The S-lemma

Michael Orlitzky

Notation

By $\mathcal{S}^n$ we denote the set of real symmetric $n \times n$ matrices. Within this set, we have the cone $\mathcal{S}^n_+$ of positive semidefinite matrices.

The nonnegative orthant in $\mathbb{R}^n$ is $\mathbb{R}^n_+$.

The set of all nonnegative linear combinations of elements of X is written $\operatorname{cone}(X)$.

Notation

Cone inequality is indicated by a subscript. For example, $x \geq_K y$ indicates that $x - y \in K$. We write $x >_K y$ when $x - y$ is in the interior of K.

The interior of K is abbreviated $\operatorname{int}(K)$.

If K is a cone, then K^* represents its dual.

Notation

If L is a linear map, then L^* is its adjoint.

We use the usual inner product on $\mathbb{R}^n$ and the trace inner product $\langle X, Y \rangle = \operatorname{trace}\left(XY^T\right)$ on matrix spaces.

We make extraordinary use of the fact that $\operatorname{trace}(XYZ) = \operatorname{trace}(YZX) = \operatorname{trace}(ZXY)$ for any conceivable argument to the trace function.

Introduction

The *S-lemma* is a quadratic programming result.

The question it answers is,

> *When is one quadratic inequality a consequence of some other quadratic inequalities?*

Introduction

This question arises in the constrained problem,

$$\begin{aligned} &\text{minimize} && x^T B x \\ &\text{subject to} && x^T A_i x \geq 0 \\ &\text{for } i = 1, 2, \ldots, m. \end{aligned}$$

Its optimal value is either $-\infty$ or zero, since we can scale any solution by a positive value.

Introduction

But which is it, negative infinity or zero?

Is there a feasible point x such that $x^T Bx < 0$?

Or does the system $x^T A_i x \geq 0$ imply that $x^T Bx \geq 0$ for all x?

Introduction

In the linear case,

$$\begin{aligned} &\text{minimize} \quad && x^T b \\ &\text{subject to} \quad && x^T a_i \geq 0 \\ &\text{for } i = 1, 2, \ldots, m, && \end{aligned}$$

we have Farkas's lemma:

$$x^T b \geq 0 \iff \exists \lambda \in \mathbb{R}^n_+ : b = \sum_{i=1}^{m} \lambda_i a_i$$

In the quadratic case, we're not so lucky.

The S-lemma for quadratic inequalities is analogous to Farkas's lemma for linear ones. But, the S-lemma only applies when there is exactly one constraint,

$$\begin{aligned} &\text{minimize} \quad x^T B x \\ &\text{subject to } x^T A_1 x \geq 0. \end{aligned}$$

Conic duality

To prove the S-lemma, we need the concept of strong duality for conic programs. The primal form of a conic program is

$$\begin{aligned} \text{minimize} \quad & \langle b, x \rangle \\ \text{subject to } L(x) &\geq_{K_2} c \\ x &\geq_{K_1} 0, \end{aligned}$$

where K_1, K_2 are cones and L is linear.

CONIC DUALITY

The dual of this conic program is

$$\begin{aligned} \text{maximize} \quad & \langle c, \mu \rangle \\ \text{subject to } L^*(\mu) & \leq_{K_1^*} b \\ \mu & \geq_{K_2^*} 0, \end{aligned}$$

where L^* is the adjoint of L, and K_1^*, K_2^* are dual to K_1, K_2.

Conic duality

Conic duality is more subtle than linear duality.

Theorem (conic duality theorem).

If the primal conic programming problem is bounded below and strictly feasible then the dual problem is solvable with the same optimal value.

Conic duality

Even the definition of *strictly feasible* is tricky.

Definition.

The primal form of a conic programming problem is said to be strictly feasible if there exists a feasible point x such that

$$\begin{cases} x \in \operatorname{int}(K_1), & \text{if } K_2 = \{0\} \\ L(x) - c \in \operatorname{int}(K_2), & \text{otherwise}. \end{cases}$$

Conic duality

Choosing $K_2 = \{0\}$ lets us have equality constraints. The associated problem is,

$$\begin{aligned} \text{minimize} \;\; & \langle b, x \rangle \\ \text{subject to} \;\; & L(x) = c \\ & x \in K_1. \end{aligned}$$

This is the form we'll need later.

Technical lemma

We also need a technical lemma that we'll prove.

Lemma. Let $P, Q \in \mathcal{S}^n$ with $\operatorname{trace}(P) \geq 0$ and $\operatorname{trace}(Q) < 0$.

Then, there exists a single vector $z \in \mathbb{R}^n$ such that both $z^T P z \geq 0$ and $z^T Q z < 0$.

Proof.

Begin by diagonalizing $Q = U \Lambda U^T$.

One way to make $z^T Q z < 0$ is to take $z = U\xi$ where $\xi \in \mathbb{R}^n$ has entries $\xi_i = \pm 1$. Then,

$$z^T Q z = \xi^T \Lambda \xi = \operatorname{trace}(\Lambda) = \operatorname{trace}(Q).$$

And by assumption, $\operatorname{trace}(Q) < 0$.

Technical lemma

Now the problem is to choose ξ so that

$$z^T P z = \xi^T U^T P U \xi \geq 0.$$

The trouble is that $U^T P U$ is not diagonal, so we don't know what to make ξ to collapse this down to something involving $\operatorname{trace}(P) \geq 0$.

Technical lemma

Ben-Tal and Nemirovski employ a clever trick: suppose ξ is a *random* vector whose entries ξ_i take the value ± 1 each with probability $1/2$.

The expectation $\mathbb{E}$ of any particular ξ_i is zero, but for a product,

$$\mathbb{E}\left(\xi_i \xi_j\right) = \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ otherwise } . \end{cases}$$

Technical lemma

We now ask, what is the expectation of $z^T P z$, or $\mathbb{E}\left(\xi^T\left(U^T P U\right)\xi\right)$? Expanding, this is

$$\mathbb{E}\left(\sum_{i=1}^{n} \xi_i \xi^T \left(\text{the } i\text{th row of } U^T P U\right)\right).$$

And since $\mathbb{E}\left(\xi_i \xi_j\right) = \delta_{ij}$, we have $\xi_i \xi^T = e_i^T$ where e_i is the ith standard basis vector.

Technical lemma

Thus,

$$\begin{aligned}
\mathbb{E}\left(z^T P z\right) &= \sum_{i=1}^{n} e_i^T \left(\text{the } i\text{th row of } U^T P U\right) \\
&= \sum_{i=1}^{n} \text{ the } i\text{th diagonal of } U^T P U \\
&= \operatorname{trace}\left(U^T P U\right) = \operatorname{trace}(P) \geq 0.
\end{aligned}$$

The expectation is nonnegative, so there must be some concrete ξ making $z^T P z \geq 0$. Use that. $\square$

S-lemma

Theorem (S-lemma).

Let $A, B \in \mathcal{S}^n$ and suppose that $\bar{x}^T A \bar{x} > 0$ for some $\bar{x} \in \mathbb{R}^n$. Then the following are equivalent:

1. The inequality $x^T A x \geq 0$ implies $x^T B x \geq 0$.
2. There exists a $\lambda \geq 0$ such that $B \geq_{\mathcal{S}^n_+} \lambda A$.

S-lemma

Proof. The second item easily implies the first. If $(B - \lambda A) \in \mathcal{S}_{+}^{n}$, then

$$x^T B x \geq \lambda x^T A x \text{ for all } x \in \mathbb{R}^n.$$

Now if $x^T A x \geq 0$, then $x^T B x \geq 0$ too, so the implication in the first item holds.

S-lemma

Next suppose that the first item holds, so that $x^T A x \geq 0$ implies $x^T B x \geq 0$. Then the solution to the optimization problem

$$\min_{x \in \mathbb{R}^n} \left\{ x^T B x \;\middle|\; x^T A x \geq 0 \right\}$$

exists and has nonnegative objective value.

S-lemma

No generality is lost by requiring $\|x\| = 1$. Any solution to the original problem can be scaled to unit norm without violating the constraint.

Thus the solution to

$$\min_{x \in \mathbb{R}^n} \left\{ x^T B x \mid x^T A x \geq 0, \ \|x\| = 1 \right\}$$

exists and has nonnegative objective value.

S-lemma

Using the fact that

$$x^T B x = \operatorname{trace}\left(x^T B x\right) = \operatorname{trace}\left(B x x^T\right),$$

we can rewrite the problem

$$\min_{x \in \mathbb{R}^n} \left\{ \left\langle B, x x^T \right\rangle \mid \left\langle A, x x^T \right\rangle \geq 0,\ \left\| x x^T \right\| = 1 \right\}.$$

S-LEMMA

The semidefinite relaxation of this problem is

$$\min_{X \in \mathcal{S}^n_+} \left\{ \langle B, X \rangle \mid \langle A, X \rangle \geq 0,\ \operatorname{trace}(X) = n \right\}.$$

This should be easier to solve, since not every $X \in \mathcal{S}^n_+$ is of the form $X = xx^T$.

The $\operatorname{trace}(X) = n$ constraint comes from considering the identity matrix.

S-lemma

In the statement of the S-lemma, we required that $\bar{x}^T A \bar{x} > 0$ for some $\bar{x} \in \mathbb{R}^n$. Here's why.

Let $\bar{X} = \bar{x}\bar{x}^T$. By scaling this appropriately, we can make $\operatorname{trace}\left(\bar{X}\right) = n$. Now,

$$\left\langle A, \bar{X} \right\rangle = \operatorname{trace}\left(A\bar{x}\bar{x}^T\right) = \operatorname{trace}\left(\bar{x}^T A \bar{x}\right) > 0.$$

S-lemma

So $\bar{X} \in \mathcal{S}^n_+$ is feasible and interior to the half-space $\{X \mid \langle A, X \rangle \geq 0\}$, which we call H_A.

From $\bar{X}$, we can move a small distance along the hyperplane $\{X \mid \operatorname{trace}(X) = n\}$ into the interior of $\mathcal{S}^n_+$. In this manner, we can find a feasible $\bar{Y}$ such that $\bar{Y} \in \operatorname{int}(\mathcal{S}^n_+ \cap H_A)$.

S-LEMMA

Also note that the problem

$$\min_{X \in \mathcal{S}^n_+} \left\{ \langle B, X \rangle \mid \langle A, X \rangle \geq 0,\ \operatorname{trace}(X) = n \right\}$$

is bounded below. The constraint $\operatorname{trace}(X) = n$ forms a hyperplane that intersects the cone $\mathcal{S}^n_+$ in a compact base. Thus the feasible set is compact.

Recall:

Theorem (conic duality theorem).

If the primal conic programming problem is bounded below and strictly feasible then the dual problem is solvable with the same optimal value.

S-lemma

To apply the duality theorem, we need to put our problem in standard form:

$$\begin{aligned} \text{minimize} \;\; & \langle B, X \rangle \\ \text{subject to} \;\; & L(X) \geq_{K_2} c \\ & X \geq_{K_1} 0. \end{aligned}$$

From the conic duality theorem, we guess that

$$K_1 := \mathcal{S}^n_+ \cap H_A \text{ and } K_2 := \{0\}\,.$$

S-lemma

By taking $L(X) \coloneqq \operatorname{trace}(X)$ and $c \coloneqq n$, we obtain the desired problem in standard form:

$$\begin{aligned} &\text{minimize} && \langle B, X \rangle \\ &\text{subject to} && \operatorname{trace}(X) = n \\ &&& X \in \mathcal{S}^n_+ \cap H_A. \end{aligned}$$

In this formulation, the $\bar{Y} \in \operatorname{int}(\mathcal{S}^n_+ \cap H_A)$ that we found a moment ago shows strict feasibility.

S-LEMMA

For the dual problem, we need the dual cones of K_1 and K_2. The dual of $K_2 = \{0\}$ is the entire space $\mathbb{R}$. The dual of K_1 is known to be

$$K_1^* = \left(\mathcal{S}_+^n \cap H_A\right)^* = \left(\mathcal{S}_+^n\right)^* + H_A^*.$$

The cone $\mathcal{S}_+^n$ is self-dual, and the dual of a half-space is the ray that defines it. So,

$$K_1^* = \mathcal{S}_+^n + \operatorname{cone}\left(\{A\}\right).$$

S-lemma

We also need the adjoint of L, defined by

$$\langle X, L^*(\mu) \rangle = \langle L(X), \mu \rangle = \mu \operatorname{trace}(X)$$

for all $X \in \mathcal{S}^n$ and $\mu \in \mathbb{R}$. By inspection, $L^*(\mu) \coloneqq \mu I$ will work:

$$\langle X, L^*(\mu) \rangle = \mu \langle I, X \rangle = \mu \operatorname{trace}(X).$$

S-lemma

Substituting, we have the dual problem:

$$\begin{aligned} \text{maximize} \quad & \langle c, \mu \rangle \\ \text{subject to } & L^{*}(\mu) \preceq_{\mathcal{S}^{n}_{+} + \operatorname{cone}(\{A\})} B \\ & \mu \in \mathbb{R}. \end{aligned}$$

Fortunately, this can be simplified a bit. First, $\langle c, \mu \rangle$ is just $n\mu$.

S-lemma

Next, the cone constraint says

$$B - \mu I \in \mathcal{S}^n_+ + \operatorname{cone}\left(\{A\}\right).$$

If λA is the component in $\operatorname{cone}\left(\{A\}\right)$, then

$$B \geq_{\mathcal{S}^n_+} \lambda A + \mu I.$$

S-lemma

Thus the dual problem simplifies to

$$\begin{aligned} \text{maximize} \;\; & n\mu \in \mathbb{R} \\ \text{subject to} \;\; & B \geq_{\mathcal{S}^n_+} \lambda A + \mu I \\ & \lambda \geq 0. \end{aligned}$$

S-lemma

Remember: we were trying to prove that there exists a $\lambda \geq 0$ such that $B \geq_{\mathcal{S}^n_+} \lambda A$.

We've just discovered that

$$B \geq_{\mathcal{S}^n_+} \lambda A + \mu I.$$

If $\mu \geq 0$ here, then $B \geq_{\mathcal{S}^n_+} \lambda A$ as desired.

S-lemma

By the conic duality theorem, this problem shares its optimal solution with the primal problem.

Let $n\bar{\mu}$ be the optimal solution to the dual:

$$\begin{aligned} \text{maximize} \quad & n\mu \in \mathbb{R} \\ \text{subject to} \quad & B \geq_{\mathcal{S}^n_+} \lambda A + \mu I \\ & \lambda \geq 0. \end{aligned}$$

S-lemma

Then $n\bar{\mu}$ is optimal for the primal as well:

$$\begin{aligned} \text{minimize} \quad & \langle B, X \rangle \\ \text{subject to} \quad & \langle A, X \rangle \geq 0 \\ & \operatorname{trace}(X) = n \\ & X \in \mathcal{S}^n_+. \end{aligned}$$

S-LEMMA

Suppose, on the contrary, that $\left\langle B, \widetilde{X} \right\rangle = n\bar{\mu} < 0$ is the optimal value of the primal, achieved by $\widetilde{X}$.

Since $\widetilde{X} \in \mathcal{S}^n_+$, it factors into $\widetilde{X} = DD^T$. Now

$$\begin{aligned} \left\langle B, \widetilde{X} \right\rangle &= \operatorname{trace}\left(BDD^T\right) \\ &= \operatorname{trace}\left(D^T BD\right) = n\bar{\mu} \\ \left\langle A, \widetilde{X} \right\rangle &= \operatorname{trace}\left(D^T AD\right) \geq 0. \end{aligned}$$

S-lemma

If $n\bar{\mu} < 0$, we let $P = D^T AD$ and $Q = D^T BD$.

Recall:

Lemma. There exists a single vector $z \in \mathbb{R}^n$ such that both $z^T Pz \geq 0$ and $z^T Qz < 0$.

S-lemma

Suppose $z^T \left(D^T A D\right) z \geq 0$ and $z^T \left(D^T B D\right) z < 0$. If we let $\bar{z} = Dz$, then $\bar{z}^T A \bar{z} \geq 0$ while $\bar{z}^T B \bar{z} < 0$.

This contradicts item one from the S-lemma:

1. The inequality $x^T A x \geq 0$ implies $x^T B x \geq 0$.

Therefore, $\mu \geq 0$.

S-lemma

Now since the optimal μ is nonnegative in

$$\begin{aligned} \text{maximize} \quad & n\mu \in \mathbb{R} \\ \text{subject to} \quad & B \geq_{\mathcal{S}^n_+} \lambda A + \mu I \\ & \lambda \geq 0, \end{aligned}$$

we have $B \geq_{\mathcal{S}^n_+} \lambda A$.

This was the desired second item from the S-lemma, and we are done. $\square$

References I

[1] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, SIAM, Philadelphia, 2001.

[2] B. Gärtner and J. Matoušek. Approximation Algorithms and Semidefinite Programming. Springer-Verlag, Berlin, 2012.