MIT 6.7220/15.084 — Nonlinear Optimization  Thu, Feb 29th 2024

# Lecture 6
# Conic optimization

Instructor: Prof. Gabriele Farina (✉ gfarina@mit.edu)*

In Lectures 2 and 4, we considered normal cones at the intersection of linear equality and inequality constraints. In Exercise 1 of Homework 1, we were also able to give closed formulas for the normal cone in certain sets, such as a ball.

In Lecture 5, we turned our attention to more general feasible sets. There, we saw how KKT conditions define necessary conditions for feasible sets defined by *functional constraints* by approximating the feasible set with the linearization of the active constraints and considering the normal cone to the intersection of those linearizations.

Today, we look at a class of convex feasible sets that are neither linear equality nor inequality constraints. Also, they are not defined via functional constraints, making the KKT machinery inapplicable.

## 1 Conic optimization

A *conic* optimization problem is a nonlinear optimization problem whose feasible set is the intersection between an affine subspace (that is, a system of linear equalities) and a nonempty closed convex cone $\mathcal{K}$:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b \\ & x \in \mathcal{K} \end{aligned}$$

This class of problems captures linear programming, semidefinite programming, second-order cone programming, copositive programming, and more, depending on the specific cone $\mathcal{K}$ that is selected.

### 1.1 Nonnegative cone ⟷ Linear programming

The first example is the nonnegative cone $\mathbb{R}^n_{\geq 0}$. In this case, the conic problem takes the more familiar form

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

which is a *linear program*. So, for the specific choice of the nonnegative cone, conic programming corresponds to linear programming.

*These notes are class material that has not undergone formal peer review. The TAs and I are grateful for any reports of typos.

## 1.2 Ice-cream (aka. Lorentz) cone $\longleftrightarrow$ Second-order conic programming

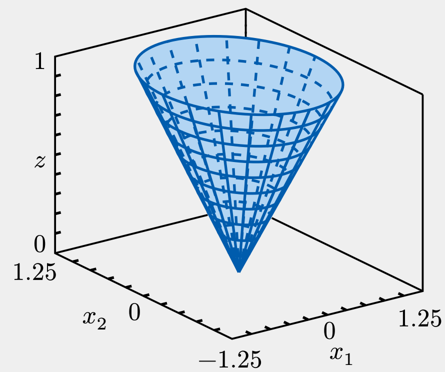
**Definition 1.1.**

The *ice-cream cone*, or *Lorentz cone*, is defined as

$$\mathcal{L}^n := \left\{(x, z) \in \mathbb{R}^{n-1} \times \mathbb{R} : z \geq \|x\|_2\right\}.$$

The figure on the right shows the shape of $\mathcal{L}^3$.

Conic problems for the specific choice of the Lorentz cone are usually called *second-order cone programs*, or *SOCP* for short. Several problems of interest in engineering and physics can be modeled as SOCP, including the design of antenna arrays, the positions of spring systems at rest, and grasp planning in robotics.




## 1.3 Semidefinite cone $\longleftrightarrow$ Semidefinite programming

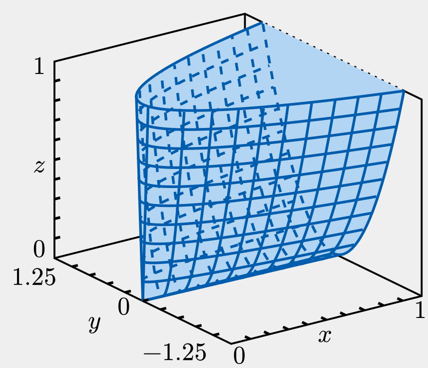
**Definition 1.2.** The *semidefinite* cone $\mathcal{S}^n$ is the set of positive semidefinite $n \times n$ matrices:

$$\begin{aligned}\mathcal{S}^n &:= \{X \in \mathbb{S}^n : X \succcurlyeq 0\} \\ &= \left\{X \in \mathbb{S}^n : a^\top X a \geq 0 \quad \forall a \in \mathbb{R}^n\right\},\end{aligned}$$

where $\mathbb{S}^n$ is the set of symmetric $n \times n$ real matrices.

The figure on the right shows the cone of values $(x, y, z)$ corresponding to the semidefinite cone of $2 \times 2$ matrices

$$\begin{pmatrix} x & y \\ y & z \end{pmatrix} \succcurlyeq 0.$$




In this simple $2 \times 2$ case, the positive semidefiniteness condition is equivalent to

$$\begin{aligned} x + z &\geq 0 \qquad \text{(the sum of the eigenvalues [trace] is nonnegative)} \\ xz - y^2 &\geq 0 \qquad \text{(the product of the eigenvalues [determinant] is nonnegative.)} \end{aligned}$$

These conditions are equivalent to $x, z \geq 0 \wedge xz \geq y^2$. So, slices of the cone with planes orthogonal to the $z$-axis are shaped like parabolas $x \geq y^2/z$. Similarly, slices with planes orthogonal to the $x$-axis are shaped like parabolas $z \geq y^2/x$. Note how the curvature of the parabolas decreases as the slicing plane gets further away from the origin.

**Remark 1.1.** Semidefinite programming subsumes both linear programming and second-order cone programming. This is because

$$x \in \mathbb{R}^n_{\geq 0} \quad \Longleftrightarrow \quad \begin{pmatrix} x_1 & & \\ & \ddots & \\ & & x_n \end{pmatrix} \succcurlyeq 0, \text{ and}$$

$$z \geq \|x\|_2 \quad \Longleftrightarrow \quad \begin{pmatrix} zI & x \\ x^\top & z \end{pmatrix} \succcurlyeq 0.$$

Semidefinite programming is an extremely powerful tool with applications in all disciplines. I have noted resources that treat semidefinite programming extensively at the end of this document.

## 1.4 **Copositive cone** $\longleftrightarrow$ Copositive programming

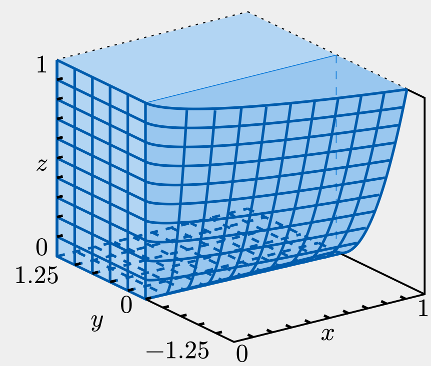
**Definition 1.3.** The *copositive* cone $\mathcal{C}^n$ is the set of symmetric $n \times n$ real matrices:

$$\mathcal{C}^n := \left\{X \in \mathbb{S}^n : a^\top X a \geq 0 \quad \forall a \in \mathbb{R}^n_{\geq 0}\right\}.$$

The difference with the positive semidefinite cone $\mathcal{S}^n$ is given by the fact that we need $a^\top X a \geq 0$ *only for nonnegative vectors* $a \in \mathbb{R}^n_{\geq 0}$, and not for all $a \in \mathbb{R}^n$.

The figure on the right shows the cone of values $(x, y, z)$ corresponding to the copositive cone of $2 \times 2$ matrices, that is

$$\begin{pmatrix} x & y \\ y & z \end{pmatrix} \in \mathcal{C}^2.$$




It follows immediately from the definition that $\mathcal{C}^n$ is a superset of both the cone of nonnegative symmetric matrices $\mathcal{N}^n := \mathbb{S}^n \cap \mathbb{R}^{n\times n}_{\geq 0}$ and the cone of semidefinite matrices. Hence,

$$\mathcal{N}^n + \mathcal{S}^n \subseteq \mathcal{C}^n \qquad \text{for all } n \in \mathbb{N}.$$

Curiously, the reverse inequality is known to hold, *but only up to dimension* $n \leq 4$ [Dia62]:

$$\mathcal{C}^n = \mathcal{S}^n + \mathcal{N}^n \qquad \text{if } n \leq 4.$$

Beyond dimension 4, some matrices are copositive, and yet they are not expressible as the sum of $\mathcal{S}^n$ with $\mathcal{N}^n$.

**Remark 1.2.** One important fact to know about the copositive cone is that *despite being a closed convex cone, optimization of even linear functions over the copositive cone is computationally intractable*!

For example, the size of the *maximum independent set* of a simple graph $G = (V, E)$, that is, the size of the largest set $S \subseteq V$ of vertices with the property that no two vertices in $S$ are connected by an edge in $E$, can be computed as the solution to the following copositive problem. Let $A \in \{0, 1\}^{V \times V}$ denote the adjacency matrix of $G$, that is,

$$A_{ij} = \begin{cases} 1 \text{ if } (i, j) \in E \\ 0 \text{ otherwise.} \end{cases}$$

The size of the maximum independent set of $G$ can then be found by solving the copositive optimization problem

$$\begin{aligned} \min_{w, t \in \mathbb{R}, Q} \quad & w \\ \text{s.t.} \quad & Q = wI + tA - \mathbb{1}\mathbb{1}^\top \\ & Q \in \mathcal{C}^n \end{aligned} \tag{MIS}$$

where $\mathbb{1} \in \mathbb{R}^V$ is the vector of all ones. More information is available in the paper by Klerk, E. de, & Pasechnik, D. V. [KP02]. Since deciding whether a graph admits an independent set of size $k$ cannot take polynomial time unless $\mathsf{P} = \mathsf{NP}$, this means that solving copositive optimization problems must take more than polynomial time in the worst case unless $\mathsf{P} = \mathsf{NP}$. The hardness applies already for linear objectives, as the previous example shows. So, we conclude—for example—that one cannot construct an efficient separation oracle for $\mathcal{C}^n$.

Several techniques are known to "relax" the copositive cone by, for example, approximating it with a (higher-dimensional) semidefinite cone. The monographs of Gärtner, B., & Matoušek, J. [GM12] and Klerk, E. de. [Kle02] do an excellent job at explaining approximation techniques that involve semidefinite cones.

## 2 Optimality conditions and conic duality

In general, the feasible set $\Omega$ of conic optimization problems cannot be easily captured via *functional* constraints. Because of that, we cannot use the KKT conditions we discussed in the previous lecture to compute the normal cone at each point of the feasible set. In this situation, we need to compute the normal cone from first principles.

To simplify the treatment and avoid the machinery of *relative* interiors (interiors in the topology induced by the smallest affine subspace that contains a set), we will assume in the following that $\mathcal{K}$ is a cone with interior. This condition is satisfied by all the cases discussed above.

In order to compute the normal cone at a point at the intersection of $\{Ax = b\}$ and $\mathcal{K}$, we will use the following general result (again based on separation), which enables us to consider the affine subspace and the cone separately and sum their normal cones.

**Theorem 2.1**. Let $H \coloneqq \{x \in \mathbb{R}^n : Ax = b, \text{with } A \in \mathbb{R}^{m\times n}\}$ be an affine subspace, and $S$ be a closed convex set (not necessarily a cone) such that $S^\circ \cap H$ is nonempty, where $S^\circ$ is the interior of $S$. For any $x \in H \cap S$,

$$\mathcal{N}_{H\cap S}(x) = \mathcal{N}_H(x) + \mathcal{N}_S(x).$$

So, using the fact (see Lecture 2 or Lecture 5) that $\mathcal{N}_H(x) = \mathrm{colspan}(A^\top)$, we have

$$\mathcal{N}_{H\cap S}(x) = \{A^\top \lambda + z : \lambda \in \mathbb{R}^m, z \in \mathcal{N}_S(x)\}.$$

**Remark 2.1**. The condition of existence of a feasible point in the interior is sometimes referred to as *Slater's condition*, *strict feasibility condition*, or *strong feasibility condition*. The term *Slater's condition* is appropriate in light of the constraint qualification condition for the KKT theory we discussed last time. In both cases, the insistence on an interior solution is required to rule out pathological behavior.

We know from Lectures 2 and 5 that Slater's condition was not necessary for the intersection of halfspaces, which are flat surfaces. The condition becomes, however, crucial when considering sets that can have curvature. For example, consider the intersection of a two-dimensional ball of radius 1 centered in $(x, y) = (1, 0)$ with the plane $x = 0$. The normal cone at the intersection is $\mathbb{R}^2$, but this is not equal to the sum between the normal cones of the sets at the intersection points (which evaluates to $\mathbb{R} \times \{0\}$).

**Remark 2.2**. The previous result can be generalized, without much effort nor substantially new ideas to the case of intersection between two generic convex sets whose relative interiors have a nonempty intersection. We do not need such a generalization for this lecture; however, you might find it useful to know that—provided mild hypotheses hold—the *normal cone to the intersection between convex sets is equal to the sum of the normal cones to the individual sets.*

## 2.1 The normal cone at a point in $\mathcal{K}$

With Theorem 2.1 in our toolbox, the only obstacle to writing first-order optimality conditions for conic problems is being able to compute the normal at points in $\mathcal{K}$. Luckily, we can do that in closed form in all the abovementioned cases.

In particular, it turns out that the normal cone at any point in $\mathcal{K}$ can be defined starting from the normal cone at 0 (remember that 0 is always part of a nonempty cone, as a straightforward consequence of the definition of cone). The normal cone for a cone $\mathcal{K}$ at $0 \in \mathcal{K}$ is such an important object, that several names for it or its related quantities are used in the literature:

- The *polar cone* of $\mathcal{K}$, often denoted $\mathcal{K}^\perp$, is exactly the normal cone at 0:

$$\mathcal{K}^\perp := \mathcal{N}_\mathcal{K}(0) = \{d : \langle d, y\rangle \le 0 \quad \forall y \in \mathcal{K}\}.$$

- The *dual cone* of $\mathcal{K}$, often denoted $\mathcal{K}^*$, is the opposite of the normal cone at 0:

$$\mathcal{K}^* := -\mathcal{K}^\perp = -\mathcal{N}_\mathcal{K}(0) = \{d : \langle d, y\rangle \ge 0 \quad \forall y \in \mathcal{K}\}.$$

For all the important cones considered earlier, we have a very precise idea of what their polar and dual cones look like:

**Theorem 2.2**.

1. The nonnegative cone, the Lorentz cone, and the semidefinite cones are *self-dual cones*, that is, $\mathcal{K}^* = \mathcal{K}$.
2. The dual cone to the copositive cone $\mathcal{C}^n$ is called the *totally positive* cone, defined as

$$\mathcal{P}^n := \left\{ \sum_{i=1}^k z_i z_i^\top : z_i \in \mathbb{R}^n_{\ge 0}, k \in \mathbb{N} \right\} = \left\{ BB^\top : B \in \mathbb{R}^{n\times m}_{\ge 0}, m \in \mathbb{N} \right\}$$

[▷ Try to prove the previous theorem.] Since for a closed convex cone $\mathcal{K}$, $(\mathcal{K}^*)^* = \mathcal{K}$, it follows that the dual to the totally positive cone is the copositive cone. [▷ You should also try to prove that $(\mathcal{K}^*)^* = \mathcal{K}$.]

Once the normal cone at 0 is established, the normal cone at any point $x \in \mathcal{K}$ can be recovered as a function of $\mathcal{N}_\mathcal{K}(0)$.

**Theorem 2.3**. Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a nonempty, closed convex cone. The normal cone at any point $x \in \mathcal{K}$ is given by

$$\mathcal{N}_\mathcal{K}(x) = \{d \in \mathbb{R}^n : \langle d, x\rangle = 0,\ d \in \mathcal{N}_\mathcal{K}(0)\}.$$

*Proof.* Fix a generic $x \in \mathcal{K}$. We break the proof into a proof of the two separate inclusions.

($\subseteq$) Let $d \in \mathbb{R}^n$ be a direction in $\mathcal{N}_{\mathcal{K}}(x)$, that is,

$$\langle d, y - x\rangle \le 0 \qquad \forall y \in \mathcal{K}.$$

Since $\mathcal{K}$ is a cone, we have that both $2x$ and $\frac{1}{2}x$ belong to $\mathcal{K}$. So,

$$\begin{aligned} \langle d, 2x - x\rangle \le 0 \qquad &\Longrightarrow \qquad \langle d, x\rangle \le 0, \text{and} \\ \left\langle d, \frac{1}{2}x - x\right\rangle \le 0 \qquad &\Longrightarrow \qquad \langle d, x\rangle \ge 0, \end{aligned}$$

showing that $\langle d, x\rangle = 0$ is necessary. But then, from the condition that $\langle d, y - x\rangle \le 0$ for all $y \in \mathcal{K}$ we deduce that necessarily $\langle d, y\rangle \le 0$ for all $y \in \mathcal{K}$, which means that $d \in \mathcal{K}^\perp = \mathcal{N}_{\mathcal{K}}(0)$.

($\supseteq$) Vice versa, let $d \in \{d \in \mathbb{R}^n : \langle d, x\rangle = 0,\ d \in \mathcal{N}_{\mathcal{K}}(0)\}$. Then, for any $y \in \mathcal{K}$,

$$\begin{aligned} \langle d, y - x\rangle &= \langle d, y\rangle - \langle d, x\rangle \\ &= \langle d, y\rangle && (\text{since } \langle d, x\rangle = 0) \\ &\le 0 && (\text{since } d \in \mathcal{K}^\perp). \end{aligned}$$

□

## 2.2 Conic duality

We now have all the ingredients to write first-order optimality conditions for a conic problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b \\ & x \in \mathcal{K}. \end{aligned} \tag{2}$$

Specifically, by combining Theorem 2.1 with Theorem 2.3 and the sufficiency of first-order optimality conditions seen in Lecture 3, we obtain the following result.

**Theorem 2.4**. If $f$ is convex and differentiable, and the conic optimization (2) is *strictly feasible*, that is, there exists $x \in \mathcal{K}^\circ$ such that $Ax = b$, then a point $x^* \in \Omega := \mathcal{K} \cap \{x : Ax = b\}$ is a minimizer of $f$ on $\Omega$ if and only if

$$-\nabla f(x^*) \in \{A^\top \lambda + z : \lambda \in \mathbb{R}^m, z \in \mathcal{N}_{\mathcal{K}}(x^*)\},$$

that is, expanding $\mathcal{N}_{\mathcal{K}}(x^*)$ using Theorem 2.3, if and only if

$$\begin{aligned} & -\nabla f(x^*) \in \{A^\top \lambda + z : \lambda \in \mathbb{R}^m, z \in \mathcal{K}^\perp, \langle z, x^*\rangle = 0\} \\ \iff \quad & -\nabla f(x^*) \in \{A^\top \lambda - z : \lambda \in \mathbb{R}^m, z \in \mathcal{K}^*, \langle z, x^*\rangle = 0\}. \end{aligned}$$

In particular, when $f(x) = \langle c, x\rangle$ is a *linear* objective, then optimality of $x^*$ for (2) is equivalent to the existence of a solution for

$$-c = -A^\top \lambda - z \qquad \text{subject to } \lambda \in \mathbb{R}^m, z \in \mathcal{K}^*, \langle z, x^*\rangle = 0.$$

This suggests that whenever the conditions of strict feasibility holds, then a duality theory for conic problems can be derived following the exact same steps as we did in Lecture 2. In particular, we find that the dual problem is

$$\begin{aligned}\max_{\lambda,z}\ & \langle b,\lambda\rangle\\ \text{s.t.}\ & z = c - A^\top\lambda\\ & \lambda\in\mathbb{R}^m\\ & z\in\mathcal{K}^*.\end{aligned}$$

In particular, after showing that the value of the dual is always upper bounded by the value of the primal (weak duality) using the same steps as in Lecture 2, we conclude the following. [▷ Try working out the details, it should only take a minute.]

**Theorem 2.5**. If the primal problem

$$\begin{aligned}\min_{x}\ & \langle c,x\rangle\\ \text{s.t.}\ & Ax = b\\ & x\in\mathcal{K}\end{aligned}$$

is strictly feasible and admits an optimal solution $x^*$, then the dual problem

$$\begin{aligned}\max_{\lambda,z}\ & \langle b,\lambda\rangle\\ \text{s.t.}\ & z = c - A^\top\lambda\\ & \lambda\in\mathbb{R}^m\\ & z\in\mathcal{K}^*\end{aligned}$$

admits an optimal solution $(\lambda^*, z^*)$ such that:

- the values of the two problems coincide: $\langle c, x^*\rangle = \langle b, \lambda^*\rangle$; and
- the solution $z^*$ satisfies the complementary slackness condition $\langle x^*, z^*\rangle = 0$.

**Failure of duality when the constraint qualification is not satisfied**. It is essential to remember that duality might fail when the strict feasibility constraint qualification is not satisfied. Specifically, the primal problem might be feasible and yet the dual might not have an optimal solution.

The failure mode can be of different kinds:

- the primal has an optimal solution, the supremum of the dual matches the value of the primal, and yet the dual does not have a maximizer (Example 2.1); or
- the primal has an optimal solution, the dual has an optimal solution, but the values of the problems differ (Example 2.2).

**Example 2.1** ([Kle02]). Consider the problem

$$\begin{aligned}\min_{X}\ & 2X_{12} + X_{22}\\ \text{s.t.}\ & X_{11} = 0\\ & X_{22} = 1\\ & X \succcurlyeq 0.\end{aligned}$$

The only point in the feasible set is $\begin{pmatrix} 0 & 0\\ 0 & 1\end{pmatrix}$; therefore, the optimal value is 1.

Since the matrix $\begin{pmatrix} 0 & 0\\ 0 & 1\end{pmatrix}$ is *not* in the interior of the semidefinite cone (one eigenvalue is 0), strict feasibility does not hold.

The dual problem in this case is

$$
\begin{aligned}
\max_{\lambda,Z}\ & \lambda_2 \\
\text{s.t.}\ & Z = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} - \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} - \lambda_2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \\
& \lambda \in \mathbb{R}^2 \\
& Z \succcurlyeq 0
\end{aligned}
$$

For $\lambda = \left(-\frac{1}{\varepsilon}, 1 - \varepsilon\right)$, $\varepsilon > 0$ we obtain the feasible point $Z = \begin{pmatrix} \frac{1}{\varepsilon} & 1 \\ 1 & \varepsilon \end{pmatrix} \succcurlyeq 0$ with objective value $1 - \varepsilon$. Hence, the supremum of the dual matches the value of the primal, but the *dual has no optimal solution.*

**Example 2.2** ([BV04]). Consider the primal problem

$$
\begin{aligned}
\min_{X}\ & x_2 \\
\text{s.t.}\ & X = \begin{pmatrix} x_2 + 1 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{pmatrix} \\
& X \succcurlyeq 0.
\end{aligned}
$$

Since the determinant must be nonnegative, it follows immediately that $x_2 = 0$ for all feasible points, and so the objective value is 0. The dual problem is (after some manipulations)

$$
\begin{aligned}
\max_{Z}\ & -Z_{11} \\
\text{s.t.}\ & Z_{11} + 2Z_{23} = 1 \\
& Z_{22} = 0 \\
& Z \succcurlyeq 0.
\end{aligned}
$$

Since $Z$ is positive semidefinite, then so must be the submatrix $\begin{pmatrix} Z_{22} & Z_{23} \\ Z_{23} & Z_{33} \end{pmatrix}$. But since $Z_{22} = 0$, then necessarily $Z_{23} = 0$. So, for any feasible point in the dual, the constraint $Z_{11} + Z_{23} = 1$ is equivalent to $Z_{11} = 1$. So, the dual attains an optimal point, but the value at optimality is different from the value at optimality of the primal problem.

# 3 Further readings and bibliography

Hiriart-Urruty, J.-B., & Seeger, A. [HS10] wrote an excellent survey of the properties of the copositive cone.

For further reading on semidefinite programming, I find the short books by Klerk, E. de. [Kle02] and Gärtner, B., & Matoušek, J. [GM12] very well-written and approachable. The book by Boyd, S., & Vandenberghe, L. [BV04] is a great reference too.

The monograph by Ben-Tal, A., & Nemirovski, A. [BN01] contains an extensive list of applications of second-order cone problems and semidefinite problems.

### Bibliography

[Dia62] P. H. Diananda, "On non-negative forms in real variables some or all of which are non-negative," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 58, no. 1, pp. 17–25, 1962, doi: 10.1017/S0305004100036185.

[KP02] E. de Klerk and D. V. Pasechnik, "Approximating of the stability number of a graph via copositive programming," *SIAM J. Optim.*, vol. 12, no. 4, pp. 875–892, 2002, doi: 10.1137/S105262340138324.

[GM12] B. Gärtner and J. Matoušek, *Approximation Algorithms and Semidefinite Programming*. Berlin, Germany: Springer, 2012. [Online]. Available: https://link.springer.com/book/10.1007/978-3-642-22015-9

[Kle02] E. de Klerk, *Aspects of Semidefinite Programming*. Springer US, 2002. [Online]. Available: https://link.springer.com/book/10.1007/b105286

[BV04] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004. [Online]. Available: https://web.stanford.edu/~boyd/cvxbook/

[HS10] J.-B. Hiriart-Urruty and A. Seeger, "A Variational Approach to Copositive Matrices," *SIAM Review*, vol. 52, no. 4, pp. 593–629, 2010, [Online]. Available: http://www.jstor.org/stable/41062014

[BN01] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.