Convex Optimization M2

Semidefinite Programming Applications

Distortion, embedding problems, . . .

Distortion, embedding problems, . . .

We cannot hope to always get low rank solutions to SDPs, unless we are willing to admit some **distortion**. . . The following result from [Ben-Tal, Nemirovski, and Roos, 2003] gives some guarantees.

Theorem

Approximate $\mathcal{S}$-lemma. *Let $A_1, \ldots, A_N \in$ **S**$_n$, $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$ and a matrix $X \in$ **S**$_n$ such that*

$$A_i,\, X \succeq 0, \quad \mathbf{Tr}(A_i X) = \alpha_i, \quad i = 1, \ldots, N$$

Let $\epsilon > 0$, there exists a matrix X_0 such that

$$\alpha_i(1-\epsilon) \leq \mathbf{Tr}(A_i X_0) \leq \alpha_i(1+\epsilon) \quad \textit{and} \quad \mathbf{Rank}(X_0) \leq 8\frac{\log 4N}{\epsilon^2}$$

Proof. Randomization, concentration results on Gaussian quadratic forms.

See [Barvinok, 2002, Ben-Tal, El Ghaoui, and Nemirovski, 2009] for more details.

Distortion, embedding problems, . . .

A particular case: Given N vectors $v_i \in \mathbb{R}^d$, construct their Gram matrix $X \in \mathbf{S}_N$, with

$$X \succeq 0, \quad X_{ii} - 2X_{ij} + X_{jj} = \|v_i - v_j\|_2^2, \quad i,j = 1, \ldots, N.$$

The matrices $D_{ij} \in \mathbf{S}_n$ such that

$$\mathbf{Tr}(D_{ij}X) = X_{ii} - 2X_{ij} + X_{jj}, \quad i,j = 1, \ldots, N$$

satisfy $D_{ij} \succeq 0$. Let $\epsilon > 0$, there exists a matrix X_0 with

$$m = \mathbf{Rank}(X_0) \leq 16 \frac{\log 2N}{\epsilon^2},$$

from which we can extract vectors $u_i \in \mathbb{R}^m$ such that

$$\|v_i - v_j\|_2^2 \, (1 - \epsilon) \leq \|u_i - u_j\|_2^2 \leq \|v_i - v_j\|_2^2 \, (1 + \epsilon).$$

In this setting, the **Johnson-Lindenstrauss** lemma is a particular case of the approximate $\mathcal{S}$ lemma. . .

Distortion, embedding problems, ...

- The problem of reconstructing an N-point Euclidean metric, given **partial** information on pairwise distances between points $v_i,\ i=1,\ldots,N$ can also be cast as an SDP, known as and **Euclidean Distance Matrix Completion** problem.

$$\begin{array}{ll} \mbox{find} & D \\ \mbox{subject to} & \mathbf{1}v^T + v\mathbf{1}^T - D \succeq 0 \\ & D_{ij} = \|v_i - v_j\|_2^2, \quad (i,j) \in S \\ & v \geq 0 \end{array}$$

  in the variables $D \in \mathbf{S}_n$ and $v \in \mathbb{R}^n$, on a subset $S \subset [1,N]^2$.

- We can add further constraints to this problem given additional structural info on the configuration.
- Applications in sensor networks, molecular conformation reconstruction etc...

Distortion, embedding problems, . . .

[Dattorro, 2005] 3D map of the USA reconstructed from pairwise distances on 5000 points. Distances reconstructed from Latitude/Longitude data.

Mixing rates for Markov chains & maximum variance unfolding

Mixing rates for Markov chains & unfolding

[Sun, Boyd, Xiao, and Diaconis, 2006]

- Let $G = (V, E)$ be an **undirected graph** with n vertices and m edges.
- We define a **Markov chain** on this graph, and let $w_{ij} \geq 0$ be the transition rate for edge $(i, j) \in V$.
- Let $\pi(t)$ be the state distribution at time t, its evolution is governed by the heat equation

$$d\pi(t) = -L\pi(t)dt$$

  with

$$L_{ij} = \begin{cases} -w_{ij} & \text{if } i \neq j,\ (i,j) \in V \\ 0 & \text{if } (i,j) \notin V \\ \sum_{(i,k)\in V} w_{ik} & \text{if } i = j \end{cases}$$

  the **graph Laplacian** matrix, which means

$$\pi(t) = e^{-Lt}\pi(0).$$

- The matrix $L \in \mathbf{S}_n$ satisfies $L \succeq 0$ and its smallest eigenvalue is zero.

Mixing rates for Markov chains & unfolding

- With
$$\pi(t) = e^{-Lt}\pi(0)$$
the **mixing rate** is controlled by the second smallest eigenvalue $\lambda_2(L)$.
- Since the smallest eigenvalue of L is zero, with eigenvector $\mathbf{1}$, we have
$$\lambda_2(L) \geq t \quad \Longleftrightarrow \quad L(w) \succeq t(\mathbf{I} - (1/n)\mathbf{1}\mathbf{1}^T),$$
- Maximizing the mixing rate of the Markov chain means solving
$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & L(w) \succeq t(\mathbf{I} - (1/n)\mathbf{1}\mathbf{1}^T) \\ & \sum_{(i,j)\in V} d_{ij}^2 w_{ij} \leq 1 \\ & w \geq 0 \end{array}$$
in the variable $w \in \mathbb{R}^m$, with (normalization) parameters $d_{ij}^2 \geq 0$.
- Since $L(w)$ is an affine function of the variable $w \in \mathbb{R}^m$, this is a semidefinite program in $w \in \mathbb{R}^m$.
- Numerical solution usually performs better than **Metropolis-Hastings**.

Mixing rates for Markov chains & unfolding

- We can also form the **dual** of the maximum MC mixing rate problem.
- The dual means solving

$$\begin{array}{ll}\mbox{maximize} & \mathbf{Tr}(X(\mathbf{I}-(1/n)\mathbf{1}\mathbf{1}^T))\\ \mbox{subject to} & X_{ii}-2X_{ij}+X_{jj}\leq d_{ij}^2, \quad (i,j)\in V\\ & X\succeq 0,\end{array}$$

  in the variable $X \in \mathbf{S}_n$.
- Here too, we can interpret X as the gram matrix of a set of n vectors $v_i \in \mathbb{R}^d$. The program above maximizes the variance of the vectors v_i

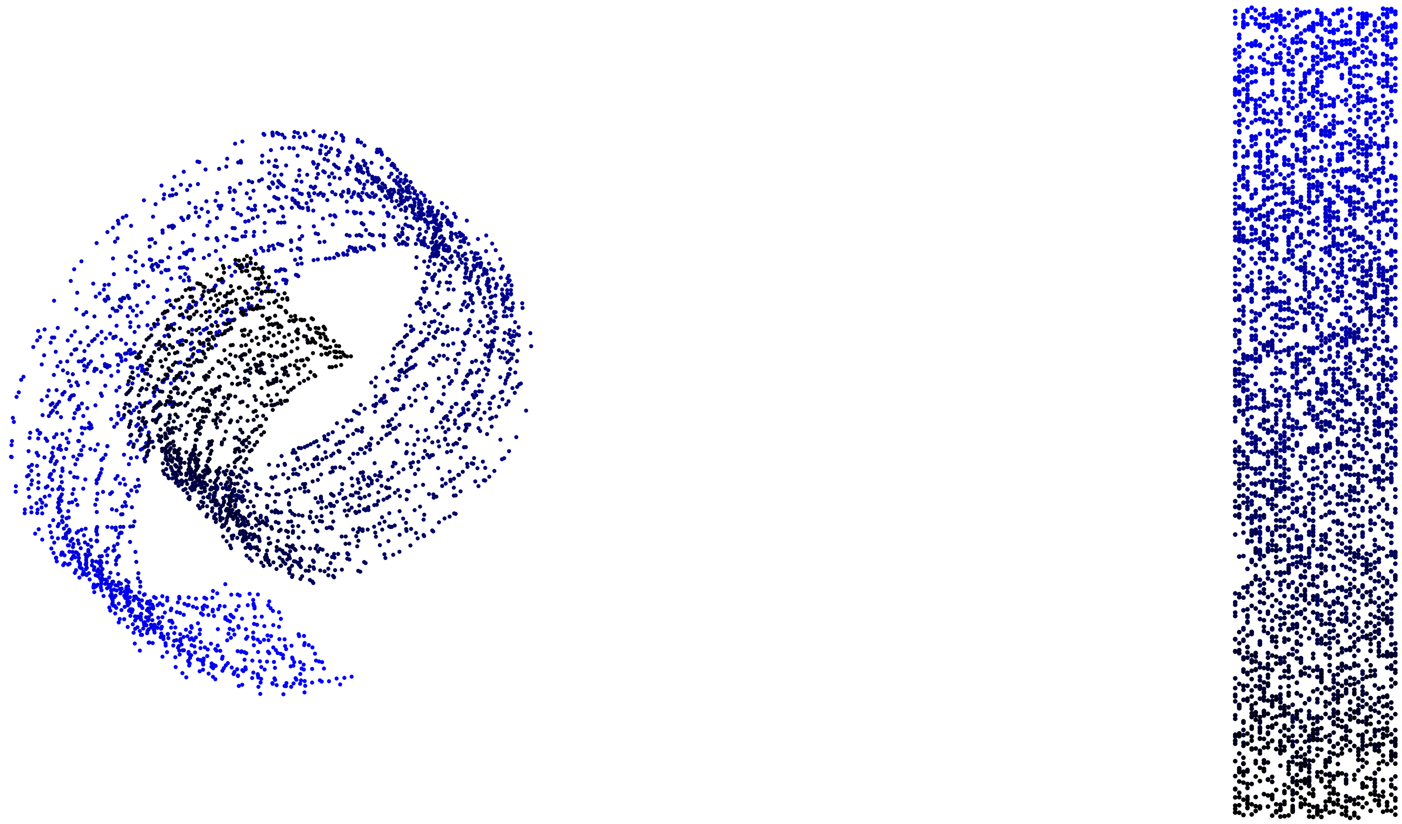
$$\mathbf{Tr}(X(\mathbf{I}-(1/n)\mathbf{1}\mathbf{1}^T)) = \textstyle\sum_i \|v_i\|_2^2 - \|\sum_i v_i\|_2^2$$

  while the constraints bound pairwise distances

$$X_{ii}-2X_{ij}+X_{jj}\leq d_{ij}^2 \quad\Longleftrightarrow\quad \|v_i-v_j\|_2^2 \leq d_{ij}^2$$

- This is a **maximum variance unfolding problem** [Weinberger and Saul, 2006, Sun et al., 2006].

Mixing rates for Markov chains & unfolding



From [Sun et al., 2006]: we are given pairwise 3D distances for k-nearest neighbors in the point set on the right. We plot the maximum variance point set satisfying these pairwise distance bounds on the right.

Moment problems & positive polynomials

Moment problems & positive polynomials

[Nesterov, 2000]. Hilbert's 17^{th} problem has a positive answer for univariate polynomials: a polynomial is nonnegative iff it is a **sum of squares**

$$p(x) = x^{2d} + \alpha_{2d-1}x^{2d-1} + \ldots + \alpha_0 \geq 0, \text{ for all } x \quad \Longleftrightarrow \quad p(x) = \sum_{i=1}^{N} q_i(x)^2$$

We can formulate this as a linear matrix inequality, let $v(x)$ be the moment vector

$$v(x) = (1, x, \ldots, x^d)^T$$

we have

$$\sum_i \lambda_i u_i u_i^T = M \succeq 0 \quad \Longleftrightarrow \quad p(x) = v(x)^T M v(x) = \sum_i \lambda_i (u_i^T v(x))^2$$

where (λ_i, u_i) are the eigenpairs of M.

Moment problems & positive polynomials

- The dual to the cone of Sum-of-Squares polynomials is the cone of moment matrices

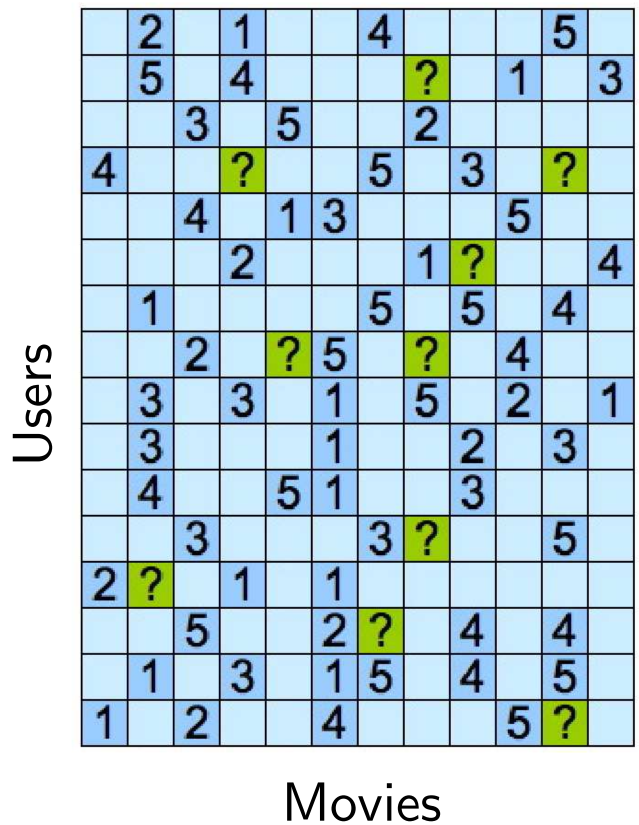
$$\mathbf{E}_\mu[x^i] = q_i,\ i = 0, \ldots, d \quad \Longleftrightarrow \quad \begin{pmatrix} q_0 & q_1 & \cdots & q_d \\ q_1 & q_2 & & q_{d+1} \\ \vdots & & \ddots & \vdots \\ q_d & q_{d+1} & \cdots & q_{2d} \end{pmatrix} \succeq 0$$

- [Putinar, 1993, Lasserre, 2001, Parrilo, 2000] These results can be extended to multivariate polynomial optimization problems over compact semi-algebraic sets.
- This forms exponentially large, ill-conditioned semidefinite programs however.

Collaborative prediction

Collaborative prediction

- Users assign **ratings** to a certain number of movies:



- Objective: make recommendations for other movies. . .

Collaborative prediction

- Infer **user preferences** and **movie features** from user ratings.
- We use a linear prediction model:

$$rating_{ij} = u_i^T v_j$$

  where u_i represents user characteristics and v_j movie features.
- This makes collaborative prediction a **matrix factorization** problem
- Overcomplete representation. . .

Collaborative prediction

- **Inputs**: a matrix of ratings $M_{ij} = \{-1, +1\}$ for $(i,j) \in S$, where S is a subset of all possible user/movies combinations.
- We look for a linear model by factorizing $M \in \mathbb{R}^{n \times m}$ as:

$$M = U^T V$$

  where $U \in \mathbb{R}^{n \times k}$ represents user characteristics and $V \in \mathbb{R}^{k \times m}$ movie features.
- **Parsimony**. . . We want k to be as small as possible.
- **Output**: a matrix $X \in \mathbb{R}^{n \times m}$ which is a low-rank approximation of the ratings matrix M.

Least-Squares

- Choose Means Squared Error as measure of discrepancy.
- Suppose S is the full set, our problem becomes:

$$\min_{\{X:\,\mathbf{Rank}(X)=k\}} \|X - M\|^2$$

- This is just a **singular value decomposition** (SVD)...

Problem: Not true when S is not the full set (partial observations). Also, MSE not a good measure of prediction performance...

Soft Margin

$$\text{minimize} \quad \mathbf{Rank}(X) + c \sum_{(i,j)\in S} \max(0, 1 - X_{ij}M_{ij})$$

non-convex and numerically hard. . .

- Relaxation result in Fazel et al. [2001]: replace $\mathbf{Rank}(X)$ by its convex envelope on the spectahedron to solve:

$$\text{minimize} \quad \|X\|_* + c \sum_{(i,j)\in S} \max(0, 1 - X_{ij}M_{ij})$$

  where $\|X\|_*$ is the **nuclear norm**, $i.e.$ sum of the singular values of X.

- Srebro [2004]: This relaxation also corresponds to multiple large margin SVM classifications.

Soft Margin

- The dual of this program:

$$\begin{array}{ll}\text{maximize} & \sum_{ij} Y_{ij}\\ \text{subject to} & \|Y \odot M\|_2 \leq 1\\ & 0 \leq Y_{ij} \leq c\end{array}$$

  in the variable $Y \in \mathbb{R}^{n \times m}$, where $Y \odot M$ is the Schur (componentwise) product of Y and M and $\|Y\|_2$ the largest singular value of Y.

- This problem is **sparse**: $Y_{ij}^* = c$ for $(i,j) \in S^c$

*

References

N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80. ACM, 2004.

A. Barvinok. *A course in convexity*. American Mathematical Society, 2002.

S. Becker, E.J. Candes, and M. Grant. Tfocs v1. 1 user guide. 2012.

A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization : analysis, algorithms, and engineering applications*. MPS-SIAM series on optimization. Society for Industrial and Applied Mathematics : Mathematical Programming Society, Philadelphia, PA, 2001.

A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560, 2003. ISSN 1052-6234.

A. Ben-Tal, L. El Ghaoui, and A.S. Nemirovski. *Robust optimization*. Princeton University Press, 2009.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

O. Bunk, A. Diaz, F. Pfeiffer, C. David, B. Schmitt, D.K. Satapathy, and JF Veen. Diffractive imaging for periodic samples: retrieving one-dimensional concentration profiles across microfluidic channels. *Acta Crystallographica Section A: Foundations of Crystallography*, 63 (4):306–314, 2007.

E. J. Candes, T. Strohmer, and V. Voroninski. Phaselift : exact and stable signal recovery from magnitude measurements via convex programming. *To appear in Communications in Pure and Applied Mathematics*, 2011a.

E.J. Candes and B. Recht. Exact matrix completion via convex optimization. *preprint*, 2008.

E.J. Candes and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *Information Theory, IEEE Transactions on*, 56(5): 2053–2080, 2010.

E.J. Candes, Y. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *Arxiv preprint arXiv:1109.0573*, 2011b.

A. Chai, M. Moscoso, and G. Papanicolaou. Array imaging using intensity-only measurements. *Inverse Problems*, 27:015005, 2011.

J. Dattorro. *Convex optimization & Euclidean distance geometry*. Meboo Publishing USA, 2005.

L. Demanet and P. Hand. Stable optimizationless recovery from phaseless linear measurements. *Arxiv preprint arXiv:1208.1803*, 2012.

M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. *Proceedings American Control Conference*, 6:4734–4739, 2001.

J.R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, 1982.

A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

Karin Gatermann and P. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. Technical Report arXiv math.AC/0211450, ETH Zurich, 2002.

R. Gerchberg and W. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35: 237–246, 1972.

M.X. Goemans and D. Williamson. Approximation algorithms for max-3-cut and other problems via complex semidefinite programming. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 443–452. ACM, 2001.

M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42:1115–1145, 1995.

D. Griffin and J. Lim. Signal estimation from modified short-time fourier transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 32(2):236–243, 1984.

R.W. Harrison. Phase problem in crystallography. *JOSA A*, 10(5):1046–1055, 1993.

C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz. An interior–point method for semidefinite programming. *SIAM Journal on Optimization*, 6:342–361, 1996.

N. K. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

L. G. Khachiyan. A polynomial algorithm in linear programming (in Russian). *Doklady Akademiia Nauk SSSR*, 224:1093–1096, 1979.

M. Kisialiou and Z.Q. Luo. Probabilistic analysis of semidefinite relaxation for binary quadratic minimization. *SIAM Journal on Optimization*, 20:1906, 2010.

J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

C. Lemaréchal and C. Sagastizábal. Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries. *SIAM Journal on Optimization*, 7(2):367–385, 1997.

L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2):166–190, 1991.

Z.Q. Luo, X. Luo, and M. Kisialiou. An efficient quasi-maximum likelihood decoder for psk signals. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 6, pages VI–561. IEEE, 2003.

P. Massart. Concentration inequalities and model selection. *Ecole d'Eté de Probabilités de Saint-Flour XXXIII*, 2007.

M. Mezard and A. Montanari. *Information, physics, and computation*. Oxford University Press, USA, 2009.

J. Miao, T. Ishikawa, Q. Shen, and T. Earnest. Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes. *Annu. Rev. Phys. Chem.*, 59:387–410, 2008.

A.S. Nemirovski. *Computation of matrix norms with applications to Robust Optimization*. PhD thesis, Technion, 2005.

A. Nemirovskii and D. Yudin. Problem complexity and method efficiency in optimization. *Nauka (published in English by John Wiley, Chichester, 1983)*, 1979.

Y. Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady*, 27(2): 372–376, 1983.

Y. Nesterov. *Global quadratic optimization via conic relaxation*. Number 9860. CORE Discussion Paper, 1998.

Y. Nesterov. Squared functional systems and optimization problems. Technical Report 1472, CORE reprints, 2000.

Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*. Society for Industrial and Applied Mathematics, Philadelphia, 1994.

P. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000.

M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *Arxiv preprint arXiv:0706.4138*, 2007.

H. Sahinoglou and S.D. Cabrera. On phase retrieval of finite-length sequences using the initial time sample. *Circuits and Systems, IEEE Transactions on*, 38(8):954–958, 1991.

N.Z. Shor. Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences*, 25:1–11, 1987.

A.M.C. So. Non–asymptotic performance analysis of the semidefinite relaxation detector in digital communications. 2010.

N. Srebro. *Learning with Matrix Factorization*. PhD thesis, Massachusetts Institute of Technology, 2004.

J. Sun, S. Boyd, L. Xiao, and P. Diaconis. The fastest mixing Markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM Review*, 48(4):681–699, 2006.

F. Vallentin. Symmetry in semidefinite programs. *Linear Algebra and Its Applications*, 430(1):360–369, 2009.

K.Q. Weinberger and L.K. Saul. Unsupervised Learning of Image Manifolds by Semidefinite Programming. *International Journal of Computer Vision*, 70(1):77–90, 2006.

Z. Wen, D. Goldfarb, S. Ma, and K. Scheinberg. Row by row methods for semidefinite programming. Technical report, Technical report, Department of IEOR, Columbia University, 2009.

S. Zhang and Y. Huang. Complex quadratic optimization and semidefinite programming. *SIAM Journal on Optimization*, 16(3):871–890, 2006.