical Risk Minimization.** In the linear case,

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n \ell(y_i, z_i) + \lambda \sum_{i=1}^p g(w_i) \\ \text{subject to} & z = Aw - b \end{array}$$

- **Multi-Task Learning.** Same format, by blocks.

- **Resource Allocation.** Aka unit commitment problem.

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^n f(x_i) \\ \text{subject to} & Ax \leq b \end{array}$$

Introduction

Minimizing finite sums.

$$\begin{array}{ll}\mbox{minimize} & \sum_{i=1}^n f_i(x_i)\\ \mbox{subject to} & x \in \mathcal{C}\end{array}$$

- **Better complexity bounds** for stochastic gradient. SAG [Schmidt et al., 2013], SVRG [Johnson and Zhang, 2013], SDCA [Shalev-Shwartz and Zhang, 2013], SAGA [Defazio et al., 2014].
- **Non convexity has a milder impact.** Weakly convex penalties for M-estimators [Loh and Wainwright, 2013, Chen and Gu, 2014].
- Equilibrium in economies where consumers have **non-convex preferences.** [Starr, 1969, Guesnerie, 1975].
- Unit commitment problem with **non-convex costs.** [Bertsekas et al., 1981].

Introduction

This talk. Focus on problems with **separable linear constraints**

$$\begin{array}{ll}\text{minimize} & \sum_{i=1}^n f_i(x_i)\\ \text{subject to} & Ax \leq b,\\ & x_i \in Y_i, \quad i=1,\ldots,n,\end{array}$$

Many results generalize to the nonlinear case,

$$\begin{array}{ll}\text{minimize} & \sum_{i=1}^n f_i(x_i)\\ \text{subject to} & \sum_{i=1}^n g_{ij}(x_i) \leq 0, \quad j=1,\ldots,m\\ & x_i \in Y_i, \quad i=1,\ldots,n,\end{array}$$

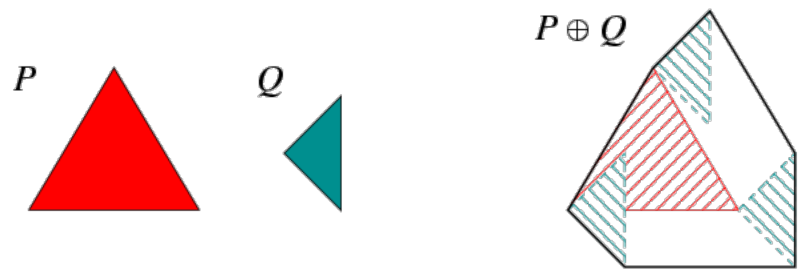
Outline

- **Shapley-Folkman theorem**
- Duality gap bounds
- Stable bounds

Introduction

Minkowski sum. Given sets $X, Y \subset \mathbb{R}^d$, we have

$$X + Y = \{x + y : x \in X,\ y \in Y\}$$

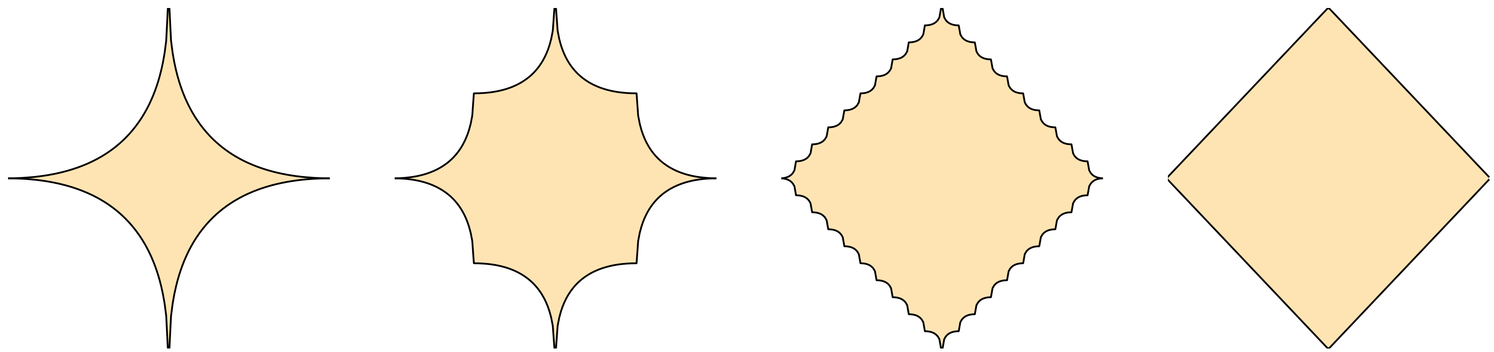


(CGAL User and Reference Manual)

Convex hull. Given subsets $V_i \subset \mathbb{R}^d$, we have

$$\mathbf{Co}\left(\sum_i V_i\right) = \sum_i \mathbf{Co}(V_i)$$

Shapley-Folkman



The $\ell_{1/2}$ ball, Minkowsi average of two and ten balls, convex hull.

Minkowsi average of five first digits (obtained by sampling).

Shapley-Folkman

Shapley-Folkman Theorem [Starr, 1969]

If $V_i \subset \mathbb{R}^d$, $i = 1, \ldots, n$, *and*

$$x \in \mathbf{Co}\left(\sum_{i=1}^{n} V_i\right) = \sum_{i=1}^{n} \mathbf{Co}(V_i)$$

then

$$x \in \sum_{[1,n]\setminus\mathcal{S}} V_i + \sum_{\mathcal{S}} \mathbf{Co}(V_i)$$

where $|\mathcal{S}| \leq d$.

Shapley-Folkman

Proof. Suppose $x \in \sum_{i=1}^n \mathbf{Co}\,(V_i)$, by Carathéodory's theorem we have

$$z = \sum_{i=1}^{n} \sum_{j=1}^{d+1} \lambda_{ij} z_{ij}$$

where $z \in \mathbb{R}^{d+n}$, $\lambda \geq 0$, and

$$z = \begin{pmatrix} x \\ \mathbf{1}_n \end{pmatrix}, \quad z_{ij} = \begin{pmatrix} v_{ij} \\ e_i \end{pmatrix}, \quad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, d+1,$$

with $e_i \in \mathbb{R}^n$ is the Euclidean basis. Conic Carathéodory on z means

$$z = \sum_{i=1}^{n} \sum_{j=1}^{d+1} \mu_{ij} z_{ij}$$

where $n + d$ nonzero coefficients μ_{ij} are spread among n sets (cf. constraints), with at least one nonzero coefficient per set.

This means $\mu_{ij} = 1$ for at least $n - d$ indices i, for which $\sum_{j=1}^{d+1} \mu_{ij} z_{ij} \in V_i$. ■

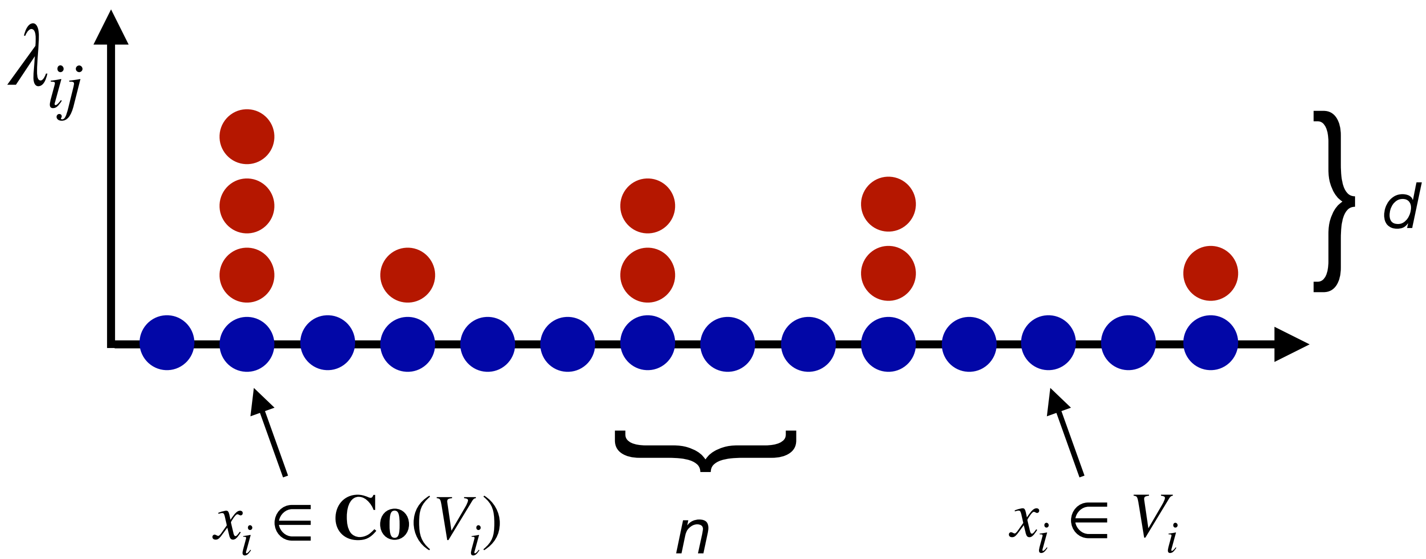
Shapley-Folkman

Proof. Write

$$x \in \sum_{[1,n]\setminus\mathcal{S}} V_i \; + \; \sum_{\mathcal{S}} \mathbf{Co}(V_i)$$

where $|\mathcal{S}| \leq d$, or

$$\begin{pmatrix} x \\ \mathbf{1}_n \end{pmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{d+1} \lambda_{ij} \begin{pmatrix} v_{ij} \\ e_i \end{pmatrix}.$$



Number of nonzero λ_{ij} controls distance to convex hull.

Shapley-Folkman: consequences

Consequences.

- If the sets $V_i \subset \mathbb{R}^d$ are uniformly bounded with $\mathrm{rad}(V_i) \leq R$, then

$$d_H\left(\left(\sum_i V_i\right), \mathbf{Co}\left(\sum_i V_i\right)\right) \leq R\sqrt{\min\{n,d\}}$$

  where $\mathrm{rad}(V) = \inf_{x\in V}\sup_{y\in V}\|x-y\|$.

- In particular, when d is fixed and $n \to \infty$

$$\left(\frac{\sum_{i=1}^n V_i}{n}\right) \to \mathbf{Co}\left(\frac{\sum_{i=1}^n V_i}{n}\right)$$

  in the Hausdorff metric.

Shapley-Folkman: consequences

In the limit.

- When $n \to \infty$, Lyapunov Theorem [Berliocchi and Lasry, 1973, Ekeland and Temam, 1999].
- Hilbert, Banach space versions [Cassels, 1975, Puri and Ralescu, 1985, Schneider and Weil, 2008]. Bound Hausdorff distance with convex hull in terms of radius.
- Strong law of large numbers for [Artstein and Vitale, 1975].

Outline

- Shapley-Folkman theorem
- **Duality gap bounds**
- Stable bounds

Nonconvex Optimization

Optimization problem. Focus on separable problem with linear constraints

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n f_i(x_i) \\ \text{subject to} & Ax \leq b, \\ & x_i \in Y_i, \quad i = 1, \ldots, n, \end{array} \tag{P}$$

in the variables $x_i \in \mathbb{R}^{d_i}$ with $d = \sum_{i=1}^n d_i$, where f_i are lower semicontinuous **(but not necessarily convex)**, $Y_i \subset \textbf{dom}\, f_i$ are compact, and $A \in \mathbb{R}^{m \times d}$.

Take the dual twice to form a **convex relaxation**,

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n (f_i + \mathbf{1}_{Y_i})^{**}(x_i) \\ \text{subject to} & Ax \leq b \end{array} \tag{CoP}$$

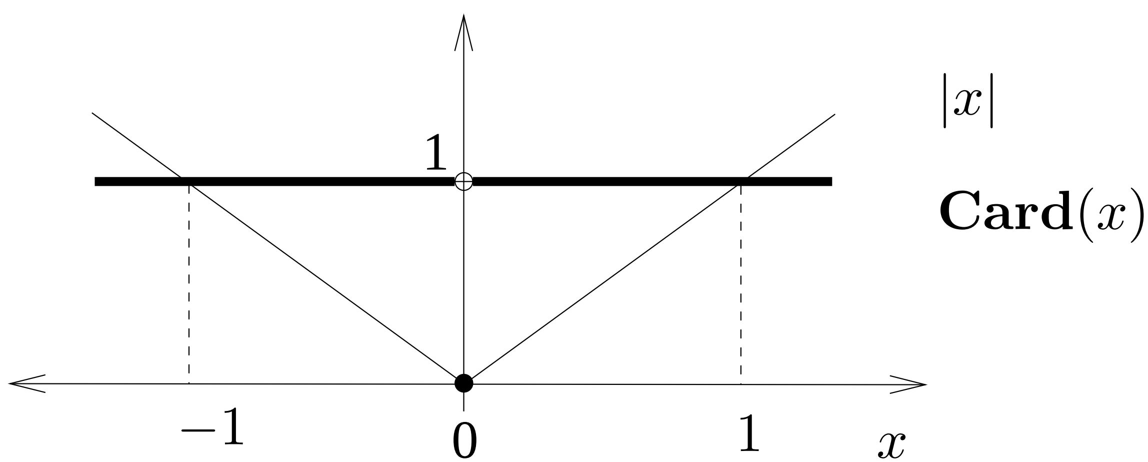
in the variables $x_i \in \mathbb{R}^{d_i}$.

Nonconvex Optimization

Convex envelope.

- Biconjugate f^{**} of f (aka convex envelope of f): pointwise supremum of all affine functions majorized by f (see e.g. [Rockafellar, 1970, Th. 12.1] or [Hiriart-Urruty and Lemaréchal, 1993, Th. X.1.3.5])
- We have $\mathbf{epi}(f^{**}) = \overline{\mathbf{Co}(\mathbf{epi}(f))}$, which means that

$f^{**}(x)$ and $f(x)$ match at extreme points x of $\mathbf{epi}(f^{**})$.



The l_1 norm is the **convex envelope** of $\mathbf{Card}(x)$ in $[-1, 1]$.

Nonconvex Optimization

Writing the epigraph of problem (P) as in [Lemaréchal and Renaud, 2001],

$$\mathcal{G} \triangleq \left\{ (x, r_0, r) \in \mathbb{R}^{d+1+m} : \sum_{i=1}^{n} f_i(x_i) + \mathbf{1}_{Y_i}(x_i) \leq r_0,\ Ax - b \leq r \right\},$$

and its projection on the last $m+1$ coordinates,

$$\mathcal{G}_r \triangleq \left\{ (r_0, r) \in \mathbb{R}^{m+1} : (x, r_0, r) \in \mathcal{G} \right\},$$

we can write the dual function of (P) as

$$\Psi(\lambda) \triangleq \inf \left\{ r_0 + \lambda^\top r : (r_0, r) \in \mathcal{G}_r^{**} \right\},$$

in the variable $\lambda \in \mathbb{R}^m$, where $\mathcal{G}^{**} = \overline{\mathbf{Co}(\mathcal{G})}$ is the closed convex hull of the epigraph $\mathcal{G}$. [Lemaréchal and Renaud, 2001, Th. 2.11]: affine constraints means the **dual functions of (P) and (CoP) are equal.** The (common) dual of (P) and (CoP) is then

$$\sup_{\lambda \geq 0} \Psi(\lambda) \tag{D}$$

in the variable $\lambda \in \mathbb{R}^m$.

Nonconvex Optimization

Epigraph. Define

$$\mathcal{F}_i = \left\{ \left((f_i + \mathbf{1}_{Y_i})^{**}(x_i), A_i x_i \right) : \; x_i \in \mathbb{R}^{d_i} \right\}$$

where $A_i \in \mathbb{R}^{m \times d_i}$ is the i^{th} block of A.

- The **epigraph** $\mathcal{G}_r^{**}$ can be written as a **Minkowski sum** of $\mathcal{F}_i$

$$\mathcal{G}_r^{**} = \sum_{i=1}^{n} \mathcal{F}_i + (0, -b) + \mathbb{R}_+^{m+1}$$

- **Lack of convexity.** Define

$$\rho(f) \triangleq \sup_{x \in \mathbf{dom}(f)} \{ f(x) - f^{**}(x) \}.$$

Bound on duality gap

A priori bound on duality gap of

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n f_i(x_i) \\ \text{subject to} & Ax \leq b, \\ & x_i \in Y_i, \quad i = 1, \ldots, n, \end{array}$$

where $A \in \mathbb{R}^{m \times d}$.

Proposition [Aubin and Ekeland, 1976, Ekeland and Temam, 1999]

A priori bounds on the duality gap *Suppose the functions f_i in* (P) *satisfy Assumption (...). There is a point $x^\star \in \mathbb{R}^d$ at which the primal optimal value of* (CoP) *is attained, such that*

$$\underbrace{\sum_{i=1}^n f_i^{**}(x_i^\star)}_{CoP} \leq \underbrace{\sum_{i=1}^n f_i(\hat{x}_i^\star)}_{P} \leq \underbrace{\sum_{i=1}^n f_i^{**}(x_i^\star)}_{CoP} + \underbrace{\sum_{i=1}^{m+1} \rho(f_{[i]})}_{\text{gap}}$$

where $\hat{x}^\star$ is an optimal point of (P) *and $\rho(f_{[1]}) \geq \rho(f_{[2]}) \geq \ldots \geq \rho(f_{[n]})$.*

Bound on duality gap

Proof sketch. Pick optimal $z^\star$ in $\mathcal{G}_r^{**}$, closed convex hull of epigraph which is a Minkowski sum,

$$\mathcal{G}_r^{**} = \sum_{i=1}^{n} \mathcal{F}_i + (0, -b) + \mathbb{R}_+^{m+1}, \text{ where } \mathcal{F}_i = \left\{ (f_i^{**}(x_i), A_i x_i) : x_i \in \mathbb{R}^{d_i} \right\} \subset \mathbb{R}^{m+1}$$

- Krein-Milman shows $\mathcal{G}_r^{**} = \sum_{i=1}^n \mathbf{Co}\left(\mathbf{Ext}(\mathcal{F}_i)\right) + (0, -b) + \mathbb{R}_+^{m+1}$.
- $\mathcal{F}_i \subset \mathbb{R}^{m+1}$ so Shapley-Folkman shows that for any $z^\star \in \mathcal{G}_r^{**}$,

  $$z^\star \in \sum_{[1,n]\setminus\mathcal{S}} \mathbf{Ext}(\mathcal{F}_i) \; + \; \sum_{\mathcal{S}} \mathbf{Co}\left(\mathbf{Ext}(\mathcal{F}_i)\right)$$

  for some index set $\mathcal{S} \subset [1, n]$ with $|\mathcal{S}| \leq m + 1$.
- Then, $f_i(x_i^\star) = f_i^{**}(x_i^\star)$ when $x_i^\star \in \mathbf{Ext}(\mathcal{F}_i)$, and $f(x_i^\star) - f^{**}(x_i^\star) \leq \rho(f_i)$ otherwise.

Bound on duality gap

Shapley-Folkman.

- A priori bound on duality gap based on tractable quantities.
- Vanishingly small if $n \to \infty$, m fixed and ρ is uniformly bounded.
- However, the bound is written in terms of **unstable quantities** which lack meaning (dimension, rank, etc.)

Significantly **tighten gap bound using stable quantities?**

Outline

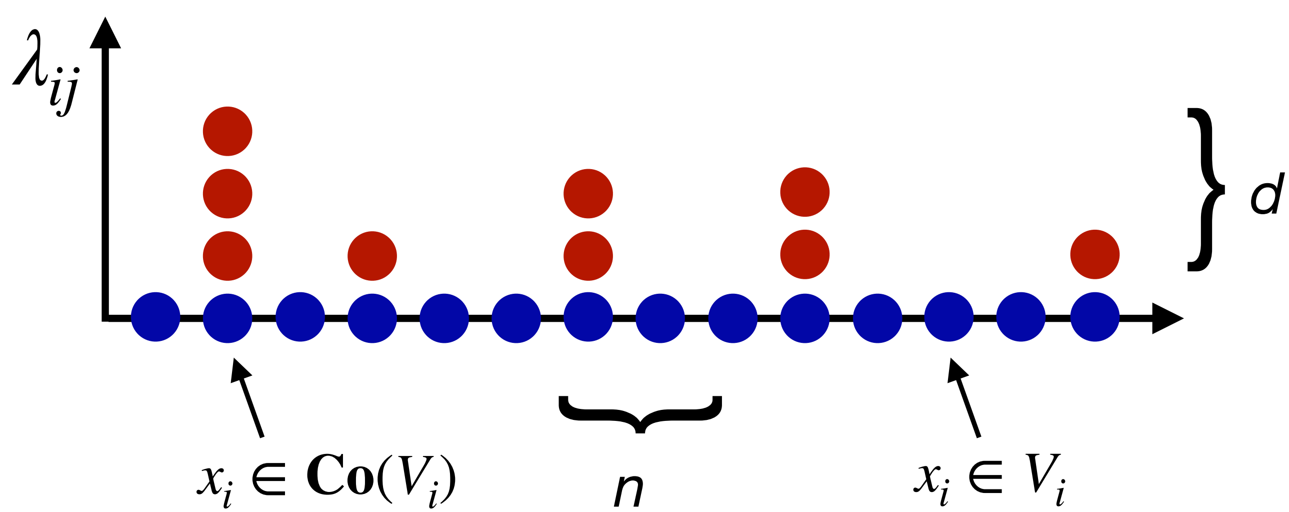
- Shapley-Folkman theorem
- Duality gap bounds
- **Stable bounds**

Stable bounds on duality gap

$\mathbf{k^{th}}$**-nonconvexity measure.** [Bi and Tang, 2016]

$$\rho_k(f) \triangleq \sup_{\substack{x_i \in \mathbf{dom}(f) \\ \alpha \in \mathbb{R}_+^{d+1}}} \left\{ f\left(\sum_{i=1}^{d+1} \alpha_i x_i\right) - \sum_{i=1}^{d+1} \alpha_i f(x_i) : \mathbf{1}^T\alpha = 1, \mathbf{Card}(\alpha) \le k \right\}$$

which restricts the number of nonzero coefficients in the formulation of $\rho(f)$.



Stable bounds on duality gap

Coupling. A priori bound on duality gap of

$$\begin{array}{ll}\text{minimize} & \sum_{i=1}^n f_i(x_i)\\ \text{subject to} & Ax \leq b,\\ & x_i \in Y_i, \quad i=1,\ldots,n,\end{array}$$

where $A \in \mathbb{R}^{m \times d}$.

- Gap bound depends on **number of coupling constraints** in $Ax \leq b$.
- The representation $Ax \leq b$ is **not unique.**

Get better bounds using shorter representations of $\mathcal{P} = \{x \in \mathbb{R}^d : Ax \leq b\}$?

Stable bounds on duality gap

Extended formulation.

Given the linear **coupling constraints**

$$\mathcal{P} = \{x \in \mathbb{R}^d : Ax \leq b\}$$

where $A \in \mathbb{R}^{m \times d}$. Write it as the projection of another, potentially simpler, polytope with

$$\mathcal{P} = \left\{x \in \mathbb{R}^d : Bx + Cu \leq d, u \in \mathbb{R}^p\right\}$$

where $B \in \mathbb{R}^{q \times d}$, $C \in \mathbb{R}^{q \times p}$ and $d \in \mathbb{R}^q$, where $q < m$.

The **extension complexity** $xc(\mathcal{P})$ is the minimum number of inequalities of an extended formulation of the polytope $\mathcal{P}$.

Stable bounds on duality gap

Extended formulation. Examples.

- The ℓ_1-**ball** $\mathcal{B}_1 = \left\{x \in \mathbb{R}^n : u^T x \leq 1,\ u \in \{-1,+1\}^n\right\}$ has 2^n inequalities. Extended formulation written

$$\mathcal{B}_1 = \left\{x \in \mathbb{R}^n : -u \leq x \leq u,\ \mathbf{1}^T u = 1,\ u \in \mathbb{R}^n\right\}$$

  has only $2n$ inequalities and one equality constraint in dimension $2n$.

- **Permutahedron** $\mathcal{P} = \mathbf{Co}(\pi(\{1,2,\ldots,n\}))$ has $2^n - 2$ facet defining inequalities.
  - Extended formulation using $O(n^2)$ inequalities in dimension $O(n^2)$ using Birkhoff polytope.
  - Optimal extended formulation by [Goemans, 2014] has only $O(n \log n)$ variables and constraints.

Stable bounds on duality gap

Extended formulation. Write S the *slack matrix* of $\mathcal{P}$, with

$$S_{ij} \triangleq b_i - (Av_j)_i \geq 0, \quad \text{where } \mathcal{P} = \mathbf{Co}(\{v_1, \ldots, v_p\}).$$

- [Yannakakis, 1991, Th. 3] shows that

$$\{x \in \mathbb{R}^d : Ax + Fy = b, y \geq 0\}$$

  is an **extended formulation of $\mathcal{P}$ iff S can be factored as $S = FV$** where F, V are nonnegative matrices.
- Smallest extended formulation of $\mathcal{P}$ from **lowest rank NMF** of S.
- Stable, approximate extended formulation using similar arguments on nested polytopes [Pashkovich, 2012, Braun et al., 2012, Gillis and Glineur, 2012].
- **Caveat:** we are counting equality constraints here, so our definition of extension complexity is different.

We can **replace m in gap bound by (modified) extension complexity.**

Stable bounds on duality gap.

Active constraints. [Udell and Boyd, 2016] show that we can replace the number of contraints m by the **number of active contraints $\tilde{m}$.**

- Write the optimal set

$$X^\star = \{M_1 \times \ldots \times M_n\} \cap \{Ax \leq b\}, \quad \text{where } M_i = \operatorname*{argmin}_{x_i \in Y_i} f_i^{**}(x_i) + \lambda^{\star T} A x_i$$

- x is an extreme point of $X^\star$ if and only if x is the only point at intersection of minimal faces F_1, F_2 of resp. $\{M_1 \times \ldots \times M_n\}$ and $\{Ax \leq b\}$ containing x [Dubins, 1962, Th. 5.1], [Udell and Boyd, 2016, Lem. 3].

- This means that $\dim F_1 + \dim F_2 \leq d$ with $d - \tilde{m} \leq \dim F_2$, so $\dim F_1 \leq \tilde{m}$.

- As faces of Cartesian products are Cartesian products of faces, the sum of dimensions of the faces of M_i containing x_i is smaller than $\tilde{m}$, hence at least $n - \tilde{m}$ points x_i of these faces are extreme points where $f_i^{**}(x_i) = f_i(x_i)$.

Approximate Shapley Folkman

Approximate Carathéodory.

- The gap bound relies on Shapley-Folkman, itself a consequence of Carathéodory.
- Approximate Carathéodory trades increased sparsity for small approx error.

Approximate Shapley-Folkman.

- In the SF proof, we start with an exact representation using $n+m$ coefficients, where $m \ll n$.
- Can we find an approximate representation using between n and $n+m$ coefficients?

We need an approximate Carathéodory theorem with **high sampling ratio.**

Approximate Shapley Folkman

Theorem [Kerdreux, Colin, and A., 2017]

Approximate Carathéodory with high sampling ratio. *Let* $x = \sum_{j=1}^{N} \lambda_j V_j$ *for* $V \in \mathbb{R}^{d \times N}$ *and some* $\lambda \in \mathbb{R}^N$ *such that* $\mathbf{1}^T \lambda = 1, \lambda \geq 0$. *Let* $\varepsilon > 0$ *and write*

$$R = \max\{R_v, R_\lambda\}, \quad \textit{where } R_v = \max_i \|\lambda_i V_i\| \textit{ and } R_\lambda = \max_i |\lambda_i|,$$

for some norm $\|\cdot\|$ *such that* $(\mathbb{R}^d, \|\cdot\|)$ *is* $(2, D)$*-smooth. Then, there exists some* $\hat{x} = \sum_{j \in \mathcal{J}} \mu_j V_j$ *with* $\mu \in \mathbb{R}^m$ *and* $\mu \geq 0$*, where* $\mathcal{J} \subset [1, N]$ *has size*

$$|\mathcal{J}| = 1 + N \frac{c(\sqrt{N}\, D\, R/\varepsilon)^2}{1 + c(\sqrt{N}\, D\, R/\varepsilon)^2}$$

for some absolute $c > 0$*, and is such that* $\|x - \hat{x}\| \leq \varepsilon$ *and* $|\sum_{j \in \mathcal{J}} \mu_j - 1| \leq \varepsilon$.

Proof. Martingale arguments for sampling without replacement as in [Serfling, 1974, Bardenet et al., 2015, Schneider, 2016].

Approximate Shapley Folkman

Approximate Shapley Folkman.

- This approximate Carathéodory yields an approximate Shapley-Folkman result.

- We get better bounds on the gap, for perturbed versions of the problem, with a much smaller number of terms in the gap bound

$$\sum_{i=1}^{m+1} \rho(f_{[i]})$$

- The quantity $R = \max\{R_v, R_\lambda\}$ in the Hoeffding bound is very conservative. We can get a Bennett-Serfling inequality instead [Kerdreux et al., 2017].

Summary

A priori gap bound on

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n f_i(x_i) \\ \text{subject to} & Ax \leq b, \\ & x_i \in Y_i, \quad i = 1, \ldots, n, \end{array}$$

where $A \in \mathbb{R}^{m \times d}$.

$$\underbrace{\sum_{i=1}^n f_i^{**}(x_i^\star)}_{CoP} \leq \underbrace{\sum_{i=1}^n f_i(\hat{x}_i^\star)}_{P} \leq \underbrace{\sum_{i=1}^n f_i^{**}(x_i^\star)}_{CoP} + \underbrace{\sum_{i=1}^{m+1} \rho(f_{[i]})}_{\text{gap}}$$

Much better than naive bound, but still **very conservative. . .**

- Replace $\rho(f_{[i]})$ by $\rho_k(f_{[i]})$.
- Replace m by the number of active contraints $\tilde{m}$ in the optimal extended formulation of the active constraint polytope.
- Use approximate Carathéodory representation to further reduce $\tilde{m}$.

Conclusion

A priori bounds on gap.

- Shapley-Folkman yields a priori bounds on duality gap of nonconvex finite sum minimization problems.
- Good but very conservative, can be significantly tightened using more stable quantities.
- Unfortunately, quantities involved are hard to bound explicitly.

Shapley-Folkman deserves a bit more limelight in Optimization, ML and statistics. . .

*

References

Zvi Artstein and Richard A Vitale. A strong law of large numbers for random compact sets. *The Annals of Probability*, pages 879–882, 1975.

Jean-Pierre Aubin and Ivar Ekeland. Estimates of the duality gap in nonconvex optimization. *Mathematics of Operations Research*, 1(3): 225–245, 1976.

Rémi Bardenet, Odalric-Ambrym Maillard, et al. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015.

Henri Berliocchi and Jean-Michel Lasry. Intégrandes normales et mesures paramétrées en calcul des variations. *Bulletin de la Société Mathématique de France*, 101:129–184, 1973.

Dimirti P Bertsekas, Gregory S Lauer, Nils R Sandell, and Thomas A Posbergh. Optimal short-term scheduling of large-scale power systems. In *Decision and Control including the Symposium on Adaptive Processes, 1981 20th IEEE Conference on*, volume 20, pages 432–443. IEEE, 1981.

Yingjie Bi and Ao Tang. Refined shapely-folkman lemma and its application in duality gap estimation. *arXiv preprint arXiv:1610.05416*, 2016.

Gabor Braun, Samuel Fiorini, Sebastian Pokutta, and David Steurer. Approximation limits of linear programs (beyond hierarchies). In *IEEE 53rd Annual Symposium on Foundations of Computer Science*, 2012.

JWS Cassels. Measures of the non-convexity of sets and the shapley–folkman–starr theorem. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 78, pages 433–436. Cambridge University Press, 1975.

Laming Chen and Yuantao Gu. The convergence guarantees of a non-convex approach for sparse recovery. *IEEE Transactions on Signal Processing*, 62(15):3754–3767, 2014.

Aaron Defazio, Francis Bach, and Simon Lacoste-Julien. Saga: A fast incremental gradient method with support for non-strongly convex composite objectives. In *Advances in Neural Information Processing Systems*, pages 1646–1654, 2014.

Lester E Dubins. On extreme points of convex sets. *Journal of Mathematical Analysis and Applications*, 5(2):237–244, 1962.

Ivar Ekeland and Roger Temam. *Convex analysis and variational problems*. SIAM, 1999.

Nicolas Gillis and François Glineur. On the geometric interpretation of the nonnegative rank. *Linear Algebra and its Applications*, 437(11): 2685–2712, 2012.

Michel X. Goemans. Smallest compact formulation for the permutahedron. *Mathematical Programming*, pages 1–7, 2014.

Roger Guesnerie. Pareto optimality in non-convex economies. *Econometrica: Journal of the Econometric Society*, pages 1–29, 1975.

Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms*. Springer, 1993.

Rie Johnson and Tong Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in Neural Information Processing Systems*, pages 315–323, 2013.

Thomas Kerdreux, Igor Colin, and Alexandre d'Aspremont. An approximate shapley-folkman theorem. *arXiv preprint arXiv:1712.08559*, 2017.

Claude Lemaréchal and Arnaud Renaud. A geometric study of duality gaps, with applications. *Mathematical Programming*, 90(3):399–427, 2001.

Po-Ling Loh and Martin J Wainwright. Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima. In *Advances in Neural Information Processing Systems*, pages 476–484, 2013.

Kanstantsin Pashkovich. *Extended formulations for combinatorial polytopes*. PhD thesis, Universitätsbibliothek, 2012.

Madan L Puri and Dan A Ralescu. Limit theorems for random compact sets in banach space. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 97, pages 151–158. Cambridge University Press, 1985.

R. T. Rockafellar. *Convex Analysis*. Princeton University Press., Princeton., 1970.

Mark Schmidt, Nicolas Le Roux, and Francis Bach. Minimizing finite sums with the stochastic average gradient. *Mathematical Programming*, pages 1–30, 2013.

Markus Schneider. Probability inequalities for kernel embeddings in sampling without replacement. In *Artificial Intelligence and Statistics*, pages 66–74, 2016.

Rolf Schneider and Wolfgang Weil. *Stochastic and integral geometry*. Springer Science & Business Media, 2008.

Robert J Serfling. Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics*, pages 39–48, 1974.

Shai Shalev-Shwartz and Tong Zhang. Stochastic dual coordinate ascent methods for regularized loss minimization. *Journal of Machine Learning Research*, 14(Feb):567–599, 2013.

Ross M Starr. Quasi-equilibria in markets with non-convex preferences. *Econometrica: journal of the Econometric Society*, pages 25–38, 1969.

Madeleine Udell and Stephen Boyd. Bounding duality gap for separable problems with linear constraints. *Computational Optimization and Applications*, 64(2):355–378, 2016.

Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3): 441–466, 1991.