# Semidefinite programming (and graph theory)

Gabriel Coutinho

September 26, 2022

These are the course notes of a (under)grad course being offered at UFMG in 2019.2.

## Contents

# 1 Introduction

## 1.1 Motivation

I will motivate this course with two problems coming from graph theory.

### Shannon capacity

The first is to compute what is known as the *Shannon capacity* of a graph.

Imagine $G$ is a graph whose vertices represent letters of an alphabet. Two vertices are made adjacent if the two letters can be confused if sent across a channel. If I simply ask the question of what is the maximum number of 1-letter words that can be sent with no ambiguity, the answer is simply the size of the largest set of vertices in the graph that span no edges — that is, the size of the largest coclique in the graph. This shall be called the *coclique or independence number* of $G$, and will be denoted by $\alpha(G)$.

What if now we are interested in sending $k$-letter words so that no two words can have all their letters equal or confusable? Certainly we can take a set of indistinguishable letters and use only those — hence giving $\alpha(G)^k$ possible words.

This is however not usually the best strategy.

**Exercise 1.1.** Find a set of five 2-letter words from the graph $C_5$ that are not confusable.

Taken graphs $G$ and $H$, we define their *strong product* $G \boxtimes H$ as the graph with vertex set $V(G) \times V(H)$, and two distinct vertices adjacent if each coordinate is either equal or adjacent in their original graph. Thus, it is immediate to see that the maximum number of $k$-letter words we can send from $G$ is $\alpha(G^k)$, where $G^k = G \boxtimes ... \boxtimes G$, $k$ times.

It follows easily from the definition that, for any graphs $G$ and $H$,

$$\alpha(G \boxtimes H) \geq \alpha(G)\alpha(H).$$

and, as a consequence, $\alpha(G^{k+\ell}) \geq \alpha(G^k)\alpha(G^\ell)$.

In a seminal paper from 1956, Claude Shannon introduced the parameter

$$\Theta(G) = \sup_k \sqrt[k]{\alpha(G^k)}.$$

Because $|V(G^k)| = |V(G)|^k$, it follows that $\Theta(G) \leq |V|$. It is actually a consequence of the inequality above and Fekete's lemma that you can replace sup by lim, meaning, $\Theta(G) = \lim_{k\to\infty} \sqrt[k]{\alpha(G^k)}$. You need not worry about this at this point.

**Exercise 1.2.** Compute $\Theta(C_4)$. Hint: draw $C_4$, $C_4^2$ and $C_4^3$. See if you get an intuition.

For some classes of graphs, it is relatively easy to find $\Theta(G)$.

**Exercise 1.3.** (Try to) show that if it is possible to cover the vertex set of $G$ with $\alpha(G)$ cliques, then $\Theta(G) = \alpha(G)$.
As a hint, note that the strong product of cliques is a clique, and the strong product of cocliques is a coclique. Then see what happens you consider a clique partition of $G$ and look at $G^2$.

Determining $\Theta(G)$ is a very hard problem in general. For instance, it took more than 20 years since Shannon's paper was published for $\Theta(C_5)$ to be settled, and computing $\Theta(C_7)$ is still an open problem, as in fact, it is not known whether $\Theta(G)$ is computable for any given graph $G$.

Let us see how to find $\Theta(C_5)$. An *orthonormal representation* of $G$ is a map

$$\rho : V(G) \to \mathcal{V},$$

where $\mathcal{V}$ is an Euclidean space, so that non-neighbours get mapped to orthogonal vectors, and the image of $\rho$ consists of unit vectors.

**Lemma 1.4.** *If $\rho_1$ and $\rho_2$ are orthogonal representations of $G_1$ and $G_2$, then the map*

$$\begin{aligned} \rho : V(G_1) \times V(G_2) &\to \mathcal{V}_1 \otimes \mathcal{V}_2 \\ (a_1, a_2) &\mapsto \rho_1(a_1) \otimes \rho_2(a_2) \end{aligned}$$

*is an orthogonal representation of $G_1 \boxtimes G_2$.*

*Proof.* Assume $(a_1, a_2)$ and $(b_1, b_2)$ are not neighbours. Without loss of generality, say $a_1$ and $b_1$ are not neighbours. Then $\rho_1(a_1)^\mathsf{T}\rho_1(b_1) = 0$, and thus

$$\rho(a_1, a_2)^\mathsf{T}\rho(b_1, b_2) = (\rho_1(a_1) \otimes \rho_2(a_2))^\mathsf{T}(\rho_1(b_1) \otimes \rho_2(b_2)) = \rho_1(a_1)^\mathsf{T}\rho_1(b_1) \otimes \rho_2(a_2)^\mathsf{T}\rho_2(b_2) = 0.$$

Also, if $\mathbf{u}$ and $\mathbf{v}$ are unit vectors, then $\mathbf{u} \otimes \mathbf{v}$ is a unit vector. □

Assume $V(G) = \{v_1, ..., v_n\}$, and if $\rho$ is a representation, let $\rho(v_i) = \mathbf{v}_1$. The value of $\rho$ is given by

$$\min_{\|\mathbf{c}\|=1} \max_{1 \le i \le n} \frac{1}{(\mathbf{v}_i^\mathsf{T}\mathbf{c})^2}.$$

The vector $\mathbf{c}$ which attains the minimum is called the handle of the representation.

The minimum value over all possible orthonormal representations of $G$ is known nowadays as the *Lovász theta parameter* of the graph $G$, denoted by $\vartheta(G)$.

A representation attaining this minimum is called optimal.

**Lemma 1.5.** *Given $G_1$ and $G_2$, we have $\vartheta(G_1 \boxtimes G_2) \le \vartheta(G_1)\vartheta(G_2)$.*

*Proof.* Let $V(G_1) = \{v_1, ..., v_n\}$, and $V(G_2) = \{u_1, ..., u_m\}$. Let $\rho_1$ and $\rho_2$ be optimal representations, with corresponding handles $\mathbf{c}_1$ and $\mathbf{c}_2$. Note that $\|\mathbf{c}_1 \otimes \mathbf{c}_2\| = 1$. As $\rho_1 \otimes \rho_2$ is a representation of $G_1 \boxtimes G_2$, we have

$$\vartheta(G_1 \boxtimes G_2) \le \max_{i,j} \frac{1}{[(\mathbf{v}_i \otimes \mathbf{u}_j)^\mathsf{T}(\mathbf{c}_1 \otimes \mathbf{c}_2)]^2} = \max_{i,j} \frac{1}{(\mathbf{v}_i^\mathsf{T}\mathbf{c}_1)^2} \frac{1}{(\mathbf{u}_j^\mathsf{T}\mathbf{c}_2)^2} = \vartheta(G_1)\vartheta(G_2).$$

□

**Lemma 1.6.** *For any graph $G$, $\alpha(G) \le \vartheta(G)$.*

*Proof.* Let $\rho$ be optimal, $\rho(v_i) = \mathbf{v}_i$, and $\mathbf{c}$ the corresponding handle. If $S$ is a coclique, then $\mathbf{v}_i^\mathsf{T}\mathbf{v}_j = 0$ for all $v_i, v_j \in S$. As a consequence

$$1 = \|c\|^2 \geq \sum_{v_i \in S} (\mathbf{v}_i^\mathsf{T}\mathbf{c})^2.$$

It follows then

$$\vartheta(G) \geq \frac{\sum_{v_j \in S}(\mathbf{v}_j^\mathsf{T}\mathbf{c})^2}{\min_i (\mathbf{v}_i^\mathsf{T}\mathbf{c})^2} \geq \alpha(G).$$

□

**Theorem 1.7** (Lovász'79)**.** *For all graphs $G$,*

$$\Theta(G) \leq \vartheta(G).$$

*Proof.* It is a simple consequence of the lemmas above:

$$\alpha(G^k) \leq \vartheta(G^k) \leq \vartheta(G)^k.$$

□

The theorem provides a way of computing $\Theta(C_5)$. There is an orthogonal representation of value $\sqrt{5}$, obtained from putting 6 unit vectors in $\mathbb{R}^3$, in a position resembling an umbrella, so that the maximum angle between the ribs is $\pi/2$ (those will correspond to non-adjacent vertices). Together with the exercise we did above, we conclude that $\Theta(C_5) = \sqrt{5}$.

Eventually in this course, we will learn that the parameter $\vartheta(G)$ has far more profound connections to graph theory, and, incredibly, can be somehow efficiently computed for all graphs.

### Max-cut

Another problem arising from graph theory that will motivate our course is that of finding the maximum sized cut of a graph. Given $\emptyset \neq S \subset V(G)$, the *cut* $\delta(S)$ is the set of edges connecting $S$ to its complement.

We are interested in the problem of finding $S$ so that $\delta(S)$ is the largest possible.

Given $G$, let $\texttt{mc}(G)$ denote the size of the maximum cut of $G$.

**Exercise 1.8.** Show that $\texttt{mc}(G) \geq n/2$ by exhibiting a very simple algorithm that constructs a cut of size at least $n/2$.

It is known that $\texttt{mc}(G)$ cannot be computed in polynomial time unless P = NP. In fact, any approximation with constant factor better than 0.94 would imply P = NP.

On the other hand, Goemans and Williamson showed in 1994 how to find a cut of size at least $0.878\texttt{mc}(G)$. They proceed as follows. Represent the vertices of $G$ as unit vectors in $\mathbb{R}^d$ so that

$$\mathcal{E} = -\frac{1}{4}\sum_{ij \in E} \|\mathbf{v}_i - \mathbf{v}_j\|^2$$

is minimized. Note that if $d = 1$, finding such representation is equivalent to finding a max-cut. For a different $d$, this value can only decrease, hence $-\mathcal{E}$ is an upper bound on $\mathtt{mc}(G)$ for all $d$. Now if $d = n$, it is possible to compute this optimal representation using semidefinite programming, as we will later see.

From such representation, one can find a cut as follows. Generate a random hyperplane through the origin of $\mathbb{R}^n$. The probability that it separates $\mathbf{v}_i$ from $\mathbf{v}_j$ is given by $\arccos \mathbf{v}_i^\mathsf{T}\mathbf{v}_j/\pi$. For $-1 \leq t \leq 1$, it is known that

$$\arccos(t) \geq 1.38005(1-t),$$

thus, the expected number of edges crossing the hyperplane is

$$\sum_{ij\in E} \frac{\arccos \mathbf{v}_i^\mathsf{T}\mathbf{v}_j}{\pi} \geq \frac{1.38005}{\pi} \sum_{ij\in E} (1 - \mathbf{v}_i^\mathsf{T}\mathbf{v}_j) = \frac{1.38005}{\pi} 2(-\mathcal{E}) \geq 0.878\mathtt{mc}(G).$$

## 1.2 Geometry of inequalities

We are in $\mathbb{R}^n$, and given a few vectors, we will concern ourselves with some special linear combinations of these vectors.

If $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ are vectors and $\alpha_1, ..., \alpha_m$ are scalars, then

$$\alpha_1\mathbf{u}_1 + ... + \alpha_m\mathbf{u}_m$$

is a linear combination of the vectors, defined to be

(i) *affine combination* if $\sum \alpha_i = 1$,

(ii) *conical combination* if $\alpha_i \geq 0$ for all $i$,

(iii) *convex combination* if $\sum \alpha_i = 1$ and $\alpha_i \geq 0$ for all $i$.

The *affine hull* of the set $\{\mathbf{u}_1, ..., \mathbf{u}_m\}$ is the set of all affine combinations, and, similarly, the *convex hull* is the set of all convex combinations.

A subset of $\mathbb{R}^n$ is called *convex* if the segment between any two points of the set is entirely contained in the set. A *cone* is a set subset $C$ of $\mathbb{R}^n$ so that if $\mathbf{v} \in C$, then $\alpha\mathbf{v} \in C$ for all $\alpha \geq 0$.

**Exercise 1.9.** Show that the convex hull of a set of vectors is precisely equal to the smallest convex set that contains those points.

A *hyperplane* $\mathcal{H}$ in $\mathbb{R}^n$ is the set of vectors that lie in the kernel of a linear functional, meaning, there is $\mathbf{a}$ so that

$$\mathcal{H} = \{\mathbf{x} : \mathbf{a}^\mathsf{T}\mathbf{x} = 0\}.$$

If instead of 0 we put a non-zero scalar $\beta$, then we call it an *affine hyperplane*. If we replace the equality sign by $\leq$, then the region defined is called a *half-space*.

**Exercise 1.10.** Verify that the intersection of convex sets is convex.

**Exercise 1.11.** Prove that a half-space is a convex set.

The intersection of a finite number of half-spaces is called a *polyhedron*, that is, any polyhedron $\mathbb{P}$ can be determined by a matrix $\mathbf{A}$ and a vector $\mathbf{b}$ as follows:

$$\mathbb{P} = \{\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}.$$

A *polytope* is a bounded polyhedron. It is not hard to believe that a set is a polytope if and only if it is the convex hull of a finite number of points, though the proof is less straightforward.

Given $\mathbf{A}$ and $\mathbf{b}$, a major and important problem is to decide whether the polyhedron determined by the inequalities $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ is empty or not. To that avail, we introduce the Theorem of Alternatives.

**Theorem 1.12.** *Given* $\mathbf{A}$ *and* $\mathbf{b}$*, exactly one of the following holds.*

*(1) The system* $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ *has a solution.*

*(2) There is* $\mathbf{z}$ *with* $\mathbf{z} \geq \mathbf{0}$*,* $\mathbf{z}^\mathsf{T}\mathbf{A} = \mathbf{0}$*, and* $\mathbf{b}^\mathsf{T}\mathbf{z} < 0$*.*

*Proof.* Clearly both cannot be true at the same time. So we must show both cannot be false at the same time. The proof is by induction on the number of columns of $\mathbf{A}$. Within the induction there is an implicit algorithm known as the Fourier-Motzkin elimination procedure.

For the base case, note that if $\mathbf{A}$ has zero columns, then (1) simply states that $\mathbf{b} \geq \mathbf{0}$. If that fails, one coordinate is negative, say $i$th, so simply choose $\mathbf{z} = \mathbf{e}_i$.

Assume the theorem holds for any matrix with $n-1$ columns. Consider now $\mathbf{A}$ a $m \times n$ matrix. Assume wlog that the inequalities were scaled so that the magnitude of the non-zero coefficients of $x_n$ is 1. Let $I_+$ be the row indices corresponding to positive coefficients of $x_n$, $I_-$ to negative, and $I_0$ to the 0 coefficients. Therefore the system writes as

$$\begin{aligned}
\sum_{k=1}^{n-1} \mathrm{A}_{ik}\mathrm{x}_k + \mathrm{x}_n &\leq \mathrm{b}_i, \quad \text{for all } i \in I_+,\\
\sum_{k=1}^{n-1} \mathrm{A}_{ik}\mathrm{x}_k - \mathrm{x}_n &\leq \mathrm{b}_i, \quad \text{for all } i \in I_-,\\
\sum_{k=1}^{n-1} \mathrm{A}_{ik}\mathrm{x}_k &\leq \mathrm{b}_i, \quad \text{for } i \in I_0.
\end{aligned}$$

This implies

$$\max_{i\in I_-}\left(\sum_{k=1}^{n-1}\mathrm{A}_{ik} - \mathrm{b}_i\right) \leq \min_{i\in I_+}\left(\mathrm{b}_i - \sum_{k=1}^{n-1}\mathrm{A}_{ik}\right),$$

so the original system has a solution if and only if the following system also does

$$\sum_{k=1}^{n-1}(\mathrm{A}_{ik}+\mathrm{A}_{jk})\mathrm{x}_k \leq \mathrm{b}_i + \mathbf{b}_j, \quad \text{for } i \in I_- \text{ and } j \in I_+,$$
$$\sum_{k=1}^{n-1}\mathrm{A}_{ik}\mathrm{x}_k \leq \mathrm{b}_i, \quad \text{for } i \in I_0.$$

If the original has no solution, the one above also is not, and, by induction, there are $w_{ij}$, $i \in I_i$ and $j \in I_+$, and , $v_i$, $i \in I_0$, so that, for all indices, $w_{ij} \geq 0$, $v_i \geq 0$, and

$$\sum_{i\in I_-, j\in I_+}(\mathrm{A}_{ik}+\mathrm{A}_{jk})w_{ij} + \sum_{i\in I_0}\mathrm{A}_{ik}v_i = 0, \quad \text{for all } k,$$

and

$$\sum_{i\in I_-, j\in I_+}(\mathbf{b}_i+\mathbf{b}_j)w_{ij} + \sum_{i\in I_0}\mathbf{b}_i v_i < 0.$$

Now, simply define the vector $\mathbf{z}$ with

$$\mathrm{z}_i = \sum_{j\in I_+} w_{ij}, \quad \text{for } i \in I_-,$$
$$\mathrm{z}_j = \sum_{i\in I_-} w_{ij}, \quad \text{for } j \in I_+,$$
$$\mathrm{z}_i = v_i, \quad \text{for } i \in I_0.$$

It satisfies $\mathbf{z} \geq \mathbf{0}$, $\mathbf{z}^\mathsf{T}\mathbf{A} = \mathbf{0}$, and $\mathbf{b}^\mathsf{T}\mathbf{z} < \mathbf{0}$. □

**Exercise 1.13.** Farka's lemma says that $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$ has no solution if and only if there is $\mathbf{y}$ so that $\mathbf{y}^\mathsf{T}\mathbf{A} \geq \mathbf{0}$ and $\mathbf{b}^\mathsf{T}\mathbf{y} < 0$. Prove this (using the Theorem of the Alternatives).

## 1.3 Linear programs

A *linear program* is an optimization problem that attempts to find the maximum or minimum of a linear functional whose viability region is a polyhedron. Upon certain operations, an equivalent formulation to any LP has the form

$$\begin{aligned} \max \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{subject to} \quad & \mathbf{Ax} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

The vector $\mathbf{x}$ corresponds to variables. The objective function is linear, and so must be the remaining constraints.

Optimization of combinatorial structures can be modelled with LPs provided integrality constraints are added to the variables $\mathbf{x}$. In this case, the optimization problem is called

an *integer program*. For example, let $\mathbf{N}$ be the incidence matrix of a graph, with rows corresponding to vertices, and columns to edges. Then

$$\begin{aligned} \max \quad & \mathbf{1}^\mathsf{T}\mathbf{x} \\ \text{subject to} \quad & \mathbf{N}\mathbf{x} \leq \mathbf{1} \\ & \mathbf{x} \geq \mathbf{0} \\ & \mathbf{x} \in \mathbb{Z}^m \end{aligned}$$

yields an optimum solution that corresponds to the edges of a matching of maximum size.

A linear program can satisfy one of three possibilities. Either there is a vector $\mathbf{x}$ that satisfies the constraints and attains a maximum, or there is not, and this happens either because there are vectors satisfying the constraints of arbitrarily large objective value, or because there is no vector at all satisfying the constraints (this is because the objective function is continuous and the region of feasibility is a closed set).

To any linear program, one can define a dual program, as follows

$$\begin{array}{cccc|cccc} & \max & \mathbf{c}^\mathsf{T}\mathbf{x} & & & \min & \mathbf{b}^\mathsf{T}\mathbf{y} & \\ \text{(P)} & \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b} & & \text{(D)} & \text{s.t.} & \mathbf{A}^\mathsf{T}\mathbf{y} \geq \mathbf{c} & \\ & & \mathbf{x} \geq \mathbf{0}. & & & & \mathbf{y} \geq \mathbf{0}. & \end{array}$$

This dual program was written carefully so that one can always guarantee that any feasible solution to the primal has objective value less or equal than any feasible solution to the dual. In fact, if $\mathbf{x}$ and $\mathbf{y}$ are a pair of primal-dual solutions, it follows that

$$\mathbf{c}^\mathsf{T}\mathbf{x} \leq (\mathbf{y}^\mathsf{T}\mathbf{A})\mathbf{x} = \mathbf{y}^\mathsf{T}(\mathbf{A}\mathbf{x}) \leq \mathbf{y}^\mathsf{T}\mathbf{b}.$$

This is known as **weak duality**, and it implies, in particular, that if any of the primal or dual is unbounded, then the other must be infeasible.

Perhaps surprisingly, if both have feasible solutions, then their respective optima have same objective value.

**Theorem 1.14** (Strong duality)**.** *Assume (P) and (D) are feasible. Then their optima solutions have the same objective values.*

*Proof.* Consider the following big system of equations

$$\begin{pmatrix} \mathbf{A} & \\ & -\mathbf{A}^\mathsf{T} \\ -\mathbf{c}^\mathsf{T} & \mathbf{b}^\mathsf{T} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \leq \begin{pmatrix} \mathbf{b} \\ -\mathbf{c} \\ 0 \end{pmatrix}, \quad \text{with } \mathbf{x}, \mathbf{y} \geq \mathbf{0}.$$

Assume the theorem is false. Then, because of weak duality, the system above has no solution. By a variant of the Theorem of the Alternatives, there is a vector $\begin{pmatrix} \mathbf{u} & \mathbf{v} & q \end{pmatrix}^\mathsf{T} \geq \mathbf{0}$ so that

$$\begin{pmatrix} \mathbf{u} & \mathbf{v} & q \end{pmatrix}^\mathsf{T} \begin{pmatrix} \mathbf{A} & \\ & -\mathbf{A}^\mathsf{T} \\ -\mathbf{c}^\mathsf{T} & \mathbf{b}^\mathsf{T} \end{pmatrix} \geq \mathbf{0} \quad \text{and} \quad \begin{pmatrix} \mathbf{u} & \mathbf{v} & q \end{pmatrix}^\mathsf{T} \begin{pmatrix} \mathbf{b} \\ -\mathbf{c} \\ 0 \end{pmatrix} < 0.$$

Thus

$$\mathbf{u}^\mathsf{T}\mathbf{A} \geq q\mathbf{c}^\mathsf{T}, \quad \mathbf{A}\mathbf{v} \leq q\mathbf{b}, \quad \mathbf{u}^\mathsf{T}\mathbf{b} - \mathbf{v}^\mathsf{T}\mathbf{c} < 0.$$

This immediately leads to a contradiction. □

**Exercise 1.15.** Prove that it is only needed to assume that either the primal or the dual is feasible to get that the other is also feasible.

**Exercise 1.16.** Prove the complementary slackness conditions, ie, if $\mathbf{x}$ and $\mathbf{y}$ are a pair of respective optima solutions for a primal-dual pair of LPs as above, then it is not possible that a variable is non-zero while the corresponding inequality is not satisfied with equality. (Hint: write extra variables $\mathbf{u}$ so that $\mathbf{Ax} + \mathbf{u} = \mathbf{b}$ and variables $\mathbf{v}$ with $\mathbf{A}^\mathsf{T}\mathbf{y} - \mathbf{v} = \mathbf{c}$.)

There are two main reasons why we are discussing linear programs. The first is that eventually I will say something like "Remember the results we had for variables whose corresponding vector lies in the positive orthant of $\mathbb{R}^n$? They work pretty much the same for this other cone over here".

### 1.3.1 Fractional chromatic number

The other reason is that we can already associate this theory to some of the concepts we saw in the first section.

We learned that $\alpha(G) \leq \Theta(G) \leq \vartheta(G)$. The definition of $\vartheta$ was however difficult to parse, and we still have no clue on how to compute it well. For the graph $G = C_5$, however, we managed to solve the problem exhibiting an orthogonal representation whose value was equal to that of a lower bound of $\Theta(C_5)$.

Computing the value of $\Theta(C_7)$ is still an open problem. Here, we will see how to use the theory of linear programs to find an upper bound.

Let $\mathbf{M}$ be a matrix whose rows are indexed by the vertices of a graph $G$, and the columns by all cliques of the graph. The integer program

$$\begin{aligned} \max \quad & \mathbf{1}^\mathsf{T}\mathbf{x} \\ \text{subject to} \quad & \mathbf{M}^\mathsf{T}\mathbf{x} \leq \mathbf{1} \\ & \mathbf{x} \geq \mathbf{0} \\ & \mathbf{x} \in \mathbb{Z}^n \end{aligned}$$

is asking for the maximum number of vertices of the graph so that no two of them belong to the same clique, ie, a maximum coclique of the graph. If we drop the integrality constraints, we obtain an LP, to which we can write the dual

$$\begin{array}{cccc|cccc} & \max & \mathbf{1}^\mathsf{T}\mathbf{x} & & & \min & \mathbf{1}^\mathsf{T}\mathbf{y} & \\ (\mathrm{P}) & \text{s.t.} & \mathbf{M}^\mathsf{T}\mathbf{x} \leq \mathbf{1} & & (\mathrm{D}) & \text{s.t.} & \mathbf{M}\mathbf{y} \geq \mathbf{0} & \\ & & \mathbf{x} \geq \mathbf{0} & & & & \mathbf{y} \geq \mathbf{0}. & \end{array}$$

Note that if we add integrality constraints to the dual, we will be finding the smallest number of cliques necessary to cover the vertices of $G$, meaning, $\chi(\overline{G})$. For this reason, the optimum of these LPs is denoted by $\chi_f(\overline{G})$, with "f" standing for fractional.

Note that

$$M(G \boxtimes H) = M(G) \otimes M(H).$$

Thus, if $\mathbf{x}$ and $\mathbf{y}$ are respective optima solutions to the formulation (D) corresponding to graphs $G$ and $H$, it follows that $\mathbf{x} \otimes \mathbf{y}$ is a feasible solution to the formulation (D) corresponding to $G \boxtimes H$. Hence

$$\chi_f(\overline{G \boxtimes H}) \leq \chi_f(\overline{G})\chi_f(\overline{H}).$$

As a consequence:

**Theorem 1.17** (Shannon)**.** *For any graph $G$, $\Theta(G) \leq \chi_f(\overline{G})$.*

**Exercise 1.18.** Verify all the details of the paragraph above, and write the explicit proof of Shannon's result.

**Exercise 1.19.** Compute $\chi_f(\overline{C_7})$. Find a lower bound to $\Theta(C_7)$ better than $\alpha(C_7)$.

Later in this course, we will actually learn that $\vartheta(G) \leq \chi_f(\overline{G})$.

## 1.4 Symmetric matrices

We shall work over the vector space $\mathbb{R}^n$. If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then $\langle \mathbf{v}, \mathbf{u} \rangle = \mathbf{v}^\mathsf{T}\mathbf{u}$ is an inner product (meaning, it is a positive-definite commutative bilinear form). A linear operator $\mathbf{M} : \mathbb{R}^n \to \mathbb{R}^n$ is self-adjoint if $\langle \mathbf{Mv}, \mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{Mu} \rangle$ for all $\mathbf{u}$ and $\mathbf{v}$, and, because $\mathbf{M}$ can (and will) be seen as a square matrix, it follows that $\mathbf{M}$ is a *self-adjoint operator* if and only if $\mathbf{M} = \mathbf{M}^\mathsf{T}$, that is, $\mathbf{M}$ is a *symmetric matrix*. Symmetric matrices enjoy two key important properties: they are diagonalizable by orthogonal eigenvectors, and all of their eigenvalues are real. We start proving both properties.

**Lemma 1.20.** *The eigenvalues of a real symmetric matrix are real numbers.*

*Proof.* Let $\mathbf{Mu} = \lambda\mathbf{u}$, with $\mathbf{u} \neq \mathbf{0}$. Some of these things could be complex numbers, so we can take the conjugate on both sides, recovering

$$\mathbf{M}\overline{\mathbf{u}} = \overline{\lambda}\overline{\mathbf{u}}.$$

Thus $\overline{\mathbf{u}}$ is an eigenvector with eigenvalue $\overline{\lambda}$. Thus

$$\lambda\mathbf{u}^\mathsf{T}\overline{\mathbf{u}} = (\mathbf{mu})^\mathsf{T}\overline{\mathbf{u}} = \mathbf{u}^\mathsf{T}(\mathbf{m}\overline{\mathbf{u}}) = \overline{\lambda}\mathbf{u}^\mathsf{T}\overline{\mathbf{u}}.$$

Because $\mathbf{u}^\mathsf{T}\overline{\mathbf{u}} \neq 0$ if $\mathbf{u} \neq 0$, then $\lambda = \overline{\lambda}$. □

Now simply assume whenever we are dealing with a symmetric matrix, its eigenvalues are real, and any eigenvector can be assumed to be real.

**Lemma 1.21.** *Let $\mathbf{M}$ be a real symmetric matrix, and assume $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors associated to different eigenvalues. Then $\mathbf{v}^\mathsf{T}\mathbf{u} = 0$, that is, they are orthogonal.*

*Proof.* Say $\mathbf{Mu} = \lambda\mathbf{u}$ and $\mathbf{Mv} = \mu\mathbf{v}$, with $\lambda \neq \mu$. It follows that

$$\lambda(\mathbf{v}^\mathsf{T}\mathbf{u}) = \mathbf{v}^\mathsf{T}\mathbf{Mu} = (\mathbf{v}^\mathsf{T}\mathbf{Mu})^\mathsf{T} = \mathbf{u}^\mathsf{T}\mathbf{M}^\mathsf{T}\mathbf{v} = \mathbf{u}^\mathsf{T}\mathbf{Mv} = \mu(\mathbf{u}^\mathsf{T}\mathbf{v}) = \mu(\mathbf{v}^\mathsf{T}\mathbf{u}).$$

As $\lambda \neq \mu$, it must be that $\mathbf{v}^\mathsf{T}\mathbf{u} = 0$. □

The lemma above already implies that if $\mathbf{M}$ is diagonalizable, then it is diagonalizable with orthogonal eigenvectors — as, in fact, we eigenvectors corresponding to distinct eigenvalues are orthogonal, and inside each eigenspace we can always find an orthogonal basis. We move forward.

A subspace $U$ of $\mathbb{R}^n$ is said to be $\mathbf{M}$-invariant if, for all $\mathbf{u} \in U$, $\mathbf{Mu} \in U$. This is a key fundamental concept in linear algebra, and several results are proven by noting that certain subspaces are invariant for certain operator.

**Lemma 1.22.** *Let $\mathbf{m}$ be a real symmetric matrix. If $U$ is $\mathbf{m}$-invariant, then $U^\perp$ is also $\mathbf{m}$-invariant.*

*Proof.* Note that $\mathbf{v} \in U^\perp$, by definition, if $\mathbf{v}^\mathsf{T}\mathbf{u} = 0$ for all $\mathbf{u} \in U$. For all $\mathbf{u} \in U$ and $\mathbf{v} \in U^\perp$, note that

$$(\mathbf{mv})^\mathsf{T}\mathbf{u} = \mathbf{v}^\mathsf{T}\mathbf{mu} = \mathbf{v}^\mathsf{T}(\mathbf{mu}) = 0,$$

because $\mathbf{u} \in U$, $U$ is $\mathbf{m}$-invariant, and so $\mathbf{mu} \in U$, and $\mathbf{v} \in U^\perp$. Thus $\mathbf{mv} \in U^\perp$, as we wanted. □

Let $\lambda$ be such that $\det(\lambda\mathbf{I} - \mathbf{M}) = 0$. Then $\lambda\mathbf{I} - \mathbf{M}$ is singular, and therefore it contains at least one non-zero vector in its kernel. This is saying that all square matrices $\mathbf{M}$ contain at least one eigenvector for each root of $\phi_\mathbf{m}(x) = \det(x\mathbf{I} - \mathbf{M})$. As $\mathbf{M}$ is symmetric, we now know that all possible roots of $\phi_\mathbf{m}$ are real.

**Lemma 1.23.** *Let $U$ be an $\mathbf{M}$-invariant subspace. Then there is one eigenvector of $\mathbf{M}$ in $U$.*

*Proof.* Let $\mathbf{P}$ be a matrix whose columns form an orthonormal basis for $U$. As $U$ is $\mathbf{m}$-invariant, it follows that there is a matrix $\mathbf{N}$ so that

$$\mathbf{MP} = \mathbf{PN}.$$

(Stop now and think carefully why this equality is true.) In particular, $\mathbf{N} = \mathbf{P}^\mathsf{T}\mathbf{MP}$, so $\mathbf{N}$ is symmetric. Let $\mathbf{u}$ be one eigenvector of $\mathbf{N}$ with eigenvalue $\lambda$. Then

$$\mathbf{MPu} = \mathbf{PNu} = \lambda\mathbf{Pu},$$

and, moreover $\mathbf{Pu} \neq \mathbf{0}$, as the columns of $\mathbf{P}$ are linearly independent. Thus $\mathbf{Pu}$ is an eigenvector for $\mathbf{m}$ in $U$. □

These four lemmas above are all you need to prove the following result by induction as an exercise.

**Theorem 1.24.** *Let $\mathbf{M}$ be a real symmetric matrix. Then $\mathbf{M}$ is diagonalizable by set of orthogonal eigenvectors, all of them corresponding to real eigenvalues.*

**Exercise 1.25.** Write the proof of this theorem as an exercise.

**Corollary 1.26.** *Let $\mathbf{v}_1, ..., \mathbf{v}_n$ be an orthonormal basis of eigenvectors for $\mathbf{m}$, each corresponding to an eigenvalue $\lambda_1, ..., \lambda_n$ (these are not necessarily distinct). Let $\mathbf{P}$ be the matrix whose ith column is $\mathbf{v}_i$, and $\Lambda$ the diagonal matrix whose ith diagonal element is $\lambda_i$. Then*

$$\mathbf{P}^\mathsf{T}\mathbf{m}\mathbf{P} = \mathbf{\Lambda},$$

*and*

$$\mathbf{M} = \lambda_1(\mathbf{v}_1\mathbf{v}_1^\mathsf{T}) + ... + \lambda_n(\mathbf{v}_n\mathbf{v}_n^\mathsf{T}).$$

*Proof.* A linear operator is defined and determined by its action on a basis. The first equality follows from the fact that both sides act equally on the canonical basis of $\mathbb{R}^n$. The second follows from

$$\mathbf{m} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^\mathsf{T},$$

and, by definition of matrix product, $\mathbf{m} = \mathbf{v}_1(\lambda_1\mathbf{v}_1^\mathsf{T}) + ... + \mathbf{v}_n(\lambda_n\mathbf{v}_n^\mathsf{T})$. □

You should recall right now that, because $\mathbf{v}_i$ is normalized, then $\mathbf{P}_i = \mathbf{v}_i\mathbf{v}_i^\mathsf{T}$ is the matrix that represents the orthogonal projection onto the line spanned by $\mathbf{v}_i$, that is, $\mathbf{P}_i$ is a projection as $\mathbf{P}_i^2 = \mathbf{P}_i$, and it is an orthogonal projection as $\mathbf{P}_i$ is symmetric. Note that $\mathbf{P}_i\mathbf{P}_j = \mathbf{0}$ whenever $i \neq j$, and so any sum of the $\mathbf{P}_i$s for distinct indices will correspond to the orthogonal projection onto the space spanned by the $\mathbf{v}_i$s of the same indices. In particular $\sum_{i=1}^n \mathbf{P}_i = \mathbf{I}$.

**Exercise 1.27.** Assume $\mathbf{P}_i$s are orthogonal projections. Show that $\mathbf{P}_1 + \mathbf{P}_2$ is an orthogonal projection if and only if $\mathbf{P}_1\mathbf{P}_2 = \mathbf{0}$.

Show now that $\mathbf{P}_1 + ... + \mathbf{P}_k$ is an orthogonal projection if and only if $\mathbf{P}_i\mathbf{P}_j = \mathbf{0}$ for $i \neq j$.

Say $\mathbf{M}$ is an $n \times n$ symmetric matrix with distinct eigenvalues $\theta_0, ..., \theta_d$. When we write the second equation from the statement of Corollary 1.26, we can collect the terms corresponding to equal eigenvalues, and have

$$\mathbf{M} = \sum_{r=0}^{d} \theta_r \mathbf{E}_r, \tag{1}$$

where, according to the discussion above, each $\mathbf{E}_r$ corresponds to the orthogonal projection onto the $\theta_r$ eigenspace. Equation (1) is usually referred to the as the *spectral decomposition* of the matrix $\mathbf{M}$.

**Exercise 1.28.** Find the spectral decomposition of

$$\mathbf{m} = \begin{pmatrix} 1+\sqrt{2} & 0 & 1-\sqrt{2} & 0 \\ 0 & 1+\sqrt{2} & 0 & 1-\sqrt{2} \\ 1-\sqrt{2} & 0 & 1+\sqrt{2} & 0 \\ 0 & 1-\sqrt{2} & 0 & 1+\sqrt{2} \end{pmatrix}$$

Hint: do not try to compute the characteristic polynomial. It is easier to simply try to look and guess which are the eigenvectors and eigenvalues.

Note that the $\mathbf{E}_r$ are symmetric matrices satisfying $\mathbf{E}_r\mathbf{E}_s = \delta_{rs}\mathbf{E}_r$, and $\sum_{r=0}^d \mathbf{E}_r = \mathbf{I}$.

**Exercise 1.29.** Prove (or at least convince yourself) that for any polynomial $p(x)$, it follows that

$$p(\mathbf{M}) = \sum_{r=0}^{d} p(\theta_r)\mathbf{E}_r.$$

**Exercise 1.30.** Let $\mathbf{M}$ be a symmetric matrix, with spectral decomposition as in (1).
vec (A) What is the minimal polynomial of $\mathbf{m}$? (B) Prove that for each $\mathbf{E}_r$, there is a polynomial $p_r$ of degree $d$ so that $p_r(\mathbf{m}) = \mathbf{E}_r$. Describe this polynomial as explicitly as you can.

**Exercise 1.31.** Prove that two symmetric matrices $\mathbf{M}$ and $\mathbf{N}$ commute if and only if they can be *simultaneously diagonalized* by the same set of orthonormal eigenvectors. Is it true that if $\mathbf{M}$ and $\mathbf{N}$ commute, then there is always a polynomial $p$ so that $p(\mathbf{M}) = \mathbf{N}$? Characterize what else you need to observe to guarantee that such polynomial exists.

**Exercise 1.32.** Let $\mathbf{A}$ and $\mathbf{B}$ be matrices (not necessarily squared shaped), so that both products $\mathbf{AB}$ and $\mathbf{BA}$ are defined. Prove that

$$\operatorname{tr}\mathbf{AB} = \operatorname{tr}\mathbf{BA},$$

and conclude that if $\mathbf{M}$ is a symmetric matrix with eigenvalues $\lambda_1, ..., \lambda_n$, then $\operatorname{tr}\mathbf{M}$ is equal to $\lambda_1 + ... + \lambda_n$. How about $\operatorname{tr}\mathbf{m}^2$?

## 1.5 A detour — interlacing

Given a symmetric matrix $\mathbf{M}$, the *Rayleigh quotient* of $\mathbf{M}$ with respect to a non-zero vector $\mathbf{v}$ is defined as

$$R_{\mathbf{M}}(\mathbf{v}) = \frac{\mathbf{v}^\mathsf{T}\mathbf{M}\mathbf{v}}{\mathbf{v}^\mathsf{T}\mathbf{v}}.$$

We will always assume vectors whose Rayleigh quotient is being taken are non-zero. If $\mathbf{v}$ is an eigenvector with corresponding eigenvalue $\theta$, then

$$R_{\mathbf{M}}(\mathbf{v}) = \theta.$$

Moreover, if $\lambda_1 \geq ... \geq \lambda_n$ are the eigenvalues of $\mathbf{M}$ with corresponding eigenprojectors $E_r$s, and assuming $\mathbf{v}$ is normalized, then

$$R_{\mathbf{M}}(\mathbf{v}) = \mathbf{v}^\mathsf{T}\left(\sum_{r=1}^{n}\lambda_r E_r\right)\mathbf{v} = \sum_{r=1}^{n}\lambda_r(\mathbf{v}^\mathsf{T}E_r\mathbf{v}) \leq \lambda_1\left(\sum_{r=0}^{d}\mathbf{v}^\mathsf{T}E_r\mathbf{v}\right) = \lambda_1$$

for all vectors $\mathbf{v}$, and equality holds if and only if $\mathbf{v}$ belongs to the $\lambda_1$ eigenspace.

**Lemma 1.33.** *Let $\mathbf{M}$ be a symmetric matrix, with largest eigenvalue $\lambda_1$ and smallest eigenvalue $\lambda_n$. Then*

$$\lambda_1 = \max_{\mathbf{v}\in\mathbb{R}^n} R_{\mathbf{M}}(\mathbf{v}) \quad \textit{and} \quad \lambda_n = \min_{\mathbf{v}\in\mathbb{R}^n} R_{\mathbf{M}}(\mathbf{v}).$$

□

Examining more carefully how we bounded the Rayleigh quotient, it is not hard to see that all eigenvalues can be defined as a max or min of the Rayleigh quotient over certain subspaces. Let $L_r$ denote the orthogonal complement to the sum of the eigenlines corresponding to the largest eigenvalues all the way to $\lambda_r$, that is

$$L_r = \text{null } (E_1 + E_1 + ... + E_{r-1}).$$

Likewise, define $S_r$ to correspond to the orthogonal complement to the sum of the eigenlines corresponding to the smallest eigenvalues all the way to $\lambda_{r+1}$, that is

$$S_r = \text{null } (E_{r+1} + E_{r+2} + ... + E_n).$$

It follows immediately that

$$\lambda_r = \max_{\mathbf{v} \in L_r} R_{\mathbf{M}}(\mathbf{v}) = \min_{\mathbf{v} \in S_r} R_{\mathbf{M}}(\mathbf{v}).$$

The expression of $\lambda_r$ can be made with the subspaces $L_r$ and $S_r$ implicitly defined, via a min-max formula.

**Lemma 1.34** (Courant–Fischer–Weyl min-max principle)**.** *Let* $\mathbf{M}$ *be a symmetric matrix, with eigenvalues* $\lambda_1 \geq ... \geq \lambda_n$*. Then*

$$\lambda_k = \min_{\substack{\textit{subspace } U \\ \dim U = n-k+1}} \max_{\mathbf{v} \in U} R_{\mathbf{M}}(\mathbf{v}) = \max_{\substack{\textit{subspace } U \\ \dim U = k}} \min_{\mathbf{v} \in U} R_{\mathbf{M}}(\mathbf{v}).$$

*Proof.* I will show the first equality only, as the second is analogous. Note that we have already seen that there is a subspace $U$ of dimension $n - k + 1$ so that

$$\lambda_k = \max_{\mathbf{v} \in U} R_{\mathbf{M}}(\mathbf{u}),$$

this subspace is simply the orthogonal complement of the sum of the eigenlines corresponding to the largest $k-1$ eigenvalues. The result will now follow if we verify that, for all subspaces $U$ of dimension $n - k + 1$, we have

$$\lambda_k \leq \max_{\mathbf{v} \in U} R_{\mathbf{M}}(\mathbf{u}).$$

To see this, let $U$ be a subspace of dimension $n - k + 1$, and let $V$ be the sum of the eigenlines corresponding to the largest $k$ eigenvalues. As $\dim U + \dim V$ exceeds $n$, it follows that $U \cap V \neq \emptyset$. Let $\mathbf{v}$ belong to this intersection. Then

$$R_{\mathbf{M}}(\mathbf{v}) \geq \lambda_k \sum_{r=1}^{k} \mathbf{v}^{\mathsf{T}} E_r \mathbf{v} \geq \lambda_k,$$

as we wanted. □

Such min-max formula provides an alternative and meaningful definition of eigenvalues. For graph theory, it is hard to find interesting applications of this formula by itself. We can use it however to prove a strong result.

**Theorem 1.35** (Cauchy's Interlacing)**.** *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix and* $\mathbf{S}$ *be an* $n \times m$ *matrix satisfying* $\mathbf{S}^\mathsf{T}\mathbf{S} = \mathbf{I}$*. Let* $\mathbf{B} = \mathbf{S}^\mathsf{T}\mathbf{A}\mathbf{S}$*. Let* $\theta_1 \geq ... \geq \theta_n$ *be the eigenvalues of* $\mathbf{A}$ *and* $\lambda_1 \geq ... \geq \lambda_m$ *be those of* $\mathbf{B}$*. Then*

*(a) For all* $k$ *with* $1 \leq k \leq m$*,*

$$\theta_{n-(m-k)} \leq \lambda_k \leq \theta_k$$

*(b) If equality holds in either of the inequalities above for some* $\lambda_k$ *eigenvalue of* $\mathbf{B}$*, then there is a* $\lambda_k$*-eigenvector* $\mathbf{v}$ *of* $\mathbf{B}$ *so that* $\mathbf{S}\mathbf{v}$ *is an eigenvector for* $\lambda_k$ *in* $\mathbf{A}$*.*

*(c) Let* $\mathbf{v}_1, ..., \mathbf{v}_m$ *be an orthogonal basis of eigenvectors of* $\mathbf{B}$*, with* $\mathbf{v}_i$ *corresponding to* $\lambda_i$*. If for some* $\ell \in \{1, ..., m\}$ *we have that* $\lambda_k = \theta_k$ *for all* $k = 1, ..., \ell$ *(or* $\lambda_k = \theta_{n-(m-k)}$ *for all* $k = \ell, ..., n$*), then* $\mathbf{S}\mathbf{v}_k$ *is an* $\theta_k$ *eigenvector for* $\mathbf{A}$ *for* $k = 1, ..., \ell$ *(respectively for* $k = \ell, ..., n$*).*

*(d) If there is an* $\ell \in \{1, ..., m\}$ *so that* $\lambda_k = \theta_k$ *for all* $k = 1, ..., \ell$*, and* $\lambda_k = \theta_{n-(m-k)}$ *for all* $k = \ell + 1, ..., n$*, then* $\mathbf{S}\mathbf{B} = \mathbf{A}\mathbf{S}$*. In this case, interlacing is called tight.*

*Proof.* Let $\mathbf{u}_1, ..., \mathbf{u}_n$ be the eigenvectors of $\mathbf{A}$ corresponding to the $\theta_k$s. The key thing now is to observe that, for all $k$, the subspace

$$\langle \mathbf{v}_1, ..., \mathbf{v}_k \rangle \cap \langle \mathbf{S}^\mathsf{T}\mathbf{u}_1, ..., \mathbf{S}^\mathsf{T}\mathbf{u}_{k-1} \rangle^\perp$$

contains at least one vector. Let $\mathbf{w}$ be such vector, which, in particular, implies $\mathbf{S}\mathbf{w} \in \langle \mathbf{u}_1, ..., \mathbf{u}_{k-1} \rangle^\perp$. Then, by Lemma 1.34, we have

$$\theta_k \geq \frac{(\mathbf{S}\mathbf{w})^\mathsf{T}\mathbf{A}(\mathbf{S}\mathbf{w})}{(\mathbf{S}\mathbf{w})^\mathsf{T}(\mathbf{S}\mathbf{w})} \geq \frac{\mathbf{w}^\mathsf{T}\mathbf{B}\mathbf{w}}{\mathbf{w}^\mathsf{T}\mathbf{w}} \geq \lambda_k.$$

If $\theta_k = \lambda_k$, then $\mathbf{w}$ and $\mathbf{S}\mathbf{w}$ are eigenvectors for $\mathbf{B}$ and $\mathbf{A}$ respectively. Item (iii) follows easily by induction. Finally, with tight interlacing, we can guarantee that $\mathbf{S}\mathbf{v}_1, ..., \mathbf{S}\mathbf{v}_m$ are all eigenvectors for $\mathbf{A}$ with the same eigenvalues they have in $\mathbf{B}$. Therefore $\mathbf{S}\mathbf{B}\mathbf{v}_k = \mathbf{A}\mathbf{S}\mathbf{v}_k$ for all $k$, and as the set of eigenvectors form a basis, the two matrices are equal. □

## 1.6 Positive semidefinite matrices

A real matrix $\mathbf{M}$ is *positive semidefinite*, denoted $\mathbf{M} \succcurlyeq \mathbf{0}$, if it satisfies the properties:

- $\mathbf{M}$ is symmetric.
- $\mathbf{v}^\mathsf{T}\mathbf{M}\mathbf{v} \geq 0$ for all $\mathbf{v} \in \mathbb{R}^n$.

If the inequality is strict for all non-zero $\mathbf{v}$, then $\mathbf{M}$ is called positive definite. The only thing we want now is a characterization.

This is probably one of the most famous "exercises" in linear algebra.

**Theorem 1.36.** *Let* $\mathbf{M}$ *be a symmetric matrix. The following are equivalent.*

*(a)* $\mathbf{M}$ *is positive semidefinite.*

*(b) The eigenvalues of* $\mathbf{M}$ *are non-negative.*

*(c) There exists a matrix* $\mathbf{B}$ *so that* $\mathbf{M} = \mathbf{B}^\mathsf{T}\mathbf{B}$.

*(d) For all positive semidefinite matrices* $\mathbf{A}$*, we have* $\langle \mathbf{M}, \mathbf{A} \rangle \geq 0$.

*Proof.* Assume (a). Let $\mathbf{Mv} = \theta\mathbf{v}$. Then $0 \leq \mathbf{v}^\mathsf{T}\mathbf{Mv} = \theta\mathbf{v}^\mathsf{T}\mathbf{v}$, thus $\theta \geq 0$. Assume (b). We diagonalize $\mathbf{M}$ as

$$\mathbf{M} = \mathbf{P}^\mathsf{T}\mathbf{D}\mathbf{P}.$$

As $\mathbf{D} \geq 0$, we have

$$\mathbf{M} = \mathbf{P}^\mathsf{T}\sqrt{\mathbf{D}}\sqrt{\mathbf{D}}\mathbf{P} = (\sqrt{\mathbf{D}}\mathbf{P})^\mathsf{T}(\sqrt{\mathbf{D}}\mathbf{P}).$$

Assume (c). Then

$$\langle \mathbf{M}, \mathbf{A} \rangle = \operatorname{tr} \mathbf{MA} = \operatorname{tr} \mathbf{B}^\mathsf{T}\mathbf{BA} = \operatorname{tr} \mathbf{BAB}^\mathsf{T}.$$

As $\mathbf{A}$ is psd, we have $\operatorname{tr} \mathbf{BAB}^\mathsf{T} \geq 0$. Finally, assume (d). Take $\mathbf{A} = \mathbf{vv}^\mathsf{T}$, which is clearly psd for any $\mathbf{v}$. We have $0 \leq \langle \mathbf{M}, \mathbf{vv}^\mathsf{T} \rangle = \mathbf{v}^\mathsf{T}\mathbf{Mv}$, as wished. □

**Exercise 1.37.** Assume $\mathbf{A} \succcurlyeq \mathbf{0}$. Show that $\mathbf{v}^\mathsf{T}\mathbf{Av} = \mathbf{0}$ if and only if $\mathbf{Av} = \mathbf{0}$. (Note that you can replace $\mathbf{v}$ by a matrix $\mathbf{U}$).

**Exercise 1.38.** Replace "semidefinite" by "definite" in (a). Come up with the correct modifications in each item.

Recall that a principal minor of a matrix $\mathbf{M}$ is the determinant of a square submatrix symmetric about the main diagonal.

**Theorem 1.39.** *We have* $\mathbf{M} \succcurlyeq \mathbf{0}$ *if and only if all of its principal minors are non-negative.* □

I leave the proof as an exercise. One side should be elementary. To prove the other, you might want to try using Cauchy's Interlacing.

**Exercise 1.40.** Let $\mathbf{M} \succcurlyeq \mathbf{0}$. Given vectors $\mathbf{u}$ and $\mathbf{v}$, and $t \in \mathbb{R}$, write $(\mathbf{u} + t\mathbf{v})^\mathsf{T}\mathbf{M}(\mathbf{u} + t\mathbf{v})$ explicitly. You know that this is $\geq 0$ for all $t$. Use this to conclude that

$$(\mathbf{u}^\mathsf{T}\mathbf{Mv})^2 \leq (\mathbf{u}^\mathsf{T}\mathbf{Mu})(\mathbf{v}^\mathsf{T}\mathbf{Mv}).$$

(This is probably our favourite proof of *Cauchy-Schwarz*'s).

Suppose the symmetric matrix $\mathbf{M}$ is written as

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix},$$

with $\mathbf{A}$ invertible. This leads to the factorization

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{CA}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

The matrix $\mathbf{S} = \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}$ is called the Schur complement of $\mathbf{A}$ in $\mathbf{M}$.

**Theorem 1.41.** *Given* $\mathbf{M}$ *in block form as above, with* $\mathbf{A}$ *invertible, then* $\mathbf{M} \succcurlyeq \mathbf{0}$ *if and only if* $\mathbf{A}$ *and* $\mathbf{S}$ *are also.*

*Proof.* Follows immediately from noting that $\mathbf{M} = \mathbf{P}^\intercal \begin{pmatrix} \mathbf{A} & \\ & \mathbf{S} \end{pmatrix} \mathbf{P}$, with a matrix $\mathbf{P}$ that is non-singular. □

This result gives the famous *Cholesky decomposition* of a matrix:

**Theorem 1.42.** *If* $\mathbf{A}$ *is positive semidefinite, there is a lower triangular matrix* $\mathbf{L}$ *so that* $\mathbf{A} = \mathbf{L}\mathbf{L}^\intercal$.

*Proof.* By induction on $n$. If $\mathrm{A}_{1,1} = 0$, then the first row and column of $\mathbf{A}$ are zero (why?), thus we can apply induction to the submatrix of $\mathbf{A}$ obtained upon deleting them. Otherwise, we have

$$\mathbf{A} = \begin{pmatrix} a & \mathbf{b}^\intercal \\ \mathbf{b} & \mathbf{A}_1 \end{pmatrix},$$

thus, there is a lower triangular $\mathbf{P}$ with

$$\mathbf{P}^{-\intercal}\mathbf{A}\mathbf{P}^{-1} = \begin{pmatrix} a & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 - a^{-1}\mathbf{b}\mathbf{b}^\intercal \end{pmatrix},$$

where both diagonal blocks are positive semidefinite. By induction, both admit a Cholesky decomposition, which regrouped with $\mathbf{P}$ gives the Cholesky decomposition of $\mathbf{A}$. □

**Exercise 1.43.** How to use this decomposition to decide (efficiently) whether a given matrix is positive semidefinite?

## 1.7 Kronecker and Schur

If $\mathbf{A}$ and $\mathbf{B}$ are matrices, their *Kronecker product* $\mathbf{A}\otimes\mathbf{B}$ is the matrix obtained from replacing the $ij$ entry of $\mathbf{A}$ by $\mathrm{A}_{i,j}\mathbf{B}$. You are invited to verify that the Kronecker product is bilinear, but, most notably, it satisfies

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D},$$

provided all products are well defined. In particular, if $\mathbf{u}$ is eigenvector of $\mathbf{A}$ and $\mathbf{v}$ is one of $\mathbf{B}$, then $\mathbf{u} \otimes \mathbf{v}$ is eigenvector of $\mathbf{A} \otimes \mathbf{B}$ — its corresponding eigenvalue is the product of the original eigenvalues. This fact alone immediately implies the following:

**Lemma 1.44.** *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are positive semidefinite, then so it is* $\mathbf{A} \otimes \mathbf{B}$.

The operator $\mathrm{vec}(\mathbf{A})$ takes $\mathbf{A}$, with $n$ columns, to the column vector

$$\begin{pmatrix} \mathbf{A}\mathbf{e}_1 \\ \vdots \\ \mathbf{A}\mathbf{e}_n \end{pmatrix}.$$

**Exercise 1.45.** Convince yourself that

$$\operatorname{vec}(\mathbf{AMB}^\mathsf{T}) = (\mathbf{A} \otimes \mathbf{B}) \operatorname{vec}(\mathbf{M}).$$

In other words, the linear map $\mathbf{M} \mapsto \mathbf{AMB}^\mathsf{T}$ is represented by $\mathbf{A} \otimes \mathbf{B}$.

The *Schur product* $\mathbf{A} \circ \mathbf{B}$ of two matrices with the same shape is defined as the entry-wise product (your high school dream).

**Exercise 1.46.** Prove that if $\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite, then $\mathbf{A} \circ \mathbf{B}$ is also. (Kronecker and Schur are together in this section and it's not a coincidence).

# 2 Semidefinite programs

Starting here, we will denote the set of all $n \times n$ real symmetric matrices by $\mathbb{S}^n$. The set of those which are positive semidefinite matrices by $\mathbb{S}^n_+$, and $\mathbb{S}^n_{++}$ for the positive definite matrices. We will use the notation $\succcurlyeq$ also to compare matrices in $\mathbb{S}^n$, having

$$\mathbf{A} \succcurlyeq \mathbf{B} \iff \mathbf{A} - \mathbf{B} \in \mathbb{S}^n_+$$

This notation comes as a natural extension of the notation $\mathbb{R}^n_+$ for vectors with non-negative entries, and $\mathbb{R}^n_{++}$ for vectors with positive entries.

## 2.1 Two examples

We are interested in optimization problems where all variables have been organized as entries of a matrix which is required to be positive semidefinite. All other constraints, and the objective function, shall be linear.

**Exercise 2.1.** Solve

$$\begin{aligned} \min \quad & \mathbf{X}_{12} \\ \text{subject to} \quad & \mathbf{X}_{11} = 1 \\ & \mathbf{X}_{22} = 2 \\ & \mathbf{X} \in \mathbb{S}^2_+. \end{aligned}$$

(Note that for LPs, a problem with integer entries would never return a non-rational solution).

**Exercise 2.2.** Contrary to the case for LPs, not all bounded problems have optimal solutions. Solve

$$\begin{aligned} \min \quad & \mathbf{X}_{11} \\ \text{subject to} \quad & \begin{pmatrix} \mathbf{X}_{11} & 1 \\ 1 & \mathbf{X}_{22} \end{pmatrix} \succcurlyeq \mathbf{0}. \end{aligned}$$

(Recall that $\mathbf{X}_{11}$ cannot be equal to 0).

## 2.2 Cones

There is a more general framework in which linear programs and semidefinite programs can be cast. An Euclidean space is a vector space over the reals of finite dimension and equipped with an inner product. For example, $\mathbb{R}^n$ and $\mathbb{S}^n$ are Euclidean spaces, and the respective inner products are given by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\mathsf{T}\mathbf{y}$, and $\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{tr} \mathbf{AB}$. A cone $\mathbb{K}$ is a subset so that for all $\mathbf{x} \in \mathbb{K}$, we have $\alpha\mathbf{x} \in \mathbb{K}$ for all $\alpha > 0$.

We introduce four properties about cones that shall be useful to us. A cone $\mathbb{K}$ is called...

- *...closed* if o point that is the limit of a sequence of points in $\mathbb{K}$ is outside of $\mathbb{K}$ — for example, $\mathbb{R}^n_+$ is closed, but $\mathbb{R}^n_{++}$ is not;

- *...pointed* if $\mathbb{K}$ contains no line about the origin, equivalently, if $\mathbf{x} \in \mathbb{K}$ and $-\mathbf{x} \in \mathbb{K}$, then $\mathbf{x} = \mathbf{0}$;
- *...convex* if $\mathbb{K} + \mathbb{K} \subseteq \mathbb{K}$, i.e., the line segment between any two points in $\mathbb{K}$ belongs entirely to $\mathbb{K}$ — for example, $\mathbb{S}^n_+$ is convex (a fact that is trivial only if you use the correct condition from Theorem 1.36...);
- *...a cone with non-empty interior* if there is at least one point in $\mathbb{K}$ so that a ball of positive radius around that point is entirely contained in $\mathbb{K}$ — for example, the vectors $\mathbf{x}$ in $\mathbb{R}^n$ with $\mathbf{x} \geq \mathbf{0}$ and $x_1 = 0$ for a cone that contains no interior point.

**Exercise 2.3.** Consider the *Lorentz cone*, defined as

$$\mathbb{K} = \{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R} : \|\mathbf{x}\| \leq t\}.$$

Verify if it satisfies each of the four properties above. Then, (try to) do the same for the cone of *copositive matrices*, defined as

$$\mathbb{P} = \{\mathbf{M} \in \mathbb{S}^n : \mathbf{x}^\mathsf{T}\mathbf{M}\mathbf{x} \geq 0 \text{ for all } \mathbf{x} \geq 0\}.$$

If $\mathbb{V}_1$ and $\mathbb{V}_2$ are Euclidean spaces, let $\mathbf{c} \in \mathbb{V}_1$, $\mathcal{A} : \mathbb{V}_1 \to \mathbb{V}_2$ be linear, and $\mathbf{b} \in \mathbb{V}_2$. Let $\mathbb{K}$ be a closed, pointed and convex cone, with non-empty interior. A *conic program* is an optimization problem that can be written in the form

$$\begin{aligned} \max \quad & \langle \mathbf{c}, \mathbf{x} \rangle \\ \text{subject to} \quad & \mathcal{A}(\mathbf{x}) = \mathbf{b} \\ & \mathbf{x} \in \mathbb{K}. \end{aligned}$$

(In some contexts, a conic program is presented with "sup" instead of "max", to highlight the fact that it can be feasible and bounded while having no optimal solution. I will present all programs with "max" and "min", understanding that those might not exist even if the program is feasible and bounded).

For a linear program, $\mathbb{K} = \mathbb{R}^n_+$. For a semidefinite program, $\mathbb{K} = \mathbb{S}^n_+$. In both cases you can view $\mathbb{V}_1$ or $\mathbb{V}_2$ as isomorphic to $\mathbb{R}^k$ for some $k$, though in the second case this has to be made explicit via an isomorphism such as vec : $\mathbb{R}^{n\times n} \to \mathbb{R}^{n^2}$, as we discussed.

In order to better understand conic programs (and their duals), let us introduce some more theory. Given $\mathcal{A} : \mathbb{V}_1 \to \mathbb{V}_2$, linear, its *adjoint* is the (unique) linear transformation $\mathcal{A}^* : \mathbb{V}_2 \to \mathbb{V}_1$ that satisfies

$$\langle \mathcal{A}(\mathbf{x}), \mathbf{y} \rangle_{\mathbb{V}_2} = \langle \mathbf{x}, \mathcal{A}^*(\mathbf{y}) \rangle_{\mathbb{V}_1}.$$

Note that when $\mathbb{V}_1 = \mathbb{R}^n$ and $\mathbb{V}_2 = \mathbb{R}^m$, with the conventional inner product, and $\mathcal{A}$ is represented by a $m \times n$ matrix $\mathbf{A}$, then $\mathcal{A}^*$ is given by $\mathbf{A}^\mathsf{T}$.

**Exercise 2.4.** Consider the linear operator diag : $\mathbb{R}^{n\times n} \to \mathbb{R}^n$, that acts as

$$\text{diag}(\mathbf{X}) = \sum_{i=1}^{n} \mathbf{X}_{ii}\mathbf{e}_i.$$

Describe its adjoint (hint: we will denote it by Diag).

Given $\mathbf{x}, \mathbf{y} \in \mathbb{V}$, we say that

$$\mathbf{x} \succcurlyeq_{\mathbb{K}} \mathbf{y} \iff \mathbf{x} - \mathbf{y} \in \mathbb{K}.$$

**Exercise 2.5.** You are invited to prove (or at least think about and convince yourself of) the following properties:

(1) $\mathbf{x} \succcurlyeq_{\mathbb{K}} \mathbf{y} \iff \mathbf{y} \succcurlyeq_{-\mathbb{K}} \mathbf{x}$

(2) $\succcurlyeq_{\mathbb{K}}$ is a partial order, meaning, for all $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in $\mathbb{V}$, we have $\mathbf{x} \succcurlyeq_{\mathbb{K}} \mathbf{x}$; $\mathbf{x} \succcurlyeq_{\mathbb{K}} \mathbf{y}$ and $\mathbf{y} \succcurlyeq_{\mathbb{K}} \mathbf{x}$ imply $\mathbf{x} = \mathbf{y}$; and $\mathbf{x} \succcurlyeq_{\mathbb{K}} \mathbf{y}$ and $\mathbf{y} \succcurlyeq_{\mathbb{K}} \mathbf{z}$ imply $\mathbf{x} \succcurlyeq_{\mathbb{K}} \mathbf{z}$.

(3) $\mathbf{x} \succcurlyeq_{\mathbb{K}} \mathbf{y}$ imply $\alpha\mathbf{x} \succcurlyeq_{\mathbb{K}} \alpha\mathbf{y}$ for all $\alpha > 0$.

(4) $\mathbf{x} \succcurlyeq_{\mathbb{K}} \mathbf{y}$ and $\mathbf{u} \succcurlyeq_{\mathbb{K}} \mathbf{v}$ imply $\mathbf{x} + \mathbf{u} \succcurlyeq_{\mathbb{K}} \mathbf{y} + \mathbf{v}$.

**Exercise 2.6.** If $\mathcal{A}$ is a linear map from $\mathbb{V}_1$ to $\mathbb{V}_2$, and $\mathbb{K}$ is a convex cone in $\mathbb{V}_1$, show that $\mathcal{A}\mathbb{K}$ is a convex cone in $\mathbb{V}_2$.

If $\mathcal{A} : \mathbb{V} \to \mathbb{V}$ is invertible and $\mathcal{A}\mathbb{K} = \mathbb{K}$, we say that $\mathcal{A}$ is an *automorphism* of $\mathbb{K}$. A cone is *homogeneous* if, for each pair of points $\mathbf{x}$ and $\mathbf{y}$ in its interior, there is an automorphism of $\mathbb{K}$ mapping $\mathbf{x}$ to $\mathbf{y}$.

**Exercise 2.7.** If $\mathbf{A}$ and $\mathbf{B}$ are invertible, show that the map from $\mathbb{S}^n_+$ to itself given by $\mathbf{M} \mapsto \mathbf{A}\mathbf{M}\mathbf{B}^\mathsf{T}$ is a cone automorphism (in fact, all automorphisms of this cone have this form, but no need to show that).

If $\mathbb{V}_1$ and $\mathbb{V}_2$ are Euclidean spaces containing cones $\mathbb{K}_1$ and $\mathbb{K}_2$, then $\mathbb{V}_1 \otimes \mathbb{V}_2$ is an Euclidean space (the inner product being the sum of the products of the coordinates), and $\mathbb{K}_1 \otimes \mathbb{K}_2$ is a cone in $\mathbb{V}_1 \times \mathbb{V}_2$. As a consequence, $(\mathbf{x}_1, \mathbf{x}_2) \succcurlyeq_{\mathbb{K}_1 \otimes \mathbb{K}_2} (\mathbf{y}_1, \mathbf{y}_2)$ if and only if $\mathbf{x}_1 \succcurlyeq_{\mathbb{K}_1} \mathbf{y}_1$ and $\mathbf{x}_2 \succcurlyeq_{\mathbb{K}_2} \mathbf{y}_2$.

The *dual of a cone* $\mathbb{K} \subseteq \mathbb{V}$ is defined as

$$\mathbb{K}^* = \{\mathbf{x} \in \mathbb{V} : \langle \mathbf{x}, \mathbf{s} \rangle \geq 0 \text{ for all } \mathbf{s} \in \mathbb{K}\}.$$

**Lemma 2.8.** *If $\mathbb{K}$ is a cone, then $\mathbb{K}^*$ is a closed convex cone.*

*Proof.* First, if $\mathbf{x} \in \mathbb{K}^*$, then $\alpha\mathbf{x} \in \mathbb{K}^*$ for all $\alpha > 0$, as $\langle \alpha\mathbf{x}, \mathbf{s} \rangle = \alpha \langle \mathbf{x}, \mathbf{s} \rangle \geq 0$ for all $\mathbf{s} \in \mathbb{K}$.

Second, $\mathbb{K}^*$ is closed because it is the intersection of closed sets (research this fact from analysis if you want).

Third, $\mathbb{K}^*$ is convex because it is the intersection of convex sets (remember that you already showed that half-spaces are convex). □

**Exercise 2.9.** Which are the cones $(\mathbb{R}^n_+)^*$ and $(\mathbb{S}^n_+)^*$? Could you also describe the duals of the two cones appearing in Exercise 2.3 ?

A cone that is equal to its dual is called a *self-dual cone*.

We end this section stating an important result about cones and their duals. We will eventually prove it later.

**Theorem 2.10.** *If $\mathbb{K}$ is a closed, convex cone, then so it is $\mathbb{K}^*$, and $\mathbb{K}^{**} = \mathbb{K}$. Moreover,*

*(i) If $K$ is pointed, then $\mathbb{K}^*$ has non-empty interior.*

*(ii) If $K$ has non-empty interior, then $\mathbb{K}^*$ is pointed.*

## 2.3 Duality

Say $\mathbb{V}_1$ and $\mathbb{V}_2$ are Euclidean spaces, let $\mathbf{c} \in \mathbb{V}_1$, $\mathcal{A} : \mathbb{V}_1 \to \mathbb{V}_2$ be linear, and $\mathbf{b} \in \mathbb{V}_2$. Let $\mathbb{K} \subseteq \mathbb{V}_1$ and $\mathbb{L} \subseteq \mathbb{V}_2$ be closed, convex cones. Consider the conic program

$$\begin{array}{rl} \max & \langle \mathbf{c}, \mathbf{x} \rangle_{\mathbb{V}_1} \\ \text{subject to} & \mathcal{A}(\mathbf{x}) \preccurlyeq_{\mathbb{L}} \mathbf{b} \\ & \mathbf{x} \in \mathbb{K}. \end{array}$$

Recall linear program duality, where from

$$\begin{array}{cccc|cccc} & \max & \mathbf{c}^\mathsf{T}\mathbf{x} & & & \min & \mathbf{b}^\mathsf{T}\mathbf{y} & \\ (\mathrm{P}) & \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b} & & (\mathrm{D}) & \text{s.t.} & \mathbf{A}^\mathsf{T}\mathbf{y} \geq \mathbf{c} & \\ & & \mathbf{x} \geq \mathbf{0} & & & & \mathbf{y} \geq \mathbf{0} & \end{array}$$

we obtained immediately that

$$\mathbf{c}^\mathsf{T}\mathbf{x} \leq \mathbf{y}^\mathsf{T}\mathbf{A}\mathbf{x} \leq \mathbf{y}^\mathsf{T}\mathbf{b}.$$

Let us now define a conic program crafted in such way that the objective value of any of its feasible solutions is an upper bound on the objective value of the conic program above.

- We know that $\mathbf{b} - \mathcal{A}(\mathbf{x}) \in \mathbb{L}$, thus $\langle \mathbf{b} - \mathcal{A}(\mathbf{x}), \mathbf{y} \rangle_{\mathbb{V}_2} \geq 0$ for all $\mathbf{y} \in \mathbb{L}^*$.
- Therefore, $\langle \mathbf{b}, \mathbf{y} \rangle_{\mathbb{V}_2} \geq \langle \mathcal{A}(\mathbf{x}), \mathbf{y} \rangle_{V_2} = \langle \mathbf{x}, \mathcal{A}^*(\mathbf{y}) \rangle_{V_1}$.
- Now, if we require $\mathcal{A}^*(\mathbf{y}) - \mathbf{c} \in \mathbb{K}^*$, it follows that, for all $\mathbf{x} \in \mathbb{K}$, we have $\langle \mathcal{A}^*(\mathbf{y}) - \mathbf{c}, \mathbf{x} \rangle_{\mathbb{V}_1} \geq 0$, hence $\langle \mathcal{A}^*(\mathbf{y}), \mathbf{x} \rangle_{\mathbb{V}_1} \geq \langle \mathbf{c}, \mathbf{x} \rangle_{\mathbb{V}_1}$.

We arrive at the following primal-dual formulation for conic programs:

$$\begin{array}{cccc|cccc} & \max & \langle \mathbf{c}, \mathbf{x} \rangle_{\mathbb{V}_1} & & & \min & \langle \mathbf{b}, \mathbf{y} \rangle_{\mathbb{V}_2} & \\ (\mathrm{P}) & \text{s.t.} & \mathcal{A}(\mathbf{x}) \preccurlyeq_{\mathbb{L}} \mathbf{b} & & (\mathrm{D}) & \text{s.t.} & \mathcal{A}^*(\mathbf{y}) \succcurlyeq_{\mathbb{K}^*} \mathbf{c} & \\ & & \mathbf{x} \in \mathbb{K}. & & & & \mathbf{y} \in \mathbb{L}^*. & \end{array}$$

The discussion above has shown the weak duality theorem

**Theorem 2.11.** *For a pair of primal-dual conic programs such as above, if* $\mathbf{x}$ *is feasible for (P) and* $\mathbf{y}$ *is feasible for (D), then* $\langle \mathbf{c}, \mathbf{x} \rangle_{\mathbb{V}_1} \leq \langle \mathbf{b}, \mathbf{y} \rangle_{\mathbb{V}_2}$.

**Exercise 2.12.** Write the dual of the conic program

$$\begin{array}{rl} \max & \langle \mathbf{c}, \mathbf{x} \rangle_{\mathbb{V}_1} \\ \text{subject to} & \mathcal{A}(\mathbf{x}) = \mathbf{b} \\ & \mathbf{x} \in \mathbb{K}. \end{array}$$

**Exercise 2.13.** Using Theorem 2.10, verify that the dual of the dual is the primal.

Let us now come back to something more concrete. Assume $\mathbb{V}_1 = \mathbb{S}^n$, and $\mathbb{V}_2 = \mathbb{R}^m$. Have $\mathbb{K} = \mathbb{S}^n_+$. How do linear functions from $\mathbb{S}^n$ to $\mathbb{R}^m$ look like?

Well, if $f : \mathbb{S}^n \to \mathbb{R}$ is a linear functional, then there exists a unique $\mathbf{M}$ so that, for all $\mathbf{X} \in \mathbb{S}^n$, $f(\mathbf{X}) = \langle \mathbf{X}, \mathbf{M} \rangle$. This can be easily obtained — in fact, if $\{\mathbf{M}_1, ..., \mathbf{M}_\ell\}$ form an orthonormal basis, then

$$\mathbf{M} = \sum f(\mathbf{M}_i)\mathbf{M}_i.$$

Therefore, if $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ is linear, there are matrices $\mathbf{A}_1$ ,..., $\mathbf{A}_m$ so that, for all $\mathbf{X}$,

$$\mathcal{A}(\mathbf{X}) = \begin{pmatrix} \langle \mathbf{A}_1, \mathbf{X} \rangle \\ \vdots \\ \langle \mathbf{A}_m, \mathbf{X} \rangle \end{pmatrix}.$$

**Exercise 2.14.** Given $\mathcal{A}$ with matrices $\mathbf{A}_i$ as above, describe who is $\mathcal{A}^*(\mathbf{y})$ for $\mathbf{y} \in \mathbb{R}^m$.

**Exercise 2.15.** Given the positive semidefinite program

$$\begin{aligned} \max \quad & \langle \mathbf{C}, \mathbf{X} \rangle \\ \text{subject to} \quad & \langle \mathbf{A}_i, \mathbf{X} \rangle \leq \mathrm{b}_i \text{ for } i = 1, ..., m \\ & \mathbf{X} \succcurlyeq \mathbf{0}, \end{aligned}$$

write its dual.

**Exercise 2.16.** Assume a positive semidefinite program (P) and its dual (D) have feasible solution with the same objective value. What would complementary slackness conditions say about them?

## 2.4 Two SDP formulations

### Lovász Theta again

We now introduce the Lovász Theta parameter with a completely different point of view. Given a graph $G$, with adjacency matrix $\mathbf{A}$, let $\mathbf{z}$ denote the characteristic vector of an independent set $S$. If $\mathbf{Z} = (1/|S|)\mathbf{z}\mathbf{z}^\mathsf{T}$, then $\operatorname{tr} \mathbf{Z} = 1$, $\mathbf{Z} \succcurlyeq \mathbf{0}$, and $\mathbf{Z}_{ij} = 0$ for all edges $ij$ of the graph. As such, it is a feasible solution to the semidefinite program

$$\begin{aligned} \max \quad & \langle \mathbf{J}, \mathbf{X} \rangle \\ \text{subject to} \quad & \mathrm{X}_{ij} = 0 \text{ for all edges } ij \text{ of the graph}, \\ & \operatorname{tr} \mathbf{X} = 1, \\ & \mathbf{X} \succcurlyeq \mathbf{0}, \end{aligned}$$

and its objective value is equal to $|S|$.

**Exercise 2.17.** You are invited to check that all constraints are of the form $\langle *, \mathbf{X} \rangle = *$.

In particular, there is a feasible solution to the program with objective value equal to $\alpha(G)$, but, unfortunately, there usually are other solutions with higher objective values. As we will later see, but for now simply assume, the program above has an optimum. Its objective value is precisely $\vartheta(G)$. Its dual also has a feasible solution attaining this value.

**Exercise 2.18.** Write the dual of the program above.

**Eigenvalues**

Recall that given $\mathbf{M} \in \mathbb{S}^n$ with largest eigenvalue $\theta$ and smallest eigenvalue $\tau$, we have

$$\theta = \max_{\mathbf{x}\in\mathbb{R}^n,\ \|\mathbf{x}\|=1} \mathbf{x}^\mathsf{T}\mathbf{M}\mathbf{x},$$

and

$$\tau = \min_{\mathbf{x}\in\mathbb{R}^n,\ \|\mathbf{x}\|=1} \mathbf{x}^\mathsf{T}\mathbf{M}\mathbf{x},$$

Consider now that SDP

$$\begin{aligned} \max \quad & \langle \mathbf{M}, \mathbf{X}\rangle \\ \text{subject to} \quad & \operatorname{tr}\mathbf{X} = 1, \\ & \mathbf{X} \succcurlyeq \mathbf{0}. \end{aligned}$$

**Exercise 2.19.** Write its dual. Conclude (trivially!) that the dual has an optimal solution with objective value $\theta$. Knowing this, conclude now that the primal also has an optimal solution with objective value $\theta$.

**Exercise 2.20.** Write an SDP formulation for $\tau$ and repeat the steps of the previous exercise.

Later in the course, we will see an interesting theorem that characterizes the sum of the $k$ largest eigenvalues of a symmetric matrix as an SDP.

## 2.5 Strong duality

To prove strong duality, we will first review a few concepts from topology, as well as some basic properties.

Let $\mathbb{V}$ be an Euclidean space — it has an inner product, and with it you can define a norm as $\|\mathbf{v}\| = \langle \mathbf{v}, \mathbf{v}\rangle$. The *unit ball* in $\mathbb{V}$ is defined as

$$\mathbb{B} = \{\mathbf{x} \in \mathbb{V} : \|\mathbf{x}\| \leq 1\}.$$

Let $S \subseteq \mathbb{V}$.

- It is *bounded* if there is an $\mu \in \mathbb{R}$ so that $S \subseteq \mu \cdot \mathbb{B}$.
- An element $\mathbf{x} \in S$ is an *interior point* of $S$ if there is $\varepsilon > 0$ with $\mathbf{x} + \varepsilon\mathbb{B} \subseteq S$. The subset of $S$ made out of its interior points is denoted by $\operatorname{int}(S)$.
- $S$ is *open* if $S = \operatorname{int}(S)$. It is *closed* if its complement is open.
- $S$ is *compact* if it is closed and bounded. (This is not the standard definition of compactness for other topological spaces, but it is what we need here.)

**Exercise 2.21.** Verify that $\operatorname{int}(\mathbb{R}^n_+) = \mathbb{R}^n_{++}$. (Meaning, show that a vector in $\mathbb{R}^n_+$ is surrounded by a ball of non-negative if and only if none of its coordinates is equal to 0.)

**Exercise 2.22.** Prove that $\operatorname{int}(\mathbb{S}^n_+) = \mathbb{S}^n_{++}$.

A function $f : \mathbb{V}_1 \to \mathbb{V}_2$ is *continuous* if for every sequence $\{x_n\}$ in $\mathbb{V}_1$ that converges to $x$, $\{f(x_n)\}$ also converges to $f(x)$. It is a standard fact in topology that the following facts are equivalent:

- $f$ is continuous.
- if $S$ is an open set, then $f^{-1}(S)$ is also open.
- if $S$ is a closed set, then $f^{-1}(S)$ is also closed.

From the fact that linear maps are continuous functions, an immediate consequence of the facts above is that hyperplanes and half-spaces are also closed sets. Moreover, arbitrary (finite) unions (intersections) of open sets are open, arbitrary (finite) intersection (unions) of closed sets are closed. Therefore polyhedra are closed sets, the same holding for the duals of cones.

Our goal in this section is to prove the following theorem.

**Theorem 2.23** (Strong duality)**.** *Let $\mathbb{V}_1$ and $\mathbb{V}_2$ be Euclidean spaces, and let $\mathbb{K} = \mathbb{K}' \oplus \mathbb{E}_1 \subseteq \mathbb{V}_1$, and $\mathbb{L} = \mathbb{L}' \oplus \mathbb{E}_2 \subseteq \mathbb{V}_2$ be cones, and unless $\mathbb{K}'$ or $\mathbb{L}'$ are empty, we have $\mathbb{K}'$ and $\mathbb{L}'$ convex, closed, pointed cones, with non-empty interior. Consider the following dual pair of conic programs*

$$\begin{array}{|lll|lll|} & \min & \langle \mathbf{b}, \mathbf{y} \rangle & & \max & \langle \mathbf{c}, \mathbf{x} \rangle \\ (P) & s.t. & \mathcal{A}^*(\mathbf{y}) \succcurlyeq_{\mathbb{K}^*} \mathbf{c} & (D) & s.t. & \mathcal{A}(\mathbf{x}) \preccurlyeq_{\mathbb{L}^*} \mathbf{b} \\ & & \mathbf{y} \in \mathbb{L} & & & \mathbf{x} \in \mathbb{K}. \end{array}$$

*Assume (D) has a Slater Point, meaning, a feasible solution $\mathbf{x}'$ with $\mathbf{x}' \in \operatorname{int}(\mathbb{K}') \oplus \mathbb{E}_1$ and $\mathbf{b} - \mathcal{A}^*(\mathbf{x}') \in \operatorname{int}(\mathbb{L}'^*) \oplus \mathbf{0}$. If the objective values of (D) are upper bounded, then (P) has an optimal solution, and the optimal values of (P) and (D) are equal.*

In order to show this theorem, we will use a Separation Theorem. We will also assume henceforth that $\mathbb{E}_1$ and $\mathbb{E}_2$ are equal to $\{0\}$ (thus we ignore them), but the proofs would go just the same without this hypothesis.

**Theorem 2.24** (Bolzano-Weierstraß)**.** *If $S \subseteq \mathbb{V}$ is non-empty and compact, and if $f : S \to \mathbb{R}$ is continuous, then $f$ attains a minimum and maximum in $S$.*

**Lemma 2.25** (Hyperplane separation)**.** *Assume $C \subseteq \mathbb{V}$ is closed and convex. Let $\overline{\mathbf{y}} \in \overline{C}$. There is $\mathbf{a} \in \mathbb{V}$, $\mathbf{a} \neq \mathbf{0}$, and $\beta \in \mathbb{R}$ so that $C \subseteq \{\mathbf{y} \in \mathbb{V} : \langle \mathbf{a}, \mathbf{y} \rangle \leq \beta\}$ and $\langle \mathbf{a}, \overline{\mathbf{y}} \rangle > \beta$.*

*Proof.* Let $\mathbf{y}'$ belong to $C$. Let $C'$ be the set of points in $C$ which are closer to $\overline{\mathbf{y}}$ than $\mathbf{y}'$ is, meaning,

$$C' = C \cap \{\overline{\mathbf{y}} + \|\mathbf{y}' - \overline{\mathbf{y}}\|\mathbb{B}\}.$$

This is a non-empty compact set. The continuous function that maps $\mathbf{z} \in C'$ to $\|\mathbf{z} - \overline{\mathbf{y}}\|$ has a point that attains a minimum, call it $\overline{\mathbf{z}}$. Let $\mathbf{a} = \overline{\mathbf{y}} - \overline{\mathbf{z}}$, and $\beta = \langle \mathbf{a}, \overline{\mathbf{z}} \rangle$. Let $\mathbf{y} \in C$. As $C$ is convex, the segment from $\mathbf{y}$ to $\overline{\mathbf{z}}$ is in there, so it holds that for all $\lambda \in (0, 1]$,

$$\|\lambda \mathbf{y} + (1 - \lambda)\overline{\mathbf{z}} - \overline{\mathbf{y}}\|^2 \geq \|\overline{\mathbf{z}} - \overline{\mathbf{y}}\|^2.$$

Thus

$$\lambda^2\|\mathbf{y}-\overline{\mathbf{z}}\|^2 \geq 2\lambda\langle \mathbf{y}-\overline{\mathbf{z}},\mathbf{a}\rangle,$$

which, making $\lambda \to 0$, gives

$$\langle \mathbf{y}-\overline{\mathbf{z}},\mathbf{a}\rangle \leq 0 \implies \langle \mathbf{a},\mathbf{y}\rangle \leq \beta.$$

On the other hand, $\langle \mathbf{a},\mathbf{a}\rangle > 0$, thus $\langle \mathbf{a},\overline{\mathbf{y}}\rangle > \beta$. □

**Exercise 2.26.** Provide a new proof of Farka's Lemma, ie, show that if $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, then if there is no $\mathbf{x} \in \mathbb{R}^n_+$ with $\mathbf{Ax} = \mathbf{b}$, then there is $\mathbf{y} \in \mathbb{R}^m$ so that $\mathbf{A}^\mathsf{T}\mathbf{y} \leq \mathbf{0}$ and $\mathbf{b}^\mathsf{T}\mathbf{y} > 0$.

Say now $S$ is a compact set, and $\{S_i\}$ is a family of closed subsets. If all finite intersections of this family are non-empty, then the intersection of the entire family is non-empty (this is actually the definition of being a compact set).

**Theorem 2.27.** *If $C \subseteq \mathbb{V}$ is non-empty and convex, not containing $\mathbf{0}$, then there is $\mathbf{a} \in \mathbb{V}$, $\mathbf{a} \neq \mathbf{0}$, with*

$$C \subseteq \{\mathbf{y} \in \mathbb{V} : \langle \mathbf{a},\mathbf{y}\rangle \geq 0\}.$$

*Proof.* The tricky thing here is that we are not assuming $C$ is closed. Let $\mathbb{B}' = \{\mathbf{y} : \|\mathbf{y}\| = 1\}$ (this is compact). To show this is result, it is enough to show the existence of $\mathbf{a} \in \mathbb{B}'$ that belongs to

$$\bigcap_{\mathbf{x}\in C} (\mathbb{R}_+ \cdot \mathbf{x})^*.$$

Let $\mathbf{x}_1, ..., \mathbf{x}_k$ be points in $C$. Their convex hull is closed and does not contain $\mathbf{0}$. Thus, there is $\mathbf{a} \in \mathbb{V}$, non-zero, and $\beta \in \mathbb{R}$, with $\langle \mathbf{a},\mathbf{x}_i\rangle \leq \beta$, and $\langle \mathbf{a},\mathbf{0}\rangle > \beta$. Thus $-\mathbf{a}/\|\mathbf{a}\|$ has unit norm and belongs to

$$\bigcap_{i=1}^{k} (\mathbb{R}_+ \cdot \mathbf{x}_i)^*.$$

Thus, by the comment before the theorem, the result is proved. □

**Lemma 2.28.** *Let $\mathbb{K}$ be a convex cone. If $\overline{\mathbf{x}} \in \mathbb{K}$ is non-zero and $\overline{\mathbf{s}}$ is in the interior of $\mathbb{K}^*$, then $\langle \overline{\mathbf{x}},\overline{\mathbf{s}}\rangle > 0$.*

*Proof.* Exercise (quite an easy one). □

**Lemma 2.29.** *If $C_1$ and $C_2$ are convex, non-empty and disjoint, then there is $\mathbf{a} \in \mathbb{V}$, non-zero, so that*

$$\inf_{\mathbf{y}\in C_1} \langle \mathbf{a},\mathbf{y}\rangle \geq \sup_{\mathbf{y}\in C_2} \langle \mathbf{a},\mathbf{y}\rangle.$$

*Proof.* Let $C = C_1 - C_2$. Then $0 \notin C$. By Theorem 2.27, it follows that there is $\mathbf{a}$, non-zero, with $\langle \mathbf{a},\mathbf{y}\rangle \geq 0$ for all $\mathbf{y} \in C$. Thus $\langle \mathbf{a},\mathbf{y}_1\rangle \geq \langle \mathbf{a},\mathbf{y}_2\rangle$ for all $\mathbf{y}_1 \in C_1$ and $\mathbf{y}_2 \in C_2$. □

We are ready to prove strong duality.

*Proof of Theorem 2.23.* Let $\mathbf{s}' = \mathbf{b} - \mathcal{A}(\mathbf{x}) \in \text{int}(\mathbb{L}^*)$. Let $\nu$ be the sup of the objective values of (D). If $\mathbf{c} = \mathbf{0}$, we are done. Assume otherwise.

Define now sets

$$C_1 = \{\mathbf{b} - \mathcal{A}(\mathbf{x}) : \langle \mathbf{c}, \mathbf{x}\rangle \geq \nu\}$$

and $C_2 = \text{int}(\mathbb{L}^*)$. They are both non-empty (take $\mathbf{x} = \nu(\mathbf{c}/\|\mathbf{c}\|^2)$), convex and with disjoint intersection, otherwise there would be $\mathbf{x}'$ with $\langle \mathbf{c}, \mathbf{x}'\rangle \geq \nu$ and $\mathbf{b} - \mathcal{A}(\mathbf{x}') \in \text{int}(\mathbb{L}^*)$ (and you can now prove this implies $\nu$ is no longer the sup).

By Lemma 2.29, there is a non-zero $\tilde{\mathbf{y}}$ with

$$\sup_{\mathbf{s}\in C_1} \langle \tilde{\mathbf{y}}, \mathbf{s}\rangle \leq \inf_{\mathbf{s}\in C_2} \langle \tilde{\mathbf{y}}, \mathbf{s}\rangle.$$

The fact that for all $\mathbf{s} \in C_2$, we also have $\mathbb{R}_+\mathbf{s} \in C_2$, implies that $\langle \tilde{\mathbf{y}}, \mathbf{s}\rangle$ cannot be negative, as $C_1$ is non-empty, and at the same time, $\langle \tilde{\mathbf{y}}, \mathbf{s}\rangle$ can be made arbitrarily small. So

$$\inf_{\mathbf{s}\in C_2} \langle \tilde{\mathbf{y}}, \mathbf{s}\rangle = 0.$$

If $\bar{\mathbf{s}} \in \mathbb{L}^*$, then $\bar{\mathbf{s}} + \epsilon\mathbf{s}' \in \text{int}(\mathbb{L}^*)$, thus $\langle \tilde{\mathbf{y}}, \bar{\mathbf{s}}\rangle \geq -\epsilon\langle \tilde{\mathbf{y}}, \mathbf{s}'\rangle \to 0$. Therefore $\tilde{\mathbf{y}} \in \mathbb{L}^{**} = \mathbb{L}$ (we still need to prove this).

For all $\mathbf{x}$ with $\langle \mathbf{c}, \mathbf{x}\rangle \geq \nu$, the displayed expressions above give that

$$\langle \mathcal{A}^*(\tilde{\mathbf{y}}), \mathbf{x}\rangle \geq \langle \mathbf{b}, \tilde{\mathbf{y}}\rangle,$$

thus the linear program

$$\begin{aligned} \min \quad & \langle \mathcal{A}^*(\tilde{\mathbf{y}}), \mathbf{x}\rangle \\ \text{subject to} \quad & \langle \mathbf{c}, \mathbf{x}\rangle \geq \nu \end{aligned}$$

is feasible and bounded, thus it admits an optimum solution. Its dual is

$$\begin{aligned} \max \quad & \nu\mu \\ \text{subject to} \quad & \mu\mathbf{c} = \mathcal{A}^*(\tilde{\mathbf{y}}) \\ & \mu \in \mathbb{R}_+, \end{aligned}$$

then feasible, hence there is $\alpha$ with $\alpha\mathbf{c} = \mathcal{A}^*(\tilde{\mathbf{y}})$, with $\alpha \geq 0$. If $\alpha = 0$, then

$$0 \geq \langle \mathbf{b}, \tilde{\mathbf{y}}\rangle = \langle \mathcal{A}(\mathbf{x}') + \mathbf{s}', \tilde{\mathbf{y}}\rangle = \langle \mathbf{x}', \mathcal{A}^*(\tilde{\mathbf{y}})\rangle + \langle \mathbf{s}', \tilde{\mathbf{y}}\rangle > 0.$$

(This was the only moment we used the existence of $\mathbf{x}'$).

Now, let $\overline{\mathbf{y}} = (1/\alpha)\tilde{\mathbf{y}}$. This $\overline{\mathbf{y}}$ is feasible in (P), and its objective value is $\nu$. Therefore $\overline{\mathbf{y}}$ is the optimum solution we have been looking for. □

## 2.6 SDPs — consequences and applications

We have already mentioned that complementary slackness conditions are naturally defined as consequences of how the dual was defined. Now we are ready to prove a "complementary slackness" theorem.

**Corollary 2.30.** *Consider a primal-dual pair of SDPs*

$$\begin{array}{rrl|rrl} & min & \langle \mathbf{b}, \mathbf{y}\rangle & & max & \langle \mathbf{C}, \mathbf{X}\rangle \\ (P) & s.t. & \mathcal{A}^*(\mathbf{y}) \succcurlyeq \mathbf{C} & (D) & s.t. & \mathcal{A}(\mathbf{X}) \leq \mathbf{b} \\ & & \mathbf{y} \in \mathbb{R}^n_+ & & & \mathbf{X} \in \mathbb{S}^n_+, \end{array}$$

*so that at least (P) or (D) contains a Slater point. Let $\overline{\mathbf{y}}$ and $\overline{\mathbf{X}}$ be feasible solutions to (P) and (D). Then these are respective optima solutions if and only if*

*(i)* $\langle \overline{\mathbf{X}}, \mathcal{A}^*(\overline{\mathbf{y}}) - \mathbf{C}\rangle = 0$.

*(ii)* $\langle \overline{\mathbf{y}}, \mathbf{b} - \mathcal{A}(\mathbf{X})\rangle = 0$.

*Proof.* Follows immediately from

$$0 \leq \langle \overline{\mathbf{X}}, \mathcal{A}^*(\overline{\mathbf{y}}) - \mathbf{C}\rangle = \langle \mathcal{A}(\overline{\mathbf{X}}), \overline{\mathbf{y}}\rangle - \langle \overline{\mathbf{X}}, \mathbf{C}\rangle \leq \langle \mathbf{b}, \overline{\mathbf{y}}\rangle - \langle \overline{\mathbf{X}}, \mathbf{C}\rangle.$$

From Theorem 2.23, $\overline{\mathbf{X}}$ and $\overline{\mathbf{y}}$ are optima if and only if equality holds throughout, which is if and only if both conditions hold. □

Recall that (i) is equivalent to $\overline{\mathbf{X}}(\mathcal{A}^*(\overline{\mathbf{y}}) - \mathbf{C}) = \mathbf{0}$, which can be quite useful information in designing algorithmic uses of SDP duality.

We come back to the two problems we have already discussed, but we also move further.

### Eigenvalues and Lovász theta

As we have seen, the maximum eigenvalue of a matrix $\mathbf{M}$ is the optima of the following primal-dual pair of SDPs:

$$\begin{array}{rrl|rrl} & \max & \langle \mathbf{M}, \mathbf{X}\rangle & & \min & y \\ (P) & s.t. & \langle \mathbf{I}, \mathbf{X}\rangle = 1 & (D) & s.t. & y\mathbf{I} - \mathbf{M} \succcurlyeq \mathbf{0} \\ & & \mathbf{X} \succcurlyeq \mathbf{0} & & & \end{array}$$

Note that both these programs contain Slater points (thus, as we already know, strong duality holds). More importantly now, one can allow some few entries of the matrix $\mathbf{M}$ to vary and the program (D) remains an SDP. For example, if we have a graph $G$ and $\mathbf{M}$ is any matrix of so that $\mathbf{M}_{ij} = 1$ whenever $i$ and $j$ are equal or correspond to non adjacent vertices, then $\mathbf{M} = \mathbf{J} - (\sum_{ij \in E(G)} \mathbf{y}_{ij}\mathbf{E}_{ij})$, where $y_{ij}$ are free variables and $\mathbf{E}_{ij}$ are the adjacency matrices of edges. Writing the programs, we encounter an old friend

$$\begin{array}{rl|rl} \max & \langle \mathbf{J}, \mathbf{X}\rangle & \min & y \\ \text{subject to} & \langle \mathbf{E}_{ij}, \mathbf{X}\rangle = \mathbf{0},\ \forall\ ij \in E(G) & \text{subject to} & y\mathbf{I} - (\mathbf{J} - \sum_{ij \in E(G)} \mathbf{y}_{ij}\mathbf{E}_{ij}) \succcurlyeq \mathbf{0} \\ & \langle \mathbf{I}, \mathbf{X}\rangle = 1 & & \\ & \mathbf{X} \succcurlyeq \mathbf{0} & & \end{array}$$

Note, again, that both of these have Slater points, and therefore their respective optima are attained. As a consequence, $\vartheta(G)$ is equal to the minimum of the largest eigenvalues of all matrices $\mathbf{M}$ whose entries corresponding to non-edges are equal to 1.

### Euclidean norm

Assume we want to guarantee $\mathbf{x} \in \mathbb{R}^n$ has norm bounded by $\mu$. Can this constraint be expressed as positive semidefiniteness? Recall Schur's complement theorem (Theorem 1.41). Note that

$$\|\mathbf{x}\| \leq \mu \iff \mu\mathbf{I} - \frac{1}{\mu}\mathbf{x}\mathbf{x}^\mathsf{T} \succcurlyeq \mathbf{0} \iff \begin{pmatrix} \mu & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & \mu\mathbf{I} \end{pmatrix} \succcurlyeq \mathbf{0}.$$

As a consequence of this fact, we are learning that all conic programs over the Lorentz cone can be formulated over $\mathbb{S}^n_+$ (however the former is more suitable to solvers than the latter).

### 2-norms of a matrix

If $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\|\cdot\|$ is a norm in $\mathbb{R}^n$, then it induces a norm $\mathbb{R}^{n\times n}$ defined as

$$\|\mathbf{A}\| = \max_{\mathbf{x}\in\mathbb{B}} \|\mathbf{A}\mathbf{x}\|.$$

The norm in $\mathbb{R}^{n\times n}$ induced by the standard Euclidean norm is called the 2-norm, denoted by $\|\cdot\|_2$.

**Exercise 2.31.** If $\mathbf{A} \in \mathbb{S}^n$, show that $\|\mathbf{A}\|_2 = \max\{|\lambda_{\min}(\mathbf{A})|, |\lambda_{\max}(\mathbf{A})|\}$. In general, for any $\mathbf{A} \in \mathbb{R}^{n\times n}$, show that

$$\|A\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^\mathsf{T}\mathbf{A})}.$$

As before, we can them easily bound the 2-norm of matrices using positive semidefiniteness:

$$\|\mathbf{X}\|_2 \leq \mu \iff \begin{pmatrix} \mu\mathbf{I} & \mathbf{X}^\mathsf{T} \\ \mathbf{X} & \mu\mathbf{I} \end{pmatrix} \succcurlyeq \mathbf{0}.$$

**Exercise 2.32.** Prove the if and only if above (use, of course, Schur's complement Theorem).

### Sum of eigenvalues

Largest (or smallest) eigenvalues are optima of SDPs. The same can be said about the sum of the $k$ largest eigenvalues. In this section, assume $\lambda_j(\mathbf{X})$ stands for the $j$th largest eigenvalue of $\mathbf{X}$.

**Theorem 2.33.** *Let $\mathbf{X} \in \mathbb{S}^n$, $\mu \in \mathbb{R}$, $k \in [n]$. The following are equivalent.*

*(1)* $\sum_{j=1}^k \lambda_j(\mathbf{X}) \leq \mu$.

*(2) There are $\mathbf{Y} \in \mathbb{S}^n_+$ and $\eta \in \mathbb{R}$ so that*

$$\mu - k\eta \geq \operatorname{tr}\mathbf{Y} \quad \textit{and} \quad \mathbf{Y} - \mathbf{X} + \eta\mathbf{I} \succcurlyeq \mathbf{0}.$$

*Proof.* From (2) to (1), we have

$$\mathbf{Y} - \mathbf{X} \succcurlyeq -\eta\mathbf{I} \implies \lambda_j(\mathbf{Y}) - \lambda_j(\mathbf{X}) \geq -\eta.$$

(This implication is not trivial — due to Ky Fan.) Thus, immediately, $\sum_{j=1}^k \lambda_j(\mathbf{X}) \leq \mu$.

From (1) to (2), have $\eta = \lambda_k(\mathbf{X})$, and

$$\mathbf{Y} = \sum_{j=1}^{k} (\lambda_j(\mathbf{X}) - \eta)\mathbf{v}_j\mathbf{v}_j^\mathsf{T},$$

where $\mathbf{v}_j$ are the eigenvectors of $\mathbf{X}$ corresponding to $\lambda_j$. Thus $\mu - k\eta - \operatorname{tr}(\mathbf{Y}) \geq 0$, and $\mathbf{Y} - \mathbf{X} + \eta\mathbf{I} \succcurlyeq \mathbf{0}$. □

**Exercise 2.34.** Given $\mathbf{M}$, show that $\lambda_1(\mathbf{M}) + ... + \lambda_k(\mathbf{M})$ is equal to

$$\begin{aligned} \max \quad & \langle \mathbf{M}, \mathbf{X} \rangle \\ \text{subject to} \quad & \langle \mathbf{I}, \mathbf{X} \rangle = k \\ & \mathbf{X} \preccurlyeq \mathbf{I} \\ & \mathbf{X} \succcurlyeq \mathbf{0}. \end{aligned}$$

(This exercise is not trivial. First, you will have to consider products of cones (don't worry, theory still applies)). Second, you will have to examine the proof above carefully).

# 3 Solving an SDP

In this section, we will discuss methods to solve a semidefinite program. Ideally, we would like to have methods that work well both in theory and practice. The first method we present satisfies only the first requirement: it was the first method developed that provably showed how to solve an SDP in polynomial time (modulo some details which we will get on).

## 3.1 Ellipsoids

An *ellipsoid* in $\mathbb{R}^n$ is a set of the form $\mathbf{c} + \mathbf{A} \cdot \mathbb{B}$, where $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{A}$ is non-singular.

**Lemma 3.1.** *Given* $\mathbf{c}$ *and* $\mathbf{A}$ *non-singular, we have*

$$\mathbf{c} + \mathbf{A}\mathbb{B} = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x} - \mathbf{c})^\mathsf{T} \mathbf{A}^{-T} \mathbf{A}^{-1} (\mathbf{x} - \mathbf{c}) \leq 1\}.$$

*Proof.* Follow immediately from

$$\mathbf{A}\mathbb{B} = \{\mathbf{A}\mathbf{x} : \|\mathbf{x}\| \leq 1\} = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y}^\mathsf{T} \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{y} \leq 1\}.$$

□

Note in particular that by taking $\mathbf{B}$ to be the square root of $\mathbf{A}\mathbf{A}^\mathsf{T}$, we may as well assume that every ellipsoid is of the form $\mathbf{c} + \mathbf{A}\mathbb{B}$ with $\mathbf{A} \in \mathbb{S}^n_{++}$.

The *volume* of $\mathbf{c} + \mathbf{A}\mathbb{B}$ is defined as

$$\operatorname{vol} \mathbf{A}\mathbb{B} = |\det(\mathbf{A})| \operatorname{vol} \mathbb{B} = |\det(\mathbf{A})| \frac{\pi^{n/2}}{\Gamma((n/2)+1)},$$

where $\Gamma(\cdot)$ is Euler's gamma function (it equals the factorial for integers).

Given a polytope $\mathcal{P} = \{\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$ (recall that a polytope is a bounded polyhedron), one could formulate the problem of finding the largest possible ellipsoid contained in $\mathbb{P}$ as follows

$$\begin{aligned} \max \quad & \mu \\ \text{subject to} \quad & \mu \leq (\det \mathbf{X})^{1/n} \\ & \mathbf{X} \in \mathbb{S}^n_{++} \\ & \mathbf{c} \in \mathbb{R}^n \\ & \|\mathbf{X}\mathbf{A}^\mathsf{T}\mathbf{e}_i\| \leq (\mathrm{b}_i - \mathbf{e}_i^\mathsf{T}\mathbf{A}\mathbf{c}) \quad \forall\ i \in [m] \end{aligned}$$

In fact, for $\mathbf{X} \in \mathbb{S}^n_{++}$, it follows that

$$\begin{aligned} \mathbf{c} + \mathbf{X}\mathbb{B} \subseteq \mathbb{P} &\iff \mathbf{e}_i^\mathsf{T}\mathbf{A}(\mathbf{c} + \mathbf{X}\mathbf{u}) \leq \mathrm{b}_i \quad \forall i \in [m] \text{ and } \forall u \in \mathbb{B} \\ &\iff \max_{\mathbf{u} \in \mathbb{B}} \mathbf{e}_i^\mathsf{T}\mathbf{A}(\mathbf{c} + \mathbf{X}\mathbf{u}) \leq \mathrm{b}_i \quad \forall i \in [m] \\ &\iff \mathbf{e}_i^\mathsf{T}\mathbf{A}\mathbf{c} + \frac{\mathbf{e}_i^\mathsf{T}\mathbf{A}\mathbf{X}\mathbf{X}\mathbf{A}^\mathsf{T}\mathbf{e}_i}{\|\mathbf{X}\mathbf{A}^\mathsf{T}\mathbf{e}_i\|} \leq \mathrm{b}_i \quad \forall i \in [m] \\ &\iff \|\mathbf{X}\mathbf{A}^\mathsf{T}\mathbf{e}_i\| \leq \mathrm{b}_i - \mathbf{e}_i^\mathsf{T}\mathbf{A}\mathbf{c}. \end{aligned}$$

(I should point out that $(\det \mathbf{X})^{1/n}$ is a concave function, therefore this optimization problem is a convex optimization problem. But we won't need to get in details).

## 3.2 Searching for a set with ellipsoids

Let us now turn to the main problem we are interested at. Let $\mathcal{C}$ be a convex compact set in $\mathbb{R}^n$, given by means of a *separation oracle*, meaning, an algorithm that, given $\mathbf{x} \in \mathbb{R}^n$, says that $\mathbf{x} \in C$ or, else, separates $\mathbf{x}$ from $\mathcal{C}$ with a hyperplane, meaning, it gives $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ with $\mathbf{a}^\mathsf{T}\mathbf{y} \geq \mathbf{a}^\mathsf{T}\mathbf{x}$ for all $\mathbf{y} \in C$.

Given a convex compact set $\mathcal{C}$, the main problem we are interested at is that of finding $\mathbf{x} \in \mathcal{C}$, given that a separation oracle is provided. To that, we will follow roughly the following steps:

(1) Maintain, at each iteration, an ellipsoid $\mathbf{c}_t + \mathbf{A}_t^{1/2}\mathbb{B} = \mathcal{E}_t$ that contains $\mathcal{C}$, with, initially, $\mathbf{A}_0 = R^2\mathbf{I}$, where $R > 0$ is so that $C \subseteq \mathbf{c}_0 + R\mathbb{B}$.

(2) If at any point $\mathbf{c}_t \in \mathcal{C}$, we are done. Otherwise, the oracle gives as back $\mathbf{a}$ with $\mathbf{a}^\mathsf{T}\mathbf{y} \geq \mathbf{a}^\mathsf{T}\mathbf{c}_t$ for all $\mathbf{y} \in \mathcal{C}$.

(3) Find a "much smaller" ellipsoid $\mathbf{c}_{t+1} + \mathbf{A}_{t+1}^{1/2}\mathbb{B}$ that contains

$$(\mathbf{c}_t + \mathbf{A}_t^{1/2}\mathbb{B}) \cap \{\mathbf{y} \in \mathbb{R}^n : \mathbf{a}\mathbf{y}^\mathsf{T} \geq \mathbf{a}^\mathsf{T}\mathbf{c}_t\}.$$

Then, go back to (1).

We hope that the volume of the ellipsoid decreases quickly, say exponentially, so that after few iterations we will either have found a point in $\mathcal{C}$, or we will know that $\mathcal{C}$ is very small. Let us now see how to perform (3).

Given $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{S}^n_{++}$ and $\mathbf{a} \in \mathbb{R}^n$, non-zero, let

$$\mathcal{E}(\mathbf{c}, \mathbf{A}) = (\mathbf{c} + \mathbf{A}^{1/2}\mathbb{B}), \quad \text{and} \quad \mathcal{E}_{1/2}(\mathbf{c}, \mathbf{A}, \mathbf{a}) = (\mathbf{c} + \mathbf{A}^{1/2}\mathbb{B}) \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\mathsf{T}\mathbf{x} \geq \mathbf{a}^\mathsf{T}\mathbf{c}\}.$$

We want to find $\mathbf{c}_+$ and $\mathbf{A}_+$ with

$$\mathcal{E}_{1/2}(\mathbf{c}, \mathbf{A}, \mathbf{a}) \subseteq \mathcal{E}(\mathbf{c}_+, \mathbf{A}_+) = \mathbf{c}_+ + \mathbf{A}_+^{1/2}\mathbf{B} = \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x} - \mathbf{c}_+)^\mathsf{T}\mathbf{A}_+^{-1}(\mathbf{x} - \mathbf{c}_+) \leq 1\}.$$

Assume first $\mathbf{c} = \mathbf{0}$ and $\mathbf{A} = \mathbf{I}$, and that $\|\mathbf{a}\| = 1$. Thinking geometrically, it makes sense to seek $\mathbf{c}_+$ of the form $\alpha\mathbf{a}$, and $\mathbf{A}_+^{-1} = \gamma(\mathbf{I} + \beta\mathbf{a}\mathbf{a}^\mathsf{T})$, with $\alpha, \beta, \gamma \in \mathbb{R}_{++}$.

Recall the definition $\mathbb{B}_= = \{\mathbf{x} \in \mathbb{B} : \|\mathbf{x}\| = 1\}$.

**Exercise 3.2.** Show that every ellipsoid is convex, compact, and has non-empty interior.

**Exercise 3.3.** Let $\mathbf{a} \in \mathbb{R}^n$, non-zero. Show that $\mathbb{B} \cap \{\mathbf{a}\}^*$ is equal to the convex hull of $\mathbb{B}_= \cap \{\mathbf{a}\}^*$.

Now it follows that $\mathcal{E}_{1/2} \subseteq \mathcal{E}_+$ if and only if

$$\begin{aligned}
&\iff \mathbb{B}_= \cap \{\mathbf{a}\}^* \subseteq \{\mathbf{x} \in \mathbb{R}^n : (\mathbf{x}-\alpha\mathbf{a})^\mathsf{T}\gamma(\mathbf{I}+\beta\mathbf{a}\mathbf{a}^\mathsf{T})(\mathbf{x}-\alpha\mathbf{a}) \leq 1\} \\
&\iff \|\mathbf{x}-\alpha\mathbf{a}\|^2 + \beta(\mathbf{a}^\mathsf{T}(\mathbf{x}-\alpha\mathbf{a}))^2 \leq 1/\gamma\ , \quad \forall\ \mathbf{x} \in \mathbb{B}_+ \cap \{\mathbf{a}\}^* \\
&\iff \|\mathbf{x}\|^2 - 2\alpha\mathbf{x}^\mathsf{T}\mathbf{a} + \alpha^2\|\mathbf{a}\|^2 + \beta(\mathbf{x}^\mathsf{T}\mathbf{a} - \alpha\|\mathbf{a}\|^2)^2 \leq 1/\gamma\ , \quad \forall\ \mathbf{x} \in \mathbb{B} \cap \{\mathbf{a}\}^* \\
&\iff 1 - 2\alpha\mathbf{x}^\mathsf{T}\mathbf{a} + \alpha^2 + \beta(\mathbf{x}^\mathsf{T}\mathbf{a} - \alpha)^2 \leq 1/\gamma\ , \quad \forall\ \mathbf{x} \in \mathbb{B}_= \cap \{\mathbf{a}\}^*
\end{aligned}$$

and now, as $0 \leq \mathbf{x}^\mathsf{T}\mathbf{a} \leq 1$ for all $\mathbf{x} \in \mathbb{B}_= \cap \{\mathbf{a}\}^*$, all values attained, we have

$$\begin{aligned}
&\iff 1 - 2\alpha\mu + \alpha^2 + \beta(\mu-\alpha)^2 \leq 1/\gamma\ , \quad \forall\ \mu \in [0,1] \\
&\iff 1 + \alpha^2(1+\beta) - \mu 2\alpha(1+\beta) + \mu^2\beta \leq 1/\gamma\ , \quad \forall\ \mu \in [0,1]
\end{aligned}$$

as the lefthand side is quadratic in $\mu$, with positive coefficient...

$$\iff 1 + \alpha^2(1+\beta) \leq 1/\gamma \quad \text{and} \quad 1 + \alpha^2(1+\beta) - 2\alpha(1+\beta) + \beta \leq 1/\gamma.$$

the first corresponding to $\mu = 0$ and the second to $\mu = 1$.

Now we can choose $\alpha = 1/(n+1)$, $\beta = 2/(n-1)$ and $\gamma = (n^2-1)/n^2$, we have

$$\begin{aligned}
&1 + \alpha^2(1+\beta) \leq 1/\gamma \\
\iff &1 + \frac{1}{(n+1)^2}\left(1 + \frac{2}{n-1}\right) \leq \frac{n^2}{n^2-1} \\
\iff &\frac{1}{(n+1)^2}\frac{n+1}{n-1} \leq \frac{1}{n^2-1}.
\end{aligned}$$

And, likewise, you can verify that

$$1 + \alpha^2(1+\beta) - 2\alpha(1+\beta) + \beta \leq 1/\gamma.$$

Phew. So we now have

$$\mathcal{E}_{1/2} \subseteq \mathcal{E}_+ = \{\mathbf{x} \in \mathbb{R} : (\mathbf{x}-\mathbf{c})^\mathsf{T}\mathbf{A}_+^{-1}(\mathbf{x}-\mathbf{c}) \leq 1\},$$

where

$$\mathbf{c} = \frac{1}{n+1}\mathbf{a}, \text{ and } \mathbf{A}_+^{-1} = \left(1 - \frac{1}{n^2}\right)\left(\mathbf{I} + \frac{2}{n-1}\mathbf{a}\mathbf{a}^\mathsf{T}\right).$$

This proves the following result.

**Lemma 3.4.** *With $\mathbf{a} \in \mathbb{B}_=$, we have*

$$\mathbb{B} \cap \{\mathbf{a}\}^* \subseteq \frac{1}{n+1}\mathbf{a} + \left[\left(1 - \frac{1}{n^2}\right)\left(\mathbf{I} + \frac{2}{n-1}\mathbf{a}\mathbf{a}^\mathsf{T}\right)\right]^{-1/2}\mathbb{B}$$

□

We can work it further a bit, using the lemma below.

**Lemma 3.5** (Sherman-Morrison)**.** *If $\mathbf{M} \in \mathbb{R}^{n\times n}$, invertible, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then*

$$(\mathbf{M} + \mathbf{u}\mathbf{v}^\mathsf{T})^{-1} = \mathbf{M}^{-1} - \frac{\mathbf{M}^{-1}\mathbf{u}\mathbf{v}^\mathsf{T}\mathbf{M}^{-1}}{1 + \mathbf{v}^\mathsf{T}\mathbf{M}^{-1}\mathbf{u}},$$

*whenever the denominator is non-zero.* □

*Proof.* Exercise. □

Therefore:

**Theorem 3.6.** *With $\mathbf{a} \in \mathbb{B}_=$, we have*

$$\mathbb{B} \cap \{\mathbf{a}\}^* \subseteq \frac{1}{n+1}\mathbf{a} + \left[\left(\frac{n^2}{n^2-1}\right)\left(\mathbf{I} - \frac{2}{n+1}\mathbf{a}\mathbf{a}^\mathsf{T}\right)\right]^{1/2} \mathbb{B}$$

*Proof.* Immediate from Lemmas 3.4 and 3.5. □

Now let us move to the general case. Given $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{S}^n_{++}$ and $\mathbf{a} \in \mathbb{R}^n$, non-zero, we want to find $\mathbf{c}_+$ and $\mathbf{A}_+$ with

$$(\mathbf{c} + \mathbf{A}^{1/2}\mathbb{B}) \cap \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\mathsf{T}\mathbf{x} \geq \mathbf{a}^\mathsf{T}\mathbf{c}\} \quad \subseteq \quad \mathbf{c}_+ + \mathbf{A}_+^{1/2}\mathbf{B}.$$

**Theorem 3.7.** *Given $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{S}^n_{++}$ and $\mathbf{a} \in \mathbb{R}^n$, non-zero, define*

$$\mathbf{c}_+ = \mathbf{c} + \frac{\mathbf{A}\mathbf{a}}{(n+1)(\mathbf{a}^\mathsf{T}\mathbf{A}\mathbf{a})^{1/2}} \quad \textit{and} \quad \mathbf{A}_+ = \frac{n^2}{n^2-1}\left[\mathbf{A} - \frac{2}{(n+1)\mathbf{a}^\mathsf{T}\mathbf{A}\mathbf{a}}\mathbf{A}\mathbf{a}\mathbf{a}^\mathsf{T}\mathbf{A}\right].$$

*Then $\mathcal{E}_{1/2}(\mathbf{c}, \mathbf{A}, \mathbf{a}) \subseteq \mathcal{E}(\mathbf{c}_+, \mathbf{A}_+)$.*

*Proof.* I will laid out the steps. You should fill in the details.

(i) First, replace $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\mathsf{T}\mathbf{x} \geq \mathbf{a}^\mathsf{T}\mathbf{c}\}$ by $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\mathsf{T}(\mathbf{x} - \mathbf{c}) \geq 0\}$.

(ii) Subtract $\mathbf{c}$ from all vectors of the space. The containment is still equivalent to the translated version.

(iii) Apply $\mathbf{A}^{-1/2}$ to all vectors.

(iv) You should have that $\mathcal{E}_{1/2}(\mathbf{c}, \mathbf{A}, \mathbf{a}) \subseteq \mathcal{E}(\mathbf{c}_+, \mathbf{A}_+)$ is equivalent to

$$\mathbb{B} \cap \{\mathbf{A}^{1/2}\mathbf{a}\}^* \quad \subseteq \quad \mathbf{A}^{-1/2}(\mathbf{c}_+ - \mathbf{c}) + \mathbf{A}^{-1/2}\mathbf{A}_+^{1/2}\mathbb{B}.$$

(v) Normalize $\mathbf{A}^{1/2}\mathbf{a}$ to $\overline{\mathbf{a}}$.

(vi) Choose $\mathbf{c}_+$ and $\mathbf{A}_+$ so that the right hand side is in the form of Theorem 3.6.

**Exercise 3.8.** Verify that $\mathbf{c}_+$ and $\mathbf{A}_+$ from the statement work exactly, and prove that $\mathbf{A}_+$ is so that

$$\mathbf{A}_+ \succcurlyeq \frac{n^2}{(n+1)^2}\mathbf{A}.$$

□

Recall that our goal, eventually, will be to show that $\mathcal{E}_+$ has "much smaller" volume. To that effect, we will need to following lemma

**Lemma 3.9.** *If* $\mathbf{A}$ *is invertible,* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$*, then*

$$\det(\mathbf{A} + \mathbf{u}\mathbf{v}^\mathsf{T}) = (1 + \mathbf{v}^\mathsf{T}\mathbf{A}^{-1}\mathbf{u}) \det \mathbf{A}.$$

**Exercise 3.10.** Prove that too (check the subsection on positive semidefinite matrices in the introduction for a killer hint).

**Exercise 3.11.** Using $\mathbf{c}_+$ and $\mathbf{A}_+$ as in the theorem, prove that

$$\frac{\operatorname{vol} \mathcal{E}(\mathbf{c}_+, \mathbf{A}_+)}{\operatorname{vol} \mathcal{E}(\mathbf{c}, \mathbf{A})} = \left(\frac{n}{n+1}\right)^{n+1} \left(\frac{n}{n-1}\right)^{n-1}.$$

**Exercise 3.12.** Prove that this thing is $< \mathrm{e}^{-1/2n}$ (you will have to use the Taylor series of the logarithm).

As a consequence, if the ellipsoids are built as we described in the beginning of this subsection, then

$$\frac{\operatorname{vol} \mathcal{E}_T}{\operatorname{vol} \mathcal{E}_0} < \mathrm{e}^{-T/2n},$$

hence, given $\epsilon$, we need only take $T \geq 2n^2 \log(R/\epsilon)$ to obtain $\operatorname{vol}(\mathcal{E}_T) \leq \operatorname{vol}(\epsilon\mathbb{B})$.

## 3.3 The ellipsoid method to optimize

First, a summary of what we had in the previous section.

**Theorem 3.13.** *Let* $\mathcal{C}$ *be convex. Let* $\mathbf{c}_0$ *and* $R \in \mathbb{R}_{++}$ *with* $\mathcal{C} \subseteq \mathbf{c}_0 + R\mathbb{B}$*. Let* $\varepsilon > 0$*, and* $\tau \in \mathbb{N}$*, with* $\tau \geq 2n \log(R/\varepsilon)$*. Let* $\mathbf{c}_1, ..., \mathbf{c}_\tau \in \mathbb{R}^n$*,* $\mathbf{A}_0, ..., \mathbf{A}_\tau \in \mathbb{S}^n_{++}$ *and* $\mathbf{a}_0, ..., \mathbf{a}_\tau \in \mathbb{R}^n$*, all non-zero, satisfying the following*

*(i)* $\mathbf{A}_0 = R^2 \cdot \mathbf{I}$.

*(ii)* $\mathbf{c}_t \notin \mathcal{C}$ *for all* $t \in \{0, ..., \tau - 1\}$.

*(iii)* $\mathcal{C} \subseteq \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}_t^\mathsf{T}\mathbf{x} \geq \mathbf{a}_t^\mathsf{T}\mathbf{c}_t\}$ *for all* $t \in \{0, ..., \tau - 1\}$.

*(iv)* $\mathbf{c}_{t+1} = \mathbf{c}_t + \dfrac{\mathbf{A}_t\mathbf{a}_t}{(n+1)(\mathbf{a}_t^\mathsf{T}\mathbf{A}_t\mathbf{a}_t)^{1/2}}$ *for all* $t \in \{0, ..., \tau - 1\}$.

*(v)* $\mathbf{A}_{t+1} = \dfrac{n^2}{n^2-1}\left[\mathbf{A}_t - \dfrac{2\mathbf{A}_t\mathbf{a}_t\mathbf{a}_t^\mathsf{T}\mathbf{A}_t}{(n+1)\mathbf{a}_t^\mathsf{T}\mathbf{A}_t\mathbf{a}_t}\right]$ *for all* $t \in \{0, ..., \tau - 1\}$.

*Then* $\operatorname{vol} \mathcal{C} \leq \varepsilon \operatorname{vol} \mathbb{B}$.

Let us now study how the procedure above leads an efficient algorithm to solve optimization problems in a more realistic scenario. First, in most cases we can only deal with rational approximations. Given $\mathcal{C}$ convex, a *weak separation oracle* to $\mathcal{C}$ is an algorithm that, given $\overline{\mathbf{x}} \in \mathbb{Q}^n$ and $\varepsilon \in \mathbb{Q}_{++}$, determines whether $\overline{\mathbf{x}} \in \mathcal{C} + \varepsilon\mathbb{B}$, and, otherwise, provides $\mathbf{a} \in \mathbb{Q}^n$, non-zero, with

$$\max_i |\mathbf{a}_i| = \|\mathbf{a}\|_\infty = 1 \quad \text{and} \quad \mathbf{a}^\mathsf{T}\overline{\mathbf{x}} \geq \mathbf{a}^\mathsf{T}\mathbf{x} - \varepsilon\,, \quad \forall\, \mathbf{x} \in \mathcal{C} + \varepsilon\mathbb{B}.$$

Now the scheme we discussed provides a method to test feasibility with a good convergence pace. Minimization of a function can be formulated as a problem of feasibility in convex sets, provided of course the function itself is convex. We can however describe in more specificity how this should work.

Let now $f : \mathbb{R}^n \to \mathbb{R}$ be a function. A *subgradient oracle* to $f$ is an algorithm that, given $\overline{\mathbf{x}} \in \mathbb{Q}^n$, provides $f(\overline{\mathbf{x}})$ and a vector $\mathbf{h} \in \mathbb{Q}^n$, non-zero, with $\|\mathbf{h}\|_\infty = 1$ and

$$f(\mathbf{x}) \geq f(\overline{\mathbf{x}}) + \mathbf{h}^\mathsf{T}(\mathbf{x} - \overline{\mathbf{x}}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

(Convex functions are precisely those which admit a subgradient oracle.)

**Exercise 3.14.** If $f(\mathbf{x}) = \mathbf{c}^\mathsf{T}\mathbf{x}$, describe a subgradient oracle for $f$.

Consider then the following procedure.

(1) Start with an ellipsoid that contains the convex set $\mathcal{C}$ where you would like to optimize $f$. Run the ellipsoid method described above until the centre of the current ellipsoid belongs to $\mathcal{C}$.

(2) Once in there, query the subgradient oracle of $f$. It provides "the half" of $\mathcal{C}$ to which $f$ only decreases. Use this direction to find a small ellipsoid that contains only this half (using, of course, the ellipsoid method).

(3) If the centre is in $\mathcal{C}$, repeat the query to the subgradient oracle of $f$. Otherwise, query the separation oracle, until you find a point in $\mathcal{C}$. Repeat.

This method yields the following result.

**Theorem 3.15.** *There exists an algorithm $A$ so that, given*

- *a weak separation oracle for a convex set $\mathcal{C}$,*
- *a subgradient oracle for $f : \mathbb{R}^n \to \mathbb{R}$,*
- *$R$, $r$, $\varepsilon$ all in $\mathbb{Q}_{++}$,*

*so that $\tilde{c} + r\mathbb{B} \subseteq \mathcal{C} \subseteq R\mathbb{B}$ for some unknown $\tilde{c}$, $A$ returns $\overline{\mathbf{x}} \in \mathcal{C}$ so that*

$$f(\overline{\mathbf{x}}) \leq \inf_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x}) + \varepsilon$$

*after $O\big(n^2[\log(R/r) + \log(\mu_0/\varepsilon)]\big)$ iterations, where $\mu_0 = \varepsilon + \sup_{\mathbf{x}\in R\mathbb{B}} f(\mathbf{x}) - \inf_{\mathbf{x}\in R\mathbb{B}} f(\mathbf{x})$. Moreover, each iteration consists in one ellipsoid upgrade, as in Theorem 3.13, and one query to each oracle.* □

## 3.4 Application to SDPs

Consider the following primal dual pair of SDPs:

$$\begin{array}{cccc|cccc} & \min & \langle \mathbf{C}, \mathbf{X}\rangle & & & \max & \mathbf{b}^\mathsf{T}\mathbf{y} & \\ (\mathrm{P}) & \text{s.t.} & \mathcal{A}(\mathbf{X}) = \mathbf{b} & & (\mathrm{D}) & \text{s.t.} & \mathcal{A}^*(\mathbf{y}) \preccurlyeq \mathbf{C} & \\ & & \mathbf{X} \in \mathbb{S}^n_+ & & & & \mathbf{y} \in \mathbb{R}^m & \end{array}$$

Assume $\mathring{\mathbf{X}}$ is a Slater point to (P), meaning, a feasible solution belonging to $\mathbb{S}^n_{++}$, and that $\mathring{\mathbf{y}}$ is also a Slater point to (D), meaning, $\mathbf{C} - \mathcal{A}^*(\mathring{\mathbf{y}}) \in \mathbb{S}^n_{++}$.

**Exercise 3.16.** Show that (P) has the same set of optimal solutions as (P') below:

$$\begin{array}{rl} \min & \langle \mathbf{C}, \mathbf{X}\rangle \\ \text{subject to} & \mathcal{A}(\mathbf{X}) = \mathbf{b} \\ & \langle \mathbf{C} - \mathcal{A}^*(\mathring{\mathbf{y}}), \mathbf{X}\rangle \leq 2\langle \mathbf{C} - \mathcal{A}^*(\mathring{\mathbf{y}}), \mathring{\mathbf{X}}\rangle \\ & \mathbf{X} \in \mathbb{S}^n_+ \end{array}$$

Moreover, show that its feasible set is bounded.

To show the first assertion, recall that for all optimal solution $\widehat{\mathbf{X}}$, $\langle \mathbf{C}, \widehat{\mathbf{X}}\rangle \leq \langle \mathbf{C}, \mathring{\mathbf{X}}\rangle$. To show the second, consider the following. Given $\mathbf{X}$, matrix, we will denote the vector in $\mathbb{R}^n$ whose entries are the eigenvalues of $\mathbf{X}$ in non-increasing order by $\lambda^{\downarrow}(\mathbf{X})$. Likewise for the non-decreasing vector given by $\lambda^{\uparrow}(\mathbf{X})$.

**Exercise 3.17.** Show that, if $\mathbf{X}$ and $\mathbf{Y}$ are symmetric matrices, with eigenvalues ordered from largest to smallest, then

$$\lambda^{\downarrow}(\mathbf{X})^\mathsf{T}\lambda^{\uparrow}(\mathbf{Y}) \leq \langle \mathbf{X}, \mathbf{Y}\rangle \lambda^{\downarrow}(\mathbf{X})^\mathsf{T}\lambda^{\downarrow}(\mathbf{Y}).$$

As a consequence, it follows that

$$\begin{aligned} \langle \mathbf{C} - \mathcal{A}^*(\mathring{\mathbf{y}}), \mathbf{X}\rangle &\geq \lambda^{\uparrow}(\mathbf{C} - \mathcal{A}^*(\mathring{\mathbf{y}}))^\mathsf{T}\lambda^{\downarrow}(\mathbf{X}) \\ &\geq \lambda_{\min}(\mathbf{C} - \mathcal{A}^*(\mathring{\mathbf{y}}))\lambda_{\max}(\mathbf{X}). \end{aligned}$$

This should be all you need.

Now in addition, we can show that the feasible region of (P') contains a ball of radius $\lambda_n(\mathring{\mathbf{X}})$ around $\mathring{\mathbf{X}}$. Thus Theorem 3.15 applies. Moreover, $\mu_0$ as in the statement of the theorem can be bounded, as

$$\sup\{\langle \mathbf{C}, \mathbf{X}\rangle : \mathbf{X} \text{ feasible}\} - \inf\{\langle \mathbf{C}, \mathbf{X}\rangle : \mathbf{X} \text{ feasible}\} \leq \frac{4n\|\mathbf{C}\|_2 \langle \mathring{\mathbf{X}}, (\mathbf{C} - \mathcal{A}^*(\mathring{\mathbf{y}}))\rangle}{\lambda_n((\mathbf{C} - \mathcal{A}^*(\mathring{\mathbf{y}})))}.$$

## 3.5 Application of separation to combinatorics

Let $G$ be a given undirected graph, $\mathbf{c} \in \mathbb{R}^E$ costs on edges. The travelling salesman problem is that of finding a Hamiltonian cycle of minimum total cost. Given $U$, let $\delta(U)$ denote the subset of edges incident to precisely one vertex in $U$. One of the possible formulations for the TSP is the following

$$\begin{aligned} \min \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{subject to} \quad & \sum_{e \in \delta(u)} \mathrm{x}_e = 2 \quad \forall\, u \in V(G) \\ & \sum_{e \in \delta(U)} \mathrm{x}_e = 2 \quad \forall\, U \subset V(G),\ 2 \leq |U| \leq |V| - 2 \\ & \mathbf{0} \leq \mathbf{x} \leq \mathbf{1} \\ & \mathbf{x} \text{ integral.} \end{aligned}$$

The second kind of constraints prevent a solution which includes several disjoint cycles. It however consists of exponentially many restrictions, yielding a possibly intractable problem even when the integrality constraint is dropped. However, this is not the case.

To see this, consider the following procedure that decides whether or not a given solution $\mathbf{x}$ satisfying the first set of constraints satisfies also the second set. It creates a new graph, replacing edges by pairs of arcs in opposing directions. It then calculates whether or not there is a cut in the graph of weight smaller than 2 (for instance, by solving a max-flow routine). Therefore the feasible set of the linear relaxation of the IP above admits a weak separation oracle (that runs efficiently).

# 4 Maxcut

Let $G$ be a graph on $n$ vértices. Given $\emptyset \neq S \subset V(G)$, the *cut* $\delta(S)$ is the set of edges connecting $S$ to its complement.

We are interested in the problem of finding $S$ so that $\delta(S)$ is the largest possible.

Given $G$, let $\mu(G)$ denote the size of the *maximum cut* of $G$.

**Exercise 4.1.** Show that $\mu(G) \geq n/2$ by exhibiting a very simple greedy algorithm that constructs a cut of size at least $n/2$.

It is known that $\mu(G)$ cannot be computed in polynomial time unless P = NP. In fact, any approximation with constant factor better than 0.94 would imply P = NP.

For the remainder of this section, if $S \subseteq V(G)$, then $\mathbf{x}_S$ denotes the 01 characteristic vector of $S$ in $\mathbb{R}^V$, and $\mathbf{x}_{\delta(S)}$ the 01 characteristic vector of $\delta(S)$ in $\mathbb{R}^E$. We shall also assume that edges of the graph have been possibly weighted by a function $\mathbf{w} : E \to \mathbb{R}_+$.

## 4.1 LP-formulation

One very natural formulation for maxcut consists of

$$\begin{aligned} \max \quad & \frac{1}{2} \sum_{ij \in E(G)} \mathrm{w}_{ij}(1 - \mathrm{x}_i \mathrm{x}_j) \\ \text{subject to} \quad & \mathrm{x}_i \in \{-1, +1\}. \end{aligned}$$

The sign determines to which side of the cut the vertex is placed, and the objective function adds to the cut the edges which precisely cross the cut, ignoring the rest. This elegant formulation is of no direct practical use, as the integrality constraint and the quadratic objective function both pose typically intractable conditions.

It is possible to formulate maxcut with a standard linear integer program, as we see below.

Given a graph $G$ and an edge-weight function, we wish to find $S \subseteq V(G)$ so that $\mathbf{1}^\mathsf{T}\mathbf{x}_{\delta(S)}$ is largest possible. Note that

$$\begin{aligned} \max_{S \subseteq V(G)} \mathbf{w}^\mathsf{T}\mathbf{x}_{\delta(S)} = \max \ & \mathbf{w}^\mathsf{T}\mathbf{x} \\ & \text{s.t. for all triangles } e, f, g, \text{ we have} \\ & \mathrm{x}_e + \mathrm{x}_f + \mathrm{x}_g \leq 2 \\ & \mathrm{x}_e - \mathrm{x}_f - \mathrm{x}_g \leq 0 \\ & -\mathrm{x}_e + \mathrm{x}_f - \mathrm{x}_g \leq 0 \\ & -\mathrm{x}_e - \mathrm{x}_f + \mathrm{x}_g \leq 0 \\ & \mathbf{x} \in \{0, 1\}^E \end{aligned}$$

This formulation enforces that for all triangles of the graph, either 0 or 2 edges cross the cut. This is clearly necessary, and with not much further effort, can be seen to be also sufficient to recover the cut from the vector $\mathbf{x}$. The linear relaxation of this formulation can hence

be used to design approximation algorithms to solve the problem. However, a result due to Erdős shows that there are graphs to which the optimum of the relaxation is about twice the size of the integer program (thus, the greedy algorithm you found is about as good as this, yet quite simpler.)

## 4.2 SDP relaxation 1

We turn to the first formulation we introduced. The *Laplacian* matrix of a graph $G$ with edge weights is defined as

$$\mathbf{L} = \sum_{ij \in E} \mathrm{w}_{ij} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^\mathsf{T}$$

where the $\mathbf{e}_i$s are the characteristic vectors of the vertices, defined in $\mathbb{R}^V$. If $\mathbf{w} \geq \mathbf{0}$, then $\mathbf{L} \succcurlyeq \mathbf{0}$. (If $\mathbf{w}$ is constant equal to 1, then $\mathbf{L}$ is simply $\mathbf{D} - \mathbf{A}$, where $\mathbf{A}$ is the adjacency matrix and $\mathbf{D}$ is the diagonal matrix of vertex degrees).

From the definition of $\mathbf{L}$, it follows immediately that

$$\mathbf{x}_S^\mathsf{T} \, \mathbf{L} \, \mathbf{x}_S = \mathbf{w}^\mathsf{T} \mathbf{x}_{\delta(S)}.$$

Hence

$$\mu(G) = \max\{\mathbf{x}^\mathsf{T} \mathbf{L} \mathbf{x} : \mathbf{x} \in \{0,1\}^V\}.$$

Given $\mathbf{x} \in \{0,1\}^V$, we can define

$$\widehat{\mathbf{x}} = \begin{pmatrix} 1 & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & \mathbf{x}\mathbf{x}^\mathsf{T} \end{pmatrix}.$$

Also, have $\widehat{\mathbf{L}} = \begin{pmatrix} 0 & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & \mathbf{L} \end{pmatrix}$. Note that

$$\langle \mathbf{L}, \mathbf{x}\mathbf{x}^\mathsf{T} \rangle = \langle \widehat{\mathbf{L}}, \widehat{\mathbf{x}} \rangle.$$

Therefore we arrive at the relaxation

$$\begin{aligned} \mu(G) \leq \max \ & \langle \widehat{\mathbf{L}}, \mathbf{X} \rangle \\ \text{s.t. } & \mathbf{X}_{00} = 1 \\ & \mathrm{diag}(\mathbf{X}) = \mathbf{X}\mathbf{e}_0 \\ & \mathbf{X} \in \mathbb{S}_+^{\{0\} \cup V(G)}. \end{aligned}$$

**Exercise 4.2.** Show that equality holds if a rank = 1 constraint is added to the right hand side.

## 4.3 SDP relaxation 2

The formulation above does not resemble in a straightforward way the $\pm 1$ formulation we introduced in the beginning. Let us now show how they relate. To say that $\alpha \in \{0,1\}$ is the same as $2\alpha - 1 \in \{-1,+1\}$. The same, of course, applies to a vector of integer coordinates,

leading to an equivalence between 01 formulations and $\pm 1$ formulations of combinatorial problems.

To see how this translates to the correspondence $\mathbf{x} \to \widehat{\mathbf{X}}$, consider the matrix

$$\mathbf{Z} = \begin{pmatrix} 1 & \mathbf{0}^\mathsf{T} \\ -\mathbf{1} & 2\mathbf{I} \end{pmatrix}.$$

Note that

$$\mathbf{Z}\begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} 1 \\ 2\mathbf{x} - \mathbf{1}. \end{pmatrix},$$

therefore

$$\mathbf{Z}\begin{pmatrix} 1 & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & \mathbf{x}\mathbf{x}^\mathsf{T} \end{pmatrix}\mathbf{Z}^\mathsf{T} = \begin{pmatrix} 1 & (2\mathbf{x}-\mathbf{1})^\mathsf{T} \\ (2\mathbf{x}-\mathbf{1}) & (2\mathbf{x}-\mathbf{1})(2\mathbf{x}-\mathbf{1})^\mathsf{T} \end{pmatrix},$$

and this resulting matrix has constant diagonal (equal to 1). This generalizes to the following exercise.

**Exercise 4.3.** Let $\mathbf{X} \in \mathbb{S}^{\{0\}\cup V}$. Show that $\mathbf{X}_{00} = 1$ and $\operatorname{diag}\mathbf{X} = \mathbf{X}\mathbf{e}_0$ if, and only if, $\mathbf{Z}\mathbf{X}\mathbf{Z}^\mathsf{T}$ has constant diagonal equal to 1s.

As a consequence, we have the reformulation

$$\begin{array}{lllllll} \max & \langle \widehat{\mathbf{L}}, \mathbf{X}\rangle & = & \max & \langle \widehat{\mathbf{L}}, \mathbf{Z}^{-1}\mathbf{Y}\mathbf{Z}^{-T}\rangle & = & \max \quad \left\langle \frac{1}{4}\mathbf{L}, \mathbf{Y}\right\rangle \\ \text{s.t.} & \operatorname{diag}(\mathbf{X}) = \mathbf{X}\mathbf{e}_0 & & \text{s.t.} & \operatorname{diag}(\mathbf{Y}) = \mathbf{1} & & \text{s.t.} \quad \operatorname{diag}\mathbf{Y} = \mathbf{1} \\ & \mathbf{X}_{00} = 1 & & & & & \\ & \mathbf{X} \in \mathbb{S}_+^{\{0\}\cup V(G)} & & & \mathbf{Y} \in \mathbb{S}_+^{\{0\}\cup V(G)}, & & \mathbf{Y} \in \mathbb{S}_+^{V(G)}. \end{array}$$

(Note that from the second to the third we simply projected down the first coordinate.)

**Exercise 4.4.** Write the dual formulation for this third presentation. Verify that $\mathbf{I}$ is a Slater point for the primal, and find a Slater point for the dual.

Also, now focusing on the third formulation, observe that if the rank of $\mathbf{Y}$ is equal to 1, then $\mathbf{Y} = \mathbf{y}\mathbf{y}^\mathsf{T}$ to some $\mathbf{y}$, which due to the first constraint must be $\mathbf{y} \in \{-1, +1\}$. Hence we would also be able to recover the cut if a rank=1 type of constraint is added (and this would correspond precisely to the first formulation we introduced in this section about maxcut).

**Exercise 4.5.** Let $\mathbf{Y} = \mathbf{B}\mathbf{B}^\mathsf{T}$. Prove that

$$\left\langle \frac{1}{4}\mathbf{L}, \mathbf{Y}\right\rangle = \frac{1}{4}\sum_{ij\in E} \mathrm{w}_{ij}\|\mathbf{B}^\mathsf{T}\mathbf{e}_i - \mathbf{B}^\mathsf{T}\mathbf{e}_j\|^2.$$

To summarize, we have

$$\mu(G) \le \max\left\{\frac{1}{4}\langle \mathbf{L}, \mathbf{Y}\rangle : \operatorname{diag}\mathbf{Y} = \mathbf{1},\ \mathbf{Y} \succcurlyeq \mathbf{0}\right\}.$$

We now wish to verify how far this upper bound really is from $\mu(G)$.

## 4.4 Approximating maxcut

Assume $\mathbf{Y}$ is a solution to the SDP relaxation of maxcut. As it is positive semidefinite, there is $\mathbf{B}$ with $\mathbf{Y} = \mathbf{B}\mathbf{B}^\mathsf{T}$. Each of the $n$ rows of $\mathbf{B}$ is a vector in $\mathbb{R}^k$ for some $k$. Brings up the question: how to map these vectors to a cut?

Very broadly, the idea is to slice $\mathbb{R}^k$ in half, and map the vertices corresponding to rows of $\mathbf{B}$ that remain on one of the sides to $S$, and have $\delta(S)$ be a candidate maximum cut. The topic of this subsection is to decide how this procedure should work, and what guarantees we can obtain.

**Lemma 4.6.** *Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$, both of norm 1. Let $\mathbf{u} \in \mathbb{R}^n$ also of norm 1 be uniformly chosen. Thus*

$$Pr\big([\mathbf{u}^\mathsf{T}\mathbf{v}_1 \geq 0] \neq [\mathbf{u}^\mathsf{T}\mathbf{v}_2 \geq 0]\big) = \frac{\cos^{-1}\langle \mathbf{v}_1, \mathbf{v}_2\rangle}{\pi}.$$

*Proof sketch.* The cases where $\langle \mathbf{v}_1, \mathbf{v}_2\rangle = \pm 1$ are trivial. Assume otherwise.

Let $\mathbf{P}$ be the orthogonal projection onto the span of $\mathbf{v}_1$ and $\mathbf{v}_2$. Note that

$$\mathbf{v}_i\mathbf{u} = \|\mathbf{P}\mathbf{u}\|\frac{\mathbf{v}_i^\mathsf{T}\mathbf{P}\mathbf{u}}{\|\mathbf{P}\mathbf{u}\|}.$$

The distribution of $\mathbf{P}\mathbf{u}/\|\mathbf{P}\mathbf{u}\|$ is rotationally invariant in the unit sphere of the image of $\mathbf{P}$. Thus, if $\theta$ is the angle between $\mathbf{v}_1$ and $\mathbf{v}_2$, we have

$$\Pr\big([\mathbf{u}^\mathsf{T}\mathbf{v}_1 \geq 0] \neq [\mathbf{u}^\mathsf{T}\mathbf{v}_2 \geq 0]\big) = \frac{\theta}{\pi}.$$

□

The following lemma is an immediate calculation.

**Lemma 4.7.** *Have $\rho = 0.878567$. With $\alpha \in [-1, 1]$, then*

$$\frac{\cos^{-1}(\alpha)}{\pi} \geq \frac{\rho}{2}(1 - \alpha).$$

□

Now we are ready to show Goemans & Williamson result, from 1995.

**Theorem 4.8.** *Let $G = (V, E)$ be a graph, $\mathbf{w} : E \to \mathbb{R}_+$, and $\mathbf{L}$ the Laplacian of $G$ with respect to $\mathbf{w}$. Let $\mathbf{Y} \in \mathbb{S}^V_+$ with $\operatorname{diag}\mathbf{Y} = \mathbf{1}$, and $\mathbf{B}$ with $\mathbf{Y} = \mathbf{B}\mathbf{B}^\mathsf{T}$. Let $\mathbf{v}_i = \mathbf{B}^\mathsf{T}\mathbf{e}_i$, for all $i \in V$. Let $\mathbf{u}$ be uniformly sampled in the unit sphere of the space where the $\mathbf{v}_i$s live. Define $S$ to be the vertices where $\mathbf{u}^\mathsf{T}\mathbf{v}_i \geq 0$, and let $\mathrm{y}_i = +1$ if $i \in S$, $-1$ otherwise. Hence*

$$\mathbb{E}\left[\frac{1}{4}\mathbf{y}^\mathsf{T}\mathbf{L}\mathbf{y}\right] \geq \rho\left\langle\frac{1}{4}\mathbf{L}, \mathbf{Y}\right\rangle.$$

*Proof.*

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{4}\mathbf{y}^\mathsf{T}\mathbf{L}\mathbf{y}\right] &= \left\langle \frac{1}{4}\mathbf{L}, \mathbb{E}\left[\mathbf{y}\mathbf{y}^\mathsf{T}\right]\right\rangle \\
&= \frac{1}{4}\sum_{ij\in E} \mathrm{w}_{ij}\mathbb{E}[\|\mathrm{y}_i - \mathrm{y}_j\|^2] \\
&= \frac{1}{4}\sum_{ij\in E} \mathrm{w}_{ij}\ 4\ \mathtt{Pr}[\mathrm{y}_i \neq \mathrm{y}_j] \\
&= \sum_{ij\in E} \mathrm{w}_{ij}\ \frac{\cos^{-1}\langle \mathbf{v}_i, \mathbf{v}_j\rangle}{\pi} \\
&\geq \sum_{ij\in E} \mathrm{w}_{ij}\ \frac{\rho}{2}(1 - \langle \mathbf{v}_i, \mathbf{v}_j\rangle) \\
&= \rho\frac{1}{2}\sum_{ij\in E} \mathrm{w}_{ij}\ \|\mathbf{v}_i - \mathbf{v}_j\|^2 \\
&= \rho\left\langle \frac{1}{4}\mathbf{L}, \mathbf{Y}\right\rangle.
\end{aligned}$$

□

As a consequence, we can proceed as follows.

1. Solve the SDP. Obtain $\mathbf{Y} = \mathbf{B}\mathbf{B}^\mathsf{T}$. Define $\mathbf{v}_i = \mathbf{B}^\mathsf{T}\mathbf{e}_i$.
2. Sample $\mathbf{u}$. Define the cut (obtaining $\mathrm{y}_i = \pm 1$). Record the value $\mathbf{y}^\mathsf{T}\mathbf{L}\mathbf{y}$. Repeat as much as desirable, or until it exceeds $\rho\langle \mathbf{L}, \mathbf{Y}\rangle$, and save the $\mathbf{y}$ giving the largest.
3. With high probability in a relatively short time, we have a $\mathbf{y}$ so that
$$\frac{1}{4}\mathbf{y}^\mathsf{T}\mathbf{L}\mathbf{y} \geq \rho\left\langle \frac{1}{4}\mathbf{L}, \mathbf{Y}\right\rangle.$$
4. As the SDP was a maxcut relaxation, we know that
$$\left\langle \frac{1}{4}\mathbf{L}, \mathbf{Y}\right\rangle \geq \mu(G).$$
5. Thus, the cut $\delta(S)$ defined by $\mathbf{y}$ is so that
$$\mathbf{w}^\mathsf{T}\mathbf{x}_{\delta(S)} = \frac{1}{4}\mathbf{y}^\mathsf{T}\mathbf{L}\mathbf{y} \geq \rho\cdot\mu(G).$$

## 4.5 Further aspects

We now discuss a few extra aspects. First, how to sample $\mathbf{u}$ uniformly in the sphere? To that, it is well known that if $n$ numbers are sampled independently and uniformly at random according to the standard normal distribution with mean zero and variance one, which has probability density

$$f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2},$$

thus forming the vector $\mathbf{x} \in \mathbb{R}^n$, then $\mathbf{x}/\|\mathbf{x}\|$ is distributed according to the rotationally invariant probability measure on the unit sphere.

Also, we can provide a quite remarkable alternative formulation. Let $\mathbf{X}^{\circ \arcsin}$ denote the matrix obtained from $\mathbf{X}$ upon applying the arcsin function to each of its coordinates.

**Corollary 4.9.**

$$\begin{aligned} \mu(G) = \max \ & \frac{2}{\pi} \left\langle \frac{1}{4}\mathbf{L}, \mathbf{X}^{\circ \arcsin} \right\rangle \\ s.t. \ & \operatorname{diag}(\mathbf{X}) = \mathbf{1} \\ & \mathbf{X} \succcurlyeq \mathbf{0}. \end{aligned}$$

*Proof.* Assume $\mathbf{y} \in \mathbb{R}^n$ is the $\pm 1$ denoting a maxcut, hence $\mu = \frac{1}{4}\mathbf{y}^\mathsf{T}\mathbf{L}\mathbf{y}$. Naturally $\mathbf{X} = \mathbf{y}\mathbf{y}^\mathsf{T}$ is feasible to the program above, and its objective value is equal to $\mu$.

On the other hand, let $\mathbf{X} = \mathbf{U}\mathbf{U}^\mathsf{T}$ be an optimal solution, and let $\mathbf{v}_i = \mathbf{U}^\mathsf{T}\mathbf{e}_i$. Upon choosing $\mathbf{u}$ uniformly at random in the sphere and defining $\mathbf{y}$ according to the sign of $\mathbf{u}^\mathsf{T}\mathbf{v}_i$, we can observe, just as above, that

$$\mathbb{E}[\mathbf{y}\mathbf{y}^\mathsf{T}] = 1 - 2\mathtt{Pr}[\mathrm{y}_i \neq \mathrm{y}_j] = \frac{2}{\pi} \arcsin(\mathbf{v}_i^\mathsf{T}\mathbf{v}_j),$$

and this is the objective value of $\mathbf{X}$. Therefore, there must be $\mathbf{y}$ attaining this value. □

With that, you can prove the following theorem due to Nesterov (by now, we already know this is not the best approximation constant).

**Theorem 4.10.** *If* $\mathbf{L} \succcurlyeq \mathbf{0}$*, then*

$$\max_{\mathbf{x} \in \{-1,+1\}^n} \mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x} \geq \frac{2}{\pi} \max\{\langle \mathbf{L}, \mathbf{X} \rangle : \operatorname{diag}(\mathbf{X}) = \mathbf{1}, \ \mathbf{X} \succcurlyeq \mathbf{0}\}.$$

**Exercise 4.11.** Prove it. You will have to use the Taylor series of arcsin, and the fact that if $\mathbf{M}$ and $\mathbf{N}$ are positive semidefinite, then so is $\mathbf{M} \circ \mathbf{N}$ (this is known as Schur's theorem; research it!).

The cut polytope is defined as

$$\mathbb{P} = \operatorname{conv}\{\mathbf{x}\mathbf{x}^\mathsf{T} : \mathbf{x} \in \{-1,+1\}^n\},$$

As usual, one wishes to provide cuts which are valid for $\mathbb{P}$. If $\mathbf{X}$ is the variable of the SDP, then it is defined by a cut if it satisfies

$$\mathbf{b}^\mathsf{T}\mathbf{X}\mathbf{b} \geq \min\{\mathbf{b}^\mathsf{T}\mathbf{x}\mathbf{x}^\mathsf{T}\mathbf{b} : \mathbf{x} \in \{-1,+1\}^n\} = \min\{(\mathbf{b}^\mathsf{T}\mathbf{x})^2 : \mathbf{x} \in \{-1,+1\}^n\},$$

with $\mathbf{b} \in \mathbb{Z}^n$. As $\mathbf{b}^\mathsf{T}\mathbf{X}\mathbf{b} \geq 0$ for all $\mathbf{b}$, we do wish to find vectors $\mathbf{b}$ so that the right hand side is positive. If $\mathbf{b} \in \{-1, 0, +1\}^n$ consists of an odd number of non-zero entries, the equalities defined are called clique inequalities. By picking $\mathbf{b}$ with only three non-zero entries, we define the triangle inequalities we had originally in the beginning:

$$\begin{aligned} x_{ij} + x_{ik} + x_{jk} &\geq -1, \\ x_{ij} - x_{ik} - x_{jk} &\geq -1, \\ -x_{ij} + x_{ik} - x_{jk} &\geq -1, \\ -x_{ij} - x_{ik} + x_{jk} &\geq -1. \end{aligned}$$

They can, therefore, be added to the SDP formulation. The polytope defined by these inequalities is called the *metric polytope*. It is exact for graphs not contractible to $K_5$ (in particular planar graphs). As a consequence, the SDP is tight for $C_5$ provided these inequalities are added. Otherwise, the gap there is around 0.88. With triangle inequalities added, the worst known case has a maxcut to relaxation ratio of 0.95. We should point however that the performance of the randomized algorithm does not seem to improve in general, as shown by Karloff. Finally, Mahajan and Ramesh showed how to de-randomize to procedure we have been discussing, but the details are perhaps too hairy for these notes. More details about all of these discussions above can be found in C. Helmberg's SDP notes.

## 4.6 Eigenvalues

There is a clear connection between maximum cuts and eigenvalues of $\mathbf{L}$. Consider the approximation

$$\begin{aligned} \mu(G) \leq \max \ & \left\langle \frac{1}{4}\mathbf{L}, \mathbf{X} \right\rangle \\ \text{s.t.} \ & \operatorname{diag}(\mathbf{X}) = \mathbf{1} \\ & \mathbf{X} \in \mathbb{S}^V_+. \end{aligned}$$

As a consequence, the trace of $\mathbf{L}$ is constant (equal to $n$), and, using the SDP formulation for the maximum eigenvalue of a matrix, we obtain

$$\begin{array}{rccccccccc} & \mu(G) & \leq & \max & \left\langle \frac{1}{4}\mathbf{L}, \mathbf{X} \right\rangle & \leq & \max & \left\langle \frac{1}{4}\mathbf{L}, n\mathbf{X} \right\rangle & = & \frac{n}{4}\lambda_{\max}(\mathbf{L}). \\ vec & & & \text{s.t.} & \operatorname{diag}(\mathbf{X}) = \mathbf{1} & & \text{s.t.} & \operatorname{tr}\mathbf{X} = 1 & & \\ & & & & \mathbf{X} \in \mathbb{S}^V_+, & & & \mathbf{X} \in \mathbb{S}^V_+, & & \end{array}$$

(And this is, of course, irregardless of any weighting in $\mathbf{L}$.)

Perhaps more interestingly, we can write the duals of the programs above, having

$$\begin{array}{rrl|rrl} & \max & \langle \mathbf{L}, \mathbf{X} \rangle & & \min & \mathbf{1}^\mathsf{T}\mathbf{u} \\ (\text{P}) & \text{s.t.} & \operatorname{diag}(\mathbf{X}) = \mathbf{1} & (\text{D}) & \text{s.t.} & \mathbf{L} + \mathbf{Z} - \operatorname{Diag}(\mathbf{u}) = \mathbf{0} \\ & & \mathbf{X} \succcurlyeq \mathbf{0} & & & \mathbf{Z} \succcurlyeq \mathbf{0}. \end{array}$$

(Omitting (1/4) for now). Have $\overline{\mathbf{u}}$ be a projection of $\mathbf{u}$ onto the orthogonal complement of $\mathbf{u}$, thus having

$$\lambda \mathbf{1} + \overline{\mathbf{u}} = \mathbf{u},$$

and therefore $\lambda \mathbf{I} + \mathrm{Diag}(\overline{\mathbf{u}}) = \mathrm{Diag}(\mathbf{u})$. Then (D) is equivalent to

$$\begin{aligned} \min \quad & n\lambda \\ \text{subject to} \quad & \lambda \mathbf{I} - (\mathbf{L} + \mathrm{Diag}(\overline{\mathbf{u}})) \succcurlyeq \mathbf{0} \\ & \mathbf{1}^\intercal \overline{\mathbf{u}} = 0. \end{aligned}$$

which is, of course, equivalent to

$$\min_{\mathbf{v} \in \mathbb{R}^n : \mathbf{1}^\intercal \mathbf{v} = 0} \ \frac{n}{4} \lambda_{\max}(\mathbf{L} + \mathrm{Diag}(\mathbf{v}))$$

This result allows us to present an interesting application. Let $G_{n,p}$ denote a random graph on $n$ vertices where each possible edge has been included with probability $p$.

**Theorem 4.12** (Juhász)**.** *For graphs $G_{n,p}$ with $0 < p < 1$ and any $\varepsilon > 0$, the eigenvalues of $\mathbf{A} = \mathbf{A}(G_{n,p})$ satisfy*

$$\lambda_{max}(\mathbf{A}) = pn + o(n^{(1/2)+\varepsilon}) \quad \textit{and} \quad \max_{i \ not \ max} |\lambda_i(\mathbf{A})| = o(n^{(1/2)+\varepsilon})$$

*with probability going to 1 as $n$ goes to infinity.*

In other words, as $n$ grows, the maximum eigenvalue is roughly the size of the expected degree, and all other eigenvalues are small in comparison. This is enough to show that for most graphs, the bound provided by the SDP is tight compared to the size of the maxcut.

**Theorem 4.13** (Delorne and Poljak)**.** *For a fixed $p$, $0 < p < 1$, and $G = G_{n,p}$,*

$$\lim_{n \to \infty} \frac{\min_{\mathbf{1}^\intercal \mathbf{v} = 0} \ \frac{n}{4} \lambda_{max}(\mathbf{L} + \mathrm{Diag}(\mathbf{v}))}{\mu(G)} = 1.$$

*Proof.* If $d$ is the average degree of $G$, we have learned that

$$\mu(G) \geq \frac{m}{2} = \frac{nd}{4}.$$

Now we choose $\mathbf{v} = d\mathbf{1} - \mathbf{A}\mathbf{1}$. Then $\mathbf{L} + \mathrm{Diag}(\mathbf{v}) = d\mathbf{I} - \mathbf{A}$, and, therefore

$$\min_{\mathbf{1}^\intercal \mathbf{v} = 0} \ \frac{n}{4} \lambda_{\max}(\mathbf{L} + \mathrm{Diag}(\mathbf{v})) \leq \frac{n}{4}(d + o(n^{(1/2)+\varepsilon})).$$

As a consequence,

$$\lim_{n \to \infty} \frac{\min_{\mathbf{1}^\intercal \mathbf{v} = 0} \ \frac{n}{4} \lambda_{\max}(\mathbf{L} + \mathrm{Diag}(\mathbf{v}))}{\mu(G)} = \lim_{n \to \infty} \frac{\frac{n}{4}(d + o(n^{(1/2)+\varepsilon}))}{\frac{nd}{4}} = 1.$$

□

**Exercise 4.14.** The max weight bissection problem asks for a cut $\delta(S)$ of maximum weight requiring that $|S| = n/2$.

(a) Come up with an equality that should be added to the $\pm 1$ formulation of maxcut that does the deed.

(b) Come up with a strong equality of the form $\langle ?, \mathbf{X} \rangle$ to be added to the SDP relaxation.

(c) Write its dual.

**Exercise 4.15.** The max $k$-cut problem asks for a partition of $V$ into $k$ parts that maximizes the weights of the edges between them. Instead of assigning values $\pm 1$ to each vertex, one could assign vectors in $\mathbb{R}^{k-1}$ so that $\mathbf{v}_i^\mathsf{T}\mathbf{v}_j = -1/(k-1)$ if $i \neq j$.

(a) Prove that such vectors exist, to any chosen partition into $k$ sets.

(b) Write the analogous of the $\pm 1$ formulation.

(c) Write the SDP relaxation.

# 5 Cocliques and colourings

In this section, we focus on perhaps the most standard application of semidefinite program to graph theory. We have already seen a couple times $\vartheta(G)$ in these notes, but this time we intend to provide a thorough presentation.

Let $G = (V, E)$ be a graph on $n$ vertices. A *coclique* (or stable set, or independent set) $S$ is a subset of $V$ that induces no edge from $E(G)$. A *clique* is the complement of a clique. A *colouring* of $G$ is a partition of $V$ into cocliques (having a colour given to each one of them). A *clique partition* is a partition of $G$ into cliques. With this:

(i) $\alpha(G)$ is the size of the largest coclique in $G$, called the *independence number* or *coclique number*.

(ii) $\omega(G)$ is the size of the largest clique, called the *clique number*.

Note that $\alpha(G) = \omega(\overline{G})$.

(iii) $\chi(G)$ is the minimum number of colours needed to colour $V(G)$ with cocliques.

(iv) $\theta(G)$ is the minimum number of cliques needed to partition $V(G)$.

Note that $\theta(G) = \chi(\overline{G})$.

## 5.1 Linear polytopes

Assume $\mathbf{M}$ has its rows indexed by vertices, and columns by cliques, and is so that entry $(i, j) = 1$ if vertex $i$ belongs to clique $j$, and $= 0$ otherwise. Likewise, $\overline{\mathbf{M}}$ is the incidence matrix of vertices and cocliques, and, finally, $\mathbf{N}$ the incidence matrix of vertices and edges in $G$. One can define weighted versions of the parameters defined above.

- $\alpha(G; \mathbf{w}) = \max\{\mathbf{w}^\mathsf{T}\mathbf{x} : \exists \mathbf{y} \text{ with } \overline{\mathbf{M}}\mathbf{y} = \mathbf{x},\ \mathbf{0} \le \mathbf{y} \le \mathbf{1},\ \mathbf{1}^\mathsf{T}\mathbf{y} = 1,\ \mathbf{y} \text{ integral}\}$. Equivalently, $\mathbf{x}$ is a column of $\overline{\mathbf{M}}$. In particular $\alpha(G) = \alpha(G; \mathbf{1})$. More importantly, note that the constraint "$\mathbf{y}$ integral" can be dropped, as the vertices of this polyhedron correspond to the cocliques. Thus, if we define

$$\texttt{STAB}(G) = \{\overline{\mathbf{M}}\mathbf{y} : \mathbf{y} \ge \mathbf{0} \text{ and } \mathbf{1}^\mathsf{T}\mathbf{y} = 1\},$$

then

$$\alpha(G; \mathbf{w}) = \max\{\mathbf{w}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \texttt{STAB}(G)\}.$$

I should remind you that it is NP-hard to approximate $\alpha(G)$ to within $n^{1-\varepsilon}$ of the optimum, for all $\varepsilon$, hence it is very hard to describe $\texttt{STAB}(G)$.

We wish then to provide useful relaxations of $\texttt{STAB}(G)$. The first:

$$\texttt{FRAC}(G) = \{\mathbf{x} \in \mathbb{R}^V_+ : \mathbf{N}^\mathsf{T}\mathbf{x} \le \mathbf{1} \text{ and } \mathbf{x} \le \mathbf{1}.\}$$

**Exercise 5.1.** Show that

$$\texttt{STAB}(G) = \text{conv}\left(\texttt{FRAC}(G) \cap \{0, 1\}^V\right).$$

**Exercise 5.2.** Show that $\alpha(K_n) = 1$, however

$$\max\{\mathbf{1}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \texttt{FRAC}(K_n)\} = \frac{n}{2}.$$

It is possible to strengthen the formulation of $\texttt{FRAC}(G)$ with tractable combinatorial inequalities (eg. odd cycles), but that is not our goal here. Instead, we introduce yet another intractable formulation, but with the goal to motivate the use of semidefinite programming. Recall $\mathbf{M}$ is the incidence matrix of vertices vs cliques. Define

$$\texttt{QSTAB}(G) = \{\mathbf{x} \in \mathbb{R}^V_+ : \mathbf{M}^\mathsf{T}\mathbf{x} \leq \mathbf{1}\}.$$

**Exercise 5.3.** Argue why $\mathbf{M}^\mathsf{T}\ \overline{\mathbf{M}} \leq \mathbf{1}$.

With the exercise above, it is hence immediate to verify that

$$\texttt{STAB}(G) \subseteq \texttt{QSTAB}(G) \subseteq \texttt{FRAC}(G).$$

**Exercise 5.4.** Show that equality holds above everywhere if $G$ is bipartite. Is this an "if and only if" ?

With the polytope QSTAB, we can provide a good description of $\chi$.

- $\chi^*(G; \mathbf{w}) = \max\{\mathbf{w}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \texttt{QSTAB}(\overline{G})\}$. From LP duality, it follows that

  $$\chi^*(G; \mathbf{w}) = \min\{\mathbf{1}^\mathsf{T}\mathbf{y} : \overline{\mathbf{M}}\mathbf{y} \geq \mathbf{w} \text{ and } \mathbf{y} \geq \mathbf{0}.\}.$$

  Upon adding integrality constraints to the program above, we have just defined $\chi(G; \mathbf{w})$. The *fractional chromatic number* of $G$ is defined as $\chi^*(G) = \chi^*(G; \mathbf{w})$. As a consequence of $\texttt{STAB} \subseteq \texttt{QSTAB}$, it follows that

  $$\omega(G) \leq \chi^*(G) \leq \chi(G).$$

Our goal now is to study what we can do between STAB and QSTAB.

## 5.2 Semidefinite relaxation

Let $\mathbf{x}_S$ be the characteristic vector of a coclique $S$. Define

$$\widehat{\mathbf{x}_S} = \begin{pmatrix} 1 & \mathbf{x}_S^\mathsf{T} \\ \mathbf{x}_S & \mathbf{X}_S \end{pmatrix},$$

where $\mathbf{X}_S = \mathbf{x}_S\mathbf{x}_S^\mathsf{T}$. Three properties unfold:

(1) $\widehat{\mathbf{x}_S} \in \mathbb{S}_+^{\{0\}\cup V}$.

(2) $\widehat{\mathbf{x}_S}\mathbf{e}_0 = \operatorname{diag}\widehat{\mathbf{x}_S}$, ie, $\mathbf{x}_S = \operatorname{diag}(\mathbf{X}_S)$.

(3) If $ij \in E(G)$, then, as $S$ is a coclique, we have $(\mathbf{X}_S)_{ij} = 0$.

This motivate the following two definitions.

(a) $\widehat{\texttt{TH}}(G) = \left\{ \begin{pmatrix} 1 & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \succcurlyeq \mathbf{0} : \text{diag}(\mathbf{X}) = \mathbf{x},\ \mathrm{X}_{ij} = 0 \ \forall\ ij \in E(G) \right\}.$

(b) $\texttt{TH}(G) = \left\{ \mathbf{x} \in \mathbb{R}^V : \exists \mathbf{X} \in \mathbb{S}^V \text{ so that } \begin{pmatrix} 1 & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \in \widehat{\texttt{TH}}(G) \right\}.$

This $\texttt{TH}(G)$ is known as the *theta body* of $G$.

**Exercise 5.5.** Show that both $\texttt{TH}(G)$ and $\widehat{\texttt{TH}}(G)$ are convex sets. As a consequence, show that

$$\texttt{STAB}(G) \subseteq \texttt{TH}(G).$$

We can now define a new function, the *theta function* of $G$.

- $\vartheta(G; \mathbf{w}) = \max\{\mathbf{w}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \texttt{TH}(G)\}$. As expected, define $\vartheta(G) = \vartheta(G; \mathbf{1})$.

Immediately,

$$\alpha(G; \mathbf{w}) \leq \vartheta(G; \mathbf{w}) \quad \text{for all } \mathbf{w} \geq \mathbf{0}.$$

Also, we can show a non-trivial result.

**Theorem 5.6** (Grötschel, Lovász, Schrijver)**.** *For any graph $G$,*

$$\textit{TH}(G) \subseteq \textit{QSTAB}(G).$$

*Proof.* Let $\mathbf{x} \in \texttt{TH}(G)$, and $\mathbf{X}$ so that $\begin{pmatrix} 1 & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \in \widehat{\texttt{TH}}(G)$. Let $\mathbf{z}$ be the characteristic vector of a clique in $G$. Then

$$0 \leq \left\langle \begin{pmatrix} 1 & -\mathbf{z}^\mathsf{T} \\ -\mathbf{z} & \mathbf{z}\mathbf{z}^\mathsf{T} \end{pmatrix}, \begin{pmatrix} 1 & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \right\rangle = 1 - 2\mathbf{z}^\mathsf{T}\mathbf{x} + \langle \mathbf{z}\mathbf{z}^\mathsf{T}, \mathbf{X} \rangle.$$

From the definition of $\mathbf{X}$ (and $\widehat{\texttt{TH}}$), it follows that

$$\langle \mathbf{z}\mathbf{z}^\mathsf{T}, \mathbf{X} \rangle = \mathbf{z}^\mathsf{T}\mathbf{x}.$$

Thus,

$$0 \leq 1 - \mathbf{z}^\mathsf{T}\mathbf{x} \implies \mathbf{x} \in \texttt{QSTAB}(G).$$

□

**Corollary 5.7.**

$$\vartheta(\overline{G}) \leq \chi^*(G).$$

Summarizing:

$$\texttt{STAB}(G) \subseteq \texttt{TH}(G) \subseteq \texttt{QSTAB}(G) \subseteq \texttt{FRAC}(G)$$

and

$$\alpha(G; \mathbf{w}) \leq \vartheta(G; \mathbf{w}) \leq \chi^*(G; \mathbf{w}) \leq \chi(G; \mathbf{w}).$$

(This last one known as the *sandwich theorem*).

## 5.3 $\vartheta$ definitions

Recall that an *orthonormal representation of a graph $G$* (with its corresponding handle) is a function $\rho : \{0\} \cup V(G) \to \mathbb{V}$ so that

(i) $\rho_i$ is a unit vector for all $i \in \{0\} \cup V(G)$.

(ii) $\langle \rho_i, \rho_j \rangle = 0$ for all $ij \in \overline{E}$.

**Exercise 5.8.** Assume $G$ has been properly coloured, and that is provided by means of a partition $\mathbb{P}$ of $V(G)$. For $a \in V(G)$, $\rho_a$ to be the vector in $\mathbb{R}^{\mathbb{P}}$ which is 1 at the class containing $a$, 0 elsewhere. Have $\rho_0$ be anything. Show that $\rho$ is a orthonormal representation of $\overline{G}$.

**Theorem 5.9.** *Let $G$ be a graph, and $\widehat{\mathbf{X}} \in \mathbb{S}_+^{\{0\}\cup V(G)}$. The following are equivalent.*

*(i)* $\widehat{\mathbf{X}} \in \widehat{TH}(G)$.

*(ii) There is an orthonormal representation $\rho$ of $\overline{G}$ with*

$$\widehat{\mathbf{X}}_{ij} = \langle \rho_0, \rho_i \rangle \langle \rho_i, \rho_j \rangle \langle \rho_j, \rho_0 \rangle$$

*for all $i, j$ in $\{0\} \cup V$.*

*Proof.* We start with (ii) implying (i). We start with an orthonormal representation $\rho$, and simply define $\widehat{\mathbf{X}}$ as in (ii). First, $\widehat{\mathbf{X}}_{00} = 1$. Second,

$$\widehat{\mathbf{X}}_{0i} = \langle \rho_0, \rho_i \rangle^2 = \widehat{\mathbf{X}}_{ii},$$

then $\widehat{\mathbf{X}}\mathbf{e}_0 = \operatorname{diag} \widehat{\mathbf{X}}$. Third, if $ij \in E$, then $\widehat{\mathbf{X}}_{ij} = 0$ because $\langle \rho_i, \rho_j \rangle = 0$. Finally, $\widehat{\mathbf{X}}$ is the Gram matrix of vectors $\{\langle \rho_0, \rho_i \rangle \rho_i\}_{i \in \{0\}\cup V(G)}$. Thus $\widehat{\mathbf{X}} \succcurlyeq \mathbf{0}$, as we wanted.

Now assume (i). Take $\mathbf{z}_i$ to be the columns of $\widehat{\mathbf{X}}^{1/2}$. Let $S \subseteq V(G)$ indexing the vertices $i$ so that $\mathbf{z}_i \neq \mathbf{0}$. Define $\rho$ as follows:

(a) $\rho_i = \mathbf{z}_i$ normalized, if $i \in S$.

(b) complete the remaining $\rho_j$, for $j \notin S$, with any orthonormal basis of the orthogonal complement of the already defined $\rho_i$s.

From construction, if $i$ or $j$ are not in $S$, then $\langle \rho_i, \rho_j \rangle = 0$. If both are in $S$, and $ij \in E$, then

$$\langle \rho_i, \rho_j \rangle = \widehat{\mathbf{X}}_{ij} = 0.$$

Thus $\rho$ is orthonormal representation of $G$. Now, note that $\mathbf{z}_i = \|\mathbf{z}_i\| \rho_i$. As $\widehat{\mathbf{X}}$ is the Gram matrix of the $\mathbf{z}_i$s, it remains to show that $\|\mathbf{z}_i\| = \langle \rho_0, \rho_i \rangle$. If $i = 0$, it is immediate, as $\widehat{\mathbf{X}}_{00} = 1$. If $i \notin S$, $\|\mathbf{z}_i\| = 0$ and $\langle \rho_i, \rho_0 \rangle = 0$. If $i \in S$, then

$$\|\mathbf{z}_i\|^2 = \|\mathbf{z}_i\|^2 \langle \rho_i, \rho_i \rangle = \langle \rho_i, \rho_i \rangle = \widehat{\mathbf{X}}_{ii} = \widehat{\mathbf{X}}_{0i} = \langle \rho_0, \rho_i \rangle = \|\mathbf{z}_0\| \|\mathbf{z}_i\| \langle \rho_0, \rho_i \rangle.$$

Thus $\|\mathbf{z}_i\| = \langle \rho_0, \rho_i \rangle$, as wished. □

**Exercise 5.10.** Let $G$ be a graph, $\mathbf{x} \in \mathbb{R}^V$. Show that $\mathbf{x} \in \mathtt{TH}(G)$ if, and only if, there is orthonormal representation $\rho$ of $\overline{G}$ with $\mathrm{x}_i = \langle \rho_0, \rho_i \rangle^2$ for all $i \in V$.

Now we can define

$$\vartheta_4(G; \mathbf{w}) = \max_{\rho \text{ o.r. of } \overline{G}} \sum_{i \in V} \mathrm{w}_i \langle \rho_0, \rho_i \rangle^2.$$

As we have shown above, we have that, for all graphs $G$ and $\mathbf{w} \geq \mathbf{0}$,

$$\vartheta(G; \mathbf{w}) = \vartheta_4(G; \mathbf{w}).$$

**Lemma 5.11.** *Let $V = V(G)$. Let $\rho : \{0\} \cup V \to \mathbb{R}^p$ be an orthonormal representation of $G$, and $\sigma : \{0\} \cup V \to \mathbb{R}^q$ an orthonormal representation of $\overline{G}$. Then, for all $i, j \in \{0\} \cup V$, $i \neq j$, we have*

$$\langle \rho_i \sigma_i^\mathsf{T}, \rho_i \sigma_i^\mathsf{T} \rangle = 1 \quad \textit{and} \quad \langle \rho_i \sigma_i^\mathsf{T}, \rho_j \sigma_j^\mathsf{T} \rangle = 0.$$

*Proof.* Follows immediately from noting that

$$\langle \rho_i \sigma_i^\mathsf{T}, \rho_j \sigma_j^\mathsf{T} \rangle = \operatorname{tr} \rho_i \sigma_i^\mathsf{T} \sigma_j \rho_j^\mathsf{T} = \langle \rho_i, \rho_j \rangle \langle \sigma_i, \sigma_j \rangle,$$

and at least one of these must be equal to 0 if $i \neq j$, and both are equal to 1 if $i = j$. □

**Theorem 5.12.** *Let $G$ be a graph. If $\mathbf{x} \in \mathtt{TH}(G)$ and $\mathbf{y} \in \mathtt{TH}(\overline{G})$, then $\mathbf{x}^\mathsf{T}\mathbf{y} \leq 1$.*

*Proof.* From Theorem 5.9, there are orthonormal representations $\rho$ and $\sigma$ of $G$ and $\overline{G}$ so that

$$\mathrm{x}_i = \langle \rho_0, \rho_i \rangle^2 \quad \text{and} \quad \mathrm{y}_i = \langle \sigma_0, \sigma_i \rangle^2.$$

Let $\mathbf{W}_i = \rho_i \sigma_i^\mathsf{T}$. We have $\{\mathbf{W}_i : i \in V(G)\}$ orthonormal, which we can complete to a basis $\{\mathbf{W}_i : i \in W\}$ of the space $\mathbb{R}^{p \times q}$. As a consequence,

$$\begin{aligned}
\mathbf{x}^\mathsf{T}\mathbf{y} &= \sum_{i \in V} \langle \rho_0, \rho_i \rangle^2 \langle \sigma_0, \sigma_i \rangle^2 \\
&= \sum_{i \in V} \langle \rho_0 \sigma_0^\mathsf{T}, \rho_i \sigma_i^\mathsf{T} \rangle^2 \\
&\leq \sum_{i \in W} \langle \mathbf{W}_0, \mathbf{W}_i \rangle^2 \\
&= \left\langle \sum_{i \in W} \langle \mathbf{W}_0, \mathbf{W}_i \rangle \mathbf{W}_i, \sum_{i \in W} \langle \mathbf{W}_0, \mathbf{W}_i \rangle \mathbf{W}_i \right\rangle \\
&= \langle \mathbf{W}_0, \mathbf{W}_0 \rangle \\
&= 1
\end{aligned}$$

□

Define now a new function:

$$\vartheta_1(G; \mathbf{w}) = \min_{\mathbf{y} \in \mathtt{TH}(\overline{G})} \max_{i \in V} \frac{\mathrm{w}_i}{\mathrm{y}_i},$$

with the understanding that if both are 0, then the ratio is 0, and if only the denominator is 0, the ratio is $\infty$.

**Theorem 5.13.** *Let $G$ be a graph, $\mathbf{w} \in \mathbb{R}^V_+$. Then $\vartheta(G;\mathbf{w}) \leq \vartheta_1(G;\mathbf{w})$.*

*Proof.* Assume there is at least one $\mathbf{y} \in \mathtt{TH}(\overline{G})$ whose entries equal to 0 also correspond to entries of $\mathbf{w}$ equal to 0 (otherwise the result is immediate). Let $\mathbf{y}$ be one of those, and $\mathbf{x} \in \mathtt{TH}(G)$ be arbitrary. Then

$$\mathbf{w}^\mathsf{T}\mathbf{x} = \sum \frac{\mathrm{w}_i}{\mathrm{y}_i}\mathrm{y}_i\mathrm{x}_i \leq \max_{i\in V} \frac{\mathrm{w}_i}{\mathrm{y}_i}(\mathbf{y}^\mathsf{T}\mathbf{x}) \leq \max_{i\in V} \frac{\mathrm{w}_i}{\mathrm{y}_i}.$$

□

Given $\mathbf{w} \in \mathbb{R}^V_+$, let $\sqrt{\mathbf{w}}$ be the vector obtained upon taking the entry-wise square root. Define now

$$\vartheta_2(G;\mathbf{w}) = \min\{\lambda_{\max}(\sqrt{\mathbf{w}}\sqrt{\mathbf{w}}^\mathsf{T} + \mathbf{A}) : \mathbf{A} \in \mathbb{S}^V,\ \mathrm{A}_{ij} = 0 \text{ if } ij \notin E\}.$$

**Theorem 5.14.** *Let $G$ be a graph, $\mathbf{w} \in \mathbb{R}^V_+$. Then $\vartheta_1(G;\mathbf{w}) \leq \vartheta_2(G;\mathbf{w})$.*

*Proof.* Let $\mathbf{A}$ be any weighted adjacency matrix, as in the definition of $\vartheta_2$. Take $\mu = \lambda_{\max}(\sqrt{\mathbf{w}}\sqrt{\mathbf{w}}^\mathsf{T} + \mathbf{A})$. Note that $\mu > 0$. Let $\mathbf{u}_0$ be one of its corresponding normalized eigenvectors. Note that

$$\mu\mathbf{I} - (\sqrt{\mathbf{w}}\sqrt{\mathbf{w}}^\mathsf{T} + \mathbf{A}) \succcurlyeq \mathbf{0},$$

thus let $\mathbf{Y}$ be so that

$$\mu\mathbf{I} - (\sqrt{\mathbf{w}}\sqrt{\mathbf{w}}^\mathsf{T} + \mathbf{A}) = \mathbf{Y}\mathbf{Y}^\mathsf{T}.$$

Therefore $\mathbf{Y}^\mathsf{T}\mathbf{u}_0 = \mathbf{0}$. As $\mathbf{Y}^\mathsf{T}$ is a square matrix, let $\mathbf{v}$ be a left normalized eigenvector, meaning, $\mathbf{v}^\mathsf{T}\mathbf{Y}^\mathsf{T} = \mathbf{0}$. Let now

$$\mathbf{Z} = \frac{1}{\sqrt{\mu}}(\mathbf{v}\sqrt{\mathbf{w}}^\mathsf{T} + \mathbf{Y}^\mathsf{T}).$$

Note that

$$\mathbf{Z}^\mathsf{T}\mathbf{Z} = \frac{1}{\mu}(\sqrt{\mathbf{w}}\mathbf{v}^\mathsf{T} + \mathbf{Y})(\mathbf{v}\sqrt{\mathbf{w}}^\mathsf{T} + \mathbf{Y}^\mathsf{T}) = \frac{1}{\mu}(\sqrt{\mathbf{w}}\sqrt{\mathbf{w}}^\mathsf{T} + \mathbf{Y}\mathbf{Y}^\mathsf{T}) = \mathbf{I} - \frac{1}{\mu}\mathbf{A}.$$

Therefore the columns of $\mathbf{Z}$ form an orthonormal representation of $G$. Choosing the handle $\mathbf{v}$, note that

$$\mathbf{Z}^\mathsf{T}\mathbf{v} \circ \mathbf{Z}^\mathsf{T}\mathbf{v} = \frac{1}{\mu}\mathbf{w}.$$

By the Exercise 5.10, it follows that $\mathbf{w}/\mu$ belongs to $\mathtt{TH}(\overline{G})$. Therefore

$$\vartheta_1(G;\mathbf{w}) \leq \max_{i\in V} \frac{\mathrm{w}_i}{\mathrm{w}_i/\mu} = \mu,$$

as we wanted. □

We keep going. Define now

$$\vartheta_3(G;\mathbf{w}) = \max\{\sqrt{\mathbf{w}}^\mathsf{T}\mathbf{X}\sqrt{\mathbf{w}} : \operatorname{tr}\mathbf{X} = 1,\ \mathrm{X}_{ij} = 0\ \forall\ ij \in E,\ \mathbf{X} \in \mathbb{S}^V_+\}.$$

(Note in particular that with $\mathbf{w} = \mathbf{1}$, this is the original definition of $\vartheta$ we provided.)

**Theorem 5.15.** *Let $G$ be a graph, $\mathbf{w} \in \mathbb{R}^V_+$. Then $\vartheta_2(G;\mathbf{w}) = \vartheta_3(G;\mathbf{w})$.*

*Proof.* This follows straight from the fact that $\vartheta_2$ and $\vartheta_3$ are optima of dual semidefinite programs, and both contain Slater points, hence strong duality holds with no duality gap. □

Actually I would like to present an alternative interpretation of the result above. Let $\mathbb{K}_G$ be the cone of symmetric matrices whose entries corresponding to edges of $G$ are equal to 0. Thus

$$\vartheta_3(G;\mathbf{w}) = \max\{\sqrt{\mathbf{w}}^\mathsf{T}\mathbf{X}\sqrt{\mathbf{w}} : \operatorname{tr}\mathbf{X} = 1,\ \mathbf{X} \in \mathbb{K}_G \cap \mathbb{S}^V_+\}.$$

Have $\mathbf{W} = \sqrt{\mathbf{w}}\sqrt{\mathbf{w}}^\mathsf{T}$. It follows that, for $\mathbf{X} \in \mathbb{K}_G \cap \mathbb{S}^V_+\}$,

$$\langle(\vartheta_3\mathbf{I} - \mathbf{W}), \mathbf{X}\rangle \geq 0.$$

Now comes the interesting part. As a consequence of hyperplane separation (or, more precisely, of $\mathbb{K}^{**} = \mathbb{K}$ for $\mathbb{K}$ closed and convex), for any two closed convex cones $\mathbb{K}$ and $\mathbb{L}$, we have

$$(\mathbb{K} \cap \mathbb{L})^* = \mathbb{K}^* + \mathbb{L}^*.$$

**Exercise 5.16.** Prove this fact (!!).

Moving forward, it means $(\vartheta_3\mathbf{I} - \mathbf{W}) \in (\mathbb{K}_G)^* + (\mathbb{S}^V_+)^*$. The dual of $\mathbb{K}_G$ is the set of symmetric matrices whose only possibly non-zero entries are indexed by edges of $G$. The dual of $\mathbb{S}^V_+$ is itself. Hence, there is a matrix $\mathbf{Y}$, with $\mathrm{Y}_{ij} = 0$ if $ij \notin E(G)$, so that

$$\vartheta_3\mathbf{I} - (\mathbf{W} + \mathbf{Y}) \succcurlyeq \mathbf{0},$$

therefore $\vartheta_3(G;\mathbf{w}) \geq \lambda_{\max}(\mathbf{W} + \mathbf{Y}) \geq \vartheta_2(G;\mathbf{w})$.

**Theorem 5.17.** *Let $G$ be a graph, $\mathbf{w} \in \mathbb{R}^V_+$. Then $\vartheta_3(G;\mathbf{w}) \leq \vartheta(G;\mathbf{w})$.*

*Proof.* Let $\overline{\mathbf{X}}$ be a feasible solution to the SDP defining $\vartheta_3$. Our goal is to show that there is $\overline{\mathbf{x}} \in \mathtt{TH}(G)$ with $\sqrt{\mathbf{w}}^\mathsf{T}\overline{\mathbf{X}}\sqrt{\mathbf{w}} \leq \mathbf{w}^\mathsf{T}\overline{\mathbf{x}}$.

We need to move from $\overline{\mathbf{X}}$ to something in $\widehat{\mathtt{TH}}(G)$. This motivates what follows below.

- Take $\mu = \sqrt{\mathbf{w}}^\mathsf{T}\overline{\mathbf{X}}\sqrt{\mathbf{w}}$.
- $\mathbf{u}_0 = \mu^{-1/2}\overline{\mathbf{X}}^{1/2}\sqrt{\mathbf{w}}$. Note that $\|\mathbf{u}_0\|^2 = 1$.
- Define $\mathbf{z}$ as $\mathrm{z}_i = \overline{\mathbf{X}}_{ii}^{-1/2}$ if $\overline{\mathbf{X}}_{ii}$ is non-zero, and $\mathrm{z}_i = 0$ otherwise.
- Define $\tilde{\mathbf{B}} = \overline{\mathbf{X}}^{1/2}\operatorname{Diag}(\mathbf{z})$. Note that $\tilde{\mathbf{B}}^\mathsf{T}\tilde{\mathbf{B}}$ has a diagonal of 0s and 1s.
- Define now $\overline{\mathbf{B}} = \tilde{\mathbf{B}}\operatorname{Diag}(\tilde{\mathbf{B}}^\mathsf{T}\mathbf{u}_0)$.

With these definition in hand, it is straightforward to verify that

$$\widehat{\mathbf{X}} = \begin{pmatrix} 1 & \mathbf{u}_0^\mathsf{T}\overline{B} \\ \overline{B}^\mathsf{T}\mathbf{u}_0 & \overline{B}^\mathsf{T}\overline{B} \end{pmatrix} \in \widehat{\mathtt{TH}}(G).$$

Let $\mathbf{d} = \operatorname{diag}(\overline{\mathbf{X}})$. Then

$$\begin{aligned}\mu &= \left(\frac{\sqrt{\mathbf{w}}^{\mathsf{T}}\overline{\mathbf{X}}^{1/2}\overline{\mathbf{X}}^{1/2}\sqrt{\mathbf{w}}}{\mu^{1/2}}\right)^2 = (\sqrt{\mathbf{w}}^{\mathsf{T}}\overline{\mathbf{X}}^{1/2}\mathbf{u}_0)^2 \\ &= (\sqrt{\mathbf{w}}^{\mathsf{T}}\operatorname{Diag}(\sqrt{\mathbf{d}})\tilde{B}^{\mathsf{T}}\mathbf{u}_0)^2 = (\sqrt{\mathbf{d}}^{\mathsf{T}}\operatorname{Diag}(\sqrt{\mathbf{w}})\sqrt{\overline{B}^{\mathsf{T}}\mathbf{u}_0})^2 \\ &\leq \|\sqrt{\mathbf{d}}\|^2\|\operatorname{Diag}(\sqrt{\mathbf{w}})\sqrt{\overline{\mathbf{x}}}\|^2 = \mathbf{w}^{\mathsf{T}}\overline{\mathbf{x}}\end{aligned}$$

□

So now we have it. Five equivalent definitions:

(0) $\vartheta(G;\mathbf{w}) = \max\{\mathbf{w}^{\mathsf{T}}\mathbf{x} : \mathbf{x} \in \mathtt{TH}(G)\}$.

(1) $\vartheta_1(G;\mathbf{w}) = \min_{\mathbf{y}\in\mathtt{TH}(\overline{G})} \max_{i\in V} \frac{\mathrm{w}_i}{\mathrm{y}_i}$.

(2) $\vartheta_2(G;\mathbf{w}) = \min\{\lambda_{\max}(\sqrt{\mathbf{w}}\sqrt{\mathbf{w}}^{\mathsf{T}} + \mathbf{A}) : \mathbf{A} \in \mathbb{S}^V,\ \mathrm{A}_{ij} = 0 \text{ if } ij \notin E\}$.

(3) $\vartheta_3(G;\mathbf{w}) = \max\{\sqrt{\mathbf{w}}^{\mathsf{T}}\mathbf{X}\sqrt{\mathbf{w}} : \operatorname{tr}\mathbf{X} = 1,\ \mathrm{X}_{ij} = 0\ \forall\ ij \in E,\ \mathbf{X} \in \mathbb{S}^V_+\}$.

(4) $\vartheta_4(G;\mathbf{w}) = \max_{\rho \text{ o.r. of } \overline{G}} \sum_{i\in V} \mathrm{w}_i\langle\rho_0,\rho_i\rangle^2$.

Moreover, all these problems admit optima solutions (something you are invited to verify).

**Corollary 5.18.** *We have $\mu = \vartheta(G;\mathbf{w})$ if and only if there are $\mathbf{x} \in \mathit{TH}(G)$ and $\mathbf{y} \in \mathit{TH}(\overline{G})$ with $\mathbf{x}^{\mathsf{T}}\mathbf{y} = 1$ and $\mu\mathbf{y} = \mathbf{w}$.*

*Proof.* Given $\mu = \vartheta$, let $\mathbf{x} \in \mathtt{TH}(G)$ with $\mu = \mathbf{w}^{\mathsf{T}}\mathbf{x}$ and $\mathbf{y} \in \mathtt{TH}(\overline{G})$ with $\vartheta_1 = \max_i(\mathrm{w}_i/\mathrm{y}_i)$. It follows that

$$\vartheta = \mathbf{w}^{\mathsf{T}}\mathbf{x} = \sum \frac{\mathrm{w}_i}{\mathrm{y}_i}\mathrm{y}_i\mathrm{x}_i \leq \max_{i\in V}\frac{\mathrm{w}_i}{\mathrm{y}_i}(\mathbf{y}^{\mathsf{T}}\mathbf{x}) \leq \max_{i\in V}\frac{\mathrm{w}_i}{\mathrm{y}_i} = \max_{i\in V}\frac{\mathrm{w}_i}{\mathrm{y}_i} = \vartheta_1.$$

As equality holds throughout, it must be that $\mathbf{y}^{\mathsf{T}}\mathbf{x} = 1$, and that $\mathbf{y}$ is a multiple of $\mathbf{w}$.

For the other direction, assume $\mathbf{x} \in \mathtt{TH}(G)$ and $\mathbf{y} \in \mathtt{TH}(\overline{G})$ with $\mathbf{x}^{\mathsf{T}}\mathbf{y} = 1$ and $\mu\mathbf{y} = \mathbf{w}$. Then

$$\mu \leq \mu\mathbf{y}^{\mathsf{T}}\mathbf{x} = \mathbf{w}^{\mathsf{T}}\mathbf{x} \leq \vartheta = \vartheta_1 \leq \max_{i\in V}\frac{\mathrm{w}_i}{\mathrm{y}_i} = \mu.$$

Thus equality holds throughout. □

We introduce yet another equivalent definition of $\vartheta$.

(6) $\vartheta_6(G;\mathbf{w}) = \max\{\lambda_{\max}(\mathbf{B}) : \mathbf{B} \in \mathbb{S}^V_+,\ \operatorname{diag}(\mathbf{B}) = \mathbf{w},\ \mathbf{B}_{ij} = 0\ \forall\ ij \in E\}$.

(I'm following D. Knuth, so don't ask me where is $\vartheta_5$.)

**Theorem 5.19.** *Let $G$ be a graph, $\mathbf{w} \in \mathbb{R}^V_+$. Then $\vartheta(G;\mathbf{w}) = \vartheta_6(G;\mathbf{w})$.*

*Proof.* Let $\mathbf{X}'$ be feasible for $\vartheta_3$, $\mathbf{d} = \operatorname{diag}(\mathbf{X}')$. If any diagonal entry of $\mathbf{X}'$ is 0, turn it into a 1, thus creating matrix $\mathbf{X}$. Let $\mathbf{D}$ be diagonal so that $\operatorname{diag}(\mathbf{DXD}) = \mathbf{1}$. Take

$$\mathbf{B} = \operatorname{Diag}(\sqrt{w})\mathbf{DXD}\operatorname{Diag}(\sqrt{\mathbf{w}}).$$

It is obviously feasible for $\vartheta_6$, and

$$\lambda_{\max}(\mathbf{B}) \geq \frac{\sqrt{\mathbf{d}}^{\mathsf{T}}\mathbf{B}\sqrt{\mathbf{d}}}{\sqrt{\mathbf{d}}^{\mathsf{T}}\sqrt{\mathbf{d}}} \geq \sqrt{\mathbf{w}}^{\mathsf{T}}\mathbf{X}'\sqrt{\mathbf{w}}.$$

Let now $\mathbf{B}$ be an optimal solution for $\vartheta_6$. It can be diagonalized as $\mathbf{B} = \mathbf{Q}\Lambda\mathbf{Q}^{\mathsf{T}}$ — assume $\lambda_{\max}$ is the first. Let

$$\mathbf{b}_i = \Lambda^{1/2}\mathbf{Q}^{\mathsf{T}}\mathbf{e}_i,$$

and, if $\mathrm{w}_i \neq 0$, have $\mathbf{u}_i = \mathrm{w}_i^{-1/2}\mathbf{b}_i$. Complete the set of $\mathbf{u}_i$s with an orthonormal basis. Note that the $\mathbf{u}$s form an orthonormal representation of $\overline{G}$, as they are normalized and, from the definition of $\mathbf{B}$, $\mathbf{u}_i$ and $\mathbf{u}_j$ are orthogonal if $ij \in \overline{E}$.

Take $\mathbf{u}_0$ to be equal to $\mathbf{e}_1$. For those $i$ with $\mathrm{w}_i \neq 0$, it follows that

$$(\mathbf{u}_0^{\mathsf{T}}\mathbf{u}_i) = \frac{1}{\mathrm{w}_i}\lambda_{\max}(\mathbf{B})(\mathbf{e}_1^{\mathsf{T}}\mathbf{Q}\mathbf{e}_i)^2,$$

thus

$$\vartheta_4 \geq \sum_{i\in V} \mathrm{w}_i(\mathbf{u}_0^{\mathsf{T}}\mathbf{u}_i)^2 = \sum_{\mathrm{w}_i\neq 0} \lambda_{\max}(\mathbf{B})(\mathbf{e}_1^{\mathsf{T}}\mathbf{Q}\mathbf{e}_i)^2 \geq \lambda_{\max}(\mathbf{B}).$$

□

As a corollary, we recover a result due to Hoffman, and one of the original characterizations of $\vartheta(G)$.

**Corollary 5.20.** *Given a matrix* $\mathbf{A}$*, recall our notation* $\tilde{\mathbf{A}} = \mathbf{A}/-\lambda_{\min}(\mathbf{A})$ *if* $\mathbf{A} \neq 0$*, and* $\tilde{\mathbf{A}} = \mathbf{A}$ *otherwise. Given* $G$*, we have*

$$\vartheta(\overline{G}) = \max\{1 + \lambda_{\max}(\tilde{\mathbf{A}}) : \mathrm{A}_{ij} = 0 \ \forall\ ij \notin E\}.$$

*Proof.* From the Theorem, we have

$$\vartheta(\overline{G}) = \vartheta_6(\overline{G}; \mathbf{1}) = \max\{\lambda_{\max}(\mathbf{I} + \mathbf{A}) : \mathbf{A} \succcurlyeq -\mathbf{I}, \mathrm{A}_{ij} = 0 \ \forall\ ij \notin E\}.$$

For any $\mathbf{A}$ feasible, just note that the best scaling so that $\succcurlyeq -\mathbf{I}$ is maintained is always obtained from dividing $\mathbf{A}$ by $-\lambda_{\min}$. Hence the result follows.

□

**Exercise 5.21.** Prove that

$$\vartheta(G) = \min\{\lambda : \mathbf{Z}_{00} = \lambda,\ \mathbf{Z}_{0i} = \mathbf{Z}_{ii} = 1 \ \forall\ i \in V(G),\ \mathbf{Z}_{ij} = 0 \text{ if } ij \in \overline{E},\ \mathbf{Z} \in \mathbb{S}_+^{\{0\}\cup V(G)}\}.$$

Further, find a way of turning this formulation into yet another equivalent definition of the theta function $\vartheta(G; \mathbf{w})$.

## 5.4 Perfect graphs

Recall the definitions: $\mathbf{N}$ is the vertex-edge incidence matrix, $\mathbf{M}$ is the incidence matrix vertex-clique, and $\overline{\mathbf{M}}$ is the incidence matrix vertex-coclique.

Recall the convex sets

- $\texttt{STAB}(G) = \text{conv}\{\mathbf{x}_S : S \text{ is a coclique}\} = \{\overline{\mathbf{M}}\mathbf{y} : \mathbf{y} \geq \mathbf{0},\ \mathbf{1}^\mathsf{T}\mathbf{y} = 1\}$.
- $\texttt{TH}(G) = \left\{\mathbf{x} \in \mathbb{R}^V : \exists \mathbf{X} \in \mathbb{S}^V \text{with} \begin{pmatrix} 1 & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \succcurlyeq \mathbf{0},\ \text{diag}(\mathbf{X}) = \mathbf{x},\ \mathrm{X}_{ij} = 0\ \forall\ ij \in E(G)\right\}$.

  Equivalently
- $\texttt{TH}(G) = \{\mathbf{x} : \mathrm{x}_i = \langle \rho_0, \rho_i\rangle^2 \text{ for some } \rho \text{ o.r. of } \overline{G}.\}$
- $\texttt{QSTAB}(G) = \{\mathbf{y} : \mathbf{y} \geq \mathbf{0},\ \mathbf{M}^\mathsf{T}\mathbf{y} \leq \mathbf{1}\}$.
- $\texttt{FRAC}(G) = \{\mathbf{y} : \mathbf{y} \geq \mathbf{0},\ \mathbf{N}^\mathsf{T}\mathbf{y} \leq \mathbf{1}\}$.

As we have seen,

$$\texttt{STAB}(G) \subseteq \texttt{TH}(G) \subseteq \texttt{QSTAB}(G) \subseteq \texttt{FRAC}(G).$$

**Theorem 5.22.** *$\texttt{STAB}(G) = \texttt{FRAC}(G)$ if and only if $G$ is bipartite.*

*Proof.* If $G$ is not bipartite, it contains an odd cycle $C$. Let $\mathbf{x}_C$ be its characteristic vector. Note that $(1/2)\mathbf{x}_C \in \texttt{FRAC}(G)$. Assume there are cocliques $S_1, ..., S_k$ with characteristic vectors $\mathbf{x}_1, ..., \mathbf{x}_k$ and

$$(1/2)\mathbf{x}_C = \sum_{i=1}^{k} \alpha_i \mathbf{x}_i,$$

having $\alpha_i > 0$ and $\sum \alpha_i = 1$. Multiplying by $\mathbf{x}_C^\mathsf{T}$, we gather

$$\frac{|C|}{2} \leq \frac{|C|-1}{2} \sum \alpha_i,$$

where the right hand side follows because an independent set intersects an odd cycle in at most $(|C|-1)/2$ vertices. Hence the contradiction.

The other side of this proof is an exercise on your assignment. □

**Exercise 5.23.** Characterize the class of graphs for which we have $\texttt{QSTAB}(G) = \texttt{FRAC}(G)$.

As opposed to the characterizations above, the gap between STAB and QSTAB leads to a much richer theory. A graph $G$ is called *perfect* if

$$\texttt{STAB}(G) = \texttt{QSTAB}(G).$$

In what follows, we will see a few equivalent definitions of perfect graphs. C. Berge originally defined a graph to be perfect using condition (b). Chavátal showed the equivalence of (a) and (b), and Lovász showed the equivalence of (b) - (d). For these, we will follow a proof due to Gasparian. Later, Grötschel, Lovász and Schrijver showed the equivalence with (f), (g)

and (h). Condition (i) was proved by Chudnovsky, Robertson, Seymour and Thomas (this result is known as the strong perfect graph theorem) — the proof of this result is a major breakthrough in structural graph theory, and certainly would not fit in these course notes.

If $G = (V, E)$ is a graph, and $S \subseteq V$, then let $G_S$ be the subgraph of $G$ with vertices $S$, and all edges of $G$ connecting vertices within $S$. This is the *induced subgraph* of $G$ on $S$.

**Theorem 5.24.** *Let $G$ be a graph on $n$ vertices. The following are equivalent.*

*(a) $G$ is perfect, ie, $\mathtt{STAB}(G) = \mathtt{QSTAB}(G)$.*

*(b) For all $S \subseteq V(G)$, $\omega(G_S) = \chi(G_S)$.*

*(c) For all $S \subseteq V(G)$, $\omega(G_S)\alpha(G_S) \geq n$.*

*(d) For all $S \subseteq V(G)$, $\alpha(G_S) = \overline{\chi}(G_S)$.*

*(e) $\mathtt{STAB}(\overline{G}) = \mathtt{QSTAB}(\overline{G})$.*

*(f) $\mathtt{TH}(G)$ is a polytope.*

*(g) $\mathtt{TH}(G) = \mathtt{STAB}(G)$.*

*(h) $\mathtt{TH}(G) = \mathtt{QSTAB}(G)$.*

*(i) $G$ has no induced subgraph isomorphic to $C_n$ or $\overline{C_n}$, for $n$ odd and $n > 3$.*

*Proof that (b) and (c) are equivalent.* Clearly, if (b) holds, then (c) follows from the obvious fact that $\chi(G)\alpha(G) \geq n$ for any graph $G$.

Now let $G$ be a minimal graph that violates (b), meaning, any proper induced subgraph of $G$ satisfies (b), but $G$ itself fails it. We will show that $\omega(G)\alpha(G) + 1 \leq n$.

Let $S_0$ be a coclique of size $\alpha$ in $G$. For each $v \in S_0$, note that $\chi(G - v) = \omega(G - v) \leq \omega$, hence $G - v$ can be partitioned into $\omega$ cocliques, for each $v$. Consider $S_0$, and $S_1, ..., S_{\alpha\omega}$ to be the collection of all such cocliques. Note that each vertex of the graph is contained in precisely $\alpha$ cocliques within the collection $\{S_i\}_{i=0}^{\alpha\omega}$.

Fix $S_i$. If $\omega(G - S_i) < \omega$, then, from the minimality of $G$, we have

$$\omega \geq \omega(G - S_i) + 1 = \chi(G - S_i) + 1 \geq \chi(G),$$

as $S_i$ can be used as a colour. Thus, $\omega(G - S_i) = \omega$, and thus, for each $S_i$, there is a clique $C_i$ of size $\omega$ with $|C_i \cap S_i| = 0$. As each vertex in $C_i$ is contained in $\alpha$ cocliques in $\{S_i\}_{i=0}^{\alpha\omega}$, but $C_i$ intersects $S_j$ in at most one vertex, we have $|C_i \cap S_j| = 1$ for all $i \neq j$.

Let $\mathbf{M}$ be the incidence matrix of vertices and cliques $\{C_i\}_{i=0}^{\alpha\omega}$, and $\mathbf{N}$ the incidence matrix of vertices and cocliques $\{S_i\}_{i=0}^{\alpha\omega}$. From the discussion above, it follows that

$$\mathbf{M}^\mathsf{T}\mathbf{N} = \mathbf{J} - \mathbf{I}.$$

This matrix has rank $\alpha\omega + 1$, thus $n \geq \alpha\omega + 1$.

This proves that (c) implies (b). □

Note that, as an immediate consequence, (d) is also equivalent to (b) and (c).

**Lemma 5.25.** *Assume $H$ arise from $G$ upon taking a vertex $v \in V(G)$, creating a new vertex $v'$, adjacent precisely to all neighbours of $v$. If $G$ satisfies (b), then so does $H$.*

*Proof.* Assume otherwise, and let $G$ be minimal so that this property fails for one of its vertices $v$. Let $u \in H$. Note that $\omega(H-u) = \chi(H-u)$, either because $u = v$ or because otherwise $G$ would not be a minimal counterexample. So it must be that $\omega(H) < \chi(H)$. However, $\omega(G) = \omega(H)$ (as any maximal clique containing $v'$ corresponds to one containing $v$), and $\chi(G) = \chi(H)$, as the colour of $v$ can be repeated in $v'$ in any proper colouring of $H$. Contradiction — hence the result follows. □

Naturally, the result above can be applied repeatedly, so replacing any vertex $v$ by a coclique, and edges at $v$ by cones of edges at the coclique, does not alter the property that a graph satisfies (b). This process is called *replicating the vertex*.

*Proof that (a) and (d) are equivalent.* Assume (a). Let $U$ be minimal so that $\alpha(G_U) < \overline{\chi}(G_U)$, and choose $\mathbf{w} = \mathbf{x}_U$, the characteristic vector of $U$. For all $S$ maximum coclique in $G_U$, note that $\mathbf{x}_S$ is an optimum for $\mathbf{w}^\mathsf{T}\mathbf{x}$ within $\mathtt{STAB}$. As (a) holds, we have

$$\begin{aligned}\alpha(G_U) &= \max\{\mathbf{w}^\mathsf{T}\mathbf{x} : \mathbf{M}^\mathsf{T}\mathbf{x} \leq \mathbf{1},\ \mathbf{x} \geq \mathbf{0}\}\\ &= \min\{\mathbf{1}^\mathsf{T}\mathbf{y} : \mathbf{M}\mathbf{y} \geq \mathbf{w},\ \mathbf{y} \geq \mathbf{0}\}.\end{aligned}$$

From the complementary slackness conditions, there is at least one row of $\mathbf{M}^\mathsf{T}\mathbf{x} \leq \mathbf{1}$ that must be met with equality for all optimum solutions. Thus, there is a clique $C$ in $G_U$ that meets all maximum cocliques of $G_U$. Hence

$$\alpha(G_{U-C}) \leq \alpha(G_U) - 1 < \overline{\chi}(G_U) - 1 \leq \overline{\chi}(G_{U-S}),$$

a contradiction to the minimality of $U$.

We now assume (d). If (a) does not hold, then there is a vertex $\mathbf{x}'$ of $\mathtt{QSTAB}(G)$ that does not belong to $\mathtt{STAB}(G)$. From hyperplane separation, there is a $\mathbf{c}$ so $\max \mathbf{c}^\mathsf{T}\mathbf{x}$ attains its maximum at $\mathbf{x}'$, and, moreover, we can assume $\mathbf{c}$ to be integral (just pick a rational point close enough to the original $\mathbf{c}$ and scale it)[1].

Hence, there is $\mathbf{c}$, integral, so that

$$\max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \mathtt{QSTAB}(G)\} > \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \mathtt{STAB}(G)\}.$$

Now simply replicate each vertex $i \in V(G)$ a total of $\max\{0, \mathbf{c}_i\}$ times, obtaining graph $H$, which still satisfies (d), thus $\alpha(H) = \overline{\chi}(H)$. However,

$$\alpha(H) = \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \mathtt{STAB}(G)\},$$

but, as cliques of $G$ and $H$ correspond, and $H$ is covered by cliques if and only if each vertex of $G$ is covered by $\max\{0, \mathbf{c}_i\}$ cliques, we have

$$\begin{aligned}\overline{\chi}(H) &= \min\{\mathbf{1}^\mathsf{T}\mathbf{y} : \mathbf{M}\mathbf{y} \geq \mathbf{c}, \mathbf{y} \geq \mathbf{0},\ \mathbf{y} \text{ integral}\}\\ &\geq \min\{\mathbf{1}^\mathsf{T}\mathbf{y} : \mathbf{M}\mathbf{y} \geq \mathbf{c}, \mathbf{y} \geq \mathbf{0}\}\\ &= \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \mathtt{QSTAB}(G)\}.\end{aligned}$$

[1] This was quite the sketchy argument, but this is a well-known fact used many times in the literature — notably by Edmonds in 1965

As a consequence,

$$\max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \mathtt{STAB}(G)\} = \max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \mathtt{QSTAB}(G)\},$$

contradicting the choice of $\mathbf{c}$. □

At this point, we know conditions (a) - (e) of Theorem 5.24 are equivalent. Let us now understand conditions (f) - (h).

## 5.5 Theta body

A set $C$ is called a convex corner (of $\mathbb{R}^n$) if

1. $C \subseteq \mathbb{R}^n_+$.
2. $C$ is closed and convex.
3. If $\mathbf{x} \in C$ and $\mathbf{0} \leq \mathbf{x}' \leq \mathbf{x}$, then $\mathbf{x}' \in C$.

Note that STAB, QSTAB and FRAC are all convex corners. We will see that TH is also. Given a convex corner $C$, the *antiblocker* of $C$ is defined as

$$\mathtt{ab}(C) = \{\mathbf{x} \in \mathbb{R}^n_+ : \mathbf{y}^\mathsf{T}\mathbf{x} \leq 1 \text{ for all } \mathbf{y} \in C\}.$$

It is quite immediate to verify that $\mathtt{ab}(C)$ is also a convex corner, and that $\mathtt{ab}(\mathtt{ab}(C)) = C$.

**Exercise 5.26.** Show that

$$\mathtt{ab}(\mathtt{STAB}(G)) = \mathtt{QSTAB}(\overline{G}).$$

Let us now recall that, as a consequence of Theorem 5.9, we had

- $\mathbf{x} \in \mathtt{TH}(G)$ if, and only if, there is orthonormal representation $\rho$ of $\overline{G}$ with $\mathrm{x}_i = \langle \rho_0, \rho_i \rangle^2$ for all $i \in V$.

With it, we could show that both definitions below are equivalent

(1) $\vartheta_1(G;\mathbf{w}) = \min\limits_{\mathbf{y} \in \mathtt{TH}(\overline{G})} \max\limits_{i \in V} \dfrac{\mathrm{w}_i}{\mathrm{y}_i} = \min\limits_{\rho \text{ o.r. of } G} \max\limits_{i \in V} \dfrac{\mathrm{w}_i}{\langle \rho_0, \rho_i \rangle^2}$,

(4) $\vartheta_4(G;\mathbf{w}) = \max\limits_{\rho \text{ o.r. of } \overline{G}} \sum\limits_{i \in V} \mathrm{w}_i \langle \rho_0, \rho_i \rangle^2 = \max\limits_{\mathbf{x} \in \mathtt{TH}(G)} \mathbf{w}^\mathsf{T}\mathbf{x} = \vartheta(G;\mathbf{w})$.

**Lemma 5.27.** *For all $\mathbf{w}, \mathbf{z} \geq \mathbf{0}$, and a graph $G$, we have*

$$\vartheta(G;\mathbf{w})\vartheta(\overline{G};\mathbf{z}) \geq \mathbf{w}^\mathsf{T}\mathbf{z}.$$

*Proof.* Let $\mathbf{w} \neq \mathbf{0}$, and $\mathbf{y} \in \mathtt{TH}(\overline{G})$ attaining the optimum in $\vartheta_1(G;\mathbf{w})$. Thus

$$\vartheta(\overline{G};\mathbf{z}) \geq \mathbf{z}^\mathsf{T}\mathbf{y} \geq \frac{1}{\vartheta_1(G;\mathbf{w})}\mathbf{z}^\mathsf{T}\mathbf{w}.$$

□

**Lemma 5.28.** *If $\mathbf{x} \in TH(G)$, then $\vartheta(\overline{G}; \mathbf{x}) \leq 1$.*

*Proof.* Follows from

$$\vartheta(\overline{G}; \mathbf{x}) = \min_{\mathbf{y} \in \mathtt{TH}(G)} \max_{i \in V} \frac{\mathrm{x}_i}{\mathrm{y}_i} \leq \max_{i \in V} \frac{\mathrm{x}_i}{\mathrm{x}_i} = 1.$$

□

**Lemma 5.29.** *For all $\mathbf{w} \geq \mathbf{0}$, $\mathbf{w} \neq \mathbf{0}$, there is $\mathbf{z} \geq \mathbf{0}$, $\mathbf{z} \neq \mathbf{0}$, with*

$$\vartheta(G; \mathbf{w})\vartheta(\overline{G}; \mathbf{z}) = \mathbf{w}^\mathsf{T}\mathbf{z}.$$

*Proof.* Simply take $\mathbf{z} \in \mathtt{TH}(G)$ that gives $\vartheta(G; \mathbf{w}) = \mathbf{w}^\mathsf{T}\mathbf{z}$, an apply the lemmas above. □

**Lemma 5.30.** *Let $\mathbf{w}, \mathbf{z} \geq \mathbf{0}$, and $\alpha \geq 0$. The following holds:*

1. *If $\mathbf{w} \geq \mathbf{z} \geq \mathbf{0}$, then $\vartheta(G; \mathbf{w}) \geq \vartheta(G; \mathbf{z})$.*
2. *$\vartheta(G; \alpha\mathbf{w}) = \alpha\vartheta(G; \mathbf{w})$.*
3. *$\vartheta(G; \mathbf{w} + \mathbf{z}) \leq \vartheta(G; \mathbf{w}) + \vartheta(G; \mathbf{z})$.*

*Proof.* All three follow immediately from the definition $\vartheta(G; \mathbf{w}) = \max_{\mathbf{x} \in \mathtt{TH}(G)} \mathbf{w}^\mathsf{T}\mathbf{x}$. □

As a consequence of these properties, $\vartheta(G; \cdot)$ is a norm on $\mathbb{R}^n_+$ (also called a gauge), which also satisfies the monotone property. In this sense, $\mathtt{TH}(G)$ is somewhat the corner of the unit ball of the norm $\vartheta(\overline{G}; \cdot)$. The following theorem presents alternative description of $\mathtt{TH}(G)$.

**Theorem 5.31.** *The following are all equivalent presentations of the convex set $TH(G)$.*

*(a) Definition:* $TH(G) = \left\{ \mathbf{x} \geq \mathbf{0} : \exists \mathbf{X} \text{ with } \begin{pmatrix} 1 & \mathbf{x}^\mathsf{T} \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \succcurlyeq \mathbf{0},\ \operatorname{diag} \mathbf{X} = \mathbf{x},\ \mathrm{X}_{ij} = 0\ \forall\ ij \in E(G) \right\}$

*(b)* $\{\langle \rho_0, \rho_i \rangle^2 : \rho \text{ is orthonormal representation of } \overline{G}\}$.

*(c)* $\{\mathbf{x} \geq \mathbf{0} : \vartheta(\overline{G}; \mathbf{x}) \leq 1\}$.

*(d)* $\{\mathbf{x} \geq \mathbf{0} : \mathbf{y}^\mathsf{T}\mathbf{x} \leq \vartheta(G; \mathbf{y})\ \forall\ \mathbf{y} \geq \mathbf{0}\}$.

*(e)* $ab(TH(\overline{G}))$.

*Proof.* We already know that the set in (b) is equal to $\mathtt{TH}(G)$. From Lemma 5.28, we have that the set in (b) is contained in (c).

Look at Lemma 5.27 now. Take $\mathbf{x}$ in the set in (c), and $\mathbf{y} \geq \mathbf{0}$, so we have

$$\mathbf{y}^\mathsf{T}\mathbf{x} \leq \vartheta(G; \mathbf{y})\vartheta(\overline{G}; \mathbf{x}) \leq \vartheta(G; \mathbf{y}).$$

Thus the set in (c) is contained in the set in (d).

If $\mathbf{x}$ is in the set in (d), then we can choose $\mathbf{y}$ so that Lemma 5.29 holds, thus giving that

$$\vartheta(\overline{G}; \mathbf{x}) \leq 1,$$

so sets in (d) and (c) are equal. Moreover, we now see

$$\vartheta(\overline{G};\mathbf{x}) = \max_{\rho \text{ o.r. of } G} \sum_{i\in V} \mathrm{w}_i \langle \rho_0, \rho_i\rangle^2 \leq 1,$$

so the set in (d) belongs to $\mathtt{ab}(\mathtt{TH}(\overline{G}))$.

Finally now, let $\mathbf{x} \in \mathtt{ab}(\mathtt{TH}(\overline{G}))$. Pick $\mathbf{z} \in \mathtt{TH}(\overline{G})$ so that $\vartheta(\overline{G};\mathbf{x}) = \mathbf{x}^\mathsf{T}\mathbf{z}$. From Lemma 5.29, we have

$$1 \geq \mathbf{x}^\mathsf{T}\mathbf{z} = \vartheta(G;\mathbf{z})\vartheta(\overline{G};\mathbf{x}),$$

which implies $\vartheta(\overline{G};\mathbf{x}) = \mu \leq 1$. We now invoke Corollary 5.18. Thus, there is $\mathbf{x}' \in \mathtt{TH}(G)$ so that

$$\mu\mathbf{x}' = \mathbf{x}.$$

If $\rho$ is the representation of $\overline{G}$ giving out $\mathbf{x}'$ as in (b), we have

$$\mathrm{x}_i = \mu\langle \rho_0, \rho_i\rangle^2.$$

Now simply replace $\rho_0$ by

$$\rho_* = \sqrt{\mu}\rho_0 + \sqrt{1-\mu}\; q,$$

where $q$ is orthogonal to $\rho_0$ and all the $\rho_i$s (if needed be, simply use an extra coordinate). As a consequence, $\|\rho_*\| = 1$, and, finally,

$$\mathrm{x}_i = \langle \rho_*, \rho_i\rangle^2,$$

thus $\mathbf{x} \in \mathtt{TH}(G)$. □

Our goal now is to show conditions (f) - (h) in the perfect graph characterization theorem. If $C$ is a convex set of dimension $n$ (meaning, containing $n$ linearly independent vectors), and $\mathbf{a}^\mathsf{T}\mathbf{x} \leq \beta$ is a valid inequality for $C$, then we say that this inequality determines a *facet* of $C$ if the affine space of all $\mathbf{x} \in C$ that satisfy the inequality with equality has dimension $n-1$.

**Lemma 5.32.** *If an inequality determines a facet of $\mathtt{TH}(G)$, then it is (a positive multiple of) $\mathrm{x}_i \geq 0$ for some $i$, or a row of $\mathbf{M}^\mathsf{T}\mathbf{x} \leq \mathbf{1}$. (Recall $\mathbf{M}$ is the vertex vs clique incidence matrix).*

*Proof.* Assume $\mathbf{a}^\mathsf{T}\mathbf{x} \leq \beta$ is satisfied for all $\mathbf{x} \in \mathtt{TH}(G)$, meaning, for all $\mathbf{x} \geq \mathbf{0}$ with $\vartheta(\overline{G};\mathbf{x}) \leq 1$, and that this inequality defines a facet. Let $\mathbf{z}$ belong to the relative interior of this facet. If $\mathrm{z}_i = 0$ for some $i$, then all other entries can be perturbed and however $\mathbf{z}$ still satisfies $\mathbf{a}^\mathsf{T}\mathbf{x} \leq \beta$ with equality. Thus this inequality is equivalent to $\mathrm{x}_i \geq 0$. Thus, assume $\vartheta(\overline{G};\mathbf{z}) = 1$. Consequently, there is a orthonormal representation $\rho$ of $G$ with

$$\sum_{i=1}^{n} \mathrm{z}_i \langle \rho_0, \rho_i\rangle^2 = 1. \tag{2}$$

Moreover, for all $\mathbf{x} \in \mathtt{TH}(G)$, including $\mathbf{z}$ as well, it follows that $\mathbf{x} \in \mathtt{ab}(\mathtt{TH}(\overline{G}))$, and as $\mathtt{TH}(\overline{G}) = \{\langle \sigma_0, \sigma_i\rangle^2 : \sigma \text{ is orthonormal representation of } G\}$, we have

$$\sum_{i=1}^{n} \mathrm{x}_i \langle \rho_0, \rho_i\rangle^2 \leq 1.$$

Thus this inequality is equivalent to $\mathbf{a}^\mathsf{T}\mathbf{x} \leq \beta$. Hence, assume $\beta = 1$, and $\mathbf{a}_i = \langle \rho_0, \rho_i\rangle^2$.

Now, observe that, for all $\mathbf{z}$ in the facet, we have

$$1 = \sum_{i=1}^{n} \mathrm{z}_i \langle \rho_0, \rho_i\rangle^2 = \rho_0^\mathsf{T} \left( \sum_{i=1}^{n} \mathrm{z}_i \rho_i \rho_i^\mathsf{T} \right) \rho_0$$

As we saw before, the right hand side is always $\leq$ than 1, thus $\rho_0$ is the unit vector that maximizes the Rayleigh quotient of the symmetric matrix $\left(\sum_{i=1}^{n} \mathrm{z}_i \rho_i \rho_i^\mathsf{T}\right)$. It is, therefore, an eigenvector, whose corresponding eigenvalue is equal to 1. Thus

$$\rho_0 = \left( \sum_{i=1}^{n} \mathrm{z}_i \rho_i \rho_i^\mathsf{T} \right) \rho_0 = \sum_{i=1}^{n} \mathrm{z}_i (\rho_0^\mathsf{T} \rho_i) \rho_i$$

Because the facet has dimension $n-1$, each equality in a row of the vector equation above is a multiple of Equation (2). Thus, for at least one index $i$, it follows that $\rho_0^\mathsf{T}\rho_i \neq 0$, and therefore,

$$\langle \rho_0, \rho_i\rangle \rho_0 = \rho_i,$$

whence, as $\|\rho_0\| = \|\rho_i\| = 1$, we may as well assume $\rho_0 = \rho_i$. If $K$ corresponds to the vertices of the graph indexed by those $i$ for which $\rho_i = \rho_0$, it follows that $K$ is a clique, and inequality $\mathbf{a}^\mathsf{T}\mathbf{x} \leq \beta$, that we saw is given by $\beta = 1$ and $\mathbf{a}_i = \langle \rho_0, \rho_i\rangle^2$, is equivalent to

$$\sum_{i \in K} \mathrm{x}_i \leq 1.$$

□

*Proof of the equivalence between (a) and (f)-(h) in Theorem 5.24.*
Clearly (a) implies (f)-(h), and (g) or (h) imply (f). We need only show now that (f) implies (a). As we saw above, if $\mathtt{TH}(G)$ is a polytope, then its facets are determined by (a subset of) the facets of $\mathtt{QSTAB}(G)$. As $\mathtt{TH}(G) \subseteq \mathtt{QSTAB}(G)$, it follows that they must be equal. Also, $\mathtt{ab}(\mathtt{TH}(G)) = \mathtt{TH}(\overline{G})$. It is not difficult to show that the antiblocker of a polytope is a polytope — thus $\mathtt{TH}(\overline{G}) = \mathtt{QSTAB}(\overline{G})$, and now, taking antiblockers, we have

$$\mathtt{TH}(G) = \mathtt{ab}(\mathtt{TH}(\overline{G})) = \mathtt{ab}(\mathtt{QSTAB}(\overline{G})) = \mathtt{STAB}(G),$$

therefore $\mathtt{STAB}(G) = \mathtt{QSTAB}(G)$. □

**Exercise 5.33.** Find an easy way to show that $\vartheta(C_5) = \sqrt{5}$.

**Exercise 5.34.** Assume $\rho$ is the orthonormal representation of $\overline{G}$ so that

$$\vartheta(G;\mathbf{x}) = \sum_{i\in V} \mathrm{x}_i \langle \rho_0, \rho_i\rangle^2.$$

Show that

$$\sum_{i\in V} \mathrm{x}_i \langle \rho_0, \rho_i\rangle \rho_i = \vartheta(G;\mathbf{x})\rho_0.$$

**Exercise 5.35.** If $C$ is a convex corner and a polytope, show that $\mathtt{ab}(C)$ is a polytope. (You are allowed to use the fact that a polytope has finitely many extreme points).

## 5.6 Optimizing on perfect graphs

In this subsection, we want to argue finding a largest clique or an optimal colouring of a perfect graph can be done in polynomial time. As we saw before, computing $\vartheta(G)$ can be done in polynomial time "up to precision". I will avoid a technical discussion on how to overcome this nuisance[2], and will simply assume we have a black box that, given a perfect graph $G$ and a rational vector $\mathbf{w}$, returns a rational vector $\mathbf{x}$ so that

$$\mathbf{w}^\mathsf{T}\mathbf{x} = \max_{\mathbf{y}\in \mathtt{TH}(G)} \mathbf{w}^\mathsf{T}\mathbf{y}.$$

The problem is: maybe $\mathbf{x}$ is not integral (yet, it will be a convex combination of optimal integral solutions).

**Exercise 5.36.** How to find $\mathbf{x}$, a coclique, so that $\mathbf{w}^\mathsf{T}\mathbf{x}$ is maximum, in polynomial time?

Finding the optimal colouring is harder. Our goal below is to describe an algorithm to achieve that.

(i) Start with a maximum sized clique of $G$ — call it $K_1$, and set $t = 1$.

(ii) Let $\mathbf{w} = \mathbf{x}_{K_1} + ... + \mathbf{x}_{K_t}$. Find an independent set maximizing $\mathbf{w}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \mathtt{TH}(G)$, call it $S$. Note that $\mathbf{w}^\mathsf{T}\mathbf{x}_S = t$ (as there is indeed an independent set that intersects all maximum cliques, as we saw in the proof of Theorem 5.24). Thus $S$ intersects $K_1$,...,$K_t$.

(iii) If indeed $\omega(G-S) < \omega(G)$, then $S$ intersects all maximum cliques. Thus we can call this a colour, a return to (i) for $G - S$.

(iv) Otherwise, $\omega(G-S) = \omega(G)$. Let $K$ be a maximum clique avoiding $S$. Return to (ii), increasing $t$ and adding it to the list.

This algorithm terminates because in each iteration, the dimension of the affine space

$$\mathbb{L}_t = \{\mathbf{y} : \mathbf{y}^\mathsf{T}\mathbf{x}_{K_i} = 1 \text{ for } i = 1, ..., t\}$$

drops, as, for the $S$ found in (ii), we have $\mathbf{x}_S$ in $\mathbb{L}_t$ but not in $\mathbb{L}_{t+1}$, and it is not a combination of the previous $\mathbf{x}_S$s as its support contains at least one new entry.

[2]Don't worry, one could do it.

# 6 Lift and project methods

We will still talk about $\alpha$ and $\chi$ in detail, but let us have a break. In this section, we will see a brief description of lift-and-project methods to combinatorial optimization — and focus on two of the methods which are related to SDP.

## 6.1 Lift-and-project

Given $\mathbf{x} \in \mathbb{R}^n$, the 1-norm is given by $\|\mathbf{x}\|_1 = \sum_{i=1}^n |\mathrm{x}_i|$. The ball relative to this norm is the $n$-dimensional generalization of the (solid) octahedron in $\mathbb{R}^3$:

$$\mathbb{P} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \leq 1\} = \{\mathbf{x} \in \mathbb{R}^n : \forall\ \mathbf{a} \in \{-1,1\}^n,\ \mathbf{a}^\mathsf{T}\mathbf{x} \leq 1\}.$$

This shows that $\mathbb{P}$ is a polytope, and also displays the inequalities that define its facets. Unfortunately we have exponentially many inequalities, suggesting that this polytope would not be tractable. We can however, making use of not so many extra variables, reduce the number of constraints drastically.

**Proposition 6.1.**

$$\mathbb{P} = \{\mathbf{x} \in \mathbb{R}^n : \exists \mathbf{y} \in \mathbb{R}^n \ \textit{with}\ -\mathbf{y} \leq \mathbf{x} \leq \mathbf{y},\ \mathbf{1}^\mathsf{T}\mathbf{y} \leq 1\}.$$

**Exercise 6.2.** Prove this as an exercise. Also, show that

$$\max_{\mathbf{x}\in\mathbb{P}} \mathbf{c}^\mathsf{T}\mathbf{x} = \|\mathbf{c}\|_\infty, \quad \forall\ \mathbf{c} \in \mathbb{R}^n.$$

Now, more generally, let

$$\mathbb{P} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{b}\}.$$

Assume that $\mathbb{P} \subseteq [0,1]^n$. We introduce the notation

$$\mathbb{P}_\mathbb{Z} = \mathrm{conv}(\mathbb{P} \cap \{0,1\}^n).$$

Ideally, we would like to have relatively small $\mathbf{B}$ and $\mathbf{c}$ so that

$$\mathbb{P}_\mathbb{Z} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{B}\mathbf{x} \leq \mathbf{c}\}.$$

One typical way of doing this consists in finding valid inequalities for $\mathbb{P}_\mathbb{Z}$ and adding them to the original formulation of $\mathbb{P}$. For instance, the Gomory-Chvátal procedure: take an inequality of the form $\sum a_i x_i \leq \alpha$, with $a_i$s integers, and add the inequality $\sum a_i x_i \leq \lfloor \alpha \rfloor$ to the formulation. Eventually $\mathbb{P}_\mathbb{Z}$ will be reached, but this procedure might just take too long.

As we saw above, sometimes it is possible to achieve that goal by adding more variables and then later projecting. There are several known methods. In this section, we will discuss the methods of Lovász & Schrijver, and Lasserre.

Let $\mathbb{Q}$ be the convex cone of $\mathbb{R}^{n+1}$ generated by all 01 vectors with the first coordinate equal to 1. Any polytope $\mathbb{P}$ of $\mathbb{R}^n$ generates a polyhedral cone via homogenization in $\mathbb{R}^{n+1}$. In particular, if $\mathbb{P} \subseteq [0,1]^n$, this polyhedral cone is contained in $\mathbb{Q}$. In other words, if

$$\mathbb{P} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{b},\ \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}\},$$

then

$$\mathbb{K} = \left\{ \begin{pmatrix} \mathrm{x}_0 \\ \mathbf{x} \end{pmatrix} \in \mathbb{R}^{n+1} : \mathbf{A}\mathbf{x} \leq \mathrm{x}_0\mathbf{b} : \mathbf{0} \leq \mathbf{x} \leq \mathrm{x}_0\mathbf{1} \right\} \subseteq \mathbb{Q}.$$

Whenever $\mathbb{P}$ is given and we understand that $\mathbb{K}$ is obtained as above, we will also use $\mathbb{K}_\mathbb{Z}$ to denote the cone obtained via homogenization of $\mathbb{P}_\mathbb{Z}$.

## 6.2 Lovász-Schrijver

Given $\mathbb{K}$ a cone in $\mathbb{R}^{n+1}$, let

$$\mathcal{M}\mathbb{K} = \{\widehat{\mathbf{X}} \in \mathbb{S}^{\{0\}\cup[n]} : \widehat{\mathbf{X}}\mathbf{e}_0 = \operatorname{diag}\widehat{\mathbf{X}},\ \widehat{\mathbf{X}}\mathbf{e}_i \in \mathbb{K},\ \widehat{\mathbf{X}}(\mathbf{e}_0 - \mathbf{e}_i) \in \mathbb{K}\},$$

and

$$\mathcal{M}_+\mathbb{K} = \{\widehat{\mathbf{X}} \in \mathbb{S}_+^{\{0\}\cup[n]} : \widehat{\mathbf{X}}\mathbf{e}_0 = \operatorname{diag}\widehat{\mathbf{X}},\ \widehat{\mathbf{X}}\mathbf{e}_i \in \mathbb{K},\ \widehat{\mathbf{X}}(\mathbf{e}_0 - \mathbf{e}_i) \in \mathbb{K}\}.$$

Note that all restrictions on $\widehat{\mathbf{X}}$ are valid if there is $\mathbf{x} \in \mathbb{K}$ with $\mathbf{x}\mathbf{x}^\mathsf{T} = \widehat{\mathbf{X}}$. This is the true lifting of $\mathbb{K}$ to a higher dimensional space, and from it we then proceed to project back. If $\mathbb{K}$ is the cone generated via homogenization from a convex set $\mathbb{C}$, then

$$\mathcal{N}\mathbb{C} = \{\mathbf{x} \in \mathbb{R}^n : \exists \widehat{\mathbf{X}} \in \mathcal{M}\mathbb{K} \text{ with } \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} = \widehat{\mathbf{X}}\mathbf{e}_0\},$$

and

$$\mathcal{N}_+\mathbb{C} = \{\mathbf{x} \in \mathbb{R}^n : \exists \widehat{\mathbf{X}} \in \mathcal{M}_+\mathbb{K} \text{ with } \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} = \widehat{\mathbf{X}}\mathbf{e}_0\},$$

Let now $J \subseteq [n]$, and define the operator

$$\mathcal{F}_J\mathbb{C} = \operatorname{conv}(\mathbb{C} \cap \{\mathbf{x} \in \mathbb{R}^n : \mathrm{x}_i \in \{0,1\}\ \forall\ i \in J\}).$$

Naturally,

**Theorem 6.3.** *Let $\mathbb{C}$ be a convex set, $\mathbb{C} \subseteq [0,1]^n$. Then, for all $i \in [n]$,*

$$\mathbb{C}_\mathbb{Z} \subseteq \mathcal{N}_+\mathbb{C} \subseteq \mathcal{N}\mathbb{C} \subseteq \mathcal{F}_i\mathbb{C} \subseteq \mathbb{C}.$$

*Proof.* The first follows from observing that, if $\mathbf{x}$ is an extreme point of $\mathbb{C}_\mathbb{Z}$, then $\begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{x} \end{pmatrix}$ is in $\mathcal{M}_+\mathbb{C}$ (check this as an exercise!). The second is immediate. For the third, let $\mathbf{x} \in \mathcal{N}\mathbb{C}$, and let $\widehat{\mathbf{X}} \in \mathcal{M}\mathbb{K}$ with $\begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} = \widehat{\mathbf{X}}\mathbf{e}_0$. For the given $i$, define

$$\begin{pmatrix} \mathrm{u}_0 \\ \mathbf{u} \end{pmatrix} = \widehat{\mathbf{X}}\mathbf{e}_i \quad \text{and} \quad \begin{pmatrix} \mathrm{v}_0 \\ \mathbf{v} \end{pmatrix} = \widehat{\mathbf{X}}(\mathbf{e}_0 - \mathbf{e}_i).$$

Both these vectors are in $\mathbb{K}$ (by definition of $\mathcal{M}\mathbb{K}$), vector $\begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix}$ is their sum, and, moreover,

$$\mathbf{u} \in \mathrm{u}_0\mathbb{C} \quad \text{and} \quad \mathbf{v} \in \mathrm{v}_0\mathbb{C},$$

therefore there are $\overline{\mathbf{u}}$ and $\overline{\mathbf{v}}$ in $\mathbb{C}$ with

$$\mathbf{u} = \mathrm{u}_0\overline{\mathbf{u}} \quad \text{and} \quad \mathbf{v} = \mathrm{v}_0\overline{\mathbf{v}}.$$

Note that $\mathrm{u}_0 + \mathrm{v}_0 = 1$, both are nonnegative, and

$$\mathbf{x} = \mathrm{u}_0\overline{\mathbf{u}} + \mathrm{v}_0\overline{\mathbf{v}}.$$

It remains to show that if $\mathrm{u}_0 > 0$, then $\overline{\mathrm{u}}_i = 1$, and that if $\mathrm{v}_0 > 0$, then $\overline{\mathrm{v}}_i = 0$. This follows immediately from the facts that

$$\mathrm{u}_i = \widehat{\mathrm{X}}_{ii} = \widehat{\mathrm{X}}_{0i} = \mathrm{u}_0,$$

and

$$\mathrm{v}_i = \widehat{\mathrm{X}}_{0i} - \widehat{\mathrm{X}}_{ii} = 0.$$

The forth containment relation is also trivial. □

**Exercise 6.4.** If $J \subseteq [n]$ and $i \in [n]\backslash J$, then $\mathcal{F}_i \circ \mathcal{F}_J = \mathcal{F}_{J\cup i}$. Prove this fact, and conclude that

$$\mathcal{F}_1 \circ \mathcal{F}_2 \circ \ldots \mathcal{F}_n\mathbb{C} = \mathbb{C}_{\mathbb{Z}}.$$

Let $\mathcal{N}^0\mathbb{C} = \mathbb{C}$, and $\mathcal{N}^k\mathbb{C} = \mathcal{N}^{k-1}(\mathcal{N}\mathbb{C})$ for $k \geq 1$. Analogously for $\mathbb{N}_+$.

**Exercise 6.5.** Prove that if $\mathbb{C} \subseteq [0,1]^n$ is a convex set, then

$$\mathbb{C}_{\mathbb{Z}} = \mathcal{N}_+^n\mathbb{C} = \mathcal{N}^n\mathbb{C}.$$

To put things in context, we did know that

$$\texttt{STAB}(G) = \mathrm{conv}(\texttt{FRAC}(G) \cap \{0,1\}^n).$$

If $\mathbb{K}$ is the polyhedral cone generated via homogenization from $\mathbb{P} = \texttt{FRAC}(G)$, let $\widehat{\mathbf{X}} \in \mathcal{M}_+\mathbb{K}$, with $\widehat{\mathrm{X}}_{00} = 1$. It follows that

(i) For all $k$, whenever $ij \in E(G)$, we have

$$\widehat{\mathrm{X}}_{ik} + \widehat{\mathrm{X}}_{jk} \leq \widehat{\mathrm{X}}_{0k}.$$

(ii) For all $k$, $\widehat{\mathrm{X}}_{0k} = \widehat{\mathrm{X}}_{kk}$.

Taking $k = i$, it follows that, for all edges $ij \in E(G)$, $\widehat{\mathrm{X}}_{ij} = 0$. Therefore

$$\mathcal{N}_+(\texttt{FRAC}(G)) \subseteq \texttt{TH}(G).$$

Note that we did not use at all the third condition defining $\mathcal{M}_+\mathbb{K}$ — it was however strictly necessary to prove the convergency of the Lovász-Schrijver procedure.

Finally, we point out that Lovász and Schrijver showed that if there is a separation oracle for $\mathbb{C}$, then the separation problem in $\mathcal{N}\mathbb{C}$ and in $\mathcal{N}_+\mathbb{C}$ can be solved in polynomial time, thus guaranteeing that for fixed $k$, optimization in $\mathcal{N}_+^k\mathbb{C}$ is (at least theoretically) possible.

## 6.3 Lasserre - preparation

Let $X$ be a finite set. By $2^X$ we denote its power set, and then the standard notation for the space of real-valued functions:

$$Ł^2(2^X) = \{f : 2^X \to \mathbb{R}\}.$$

With $X$ finite, the standard inner product is

$$(f, g) = \sum_{A \in 2^X} f(A)g(A).$$

Consider the basis $\{\chi_A : A \in 2^X\}$ defined as

$$\chi_A(B) = 1 \text{ if } B \subseteq A, \text{ and } = 0 \text{ otherwise.}$$

It is straightforward to verify that it forms a basis (for instance, a correct arrangement of these functions as columns of a matrix will leave the matrix upper triangular).

**Exercise 6.6.** Verify that for all $B$ and $C$ subsets of $X$, we have

$$\chi_A(B \cup C) = \chi_A(B)\chi_A(C).$$

The dual basis $\{\chi_A^* : A \in 2^X\}$ is defined as

$$(\chi_A, \chi_B^*) = 1 \text{ if } A = B, \text{ and } = 0 \text{ otherwise.}$$

**Exercise 6.7.** Verify that this is well defined (think of matrices?) — actually, prove that

$$\chi_A^*(B) = (-1)^{|A|-|B|} \text{ if } B \subseteq A, \text{ and } = 0 \text{ otherwise.}$$

Naturally, in having two basis, we would like to write a given function $f$ in terms of both basis:

$$f = \sum_{A \in 2^X} \hat{f}(A)\chi_A = \sum_{A \in 2^X} \check{f}(A)\chi_A^*.$$

(At this point, if you know anything about harmonic analysis, you should have noted the analogy.) With no difficulty, we get that

$$(f, g) = \sum_{A \in 2^X} \hat{f}(A)\check{g}(A) = \sum_{A \in 2^X} \check{f}(A)\hat{g}(A)$$

Given $A \in 2^X$ and $f \in Ł^2(2^X)$, the shifted function $f_A$ is defined as

$$f_A(B) = f(A \cup B).$$

**Exercise 6.8.** Verify that

$$\hat{f_A}(B) = \hat{f}(B)\chi_B(A).$$

The *convolution* between functions $f$ and $g$ is

$$(f * g)(A) = (f, g_A) = \sum_{B \in 2^X} \check{f}(B)\hat{g}(B)\chi_B(A),$$

thus

$$\widehat{f * g}(B) = \check{f}(B)\hat{g}(B).$$

We say that $f \in \text{Ł}^2(2^X)$ is of *positive type* if the symmetric matrix $\mathbf{M}_f \in \mathbb{S}^{2^X}$ defined by

$$(\mathrm{M}_f)_{A,B} = f(A \cup B)$$

is positive semidefinite. Such matrices are called the moment matrices. The following theorem says that the cone of positive semidefinite moment matrices is polyhedral.

**Theorem 6.9.** *A function $f \in L^2(2^X)$ is of positive type if and only if $\hat{f}(B) \geq 0$ for all $B$.*

*Proof.* Fix a function $g$. Then

$$\begin{aligned}(g, \mathbf{M}_f\ g) &= \sum_{A,B,C} g(A)\hat{f}(C)\chi_C(A)\chi_C(B)g(B) \\ &= \sum_C \hat{f}(C) \left( \sum_A g(A)\chi_C(A) \right)^2.\end{aligned}$$

One side therefore follows immediately. To see the other, note that if $g = \chi_B^*$ for some $B$, then

$$(g, \mathbf{M}_f\ g) = \hat{f}(B),$$

thus the condition is also necessary. □

## 6.4 Lasserre hierarchy

From here on, $X = [n]$, and any set $A \in 2^X$ is identified with a vector $\mathbf{a} \in \{0,1\}^n$.

We now want to concern ourselves with problems of the form

$$\max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \{0,1\}^n,\ p_j(\mathbf{x}) \geq 0\},$$

where $p_j \in \mathbb{R}[x_1, ..., x_n]$ are square free polynomials.

We identify a square free monomial $m = x_{i_1}...x_{i_k} \in \mathbb{R}[x_1, ..., x_n]$ with

$$e_{i_1,...,i_k} = \sum_{A : \{i_1,...,i_k\} \subseteq A} \chi_A^*.$$

This definition is then extended by linearity to the polynomials where all terms are square free. Thus, given a set $A \in 2^X$ (and corresponding vector $\mathbf{a}$), and $p \in \mathbb{R}[x_1, ..., x_n]$, note that

$$p(\mathbf{a}) = (p, \chi_A) = \check{p}(A),$$

because

$$(e_{i_1,...,i_k}, \chi_A) = 1 \text{ if } \{i_1, ..., i_k\} \subseteq A, \quad \text{and } = 0 \text{ otherwise.}$$

**Theorem 6.10.** *Let $\mathbb{P} = \{\mathbf{x} \in \{0,1\}^n : p_1(\mathbf{x}) \geq 0, ..., p_m(\mathbf{x}) \geq 0\}$. Let $f \in L^2(2^{[n]})$. Then $f$ is of positive type and $p_j * f$ is of positive type for all $j = 1, ..., m$ if, and only if,*

$$\hat{f}(B) \geq 0 \text{ for all } B \subseteq X,$$

*having*

$$\hat{f}(B) = 0 \text{ when } p_j(\mathbf{b}) < 0 \text{ for some } j.$$

*Proof.* We have that

$$(p_j * f) = \sum_{B \in 2^X} \check{p}_j \hat{f}(B) \chi_B.$$

Recall Theorem 6.9. This is of positive type if and only if

$$\widehat{p_j * f}(B) = \check{p}_j(B)\hat{f}(B) = p_j(\mathbf{b})\hat{f}(B) \geq 0.$$

This condition is equivalent to having $\hat{f}(B) = 0$; or $\hat{f}(B) > 0$ and $p_j(\mathbf{b}) \geq 0$ for all $j$, as we wanted. □

Consider the original problem

$$\max\{\mathbf{c}^\mathsf{T}\mathbf{x} : \mathbf{x} \in \{0,1\}^n,\ p_j(\mathbf{x}) \geq 0\}.$$

Following, consider the linear optimization problem

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} \mathrm{c}_i f(\{i\}) \\ \text{subject to} \quad & f \in \mathrm{L}^2(2^X), \\ & f \text{ of positive type }, \\ & f(\emptyset) = 1, \\ & p_j * f \text{ of positive type for all } j. \end{aligned}$$

Note that any $f$ satisfying the restrictions is of the form

$$f = \sum_{\mathbf{b} \in \mathbb{P}} \hat{f}(B) \chi_B.$$

Moreover,

$$1 = f(\emptyset) = \sum_{\mathbf{b} \in \mathbb{P}} \hat{f}(B).$$

So the optimum of the linear optimization problem is of the form $\overline{f} = \chi_B$, where $\chi_B$ maximizes the linear function over all $B$ with $\mathbf{b} \in \mathbb{P}$. Naturally, such $\mathbf{b}$ is an optimum for the original problem.

The linear optimization problem is clearly equivalent to the SDP

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} \mathrm{c}_i (\mathrm{M}_f)_{\{i\},\{i\}} \\ \text{subject to} \quad & \mathbf{M}_f \in \mathbb{S}_+^{2^X}, \\ & (\mathrm{M}_f)_{\emptyset,\emptyset} = 1, \\ & \mathbf{M}_{p_j * f} \in \mathbb{S}_+^{2^X} \text{ for all } j. \end{aligned}$$

This SDP formulation above, as stated, is useless for practical purposes. The Lasserre hierarchy consists of, at step $t$, truncating $\mathbf{M}_f$ to the rows and columns indexed by subsets of cardinality at most $t+1$, and doing the same for $\mathbf{M}_{p_j * f}$ (actually for this one you can truncate before, but it is not really relevant here).

**Exercise 6.11.** We want to use Lasserre hierarchy to approximate $\texttt{STAB}(G)$. For that, we will use polynomials $p_{ab}(x) = 1 - x_a - x_b \in \mathbb{R}[x_v : v \in V(G)]$, for all $ab \in E(G)$. Of course, having $p_{ab}(x) \geq 0$ for all $ab \in E(G)$ is equivalent to defining $\texttt{FRAC}(G)$, which is where we start.

Given $f : 2^V \to \mathbb{R}$, let $\mathbf{M}_f^t$ and $\mathbf{M}_{p_{ab} * f}^t$ be the truncations of $\mathbf{M}_f$ and $\mathbf{M}_{p_{ab} * f}$ to the subsets of at most $t$ elements.

Show that the following are all equivalent, for $t \geq 1$.

(a) $\mathbf{M}_f^{t+1} \succcurlyeq \mathbf{0}$, and $\mathbf{M}_{p_{ab} * f}^t \succcurlyeq \mathbf{0}$ for all $ab \in E(G)$.

(b) $\mathbf{M}_f^{t+1} \succcurlyeq \mathbf{0}$, sand $f(\{a,b\}) = 0$ for all $ab \in E(G)$.

(c) $\mathbf{M}_f^{t+1} \succcurlyeq \mathbf{0}$, and $f(U) = 0$ for all $U \subseteq V(G)$ which is not independent, with $|U| \leq 2t+2$.

Start by noting that $\mathbf{M}_f^{t+1} \succcurlyeq \mathbf{0}$ implies $f(U) \geq 0$ for all $U$ subset of at most $2t+2$ elements.

# 7 Conic optimization for cocliques and colourings

In this section, we introduce more aspects of the theory of conic programming related to the optimization of cocliques and colourings. Whereas in the previous section we focused on $\vartheta(G;\cdot)$, and later on the relationship between $\texttt{TH}(G)$, and $\texttt{STAB}(G)$ and $\texttt{QSTAB}(G)$, in this section the topics will be more independent. First, we will see how to formulate the problems of finding $\alpha$ and $\chi$ as conic programs. Second, we will introduce more parameters between $\alpha$ and $\chi$. They will be motivated by the well known vector colourings. Then, we will see how this can lead to a nice approximation algorithm to 3-colourable graphs. Finally, I intend to discuss connections to symmetries of a graph.

## 7.1 Co(mpletely )positive cones

A matrix $\mathbf{M}$ is called *completely positive* if, for some $k$, there are nonnegative vectors $\mathbf{x}_1, ..., \mathbf{x}_k \in \mathbb{R}^n_+$ so that

$$\mathbf{M} = \mathbf{x}_1\mathbf{x}_1^\mathsf{T} + \ldots + \mathbf{x}_k\mathbf{x}_k^\mathsf{T}$$

Clearly, completely positive matrices are positive semidefinite, but the converse does not hold.

**Exercise 7.1.** Why not?

We will denote

$$\mathbb{S}^n_{\text{clyp}} = \{\mathbf{M} \in \mathbb{S}^n : \mathbf{M} \text{ is completely positive.}\}$$

The completely positive rank of a completely positive matrix $\mathbf{M}$ is the minimum $k$ so that $\mathbf{M}$ writes as a sum of $k$ nonnegative rank-1 projectors, just as above.

**Exercise 7.2.** If $\mathbf{M}$ is a $n \times n$ completely positive, show that its completely positive rank is upper bounded by $\binom{n+1}{2}$. (Hint: this is actually the dimension of the space of symmetric $n \times n$ matrices.)

A symmetric matrix $\mathbf{M}$ is called *copositive* if, for all $\mathbf{x} \in \mathbb{R}^n_+$, we have

$$\mathbf{x}^\mathsf{T}\mathbf{M}\mathbf{x} \geq 0.$$

Clearly, all positive semidefinite matrices are copositive, but the converse does not hold.

**Exercise 7.3.** Why not?

We will denote

$$\mathbb{S}^n_{\text{copo}} = \{\mathbf{M} \in \mathbb{S}^n : \mathbf{M} \text{ is copositive.}\}$$

**Theorem 7.4.** *The sets $\mathbb{S}^n_{copo}$ and $\mathbb{S}^n_{clyp}$ are closed, convex conves, and duals of each other.*

*Proof.* I leave it as an exercise to show that both sets are convex cones. If $\mathbf{M} \notin \mathbb{S}^n_{\text{copo}}$, then there is $\mathbf{x} \geq \mathbf{0}$ with $\mathbf{x}^\mathsf{T}\mathbf{M}\mathbf{x} < 0$. Any small variation around $\mathbf{M}$ will not change this fact, thus $\mathbb{S}^n_{\text{copo}}$ is closed. To see that $\mathbb{S}^n_{\text{clyp}}$ is closed, we will show that the limit of any convergent

sequence in it belongs to it. Let $(\mathbf{M}^{(k)})_{k\geq 0} \in \mathbb{S}^n_{\text{clyp}}$ be a convergent sequence, converging to $\mathbf{M}$. Let $\mathbf{A}^{(k)} \geq \mathbf{0}$ be so that

$$\mathbf{M}^{(k)} = \mathbf{A}^{(k)}(\mathbf{A}^{(k)})^\mathsf{T}.$$

Thus the $i$th row of $\mathbf{A}^{(k)}$, for $k \to \infty$, for a bounded sequence of vectors. Thus this sequence has a convergent subsequence, say to $\mathbf{a}_i$, with $\mathbf{a}_i \geq \mathbf{0}$. If $\mathbf{A}$ has these as its columns, it follows that

$$\mathbf{M} = \mathbf{A}\mathbf{A}^\mathsf{T}.$$

Finally, if $\mathbf{M} \in \mathbb{S}^n_{\text{copo}}$, and $\mathbf{N} = \sum_{i=1}^k \mathbf{x}_k\mathbf{x}_k^\mathsf{T}$ with $\mathbf{x}_i \geq \mathbf{0}$, then

$$\langle \mathbf{M}, \mathbf{N}\rangle = \sum_{i=1}^k \langle \mathbf{M}, \mathbf{x}_i\mathbf{x}_i^\mathsf{T}\rangle \geq 0,$$

hence $\mathbf{M} \in (\mathbb{S}^n_{\text{clyp}})^*$. For the converse, if $\mathbf{M} \notin \mathbb{S}^n_{\text{copo}}$, there is $\mathbf{x} \geq \mathbf{0}$ with $\mathbf{x}^\mathsf{T}\mathbf{M}\mathbf{x} < 0$. Consequently, $\mathbf{M} \notin (\mathbb{S}^n_{\text{clyp}})^*$. Because they are closed convex cones, we also get

$$(\mathbb{S}^n_{\text{copo}})^* = \mathbb{S}^n_{\text{clyp}}.$$

□

**Exercise 7.5.** Describe as best as you can the cones $\mathbb{S}^2_{\text{copo}}$ and $\mathbb{S}^2_{\text{clyp}}$.

## 7.2 Conic program for the independence number

We now consider the following primal-dual pair of conic programs.

$$\text{(P)} \quad \begin{array}{ll} \max & \langle \mathbf{J}, \mathbf{X}\rangle \\ \text{s.t.} & \mathrm{X}_{ij} = 0 \ \forall\ ij \in E(G) \\ & \operatorname{tr}\mathbf{X} = 1 \\ & \mathbf{X} \in \mathbb{S}^n_{\text{clyp}} \end{array} \qquad \text{(D)} \quad \begin{array}{ll} \min & \lambda \\ \text{s.t.} & \lambda\mathbf{I} - \mathbf{J} - \mathbf{Y} \in \mathbb{S}^n_{\text{copo}} \\ & \mathrm{Y}_{ij} = 0 \ \forall\ ij \notin E(G). \end{array}$$

We will now show that the optimum of both programs above is $\alpha(G)$.

**Lemma 7.6.** *The optimum of (P) is $\geq \alpha(G)$.*

*Proof.* Let $S$ be a maximum sized coclique, with characteristic vector $\mathbf{x}$. Simply note now that

$$\mathbf{X} = \frac{1}{\alpha(G)}\mathbf{x}\mathbf{x}^\mathsf{T}$$

is feasible for (P), with objective value $\alpha(G)$. □

It remains to show that the optimum of (D) is $\leq \alpha(G)$. To achieve this, we will use an influential result due to Motzkin and Straus.

**Theorem 7.7.** *For any graph $G$,*

$$\frac{1}{\alpha(G)} = \min\{\mathbf{x}^\mathsf{T}(\mathbf{A}+\mathbf{I})\mathbf{x} : \mathbf{1}^\mathsf{T}\mathbf{x} = 1,\ \mathbf{x} \geq \mathbf{0}\}.$$

*Proof.* Let $f(G)$ be the optimum of this program. If $S$ is a coclique, with characteristic vector $\mathbf{x}$, then $(1/|S|)\mathbf{x}$ is feasible, and

$$\frac{1}{|S|^2}\mathbf{x}^\mathsf{T}(\mathbf{A}+\mathbf{I})\mathbf{x} = \frac{1}{|S|},$$

thus $f(G) \leq \alpha^{-1}$.

The other direction is more involved, and goes by induction on $n = |V(G)|$. If $n = 1$, the result is trivial. Assume now $f(H) \geq \alpha(G)^{-1}$ for all proper induced subgraphs $H$ of $G$. Let $\mathbf{y}$ be an optimum giving $f(G)$. If one entry $\mathrm{y}_i = 0$, then we are lucky, and setting $H = G - i$, and $\overline{\mathbf{y}}$ the restriction of $\mathbf{y}$ to $H$, we have

$$\alpha(G)^{-1} = \alpha(H)^{-1} \leq f(H) \leq \overline{\mathbf{y}}^\mathsf{T}(\mathbf{A}(H)+\mathbf{I})\overline{\mathbf{y}} = \mathbf{y}^\mathsf{T}(\mathbf{A}(G)+\mathbf{I})\mathbf{y} = f(G).$$

Let $\mathbf{y}$ be a minimizer for $f(G)$, and $\mathbf{y} > \mathbf{0}$. Pick $ij \in E(G)$ and define $\mathbf{z}$ with $\mathrm{z}_i = \mathrm{y}_i + \varepsilon$, $\mathrm{z}_j = \mathrm{y}_j - \varepsilon$, and otherwise $\mathbf{z}_\ell = \mathbf{y}_\ell$. Clearly

$$\mathbf{z}^\mathsf{T}(\mathbf{A}+\mathbf{I})\mathbf{z} = f(G) + \varepsilon\left[\mathbf{y}^\mathsf{T}(\mathbf{A}+\mathbf{I})(\mathbf{e}_i-\mathbf{e}_j) + (\mathbf{e}_i-\mathbf{e}_j)^\mathsf{T}(\mathbf{A}+\mathbf{I})\mathbf{y}\right] + \varepsilon^2(\mathbf{e}_i-\mathbf{e}_j)^\mathsf{T}(\mathbf{A}+\mathbf{I})(\mathbf{e}_i-\mathbf{e}_j)$$

The term on $\varepsilon^2$ vanishes, as $ij \in E(G)$. The term on $\varepsilon$ vanishes because $\mathbf{y}$ is a minimizer. Thus, we can choose $\varepsilon$ so that $\mathbf{z}$ becomes 0 at a coordinate, and apply induction on the graph with the vertex corresponding to this coordinate removed. □

**Exercise 7.8.** Show that

$$\max\{\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} : \mathbf{1}^\mathsf{T}\mathbf{x} = 1,\ \mathbf{x} \geq \mathbf{0}\} = \frac{1}{2}\left(1 - \frac{1}{\omega(G)}\right).$$

**Exercise 7.9.** In the program

$$\begin{aligned} \min \quad & \lambda \\ \text{subject to} \quad & \lambda\mathbf{I} - \mathbf{J} - \mathbf{Y} \in \mathbb{S}^n_{\text{copo}} \\ & \mathrm{Y}_{ij} = 0 \ \forall\ ij \notin E(G), \end{aligned}$$

argue why there is an optimum solution with $\mathbf{Y}$ having all entries corresponding to edges equal to a constant.

**Theorem 7.10.** *The optimum of (D) is at most $\alpha(G)$.*

*Proof.* From the exercise above, we need to show that

$$\alpha(G) \geq \min\{\lambda : \lambda\mathbf{I} - \mathbf{J} - z\mathbf{A} \in \mathbb{S}^n_{\text{copo}},\ \lambda, z \in \mathbb{R}\}.$$

From Motzkin Straus, it follows that, for all $\mathbf{x}$ with $\mathbf{1}^\mathsf{T}\mathbf{x} = 1$ and $\mathbf{x} \geq \mathbf{0}$, we have

$$\alpha(G)\mathbf{x}^\mathsf{T}(\mathbf{A}+\mathbf{I})\mathbf{x} \geq 1 = \mathbf{x}^\mathsf{T}\mathbf{J}\mathbf{x}.$$

Therefore, for all $\mathbf{x} \geq \mathbf{0}$, we have

$$\mathbf{x}^\mathsf{T}(\alpha\mathbf{A} + \alpha\mathbf{I} - \mathbf{J})\mathbf{x} \geq 0.$$

Thus the matrix $(\alpha\mathbf{A} + \alpha\mathbf{I} - \mathbf{J})$ is copositive, and thus a feasible solution to the program above, with value $\alpha(G)$. □

**Exercise 7.11.** Show that

$$\alpha(G)^{-1} = \min\{\langle \mathbf{A}+\mathbf{I}, \mathbf{X}\rangle : \langle \mathbf{J}, \mathbf{X}\rangle = 1,\ \mathbf{X} \in \mathbb{S}^n_{\text{clyp}}\}.$$

## 7.3 Conic programming for the chromatic number

We now consider the following primal-dual pair of conic programs.

$$\text{(P)}\quad \begin{array}{ll} \max & \langle \mathbf{J}, \mathbf{X}\rangle \\ \text{s.t.} & \mathrm{X}_{ij} + \mathrm{Z}_{ij} = 0 \ \forall\ ij \in E(G) \\ & \mathrm{tr}(\mathbf{X}+\mathbf{Z}) = 1 \\ & \mathbf{X} \in \mathbb{S}^n_+,\ \mathbf{Z} \in \mathbb{S}^n_{\text{copo}} \end{array} \qquad \text{(D)}\quad \begin{array}{ll} \min & \lambda \\ \text{s.t.} & \lambda\mathbf{I} - \mathbf{J} - \mathbf{Y} \in \mathbb{S}_+ \\ & \lambda\mathbf{I} - \mathbf{Y} \in \mathbb{S}_{\text{clyp}} \\ & \mathrm{Y}_{ij} = 0 \ \forall\ ij \notin E(G). \end{array}$$

We will now show that the optimum of both programs above is $\chi_f(\overline{G})$.

**Lemma 7.12.** *The optimum of (D) is* $\leq \chi_f(\overline{G})$.

*Proof.* Assume $\mathbf{y}$ is an optimum for

$$\min\{\mathbf{1}^T\mathbf{y} : \mathbf{M}\mathbf{y} = \mathbf{1},\ \mathbf{y} \geq \mathbf{0}\}.$$

Recalling that $\chi_f = \mathbf{1}^\mathsf{T}\mathbf{y}$, consider the matrix

$$\mathbf{W} = \chi_f \sum_i \mathrm{y}_i \mathbf{x}_{K_i}\mathbf{x}_{K_i}^\mathsf{T} = \sum_{i,j} \mathrm{y}_i\mathrm{y}_j \mathbf{x}_{K_i}\mathbf{x}_{K_i}^\mathsf{T},$$

where the sums run over all cliques of the graph. This matrix is clearly completely positive, and, moreover, its diagonal is constant equal to $\chi_f$. We also have $\mathrm{W}_{ij} = 0$ for all $ij \in E(\overline{G})$. So it remains to show that $\mathbf{W} - \mathbf{J} \succcurlyeq \mathbf{0}$. Recalling that $\mathbf{M}\mathbf{y} = \mathbf{1}$, it follows that

$$\mathbf{W} - \mathbf{J} = \sum_{i,j} \mathrm{y}_i\mathrm{y}_j \mathbf{x}_{K_i}\mathbf{x}_{K_i}^\mathsf{T} - \sum_{i,j} \mathrm{y}_i\mathrm{y}_j \mathbf{x}_{K_i}\mathbf{x}_{K_j}^\mathsf{T}.$$

Thus, given any $\mathbf{x} \in \mathbb{R}^n$, and having $a_i = \mathbf{x}_{K_i}^\mathsf{T}\mathbf{x}$, it follows that

$$\mathbf{x}^\mathsf{T}(\mathbf{W} - \mathbf{J})\mathbf{x} = \sum_{i,j} \mathrm{y}_i\mathrm{y}_j(a_i^2 - a_i a_j) = \sum_{i<j} \mathrm{y}_i\mathrm{y}_j(a_i^2 - 2a_i a_j + a_j^2) \geq 0.$$

□

**Lemma 7.13.** *The optimum of (P) is* $\geq \chi_f(\overline{G})$.

**Exercise 7.14.** Prove this Lemma by exhibiting a feasible solution for (P) with objective value $\chi_f(\overline{G})$.

## 7.4 Vector colourings and theta variants

A graph $G$ has an $\alpha$-vector colouring if there is an assignment of unit vectors $\{\mathbf{v}_a : a \in V(G)\} \subseteq \mathbb{R}^d$ so that, if $a \sim b$, then

$$\langle \mathbf{a}, \mathbf{b}\rangle \leq \frac{1}{1-\alpha}.$$

The smallest $\alpha > 1$ so that $G$ has an $\alpha$-vector colouring is called the *vector chromatic number* of $G$, to be denoted by $\chi_{\text{vec}}(G)$ (if $G$ has no edge, this is defined to be equal to 1). The existence of an $\alpha$-vector colouring is equivalent to the existence of a matrix $\mathbf{W}$ so that

(i) $\mathbf{W} \succcurlyeq \mathbf{0}$.

(ii) $\operatorname{diag} \mathbf{W} = \mathbf{1}$.

(iii) $W_{ij} \leq 1/(1-\alpha)$ for all $ij \in E(G)$.

The existence of such $\mathbf{W}$ is equivalent to the existence of $\mathbf{Y}$ so that

$$(\alpha - 1)\mathbf{W} = \alpha\mathbf{I} - \mathbf{Y} - \mathbf{J},$$

where $Y_{ij} \geq 0$ if $ij \in E(G)$, and $Y_{ii} = 0$. Given that $\alpha > 1$, it follows that the vector chromatic number of $G$ is the optimum of the SDP

$$\begin{aligned} \min \quad & \alpha \\ \text{subject to} \quad & \alpha\mathbf{I} - \mathbf{Y} - \mathbf{J} \succcurlyeq \mathbf{0} \\ & Y_{ij} \geq 0 \quad \text{for all } ij \notin E(\overline{G}). \end{aligned}$$

This formulation is dual to

$$\begin{aligned} \max \quad & \langle \mathbf{J}, \mathbf{X} \rangle \\ \text{subject to} \quad & \operatorname{tr} \mathbf{X} = 1 \\ & X_{ij} = 0 \quad \text{for all } ij \in E(\overline{G}) \\ & \mathbf{X} \geq \mathbf{0},\ \mathbf{X} \succcurlyeq \mathbf{0}. \end{aligned}$$

Given both formulations above, and recalling the formulation for $\alpha(G)$ as a conic program over $\mathbb{S}^n_{\text{clyp}}$, which in particular belongs to the set of matrices which are non-negative and positive semidefinite, the corollary below follows immediately:

**Corollary 7.15.** *For any graph $G$,*

$$\omega(G) \leq \chi_{vec}(G) \leq \vartheta(\overline{G}).$$

Note that $\chi_{\text{vec}}(G)$ can be computed efficiently.

**Exercise 7.16.** Prove in an elementary way that $\omega(G) \leq \chi_{\text{vec}}(G)$. Hint: consider an optimal $\alpha$-vector colouring, and sum the vectors corresponding to a maximum clique. Then consider the norm of this resulting vector.

If we had enforced that the vector colouring satisfies inner product inequality with an equality, we would have arrived at what is known as the *strict vector chromatic number*. This is denoted by $\chi_{\text{svec}}(G)$, and formally defined as the lest $\alpha > 1$ so that there is an assignment of unit vectors $\{\mathbf{v}_a : a \in V(G)\} \subseteq \mathbb{R}^d$ so that, if $a \sim b$, then

$$\langle \mathbf{a}, \mathbf{b} \rangle = \frac{1}{1-\alpha}.$$

**Exercise 7.17.** Verify that $\chi_{\text{svec}}(G) = \vartheta(\overline{G})$. Hint: look at the SDP formulations above.

An $\alpha$-strict vector colouring which further satisfies

$$\langle \mathbf{a}, \mathbf{b} \rangle \geq \frac{1}{1-\alpha}.$$

for all $a, b \in V(G)$ is called a *rigid vector colouring*. The least $\alpha$ so that $G$ has a rigid vector colouring is denoted by $\chi_{\text{rvec}}(G)$. Note that

$$\chi_{\text{vec}}(G) \leq \chi_{\text{svec}}(G) \leq \chi_{\text{rvec}}(G),$$

as these are all minimization problems with increasing degrees of restrictions. Moreover, it is immediate to verify that $\chi_{\text{rvec}}(G)$ is defined by the following pair of primal-dual SDPs:

$$\text{(P)} \quad \begin{array}{ll} \min & \alpha \\ \text{s.t.} & \alpha \mathbf{I} - \mathbf{Y} - \mathbf{J} \succcurlyeq \mathbf{0} \\ & Y_{ij} = 0 \quad \text{for all } ij \notin E(\overline{G}) \\ & \mathbf{Y} \leq \mathbf{0}. \end{array} \qquad \text{(D)} \quad \begin{array}{ll} \max & \langle \mathbf{J}, \mathbf{X} \rangle \\ \text{s.t.} & \operatorname{tr} \mathbf{X} = 1 \\ & X_{ij} \leq 0 \quad \text{for all } ij \in E(\overline{G}) \\ & \mathbf{X} \succcurlyeq \mathbf{0}. \end{array}$$

**Exercise 7.18.** Check the definition of $\chi_f(\overline{G})$ as a conic program, and prove that $\chi_f(G) \geq \chi_{\text{rvec}}(G)$.

Historically, the vector chromatic number was introduced as a variant of $\vartheta$ by Schrijver, and it was originally denoted by $\vartheta'$ (and called, later, by Schrijver's Theta). The parameter $\chi_{\text{rvec}}$ was introduced by Szegedy also as a $\vartheta$ variant. If one decides to denote $\chi_{\text{vec}}(G) = \vartheta^-(\overline{G})$ and $\chi_{\text{rvec}}(G) = \vartheta^+(\overline{G})$, the inequalities seen above take the pleasant form

$$\alpha(G) \leq \vartheta^-(G) \leq \vartheta(G) \leq \vartheta^+(G) \leq \chi_f(\overline{G}).$$

Perhaps remarkably, all these inequalities can be very naturally extended to the definition of functions from $\mathbb{R}^n_+$ to $\mathbb{R}$, as we did for $\vartheta$, having these parameters defined above taken to be the evaluation of such functions at $\mathbf{1}$.

**Exercise 7.19.** Let $G$ be a graph on $n$ vertices and $m$ edges, and $\mathbf{A}$ be its adjacency matrix with largest eigenvalue $\theta$ and smallest eigenvalue $\tau$. Show that

$$1 - \frac{\theta}{\tau} \leq \chi_{\text{vec}}(G).$$

Hint: you want to find a feasible solution to one of its SDP formulations. Start by finding one that gives objective value $1 - \frac{2m/n}{\tau}$, and think on how to improve it.

**Exercise 7.20.** If $G$ is a $k$-regular graph, show that

$$\chi_{\text{rvec}}(\overline{G}) \leq \frac{n(-\tau)}{k - \tau}.$$

# Index