Optimal-Storage Semidefinite Programming using Approximate Complementarity

Madeleine Udell

Operations Research and Information Engineering
Cornell University

Based on joint work with
Lijun Ding (Cornell), Alp Yurtsever (MIT),
Volkan Cevher (EPFL), and Joel Tropp (Caltech)

March 26, 2019

Outline

Motivation

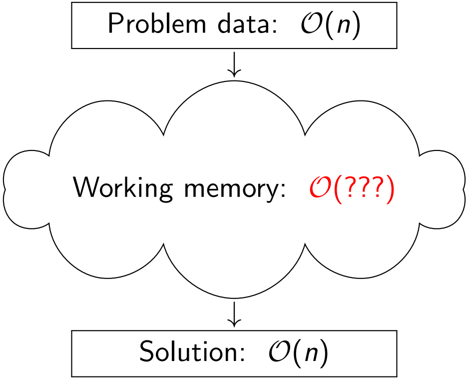
Exact complementarity

Approximate complementarity

Numerics

Goal

Can we develop algorithms that provably solve a problem using *storage* controlled by the size of the *problem data* and the size of the *solution*?



Model problem: semidefinite program (SDP)

consider primal SDP with decision variable $X \in \mathbf{S}^n$:

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CX) \\ \text{subject to} & \mathcal{A}X = b \\ & X \succeq 0, \end{array} \qquad (\mathcal{P})$$

problem data:

- cost matrix $C \in \mathbf{S}^n$
- linear map $\mathcal{A} : \mathbf{R}^{n \times n} \to \mathbf{R}^m$
- righthand side $b \in \mathbf{R}^m$

Are desiderata achievable?

suppose $(\mathcal{P})$ has

- *compact specification*: problem data use $\mathcal{O}(n)$ storage
- *compact solution*: solution $X_\star$ has constant rank $r_\star$
  $\Longrightarrow$ solution uses $\mathcal{O}(n)$ storage

$(\mathcal{P})$, using any first order method:

Problem data: $\mathcal{O}(n)$

$\downarrow$

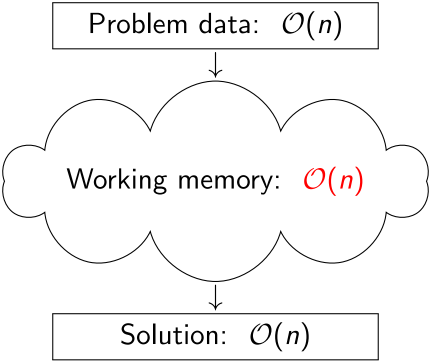
Working memory: $\mathcal{O}(n^2)$

$\downarrow$

Solution: $\mathcal{O}(n)$

Are desiderata achievable?

$(\mathcal{P})$, using approximate complementarity:



Motivation: large scale SDP

Max-Cut [Goemans and Williamson 1995]	**Matrix Completion** [Srebro and Shraibman 2005]
$\begin{array}{ll} \text{minimize} & \mathbf{tr}(-LX) \\ \text{subject to} & \mathbf{diag}(X) = \mathbf{1} \\ & X \succeq 0 \end{array}$	$\begin{array}{ll} \text{minimize} & \mathbf{tr}(W_1) + \mathbf{tr}(W_2) \\ \text{subject to} & X_{ij} = \bar{X}_{ij},\ (i,j) \in \Omega \\ & \begin{bmatrix} W_1 & X \\ X^* & W_2 \end{bmatrix} \succeq 0 \end{array}$

- Matrix completion:
  - 10^9 users, 10^9 products
  - $\implies$ SDP with 10^{18} variables
- MaxCut:
  - 10^9 people in social network
  - $\implies$ SDP with 10^{18} variables
- Phase retrieval:
  - $10^3 \times 10^3$ discretization of sample
  - $\implies$ SDP with 10^{12} variables

Optimal Storage

What kind of storage bounds can we hope for?

- Assume black-box implementation of

$$u \mapsto Cu, \qquad (u, v) \mapsto \mathcal{A}(uv^*), \text{and} \qquad (u, y) \mapsto (\mathcal{A}^* y)u,$$

  where u, $v \in \mathbf{R}^n$, $y \in \mathbf{R}^m$.
- Need $\Omega(m+n)$ storage to apply problem data.
- Need $\Theta(rn)$ storage for a rank-r approximate solution.

Definition. An algorithm to return a rank r (approximate) solution to $(\mathcal{P})$ has **optimal storage** if it uses working storage

$$\Theta(m + rn).$$

Related work: convex methods

- **Interior point methods**: storage $\Theta(n^2 + m^2)$
  - [Nesterov and Nemirovski 1989, Nesterov and Nemirovskii 1994, Alizadeh 1991; 1995]; ...
- **Matrix-free first order methods**: storage $\Theta(n^2 + m)$
  - [O'Donoghue et al. 2016]; ...
- **Storage-efficient first order methods**: storage ?
  - spectral bundle method [Helmberg and Rendl 2000]; truncated CGM [Rao et al. 2013]; spectral low rank optimization [Friedlander and Macedo 2016]; ...
  - control storage or guarantee convergence
  - not both
- **Storage-optimal first order methods**: storage $\Theta(nr + m)$
  - SketchyCGM [Yurtsever et al. 2017]; ...
  - controls storage and guarantees convergence
  - only solves smooth nuclear-norm-constrained SDP

Related work: non-convex methods

Burer-Monteiro approach: Solve

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CFF^*) \\ \text{subject to} & \mathcal{A}(FF^*) = b \end{array} \qquad \text{(BM)}$$

with variable $F \in \mathbf{R}^{n \times r}$

- Advantage: when
$$\frac{r(r+1)}{2} > m,$$
any second order stationary point is globally optimal [Burer and Monteiro 2003, Boumal et al. 2016].
- Limitation: [Waldspurger and Waters 2018] show that (for some $\mathcal{A}, b, C$)) (BM) admits bad local minima when
$$\frac{r(r+1)}{2} + r \leq m$$

$\implies$ BM approach is not storage optimal

Storage optimal methods for SDP

two approaches (so far) to storage-optimal SDP:

- **SketchySDP.** [Yurtsever et al. 2017]; ...
  1. Use primal-dual solver to solve dual problem
  2. Insight: can sketch primal during iteration
  3. Recover from sketch
- **Complementary slackness.** [Ding et al. 2019]
  1. Use any dual solver to solve dual problem
  2. Insight: optimality conditions identify range of primal
  3. Recover primal by solving smaller SDP

Outline

Motivation

Exact complementarity

Approximate complementarity

Numerics

Dual problem

consider dual SDP with decision variable $y \in \mathbf{R}^m$:

$$\begin{array}{ll} \text{maximize} & b^*y \\ \text{subject to} & C - \mathcal{A}^*y \succeq 0 \end{array} \qquad (\mathcal{D})$$

for sufficiently large $\alpha > 0$, equivalent to *penalized dual*

$$\text{maximize} \quad b^*y + \alpha \min(\lambda_{\min}(C - \mathcal{A}^*y), 0)$$

- any first order method for $(\mathcal{D})$ uses optimal storage
- (just compute minimal eigenvalue with iterative method)
- e.g., subgradient method, AdaGrad [Duchi et al. 2011], AdaNGD [Levy 2017], AccelGrad [Levy et al. 2018], ...

Assumptions

Assumption (Genericity)

- primal SDP attains its solution(s)
- dual SDP has a unique solution $y_\star$
- primal and dual SDP satisfy strong duality

$$0 = \mathbf{tr}(CX_\star) - b^* y_\star = \mathbf{tr}(X_\star(C - \mathcal{A}^* y_\star)) = X_\star(C - \mathcal{A}^* y_\star)$$

- (for storage optimality) strict complementary slackness

$$\mathbf{rank}(X_\star) + \mathbf{rank}(C - \mathcal{A}^* y_\star) = n$$

$\implies$ solution $X_\star$ of primal SDP is unique [Lemon et al. 2016]

Note: these conditions hold generically [Alizadeh et al. 1997]

Primal recovery via exact complementarity

suppose $y_\star \in \mathbf{R}^m$ solves $(\mathcal{D})$

- define dual slack matrix $Z_\star = C - \mathcal{A}^* y_\star$
- strict complementarity holds between $X_\star$ and $Z_\star$

$$\mathbf{rank}(X_\star) + \mathbf{rank}(Z_\star) = n \quad \Longrightarrow \quad \mathbf{range}(X_\star) \in \mathbf{null}(Z_\star)$$

- let $V_\star$ be a basis for nullspace of dual slack matrix $Z_\star$
- constrain $X = V_\star S V_\star^*$ in primal SDP for some $S \in \mathbf{S}_\star^{r_\star}$.
  $\Longrightarrow$ solution is preserved!

SDP via exact complementarity: algorithm

Algorithm SDP via exact complementarity

Given: problem data C, $\mathcal{A}$, b

1. compute solution $y_\star \in \mathbf{R}^m$ to $(\mathcal{D})$
2. compute basis $V_\star$ for nullspace of dual slack matrix $Z_\star = C - \mathcal{A}^* y_\star$
3. solve **reduced primal** $(\mathcal{P}_{V_\star})$ with variable $S \in \mathbf{S}_+^{r_\star}$

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CV_\star S V_\star^*) \\ \text{subject to} & \mathcal{A}(V_\star S V_\star^*) = b \\ & S \succeq 0 \end{array} \qquad (\mathcal{P}_{V_\star})$$

   to find primal solution $X_\star = V_\star S V_\star^*$

Reduced primal is easy: e.g., when $r_\star$=1, $(\mathcal{P}_{V_\star})$ solves a 1D problem over $S \in \mathbf{R}_+$

SDP via exact complementarity: storage complexity

SDP via exact complementarity uses optimal storage:

1. compute $y_\star$ with any subgradient method
   ($\mathcal{O}(m+n)$ storage)
2. compute $V_\star$ using randomized linear algebra (e.g., [Halko et al. 2011]) ($\Theta(nr_\star)$ storage)
3. solve ($\mathcal{P}_{V_\star}$) via a matrix-free method (e.g., [O'Donoghue et al. 2016]) ($\Theta(m+n+r_\star^2)$ storage)

$= (\mathcal{O}(m+nr_\star)$ storage

(in practice, use an IPM to solve ($\mathcal{P}_{V_\star}$); it's tiny!)

Primal recovery via exact complementarity: picture

Approximate complementarity?

Problem: subgradient methods do not (exactly) compute $y^\star$

try the obvious fix:

- $(y \approx y_\star)$: compute approximate dual solution y
- $(V \approx V_\star)$: compute r-dimensional eigenspace V of $C - \mathcal{A}^* y$ with smallest eigenvalues
- $(X \approx X_\star)$? solve reduced SDP $(\mathcal{P}_V)$ with variable $S \in \mathbf{S}^r_+$

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CVSV^*) \\ \text{subject to} & \mathcal{A}(VSV^*) = b \\ & S \succeq 0 \end{array} \qquad (\mathcal{P}_V)$$

to find primal solution $X = VSV^*$

```
Status:  Infeasible
Optimal value (cvx_optval):  +Inf
```

Primal recovery via approximate complementarity: picture

Outline

Motivation

Exact complementarity

Approximate complementarity

Numerics

SDP via approximate complementarity

Algorithm SDP via approximate complementarity

Given: problem data C, $\mathcal{A}$, b; target rank $r \geq r_\star$

1. Compute approximate dual solution y.
2. Compute basis V for eigenspace of $C - \mathcal{A}^* y$ with r smallest eigenvalues.
3. To find primal solution $X_\star = V_\star S V_\star^*$, $S \in \mathbf{S}_r^+$, minimize infeasibility

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\mathcal{A}(VSV^*) - b\|^2 \\ \text{subject to} & S \succeq 0 \end{array} \qquad \text{(MinFeasSDP)}$$

or accept infeasibility δ and minimize loss

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(CVSV^*) \\ \text{subject to} & \|\mathcal{A}(VSV^*) - b\| \leq \delta \\ & S \succeq 0. \end{array} \qquad \text{(MinObjSDP)}$$

Exact vs approximate complementarity

Exact complementarity	**Approximate complementarity**
Compute dual solution $y_\star$	Compute approximate dual solution y
Compute basis $V_\star$ for $\textbf{null}(C - \mathcal{A}^* y_\star)$	Compute basis V for eigenspace of $C - \mathcal{A}^* y$ with r smallest eigenvalues
Solve $\mathcal{P}_{V_\star}$	Solve MinFeasSDP or MinObjSDP.

Theoretical guarantees: approximate complementarity

Theorem (Recovery guarantees)

Suppose $(\mathcal{P})$ and $(\mathcal{D})$ satisfy genericity assumptions.
If $r = r_\star$ (MinFeasSDP) or $r \geq r_\star$ (MinObjSDP),
primal recovery via approximate complementarity from an
ϵ-suboptimal dual solution y produces a $\sqrt{\epsilon}$-suboptimal primal solution X:

	MinFeasSDP	MinObjSDP
suboptimality $\mathbf{tr}(CX) - \mathbf{tr}(CX_\star)$	$\mathcal{O}(\kappa\sqrt{\epsilon})$	$\mathcal{O}(\sqrt{\epsilon})$
infeasibility $\|\mathcal{A}X - b\|_2$	$\mathcal{O}(\kappa\sqrt{\epsilon})$	$\mathcal{O}(\sqrt{\epsilon})$
distance to solution $\|X - X_\star\|_F$	$\mathcal{O}(\kappa\sqrt{\epsilon})$	$\mathcal{O}(\epsilon^{\frac{1}{4}})$

where the condition number $\kappa = \frac{\sigma_{\max}(\mathcal{A})}{\sigma_{\min}(\mathcal{A}|_{V_\star})}$.

Core Lemma

Lemma (Projected solution)

Suppose

- $(\mathcal{P})$ and $(\mathcal{D})$ admit solutions and satisfy strong duality;
- $y \in \mathbf{R}^m$ is feasible and ϵ-suboptimal for $(\mathcal{D})$;
- and the threshold $T := \lambda_{n-r}(C - \mathcal{A}^* y) > 0$.

Then for any solution $X_\star$ of the primal SDP $(\mathcal{P})$,

$$\left\| X_\star - V \underbrace{V^* X_\star V}_{\text{feasible } S \text{ for MinFeasSDP}} V^* \right\| \leq \frac{\epsilon}{T} + \sqrt{\frac{2r\epsilon}{T}} \left\| X_\star \right\|_{\text{op}}$$

where $\|\cdot\|$ is either the nuclear or Frobenius norm.

A practical algorithm for SDP

how to choose parameters?

- infeasibility tolerance δ
  - solve MinFeasSDP first, then solve MinObjSDP
- rank target r
  - bigger is better
  - use spectrum of slack matrix $C - \mathcal{A}^* y$:
    need $T = \lambda_{n-r}(C - \mathcal{A}^* y) > 0$

Algorithm Primal recovery via approximate complementarity

Given: approximate dual solution y, rank target r

- compute basis V for eigenspace of $C - \mathcal{A}^* y$ with r smallest eigenvalues
- solve (MinFeasSDP) to obtain a solution X_{infeas},
- then solve (MinObjSDP) with $\delta = 1.1 \|\mathcal{A} X_{\text{infeas}} - b\|_2$

Outline

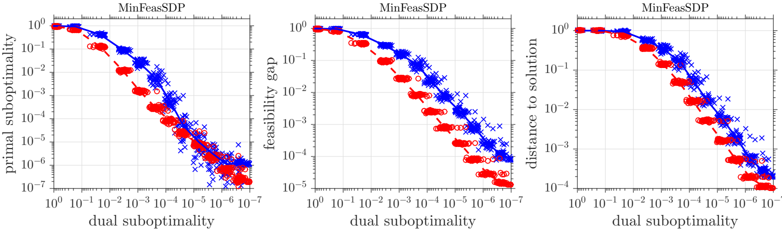
Motivation

Exact complementarity

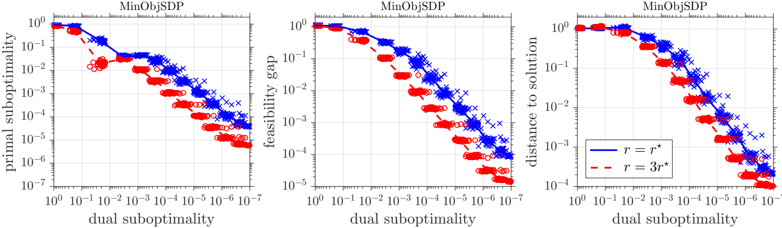
Approximate complementarity

Numerics

Primal recovery works



Max-Cut



Matrix Completion

Comments: (dashed) $r = r_\star$ (solid) $r = 3r_\star$

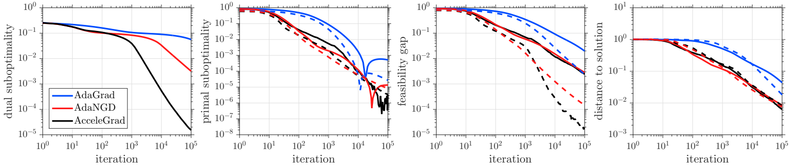
Don't try this at home

Algorithm for plot:

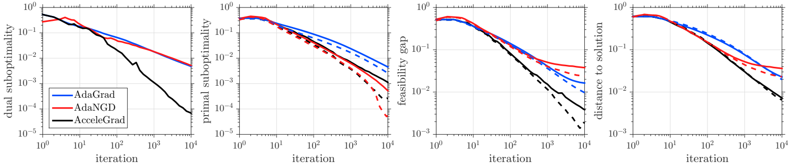
For $k = 1, 2, \ldots,$

- compute k-th dual iterate y_k using dual solver (here, AdaGrad [Duchi et al. 2011], AdaNGD [Levy 2017], and AccelGrad [Levy et al. 2018])
- recover a primal iterate X_k from y_k using RobustPrimalRecovery

Approximate complementarity solves primal SDP



Max-Cut



Matrix Completion

Comments: (dashed) $r = r_\star$ (solid) $r = 3r_\star$

Conclusion

Approximate complementarity approach provably solves (generic) SDP with optimal storage

- (+) uses $\mathcal{O}(m + nr)$ storage to find rank r solution
- (+) recovers approximate primal from approximate dual
- (+) parameters are easy to choose
- (+/-) relies on any subgradient solver for dual

References:

- L. Ding, A. Yurtsever, V. Cevher., J. A. Tropp, and M. Udell. An Optimal-Storage Approach to Semidefinite Programming using Approximate Complementarity. `http://www.arxiv.org/abs/1902.03373`
- A. Yurtsever, M. Udell, J. A. Tropp, and V. Cevher. Sketchy Decisions: Convex Optimization with Optimal Storage. AISTATS 2017.

Core Lemma

Lemma (Quadratic Growth)

Suppose genericity assumptions hold. Define the threshold $T = \lambda_{\min>0}(X_\star)$. For any dual feasible y with slack matrix $Z(y) := C - \mathcal{A}^* y$ and suboptimality $\epsilon = \epsilon_d(y) = d_\star - b^* y$,

$$\|(Z(y), y) - (Z(y_\star), y_\star)\| \leq \frac{1}{\sigma_{\min}(\mathcal{D})}\left(\frac{\epsilon}{T} + \sqrt{\frac{2\epsilon}{T}\|Z(y)\|_{\text{op}}}\right),$$

where the linear operator $\mathcal{D} : \mathbf{S}^n \times \mathbf{R}^m \to \mathbf{S}^n \times \mathbf{S}^n$ is defined by

$$\mathcal{D}(Z, y) := (Z - (U_\star U_\star{}^*) Z (U_\star U_\star{}^*), Z + \mathcal{A}^* y).$$

The orthonormal matrix $U_\star$ is the orthogonal complement of $V_\star$.

Core Lemma

Lemma (Projected solution)

Suppose

- $(\mathcal{P})$ and $(\mathcal{D})$ admit solutions and satisfy strong duality;
- $y \in \mathbf{R}^m$ is feasible and ϵ-suboptimal for $(\mathcal{D})$;
- and the threshold $T := \lambda_{n-r}(C - \mathcal{A}^* y) > 0$.

Then for any solution $X_\star$ of the primal SDP $(\mathcal{P})$,

$$\|X_\star - VV^* X_\star VV^*\| \le \frac{\epsilon}{T} + \sqrt{2\frac{\epsilon}{T}\|X_\star\|_{\text{op}}}$$

where $\|\cdot\|$ is either the nuclear or Frobenius norm.

Core Lemma

Lemma

Assume the same as Lemma 4.1. Then $\mathbf{null}(\mathcal{A}_{V^\star}) = \{0\}$.

Outline

*

References

Alizadeh, F. (1991). Combinatorial optimization with interior point methods and semi-definite matrices. *Ph. D. thesis, University of Minnesota*.

Alizadeh, F. (1995). Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM journal on Optimization*, 5(1):13–51.

Alizadeh, F., Haeberly, J.-P. A., and Overton, M. L. (1997). Complementarity and nondegeneracy in semidefinite programming. *Mathematical programming*, 77(1):111–128.

Boumal, N., Voroninski, V., and Bandeira, A. (2016). The non-convex burer-monteiro approach works on smooth semidefinite programs. In *Advances in Neural Information Processing Systems*, pages 2757–2765.

Burer, S. and Monteiro, R. D. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357.

Ding, L., Yurtsever, A., Cevher, V., Tropp, J. A., and Udell, M. (2019). An optimal-storage approach to semidefinite programming using approximate complementarity. *Submitted*.

Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159.

Friedlander, M. P. and Macedo, I. (2016). Low-rank spectral optimization via gauge duality. *SIAM Journal on Scientific Computing*, 38(3):A1616–A1638.

Goemans, M. X. and Williamson, D. P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145.

Halko, N., Martinsson, P. G., and Tropp, J. A. (2011). Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.*, 53(2):217–288.

Helmberg, C. and Rendl, F. (2000). A spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10(3):673–696.

Lemon, A., So, A. M.-C., Ye, Y., et al. (2016). Low-rank semidefinite programming: Theory and applications. *Foundations and Trends® in Optimization*, 2(1-2):1–156.

Levy, K. (2017). Online to offline conversions, universality and adaptive minibatch sizes. In *Advances in Neural Information Processing Systems*, pages 1613–1622.

Levy, K. Y., Yurtsever, A., and Cevher, V. (2018). Online adaptive methods, universality and acceleration. *arXiv preprint arXiv:1809.02864*.

Nesterov, Y. and Nemirovski, A. (1989). Self-concordant functions and polynomial time methods in convex programming, ussr acad. *Sci., Central Economic&Mathematical Institute, Moscow*.

Nesterov, Y. and Nemirovskii, A. (1994). *Interior-point polynomial algorithms in convex programming*, volume 13. Siam.

O'Donoghue, B., Chu, E., Parikh, N., and Boyd, S. (2016). Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068.

Rao, N., Shah, P., and Wright, S. (2013). Conditional gradient with enhancement and truncation for atomic-norm regularization. In *NIPS workshop on Greedy Algorithms*.

Srebro, N. and Shraibman, A. (2005). Rank, trace-norm and max-norm. In *International Conference on Computational Learning Theory*, pages 545–560. Springer.

Waldspurger, I. and Waters, A. (2018). Rank optimality for the burer-monteiro factorization. *arXiv preprint arXiv:1812.03046*.

Yurtsever, A., Udell, M., Tropp, J., and Cevher, V. (2017). Sketchy decisions: Convex low-rank matrix optimization with optimal storage. In *Artificial Intelligence and Statistics*, pages 1188–1196.