Bad semidefinite programs: they all look the same

Gábor Pataki

Department of Statistics and Operations Research

UNC Chapel Hill

Talk at Jon Fest

A conference in honour of Jonathan Borwein's 60th birthday

A Semidefinite Program (SDP)

$$\begin{array}{ll} \sup_x & c^T x \\ s.t. & \sum_{i=1}^m x_i A_i \preceq B. \end{array} \qquad (SDP)$$

Here

- A_i, B are symmetric matrices, $c, x \in \mathbb{R}^m$.
- $A \preceq B$ means that $B - A$ is symmetric positive semidefinite (psd).
- An $n \times n$ matrix Y is positive semidefinite, if all principal subdeterminants are nonnegative.
- Equivalently, if $v^T Y v \geq 0 \, \forall v \in \mathbb{R}^n$.

SDP in a different shape

$$\begin{aligned} \inf_Y \; & B \bullet Y \\ s.t. \; & Y \succeq 0 \\ & A_i \bullet Y = c_i \, (i = 1, \ldots, m). \end{aligned}$$

Here

- A_i, B are symmetric matrices, $c \in \mathbb{R}^m$.
- $A \bullet B = \sum_{i,j} a_{ij} b_{ij}$
- **Example:** $\{Y \succeq 0 \,|\, y_{ii} = 1\}$ the set of **correlation matrices.**

History of SDP and conic programming

• **Common framework for LP and SDP:** both $\mathbb{R}^n_+$ and **psd matrices are closed convex cones.**

• **A set C is a cone, if $x \in C, \lambda \geq 0 \Rightarrow \lambda x \in C$.**

• **Linear objective, conic constraint both in LP and SDP, and many other interesting problems.**

History of SDP and conic programming

Early duality theory for conic and semi-infinite problems:

- **Duffin '56**
- **Bellman-Fan '63**
- **Ben-Israel '69-'70**
- **Ben-Israel-Charnes-Kortanek '69-'70**
- **Berman '70-'73**
- **Duffin-Jeroslow-Karlovitz '83**
- **. . .**

Later duality theory:

- **Shapiro '85, '97**
- **Borwein-Wolkowicz '81-'86**
- **Bot-Wanka '06**
- **Jeyakumar, Dinh, Lee '04**
- **. . .**

History of SDP and conic programming

Surveys and textbooks (on SDP in general, and on duality theory):

- Shapiro '00
- Wolkowicz-Vandenberghe-Saigal, '00;
- Bonnans-Shapiro '00
- Renegar '01;
- Vandenberghe-Boyd '96, '04;
- Todd '01;
- Luo-Sturm-Zhang '97;
- Ben-Tal-Nemirovskii '01;
- Burer (talk) '07
- ...

History of SDP and conic programming

- **Grötschel-Lovász-Schrijver '84: Polynomial time solvability for important classes of conic (and other convex) optimization problems, including SDP via the ellipsoid method of Shor-Nemirovskii-Yudin.**

- **Nesterov-Nemirovskii '89: General theory of interior point methods to solve conic problems in polynomial time.**

- **More than 1200 references in SDP since 1995.**

An important, related question: when is the linear image of a closed convex cone closed?

- Classic: Theorem 9.1 in Rockafellar;
- Waksman-Epelman, 1976;
- Auslender, 1996;
- Bauschke-Borwein, 1999;
- Pataki, 2007;
- Borwein-Moors, 2009-11;

SDP duality

The primal-dual pair of SDPs:

$$\sup_x \; c^T x \qquad \text{s.t.} \; \sum_{i=1}^m x_i A_i \preceq B$$

$$\inf_Y \; B \bullet Y \qquad Y \succeq 0, \quad A_i \bullet Y = c_i \; (i = 1, \ldots, m).$$

SDP duality

The primal-dual pair of SDPs:

$$\sup_x \; c^T x \qquad\qquad \inf_Y \; B \bullet Y$$

$$s.t. \; \sum_{i=1}^m x_i A_i \preceq B \qquad\qquad Y \succeq 0$$

$$A_i \bullet Y = c_i \, (i = 1, \ldots, m).$$

Easy: If x and Y are feasible, then $c^T x \leq B \bullet Y$.

SDP duality

The primal-dual pair of SDPs:

$$\sup_x \; c^T x \qquad\qquad \inf_Y \; B \bullet Y$$

$$s.t. \; \sum_{i=1}^m x_i A_i \preceq B \qquad\qquad Y \succeq 0$$

$$A_i \bullet Y = c_i \; (i = 1, \ldots, m).$$

Easy: If x and Y are feasible, then $c^T x \leq B \bullet Y$.

Ideal situation: $\exists \bar{x}, \exists \bar{Y} : c^T \bar{x} = B \bullet \bar{Y}$.

SDP duality

The primal-dual pair of SDPs:

$$\sup_x \; c^T x \qquad\qquad \inf_Y \; B \bullet Y$$

$$s.t. \; \sum_{i=1}^m x_i A_i \preceq B \qquad\qquad Y \succeq 0$$

$$A_i \bullet Y = c_i \; (i = 1, \ldots, m).$$

Easy: If x and Y are feasible, then $c^T x \leq B \bullet Y$.

Ideal situation: $\exists \bar{x}, \exists \bar{Y} : c^T \bar{x} = B \bullet \bar{Y}$.

But: in SDP, unlike in LP pathological phenomena occur: nonattainment, positive gaps.

Previous approaches to guarantee strong duality

- **Duffin '56:** **Assume strict feasibility in primal $\Rightarrow$ dual has same value, and attains, and vice versa.**

- **Borwein-Wolkowicz '81-'86:** **Theoretical algorithm to reduce the primal system to an equivalent strictly feasible system.**

- **Ramana '95:** **Gapfree dual for SDP, with polynomially many extra variables.**

Previous approaches to guarantee strong duality

- **Duffin '56:** Assume strict feasibility in primal $\Rightarrow$ dual has same value, and attains, and vice versa.

- **Borwein-Wolkowicz '81-'86:** Theoretical algorithm to reduce the primal system to an equivalent strictly feasible system.

- **Ramana '95:** Gapfree dual for SDP, with polynomially many extra variables.

- **Ramana-Tuncel-Wolkowicz '97, Pataki '00:** The two are related.

What are the pathologies?

Pathology # 1: nonattainment in dual

Primal:

$$\sup \; 2x_1 \quad \text{s.t.} \; x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \Leftrightarrow \quad \sup \; 2x_1 \quad \text{s.t.} \begin{pmatrix} 1 & -x_1 \\ -x_1 & 0 \end{pmatrix} \succeq 0$$

Pathology # 1: nonattainment in dual

Primal:

$$\sup \ 2x_1 \quad \text{s.t.} \ x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \Leftrightarrow \quad \sup \ 2x_1 \quad \text{s.t.} \ \begin{pmatrix} 1 & -x_1 \\ -x_1 & 0 \end{pmatrix} \succeq 0$$

Only feasible x_1 **is** $x_1 = 0$.

Pathology # 1: nonattainment in dual

Primal:

$$\begin{array}{llll} \sup & 2x_1 & \Leftrightarrow & \sup \ 2x_1 \\ s.t. & x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & & s.t. \ \begin{pmatrix} 1 & -x_1 \\ -x_1 & 0 \end{pmatrix} \succeq 0 \end{array}$$

Only feasible x_1 **is** $x_1 = 0$**.**

Dual: Dual variable is $Y \succeq 0$**.**

$$\begin{array}{ll} \inf & y_{11} \\ s.t. & \begin{pmatrix} y_{11} & 1 \\ 1 & y_{22} \end{pmatrix} \succeq 0 \end{array}$$

Pathology # 1: nonattainment in dual

Primal:

$$\sup \; 2x_1 \quad \Leftrightarrow \quad \sup \; 2x_1$$

$$s.t. \; x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad s.t. \; \begin{pmatrix} 1 & -x_1 \\ -x_1 & 0 \end{pmatrix} \succeq 0$$

Only feasible x_1 is $x_1 = 0$.

Dual: Dual variable is $Y \succeq 0$.

$$\inf \; y_{11}$$

$$s.t. \; \begin{pmatrix} y_{11} & 1 \\ 1 & y_{22} \end{pmatrix} \succeq 0$$

Here $\inf = 0$, but not attained: Any $y_{11} > 0$, $y_{22} = 1/y_{11}$ is feasible, but $y_{11} = 0$ is not.

Pathology # 2: positive duality gap

Primal:

$$\begin{aligned} \sup\ & x_2 \\ s.t.\ & x_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Pathology # 2: positive duality gap

Primal:

$$\sup \; x_2$$
$$s.t. \; x_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Only feasible x_2 **is** $x_2 = 0$**.**

Pathology # 2: positive duality gap

Primal:

$$\sup \ x_2$$

$$s.t. \ x_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Only feasible x_2 **is** $x_2 = 0$.

Dual value is 1, **and it is attained.**

What to do with the pathologies? Our goal

Let us find a characterization of bad SDPs, which is

- **exact**
- **efficiently verifiable**
- **aesthetic**

Terminology

Definition:

- The system $P = \{ x \mid \sum_{i=1}^m x_i A_i \preceq B \}$ is **well-behaved**, if for all c such that

  $$\sup\{ c^T x \mid x \in P \} \text{ is finite,}$$

  the dual program has the same value, and it attains.
- **Badly behaved**, otherwise.

- We would like to understand badly behaved semidefinite systems.

Motivation

The systems

$$x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$x_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

are both badly behaved.

Motivation

The systems

$$x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$x_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

are both badly behaved.

Curious similarity:

- if we delete 2nd row and 2nd column in all matrices in the second system, and
- delete the first matrix,
- we get back the first system!

Motivation

• **In fact, all badly behaved systems appearing in the literature look similar.**

Question:

- **Do all bad SDPs "look the same"?**
- **Is the first. minimal system "contained" in all of them?**

The answer is yes to both.

Technicalities

Definition: A slack matrix in P is a matrix

$$Z := B - \sum_{i=1}^{m} x_i A_i \succeq 0.$$

Fact: There is a slack matrix with maximum rank. E.g. the maximum rank slack in

$$x_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ is } \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Assumption: We can replace all A_i by $T^T A_i T$ and B by $T^T B T$, where T is invertible.

Main Theorem

Assume that in P the maximum rank slack is of the form

$$Z = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Then P is badly behaved $\Leftrightarrow$ $\exists V$ which is a linear combination of the A_i and B of the form

$$V = \begin{pmatrix} V_{11} & e_1 & \dots & \\ e_1^T & 0 & \dots & \\ \vdots & & & \ddots \end{pmatrix},$$

where V_{11} is r by r, e_1 is first unit vector.

Main Theorem

Assume that in P the maximum rank slack is of the form

$$Z = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Then P is badly behaved $\Leftrightarrow$ $\exists V$ which is a linear combination of the A_i and B of the form

$$V = \begin{pmatrix} V_{11} & e_1 & \dots \\ e_1^T & 0 & \dots \\ \vdots & & \ddots \end{pmatrix},$$

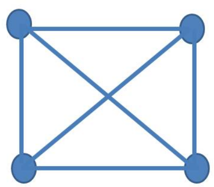
where V_{11} is r by r, e_1 is first unit vector.

Explanation: The Z and V are "easy to visualize" certificates of the bad behavior. E.g. in the first system they are

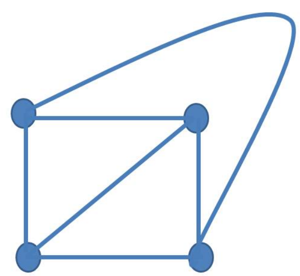
$$Z = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

More motivation: excluded minors

Unrelated question: Given undirected graph, is it planar, i.e. can we draw the edges on the plane, so they only meet at nodes? E.g. graph below is planar,

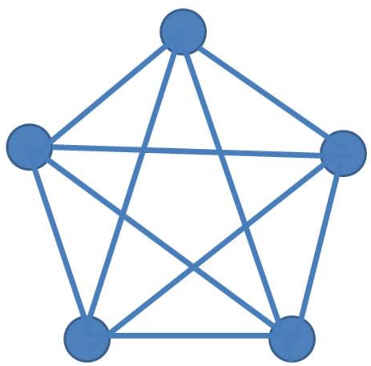


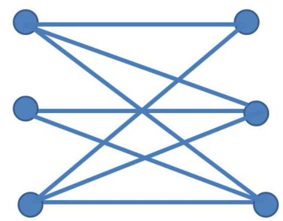
since it can be redrawn as



More motivation: excluded minors

Theorem (Kuratowski): **A graph is not planar, iff by deleting and contracting edges it can be reduced to one of the two graphs below:**





Corollary to Main Theorem

Consider the elementary operations performed on P:

- Rotation: $A_i \leftarrow T^T A_i T$ for all i and $B \leftarrow T^T BT$, where T is invertible.
- Contraction: $A_i \leftarrow \sum_{j=1}^m \lambda_j A_j$, for some i, where $\lambda_i \neq 0$, and $B \leftarrow B + \sum_{j=1}^m \mu_j A_j$.
- Deletion:
  - Delete a row and column with same index from all matrices,
  - Delete a matrix A_i.

Corollary to Main Theorem

Consider the elementary operations performed on P:

- Rotation: $A_i \leftarrow T^T A_i T$ for all i and $B \leftarrow T^T BT$, where T is invertible.
- Contraction: $A_i \leftarrow \sum_{j=1}^m \lambda_j A_j$, for some i, where $\lambda_i \neq 0$, and $B \leftarrow B + \sum_{j=1}^m \mu_j A_j$.
- Deletion:
  - Delete a row and column with same index from all matrices,
  - Delete a matrix A_i.

P badly behaved $\Rightarrow$ using these we can get

$$x_1 \begin{pmatrix} \alpha & 1 \\ 1 & 0 \end{pmatrix} \preceq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

where α is some real number.

Complexity implications

We use the real number model of computing: (see e.g. Blum, Cucker, Shub, Smale '98), in which one can store, and do operations on real numbers in unit time.

Reason: SDP can have irrational solutions, or solutions with exponentially many digits.

Corollary: In this model, the question "is a semidefinite system well-behaved?" is in NP $\cap$ co-NP.

I.e., we can verify in polynomial time that a system is well-behaved, or badly behaved.

Other conic LPs

• A conic linear system is

$$P = \{ x \,|\, Ax \leq_K b \} = \{ x \,|\, b - Ax \in K \},$$

where K is a closed, convex cone.

• Well-behaved, badly-behaved notions are defined analogously.

• Dual problem of $\sup \{ c^T x : x \in P \}$ involves K^*, the dual cone of K.

A result on general conic LPs

Let

- z be a maximum slack in $P = \{ x \mid Ax \leq_K b \}$, and
- $\text{dir}(z, K)$ the set of feasible directions at z in K.

A result on general conic LPs

Let

- z be a maximum slack in $P = \{ x \mid Ax \leq_K b \}$, and
- $\mathrm{dir}(z, K)$ the set of feasible directions at z in K.

Then P well-behaved $\Rightarrow$

$$R(A, b) \cap (\mathrm{cl}\ \mathrm{dir}(z, K) \setminus \mathrm{dir}(z, K)) = \emptyset.$$

A result on general conic LPs

Let

- z be a maximum slack in $P = \{ x \mid Ax \leq_K b \}$, and
- $\mathrm{dir}(z, K)$ the set of feasible directions at z in K.

Then P well-behaved $\Rightarrow$

$$R(A, b) \cap (\mathrm{cl}\ \mathrm{dir}(z, K) \setminus \mathrm{dir}(z, K)) = \emptyset.$$

If $K^* + \mathrm{face}(z, K)^\perp$ is closed, then $\Leftrightarrow$.

A result on general conic LPs

Let

- z be a maximum slack in $P = \{\, x \,|\, Ax \leq_K b \,\}$, and
- $\text{dir}(z, K)$ the set of feasible directions at z in K.

Then P well-behaved $\Rightarrow$

$$R(A, b) \cap (\text{cl dir}(z, K) \setminus \text{dir}(z, K)) = \emptyset.$$

If $K^* + \text{face}(z, K)^{\perp}$ is closed, then $\Leftrightarrow$.

Corollaries

K polyhedral $\Rightarrow$ P well-behaved.

A result on general conic LPs

Let

- z be a maximum slack in $P = \{ x \mid Ax \leq_K b \}$, and
- $\mathrm{dir}(z, K)$ the set of feasible directions at z in K.

Then P well-behaved $\Rightarrow$

$$R(A, b) \cap (\mathrm{cl}\ \mathrm{dir}(z, K) \setminus \mathrm{dir}(z, K)) = \emptyset.$$

If $K^* + \mathrm{face}(z, K)^\perp$ is closed, then $\Leftrightarrow$.

Corollaries

K polyhedral $\Rightarrow$ P well-behaved.

P strictly feasible, i.e. $z \in \mathrm{ri} K \Rightarrow P$ well-behaved.

A result on general conic LPs

Let

- z be a maximum slack in $P = \{\, x \mid Ax \leq_K b \,\}$, and
- $\mathrm{dir}(z, K)$ the set of feasible directions at z in K.

Then P well-behaved $\Rightarrow$

$$R(A, b) \cap (\mathrm{cl}\ \mathrm{dir}(z, K) \setminus \mathrm{dir}(z, K)) = \emptyset.$$

If $K^* + \mathrm{face}(z, K)^\perp$ is closed, then $\Leftrightarrow$.

Corollaries

K polyhedral $\Rightarrow$ P well-behaved.

P strictly feasible, i.e. $z \in \mathrm{ri} K \Rightarrow P$ well-behaved.

K "nice" (i.e. $K^* + F^\perp$ is closed for all F faces of K) $\Rightarrow$
z and $v \in \mathrm{cl}\ \mathrm{dir}(z, K) \setminus \mathrm{dir}(z, K)$ are certificates that P is badly behaved, when it is so.

Explanation

- Unification of two unrelated conditions: Slater's, and polyhedrality through a simple concept (feasible directions).
- When P is homogeneous, P is well-behaved $\Leftrightarrow$ A^*K^* closed.
- $\rightarrow$ we get back characterization of closedness of A^*K^* in Pataki 2007.

Conclusion

Conclusion

- **Duality in SDP: similar to LP, and similarly important: a dual solution gives a certificate of optimality.**

Conclusion

- Duality in SDP: similar to LP, and similarly important: a dual solution gives a certificate of optimality.

- However: pathologies occur: nonattainment, duality gaps, etc.

Conclusion

• Duality in SDP: similar to LP, and similarly important: a dual solution gives a certificate of optimality.

• However: pathologies occur: nonattainment, duality gaps, etc.

• Main result: all pathologies have a very simple underlying structure, i.e. "all bad SDPs look the same."

Conclusion

- **Duality in SDP: similar to LP, and similarly important: a dual solution gives a certificate of optimality.**

- **However: pathologies occur: nonattainment, duality gaps, etc.**

- **Main result: all pathologies have a very simple underlying structure, i.e. "all bad SDPs look the same."**

- **Also, an "excluded minor" type theorem for SDPs, and a general result for conic LPs.**

Thank you!