Chapter 4

Interior-Point Methods

In this chapter we are concerned with the problem of solving a semidefinite program (1.2) in equational form:

$$\begin{array}{ll} \text{Maximize} & \mathrm{Tr}(C^T X) \\ \text{subject to} & A(X) = \mathbf{b} \\ & X \succeq 0. \end{array} \tag{4.1}$$

The main idea behind all *central path interior point methods* is to get rid of the "difficult" constraint $X \succeq 0$, by adding a "barrier term" to the objective function that tends to $-\infty$ as we approach the boundary of the set $\mathcal{S}_n^+ = \{X \in S_n : X \succeq 0\}$ from the interior. Having done this, we can drop the constraints $X \succeq 0$ as they will be "non-binding" at optimality.

4.1 The Auxiliary Problem

Here is a concrete realization of this idea. For any real number $\mu > 0$, we consider the auxiliary problem

$$\begin{array}{ll} \text{Maximize} & f_\mu(X) := \mathrm{Tr}(C^T X) + \mu \ln\det X \\ \text{subject to} & A(X) = \mathbf{b} \\ & X \succ 0, \end{array} \tag{4.2}$$

where $X \succ 0$ means that X is positive definite (all eigenvalues are strictly positive). As all matrices on the boundary of $\mathcal{S}_n^+$ have at least one eigenvalue equal to 0, they are singular and satisfy $\det X = 0$. Thus, $\mu \ln\det X$ is indeed a barrier term in the above sense, by continuity of $\ln\det X$.

We would like to claim that under suitable conditions, the auxiliary problem has a unique optimal solution $X^*(\mu)$ for every $\mu > 0$, and that $\mathrm{Tr}(C^T X^*(\mu))$ converges to the value of (4.1) as $\mu \to 0$. Obviously, we need to assume that there is a feasible $X \succ 0$, but other conditions will be needed as well.

To justify the claim, we proceed in the following steps, always assuming that μ is strictly positive.

(i) We show that *if* the problem (4.2) has an optimal solution, then it has a *unique* optimal solution $X^*(\mu)$.

(ii) We derive necessary conditions for the existence of $X^*(\mu)$, in the form of a system of equations and inequalities that $X^*(\mu)$ has to satisfy. These conditions will also imply the desired convergence of $\mathrm{Tr}(C^T X^*(\mu))$.

(iii) We prove that the necessary conditions are also sufficient: the system derived in (ii) has a unique solution $X = X^*(\mu)$.

4.2 Uniqueness of Solution

Here is the first step in the plan.

4.2.1 Lemma. *If f_μ attains a maximum over the feasible region of (4.2), then f_μ attains a unique maximum.*

Proof. This easily follows from the fact that f_μ is *strictly concave* over the interior of $\mathcal{S}_n^+$ (see next lemma), meaning that for all $X, Y \succ 0$ with $X \neq Y$,

$$f_\mu((1-t)X + tY) > (1-t)f_\mu(X) + tf_\mu(Y), \quad 0 < t < 1.$$

Indeed, if the maximum would be attained for two different matrices X^* and Y^*, then strict concavity would imply that $(X^* + Y^*)/2$ has even higher f_μ-value, a contradiction. □

4.2.2 Lemma. *The function $X \mapsto \ln\det X$ is strictly concave over the interior of $\mathcal{S}_n^+$.*

Since $\mathrm{Tr}(C^T X)$ is linear in X, this also implies strict concavity of f_μ for any $\mu > 0$.

Proof. Let us first recall how we can prove strict concavity of a *twice differentiable* function in one real variable over an open set U. A sufficient condition from analyis is that the second derivative is negative throughout U. For example, to prove that $f(x) = \ln(x)$ is strictly concave over the positive numbers, we compute $f'(x) = 1/x$ and $f''(x) = -1/x^2 < 0$ for all $x > 0$.

For fixed $X, Y \succ 0, X \neq Y$, we define

$$g(t) = \ln\det((1-t)X + tY).$$

If we can prove that g is strictly concave over $[0, 1]$, we are done, since then

$$\underbrace{\ln\det((1-t)X + tY)}_{g((1-t)0+t1)} > \underbrace{(1-t)\ln\det X + t\ln\det Y}_{(1-t)g(0)+tg(1)}.$$

By $X, Y \succ 0$, g is actually defined on some open interval containing $[0,1]$, so we can indeed compute derivatives throughout $[0,1]$. Since the determinant is a polynomial, g is infinitely often differentiable.

Let us write $g(t)$ as

$$g(t) = \ln\det(X + tZ),\ \ Z = Y - X \in \mathcal{S}_n.$$

Since $X \succ 0$, there is a regular matrix U such that $X = U^T U$, and we can write

$$X + tZ = U^T(I_n + t(U^{-1})^T Z U^{-1})U.$$

Using $\det U = \det U^T$ together with multiplicativity of the determinant and the rules of logarithms, this gives

$$\begin{aligned} \ln\det(X + tZ) &= 2\ln\det U + \ln\det(I_n + t(U^{-1})^T Z U^{-1}) \\ &= \ln\det(X) + \ln\det(I_n + tZ'), \end{aligned}$$

with $Z' := (U^{-1)^T} Z U^{-1} \in \mathcal{S}_n$. Let Z' have eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $I_n + tZ'$ has eigenvalues $1 + t\lambda_1, \ldots, 1 + t\lambda_n$, and since the determinant of a symmetric matrix is the product of its eigenvalues (another fact that we get from diagonalization), we further have

$$g(t) = \ln\det(X + tZ) = \ln\det X + \sum_{i=1}^{n} \ln(1 + t\lambda_i).$$

This yields

$$g'(t) = \sum_{i=1}^{n} \frac{\lambda_i}{1 + t\lambda_i},$$

and using the quotient rule we obtain

$$g''(t) = -\sum_{i=1}^{n} \frac{\lambda_i^2}{(1 + t\lambda_i)^2} < 0$$

since $X \neq Y$ implies that $Z' \neq 0$, so least one of the λ_i is nonzero. □

4.3 Necessary Conditions for Optimality

According to the previous section, we know that if the auxiliary problem has an optimal solution at all, then it has a unique optimal solution $X^*(\mu)$. Now we use the method of *Lagrange multipliers* from analysis to derive a system of equations and inequalities that $X^*(\mu)$ has to satify if it exists at all.

4.3.1 The Method of Lagrange Multipliers

We recall that this is a general method for finding a (local) maximum of $f(\mathbf{x})$ subject to m constraints $g_1(\mathbf{x}) = 0$, $g_2(\mathbf{x}) = 0, \ldots,$ $g_m(\mathbf{x}) = 0$, where f and $g_1, \ldots, g_m$ are functions from $\mathbb{R}^n$ to $\mathbb{R}$. It can be seen as a generalization of the basic calculus trick for maximizing a univariate function by seeking a zero of its derivative. It introduces the following system of equations with unknowns $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ (the y_i are auxiliary variables called the *Lagrange multipliers*):

$$g_1(\mathbf{x}) = g_2(\mathbf{x}) = \cdots = g_m(\mathbf{x}) = 0 \quad \text{and} \quad \nabla f(\mathbf{x}) = \sum_{i=1}^{m} y_i \nabla g_i(\mathbf{x}). \tag{4.3}$$

Here ∇ denotes the gradient (which by convention is a *row* vector):

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right).$$

That is, ∇f is a vector function from $\mathbb{R}^n$ to $\mathbb{R}^n$ whose ith component is the partial derivative of f with respect to x_i. Thus, the equation $\nabla f(\mathbf{x}) = \sum_{i=1}^m y_i \nabla g_i(\mathbf{x})$ stipulates the equality of two n-component vectors. The method of Lagrange multipliers tells us that if $\tilde{\mathbf{x}}$ is a local maximum of $f(\mathbf{x})$ subject to $g_1(\mathbf{x}) = g_2(\mathbf{x}) = \cdots = g_m(\mathbf{x}) = 0$, then $\tilde{\mathbf{x}}$ satisfies (4.3); that is, there exists $\mathbf{y}$ such that $\tilde{\mathbf{x}}$ and this $\mathbf{y}$ together fulfill (4.3). For this to hold, we need some conditions on f, the g_i and $\tilde{\mathbf{x}}$. It suffices to require that f and the g_i be defined on a nonempty open subset of $\mathbb{R}^n$ and have continuous first partial derivatives there, and this will be obviously satisfied in our simple application. For $\tilde{\mathbf{x}}$, we need to require that it is a *regular point*, meaning that the vectors $\nabla g_i(\tilde{\mathbf{x}})$ are linearly independent [14]. However, if the constraint functions g_i are *linear* (as in our application), this requirement on $\tilde{\mathbf{x}}$ can be dropped (Exercise 4.7.1).

4.3.2 Application to the Auxiliary Problem

For the auxiliary problem (4.2), we want to work with

$$f(X) = f_\mu(X), \quad g_i(X) = A(X)_i - b_i, \; i = 1, \ldots, m.$$

Indeed, if $X^*(\mu)$ is a maximum of (4.2), it is also a (local) maximum subject to only the constraints $g_i(X) = 0$, because of $X^*(\mu) \succ 0$. Since the g_i are linear, we do not have to worry about regularity of $X^*(\mu)$.

A small technical problem is that the functions f and g_i are in our setting not defined on $\mathbb{R}^n$ but rather on $\mathcal{S}_n$, the vector space of symmetric $n \times n$ matrices. The problem can be resolved for example by identifying a matrix $X \in \mathcal{S}_n$ with the $\binom{n+1}{2}$-vector of matrix entries on and above the diagonal, written row by row.

To simplify the computations, we will for the moment forget about the symmetry of X and compute all derivatives over the space $\mathbb{R}^{n \times n}$ of $n \times n$ matrices which we identify with $\mathbb{R}^{n^2}$ by writting the n^2 entries as one long vector, row by row. Under this identification, $\mathrm{Tr}(X^TY)$ corresponds to the "normal" scalar product over $\mathbb{R}^{n^2}$.

Here is the matrix analog of the fact that $\ln'(x) = 1/x$.

4.3.1 Lemma. *For* $X \in \mathbb{R}^{n\times n}$ *such that* $\det X > 0$,

$$\nabla \ln\det X = (X^T)^{-1}.$$

Formally, $\nabla \ln\det X$ is a (row) vector in $\mathbb{R}^{n^2}$, but for the statement of the lemma, we reinterpret it as matrix in $\mathbb{R}^{n\times n}$.

Proof. By the chain rule, we have

$$\frac{\partial \ln\det X}{\partial x_{ij}} = \frac{1}{\det X} \cdot \frac{\partial \det X}{\partial x_{ij}}.$$

To compute the latter partial derivative, we write X as a sequence of columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and directly apply the definition of the derivative:

$$\frac{\partial \det X}{\partial x_{ij}} = \lim_{h\to 0} \frac{\det(\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_j + h\mathbf{e}_i, \mathbf{x}_{j+1}, \ldots, \mathbf{x}_n) - \det(\mathbf{x}_1, \ldots, \mathbf{x}_n)}{h}.$$

Using that the determinant is linear in every argument, this further yields

$$\frac{\partial \det X}{\partial x_{ij}} = \lim_{h\to 0} \frac{h \det(\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{e}_i, \mathbf{x}_{j+1}, \ldots, \mathbf{x}_n)}{h} = \det X^{ij},$$

where X^{ij} is the matrix obtained from X by replacing the j-th column with $\mathbf{e}_i$. We thus have

$$\frac{\partial \ln\det X}{\partial x_{ij}} = \frac{\det X^{ij}}{\det X},$$

and it is known (Cramer's rule) that this is the entry in row j and column i of X^{-1}. The statement follows. □

This lemma allows us to compute the left-hand side of the Lagrange equation in (4.3) for our application:

$$\nabla f_\mu(X) = C^T + \mu(X^T)^{-1}, \tag{4.4}$$

since $\nabla\,\mathrm{Tr}(C^TX) = C^T$.

For the right-hand side, we have $g_i(X) = A(X)_i - b_i$ which we can also write as $\mathrm{Tr}(A_i^TX) - b_i$ for a suitable $n \times n$ matrix A_i. This yields $\nabla g_i(X) = A_i^T$, so that (4.3) becomes

$$A(X) = \mathbf{b}, \quad C^T + \mu(X^T)^{-1} = \sum_{i=1}^{m} y_i A_i^T = (A^T(\mathbf{y}))^T,$$

or (after transposing)

$$A(X) = \mathbf{b}, \quad C + \mu X^{-1} = \sum_{i=1}^{m} y_i A_i = A^T(\mathbf{y}), \tag{4.5}$$

The last equality in (4.5) follows from

$$\mathbf{y}^T A(X) = \sum_{i=1}^{m} y_i A_i(X) = \sum_{i=1}^{m} y_i \operatorname{Tr}(A_i^T X) = \operatorname{Tr}(\sum_{i=1}^{m} y_i A_i^T X) = \operatorname{Tr}((\sum_{i=1}^{m} y_i A_i)^T X),$$

so $A^T(\mathbf{y}) := \sum_{i=1}^m y_i A_i$ indeed defines the adjoint of A.

4.3.3 The Symmetric Case

We have now derived the following: if $X^* \in \mathbb{R}^{n\times n}$ is a local maximum of the problem

$$\begin{array}{ll} \text{Maximize} & f_\mu(X) := \operatorname{Tr}(C^T X) + \mu \ln\det X \\ \text{subject to} & A(X) = \mathbf{b} \\ & X \in \mathbb{R}^{n\times n}, \end{array}$$

then X^* satisfies (4.5). But we are actually interested in the "symmetric" version

$$\begin{array}{ll} \text{Maximize} & f_\mu(X) := \operatorname{Tr}(C^T X) + \mu \ln\det X \\ \text{subject to} & A(X) = \mathbf{b} \\ & X \in \mathcal{S}_n \end{array} \tag{4.6}$$

since the maximum $X^*(\mu)$ of the auxilary problem (4.2) can be found among the local maxima of the latter.

We claim that both problems lead to the same Lagrange equations (4.5). To argue formally, we identify a symmetric matrix with the $\binom{n+1}{2}$-vector of entries on and above the diagonal, and we define a function $X : \mathbb{R}^{\binom{n+1}{2}} \to \mathbb{R}^{n^2}$ that supplements the missing entries below the diagonal:

$$X(\mathbf{x})_{ij} = \begin{cases} x_{ij} & \text{if } i \le j \\ x_{ji} & \text{otherwise} \end{cases}.$$

We can alternatively describe X as the linear function $\mathbf{x} \mapsto \mathcal{D}\mathbf{x}$, where $\mathcal{D}$ is the matrix whose (i,j)-th column, $i \le j$, has a 1 at rows (i,j) and (j,i), and 0's otherwise.

For (4.6), we need to compute all derivatives over $\mathbb{R}^{\binom{n+1}{2}}$, such that the Lagrange equation in (4.3) becomes

$$\nabla(f_\mu \circ X)(\mathbf{x}) = \sum_{i=1}^{m} y_i \nabla(g_i \circ X)(\mathbf{x}).$$

By the chain rule for higher-dimensional derivatives, this is equivalent to

$$\nabla f_\mu(X(\mathbf{x}))\mathcal{D} = \sum_{i=1}^{m} y_i \nabla g_i(X(\mathbf{x}))\mathcal{D}. \tag{4.7}$$

We want to argue that we can cancel $\mathcal{D}$ on both sides. To this end, let us analyze what happens when a row vectors in $\mathbb{R}^{n^2}$ corresponding to a *symmetric* matrix is postmultiplied by $\mathcal{D}$. By definition of $\mathcal{D}$, we get the $\binom{n+1}{2}$ entries on and above the diagonal, where the off-diagonal entries are multiplied by 2:

$$\begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{1,2} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,n-1} & x_{2,n-1} & \cdots & x_{n-1,n} \\ x_{1,n} & x_{2,n} & \cdots & x_{n,n} \end{pmatrix} \stackrel{\mathcal{D}}{\mapsto} \begin{pmatrix} x_{1,1} & \mathbf{2}x_{1,2} & \mathbf{2}x_{1,3} & \cdots & \mathbf{2}x_{1,n} \\ & x_{2,2} & \mathbf{2}x_{2,3} & \cdots & \mathbf{2}x_{2,n} \\ & & \ddots & & \\ & & & x_{n-1,n-1} & \mathbf{2}x_{n-1,n} \\ & & & & x_{n,n} \end{pmatrix}$$

This is a bijection between the symmetric matrices and $R^{\binom{n+1}{2}}$.

Now, since the row vectors

$$\nabla f_\mu(X(\mathbf{x})) = C^T + \mu(X(\mathbf{x})^T)^{-1}$$

and

$$\nabla g_i(X(\mathbf{x})) = A_i, \; i = 1, \ldots, m,$$

do correspond to symmetric $n \times n$ matrices (we may choose A_i such that $\mathrm{Tr}(A_i^T X(\mathbf{x})) = A(X(\mathbf{x}))_i$ to be symmetric since $X(\mathbf{x})$ is symmetric), we can indeed cancel $\mathcal{D}$ on both sides of (4.7) and recover (4.5).

Let us summarize our findings as follows.

4.3.2 Lemma. *If $X^*(\mu)$ is an optimal solution of the auxiliary problem (4.2), then $X^*(\mu)$ is a local maximum of (4.6). By the method of Lagrange multipliers, there is a vector $\tilde{\mathbf{y}} \in \mathbb{R}^m$ such that $X^*(\mu)$ and $\tilde{\mathbf{y}}$ satisfy the equations*

$$A(X) = \mathbf{b}, \quad C + \mu X^{-1} = A^T(\mathbf{y}).$$

Let us rewrite this into a more convenient form, by introducing a "slack" matrix $S = A^T(\mathbf{y}) - C = \mu X^{-1}$. Then, $X^*(\mu)$ satisfies the system

$$\begin{array}{rcl} A(X) & = & \mathbf{b} \\ A^T(\mathbf{y}) - S & = & C \\ SX & = & \mu I_n \\ S, X & \succ & 0 \end{array} \tag{4.8}$$

for suitable $\mathbf{y}, S$.

4.3.4 A Primal-Dual Interpretation

From the fact that $X^*(\mu)$ must satisfy (4.8), if it exists, we can already deduce that the objective function value $\mathrm{Tr}(C^T X^*(\mu))$ converges to the value of our original semidefinite program (4.1). Even more is true: We get a pair of primal and dual solutions whose *duality gap* depends on μ.

4.3.3 Corollary. *If $\tilde{X}, \tilde{\mathbf{y}}, \tilde{S}$ satisfy (4.8), then $\tilde{X}$ is a strictly feasible solution of the primal semidefinite program*

$$\begin{array}{ll} \textit{Maximize} & \mathrm{Tr}(C^T X) \\ \textit{subject to} & A(X) = \mathbf{b} \\ & X \succeq 0, \end{array} \tag{4.9}$$

$\tilde{\mathbf{y}}$ is a strictly feasible solution of the dual semidefinite program

$$\begin{array}{ll} \textit{Minimize} & \mathbf{b}^T \mathbf{y} \\ \textit{subject to} & A^T(\mathbf{y}) - C \succeq 0, \end{array} \tag{4.10}$$

and the duality gap is

$$\mathbf{b}^T \tilde{\mathbf{y}} - \mathrm{Tr}(C^T \tilde{X}) = n\mu.$$

Here, strict feasibility means that $X \succ 0$ and $A^T(\mathbf{y}) - C \succ 0$, respectively.

Proof. $\tilde{S}, \tilde{X} \succ 0$ immediately implies that $\tilde{X}$ is strictly feasible for the primal, and $\tilde{\mathbf{y}}$ is strictly feasible for the dual. For the duality gap, we compute

$$\begin{aligned} \mathrm{Tr}(C^T \tilde{X}) &= \mathrm{Tr}((A^T(\tilde{\mathbf{y}}) - \tilde{S})^T \tilde{X}) \\ &= \mathrm{Tr}(A^T(\tilde{\mathbf{y}})^T \tilde{X}) - \mathrm{Tr}(\tilde{S}^T \tilde{X}) \\ &= \tilde{\mathbf{y}}^T A(\tilde{X}) - \mathrm{Tr}(\tilde{S}\tilde{X}) \quad \text{(adjoint; symmetry)} \\ &= \mathbf{b}^T \tilde{\mathbf{y}} - n\mu. \end{aligned}$$

□

This is good news: if we were able to compute $X^*(\mu)$ for small μ, then we would have an almost optimal solution of our semidefinite program (4.1). Indeed, since $\mathrm{Tr}(C^T X) \leq \mathbf{b}^T \tilde{\mathbf{y}}$ for all feasible solutions X by weak duality (Theorem 2.5.1), $\mathrm{Tr}(C^T X^*(\mu))$ comes to within $n\mu$ of (4.1)'s value.

4.4 Sufficient Conditions for Optimality

So far we have shown that *if* the problem of maximizing $f_\mu(X)$ subject to $A(X) = \mathbf{b}$ and $X \succ 0$ has a maximum at X^*, then there exist $S^* \succ 0$ and $\mathbf{y}^* \in \mathbb{R}^m$ such that $X^*, \mathbf{y}^*, S^*$ satisfy (4.8). Next, we formulate conditions for the existence of the maximum, and we show that under these conditions, the maximum is characterized by (4.8).

4.4.1 Lemma. *Let us suppose that the primal program (4.9) has a feasible solution* $\tilde{X} \succ 0$ *and that the dual program (4.10) has a feasible solution* $\tilde{\mathbf{y}}$ *such that the slack matrix* $\tilde{S} = A^T(\tilde{\mathbf{y}}) - C$ *satisfies* $\tilde{S} \succ 0$. *(Less formally, both the primal and dual programs have an interior feasible point.) Moreover, let us assume that the matrices* A_i *for which* $A(X)_i = \mathrm{Tr}(A_i^T X)$ *are linearly independent (an assumption that can be made without loss of generality, see also Exercise 4.7.1).*

Then for every $\mu > 0$ *the system (4.8) has a unique solution* $X^* = X^*(\mu)$, $\mathbf{y}^* = \mathbf{y}^*(\mu)$, $S^* = S^*(\mu)$, *and* $X^*(\mu)$ *is the unique maximizer of* f_μ *subject to* $A(X) = \mathbf{b}$ *and* $X \succ 0$.

Proof. Let $\mu > 0$ be fixed. We begin with the following claim.

Claim. Under the assumptions of the lemma, the set $Q = \{X \in \mathcal{S}_n : A(X) = \mathbf{b}, X \succ 0, f_\mu(X) \geq f_\mu(\tilde{X})\}$ *is closed and bounded, hence compact (when interpreted as a subset of* $\mathbb{R}^{n^2}$*).*

Proof of the claim. Closedness is the easy part (Exercise 4.7.2), the main thing to prove is boundedness. We have

$$\begin{aligned} f_\mu(X) &= \mathrm{Tr}(C^T X) + \mu \ln\det X \\ &= \mathrm{Tr}(C^T X) + \tilde{\mathbf{y}}^T(\mathbf{b} - A(X)) + \mu \ln\det X \quad (\text{since } A(X) = \mathbf{b}) \\ &= \tilde{\mathbf{y}}^T\mathbf{b} + \mathrm{Tr}(C^T X) - \mathrm{Tr}((A^T(\tilde{\mathbf{y}}))^T X) + \mu \ln\det X \quad (\text{adjoint}) \\ &= \tilde{\mathbf{y}}^T\mathbf{b} + \mathrm{Tr}((C - A^T(\tilde{\mathbf{y}}))^T X) + \mu \ln\det X \\ &= \tilde{\mathbf{y}}^T\mathbf{b} - \mathrm{Tr}(\tilde{S}^T X) + \mu \ln\det X \quad (\text{since } A^T(\tilde{\mathbf{y}}) - C = \tilde{S}) \end{aligned}$$

We thus have $f_\mu(X) \geq f_\mu(\tilde{X})$ if and only if

$$\mu \ln\det X - \mathrm{Tr}(\tilde{S}^T X) \geq \mu \ln\det \tilde{X} - \mathrm{Tr}(\tilde{S}^T \tilde{X}) =: c. \tag{4.11}$$

We want to show next that this lower bound implies an *upper* bound on the eigenvalues of any matrix $X \in Q$. Let $\sigma > 0$ be the smallest eigenvalue of $\tilde{S}$ and $\lambda_1(X), \ldots, \lambda_n(X)$ the eigenvalues of X. Since $\tilde{S} - \sigma I_n \succeq 0$, we have $\mathrm{Tr}((\tilde{S} - \sigma I_n)^T X) \geq 0$ (a consequence of self-duality of $\mathcal{S}_n^+$, see Lemma 3.1.1), hence

$$\begin{aligned} \mu \ln\det X - \mathrm{Tr}(\tilde{S}^T X) &= \mu \ln \prod_{j=1}^n \lambda_j(X) - \mathrm{Tr}(\tilde{S}^T X) \\ &\leq \mu \ln \prod_{j=1}^n \lambda_j(X) - \sigma\, \mathrm{Tr}(I_n^T X) \\ &= \mu \sum_{j=1}^n \ln \lambda_j(X) - \sigma \sum_{j=1}^n \lambda_j(X), \end{aligned}$$

see Exercise 3.4.1 for the latter equality. Putting this together with (4.11), we have

$$c \le \mu \sum_{j=1}^{n} \ln \lambda_j(X) - \sigma \sum_{j=1}^{n} \lambda_j(X), \quad X \in Q. \tag{4.12}$$

The right-hand side is of the form $\sum_{j=1}^{n} h(\lambda_j(X))$, where h is the univariate function $h(x) = \mu \ln x - \sigma x$. Elementary calculus shows that $h(x)$ attains a unique maximum at $x = \mu/\sigma$, and in particular, $h(x)$ is bounded from above. Since (4.12) further yields

$$\sigma\lambda_i(X) - \mu \ln \lambda_i(X) \le \sum_{j \ne i} h(\lambda_j(X)) - c, \quad i = 1, \ldots, n,$$

we thus know that $\sigma\lambda_i(X) - \mu \ln \lambda_i(X)$ is bounded from above over Q, and since the first term asymptotically dominates the second, $\lambda_i(X)$ itself is bounded from above over Q, for all i. According to Exercise 4.7.3, Q is then bounded as well.

So the set Q is compact. Hence the continuous function f_μ attains a maximum on it, which, as we know, is unique. This shows that f_μ attains a maximum over $A(X) = \mathbf{b}, X \succ 0$ under the assumptions of the lemma, and by means of Lagrange multipliers we have shown that this maximum yields a solution of (4.8). It remains to verify that this is the only solution of (4.8).

What we do is to show that for every solution $X, \mathbf{y}, S$ of (4.8), X also maximizes f_μ, hence $X = X^*(\mu)$. We note that S and $\mathbf{y}$ are uniquely determined by X through the relations $SX = \mu I_n$ and $A^T(\mathbf{y}) - S = \sum_{i=1}^{m} y_i A_i - S = C$ from (4.8), using the assumption that the A_i are linearly independent. Thus, unique solvability of (4.8) follows.

To show that every solution X of(4.8) maximizes f_μ in the auxiliar problem, we use the fact that the method of Lagrange multipliers as outlined in Section 4.3.1 not only provides necessary but actually *sufficient* conditions for a local maximum at $\tilde{\mathbf{x}}$ in the case where the function f is concave in a small neighborhood of $\tilde{\mathbf{x}}$ and the g_i are linear (Exercise 4.7.4). To apply this, we recall that if $X, \mathbf{y}, S$ solve (4.8), then X is in particular a feasible solution of (4.6), and $y_1, \ldots, y_m$ are Lagrange multipliers for X w.r.t. (4.6). This is how we derived (4.8).

Since $X \succ 0$, the function f_μ is concave in a small neighborhood of X (Lemma 4.2.2), hence X is a local maximum of f_μ over the set of positive definite matrices, by sufficiency of Lagrange multipliers. But strict concavity of f_μ also implies that every local maximum coincides with the unique global maximum (the argument is similar to the one that we have used to establish uniqueness of the global maximum). Hence X indeed maximizes f_μ in the auxiliary problem, and the lemma is proved. □

4.5 Central Path and The Algorithm

In this section we address the question how system (4.8) can be solved for small μ, since this is what we need in order to get good primal and dual solutions, see Corollary 4.3.3. The idea is to start at some $(\tilde{X}, \tilde{y}, \tilde{S})$ that is close to some $(X^*(\mu), \mathbf{y}^*(\mu), S^*(\mu))$, and from this point on (approximately) follow the **primal–dual central path** of the semidefinite program (4.9), which is the set

$$\Big\{(X^*(\mu), \mathbf{y}^*(\mu), S^*(\mu)) \in \mathbb{R}^{2n^2+m} : \mu > 0\Big\},$$

until μ is small enough. We can indeed call this a path, since $X^*(\mu), \mathbf{y}^*(\mu)$ and $S^*(\mu)$ are continuous functions of μ (Exercise 4.7.5).

Let us introduce the function $F_\mu : \mathbb{R}^{2n^2+m} \to \mathbb{R}^{2n^2+m}$, defined by

$$F_\mu(X, \mathbf{y}, S) = \begin{pmatrix} A(X) - \mathbf{b} \\ A^T(\mathbf{y}) - S - C \\ SX - \mu I_n \end{pmatrix} =: \begin{pmatrix} F_\mu^1(X, \mathbf{y}, S) \\ F_\mu^2(X, \mathbf{y}, S) \\ F_\mu^3(X, \mathbf{y}, S) \end{pmatrix}.$$

We know that $F_\mu(X^*(\mu), \mathbf{y}^*(\mu), S^*(\mu)) = \mathbf{0}$, and that this is the only zero of F_μ subject to $X, S \succ 0$. Furthermore, we would like to compute this zero for small μ, in order to obtain almost optimal solutions of (4.9) and (4.10) as guaranteed by Corollary 4.3.3.

Directly solving the system $F_\mu(X, \mathbf{y}, S) = \mathbf{0}$ is difficult, since it contains the n^2 nonlinear equations $SX - \mu I_n = 0$. But there is a well-known stepwise method for computing the zero of a function that is based on solving systems of *linear* equations only. This method is *Newton's Method.*

4.5.1 Newton's Method

Let us first recall Newton's Method for finding the zero of a differentiable univariate function f. We start with some initial guess $x^{(0)}$, and given $x^{(i)}, i \geq 0$, we set

$$x^{(i+1)} := x^{(i)} - \frac{f(x^{(i)})}{f'(x^{(i)})},$$

equivalently,

$$f'(x^{(i)})\left(x^{(i+1)} - x^{(i)}\right) = -f(x^{(i)}). \tag{4.13}$$

Under suitable conditions on f and the initial guess, the sequence $(x^{(i)})_{i\in\mathbb{N}}$ quickly converges to a zero of f.

Newton's Method generalizes to higher dimensions. If $f : \mathbb{R}^n \to \mathbb{R}^n$, then (4.13) becomes

$$Df(\mathbf{x}^{(i)})\left(\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\right) = -f(\mathbf{x}^{(i)}), \tag{4.14}$$

where $Df(x)$ is the *Jacobian* of f at $\mathbf{x}$, the $n \times n$ matrix satisfying

$$(Df(\mathbf{x}))_{ij} = \frac{\partial f(\mathbf{x})_i}{\partial x_j}.$$

For this to work, the matrix $Df(\mathbf{x}^{(i)})$ must be invertible, otherwise the next iterate $\mathbf{x}^{(i+1)}$ is not well-defined.

4.5.2 Application to the Central Path Funtion

Let us now derive the formulas for one step of Newton's Method, applied to the central path function F_μ. As in Lemma 4.4.1, we assume that we have $X^{(i)} = \tilde{X}, \mathbf{y}^{(i)} = \tilde{\mathbf{y}}, S^{(i)} = \tilde{S}$ such that

$$A(\tilde{X}) = \mathbf{b}, \quad A^T(\tilde{\mathbf{y}}) - \tilde{S} = C, \quad \tilde{S}, \tilde{X} \succ 0.$$

Now we want to compute the next iterate $X^{(i+1)} = \tilde{X}', \mathbf{y}^{(i+1)} = \tilde{\mathbf{y}}', S^{(i+1)} = \tilde{S}'$ according to the general recipe in (4.14). Let us write

$$\Delta X = \tilde{X}' - \tilde{X}, \quad \Delta \mathbf{y} = \tilde{\mathbf{y}}' - \tilde{\mathbf{y}}, \quad \Delta S = \tilde{S}' - \tilde{S}. \tag{4.15}$$

Then (4.14) is

$$DF_\mu(\tilde{X}, \tilde{\mathbf{y}}, \tilde{S}) \begin{pmatrix} \Delta X \\ \Delta \mathbf{y} \\ \Delta S \end{pmatrix} = -F_\mu(\tilde{X}, \tilde{\mathbf{y}}, \tilde{S}) = \begin{pmatrix} \mathbf{0} \\ 0 \\ \tilde{S}\tilde{X} - \mu I_n \end{pmatrix}. \tag{4.16}$$

The matrix $DF_\mu(X, \mathbf{y}, S)$ has the following block structure.

$$DF_\mu(X, \mathbf{y}, S) = \left(\begin{array}{c|c|c} DF^1_{\mu,\mathbf{y},S}(X) & 0 & 0 \\ \hline 0 & DF^2_{\mu,X,S}(\mathbf{y}) & DF^2_{\mu,X,\mathbf{y}}(S) \\ \hline DF^3_{\mu,\mathbf{y},S}(X) & 0 & DF^3_{\mu,X,\mathbf{y}}(S), \end{array}\right)$$

where the additional subscripts mean that the corresponding arguments are being fixed. After dropping constant terms (which does not change derivatives), all the five "single-argument" functions that we need to differentiate in the blocks are linear. Exercise 4.7.6 shows that for a linear function F, $DF(\mathbf{x})\mathbf{y} = F(\mathbf{y})$, and we can use this formula to compute the left-hand side of (4.16) componentwise:

$$\begin{aligned} DF^1_{\mu,\mathbf{y},S}(X)\Delta X &= A(\Delta X), \\ DF^2_{\mu,X,S}(\mathbf{y})\Delta \mathbf{y} + DF^2_{\mu,X,\mathbf{y}}(S)\Delta S &= A^T(\Delta \mathbf{y}) - \Delta S \\ DF^3_{\mu,\mathbf{y},S}(X)\Delta X + DF^3_{\mu,X,\mathbf{y}}(S)\Delta S &= S\Delta X + \Delta S X. \end{aligned}$$

Hence, (4.16) is the following system of *linear* equations for $\Delta X, \Delta \mathbf{y}, \Delta S$.

$$\begin{aligned} A(\Delta X) &= \mathbf{0} && (4.17)\\ A^T(\Delta \mathbf{y}) - \Delta S &= 0 && (4.18)\\ \tilde{S}\Delta X + \Delta S\tilde{X} &= \mu I_n - \tilde{S}\tilde{X} && (4.19) \end{aligned}$$

We claim that this system has a unique solution $(\Delta X, \Delta \mathbf{y}, \Delta S)$. To see this, we start by solving the last equation (4.19) for ΔX:

$$\Delta X = \tilde{S}^{-1}(\mu I_n - \tilde{S}\tilde{X} - \Delta S\tilde{X}) = \tilde{S}^{-1}(\mu I_n - \tilde{S}\tilde{X} - A^T(\Delta \mathbf{y})\tilde{X}), \qquad (4.20)$$

using the second equation (4.18). Subsituting this into the first equation (4.17) yields a system only for $\Delta \mathbf{y}$:

$$A(\tilde{S}^{-1}(\mu I_n - \tilde{S}\tilde{X} - A^T(\Delta \mathbf{y})\tilde{X})) = \mathbf{0}.$$

Equivalently,

$$A(\tilde{S}^{-1}A^T(\Delta \mathbf{y})\tilde{X}) = A(\mu\tilde{S}^{-1} - \tilde{X}) = \mu A(\tilde{S}^{-1}) - \mathbf{b}.$$

Exercise 4.7.7 asks you to prove that the left-hand side is of the form

$$M\Delta \mathbf{y}, \quad M \succ 0,$$

so M is in particular invertible. It follows that $\Delta \mathbf{y}$ is uniquely determined, and substituting back into (4.18) yields ΔS, from which we in turn obtain ΔX through (4.20).

4.5.3 Making ΔX Symmetric

Recall that our Ansatz was (4.15), so we would like to obtain the next iterate in Newton's Method (which hopefully brings us closer to the zero of $F_\mu(X, \mathbf{y}, S)$) via $\tilde{X}' = \tilde{X} + \Delta X, \tilde{\mathbf{y}}' = \tilde{\mathbf{y}} + \Delta \mathbf{y}, \tilde{S}' = \tilde{S} + \Delta S$. The only problem is that ΔX may not be symmetric, in which case $(\tilde{X}', \tilde{\mathbf{y}}', \tilde{S}')$ is not a valid next iterate. We note that ΔS will be symmetric as a consequence of (4.18) and $A^T : \mathbb{R}^m \rightarrow \mathcal{S}_n$.

The problem with ΔX is due to the fact that we have (as in Section 4.3.3) ignored the symmetry constraints and solved the system (4.16) for $X, S \in \mathbb{R}^{n\times n}$ instead of $X, S \in \mathcal{S}_n$.

The simple fix is to update according to a "symmetrized" ΔX:

$$\tilde{X}' = \tilde{X} + \frac{1}{2}\left(\Delta X + \Delta X^T\right).$$

It can be shown that this *modified Newton step* also leads to theoretical convergence and good practical performance [9].

But to get polynomial runtime bounds, we need to proceed differently. It is possible (actually in various ways) to define a function $F'_\mu(X, \mathbf{y}, S)$ for which the Newton step yields a symmetric update matrix ΔX. This is discussed in detail in [13, Section 10.3].

One possible choice is the following. Let

$$S_{\tilde{X}}(M) = \frac{1}{2}\left(\tilde{X}^{-1/2}M\tilde{X}^{1/2} + (\tilde{X}^{-1/2}M\tilde{X}^{1/2})^T\right),$$

where $\tilde{X}^{1/2}$ is the square root of $\tilde{X}$, the unique positive definite matrix whose square is $\tilde{X}$ (the existence of $\tilde{X}^{1/2}$ follows from diagonalization: if $\tilde{X} = U^T DU$ with U orthogonal and D diagonal, we can set $\tilde{X}^{1/2} = U^T\sqrt{D}U$. where $\sqrt{D}$ is the diagonal matrix obtained from D by taking the square roots of all diagonal elements).

Instead of F_μ, we now use the function

$$F_\mu^{\tilde{X}}(X, \mathbf{y}, S) = \begin{pmatrix} A(X) - \mathbf{b} \\ A^T(\mathbf{y}) - S - C \\ S_{\tilde{X}}(XS) - \mu I_n \end{pmatrix}.$$

When we compute the Newton system (4.16) for $F_\mu^{\tilde{X}}$, we again arrive at (4.17) and (4.18), but (4.19) gets replaced with

$$\tilde{X}^{-1/2}(\tilde{X}\Delta S + \Delta X\tilde{S})\tilde{X}^{1/2} + \tilde{X}^{1/2}(\Delta S\tilde{X} + \tilde{S}\Delta X)\tilde{X}^{-1/2} = 2\mu I_n - \tilde{X}^{1/2}\tilde{S}\tilde{X}^{1/2}. \tag{4.21}$$

This yields a symmetric ΔX [12], but we had to change the function F_μ for that. Actually, $F_\mu^{\tilde{X}}$ even depends on the current iterate $\tilde{X}$, so the next Newton step will be with respect to a different function. Still, let us suppose that we somehow manage to find a zero of $F_\mu^{\tilde{X}}$ for some $\tilde{X}$. What does this tell us about our original problem of solving $F_\mu(X, \mathbf{y}, S) = \mathbf{0}$? It is easy to see that the zeros of F_μ are among the zeros of $F_\mu^{\tilde{X}}$, but not the other way around. Thus, if $(X^*, \mathbf{y}^*, S^*)$ is a zero of $F_\mu^{\tilde{X}}$, we cannot guarantee $S^*X^* = \mu I_n$. But Exercise 4.7.8 implies that we still have

$$\mathrm{Tr}(S^*X^*) = \mu n,$$

and if you reexamine the proof of Corollary 4.3.3, small *trace* is all we need in order to get small duality gap.

4.5.4 The Algorithm

Given that we know how to perform one step of Newton's Method, the following seems to be a natural way of simultaneously solving the semidefinite program (4.9) and its dual (4.10) up to duality gap $\varepsilon > 0$. Choose $\mu = \varepsilon/n$ and then perform Newton steps on $F_\mu^{\tilde{X}}$ (with $\tilde{X}$ always being the current iterate), until

$F_\mu^{\tilde{X}} \approx \mathbf{0}$. Then the current $\tilde{X}, \tilde{\mathbf{y}}$ are almost optimal solutions of (4.9) and (4.10) with duality gap ε.

But even if this method converges, it does not solve our problem. Remember that we are attempting to find the unique solution of (4.8) **subject to** $X, S \succ 0$. But Newton's Method does not know obout the latter constraints and cannot guarantee $X^{(i)}, S^{(i)} \succ 0$ even if this holds for $i = 0$. Moreover, *fast* convergence can be proved only if we start sufficiently close to the central path. Our initial solution $(\tilde{X}, \tilde{\mathbf{y}}, \tilde{S})$ could be too far away.

Therefore, the idea is to *choose* μ such that $(\tilde{X}, \tilde{\mathbf{y}}, \tilde{S})$ is close to the central path at μ, and then perform *just one* Newton step with respect to a slightly smaller μ. The intention is to bring the iterate closer to the central path point at the smaller μ, and then to repeat the process until the current iterate is close to the central path for sufficiently small μ.

Here is a generic step of the algorithm:

1. Given the current iterate $X^{(i)}, \mathbf{y}^{(i)}, S^{(i)}$, choose

$$\mu = \frac{\mathrm{Tr}(S^{(i)} X^{(i)})}{n}.$$

If $(X^{(i)}, \mathbf{y}^{(i)}, S^{(i)})$ happens to be on the central path, this is indeed the best μ, since it yields $X^{(i)} = X^*(\mu)$. We actually need to assume that $(X^{(i)}, \mathbf{y}^{(i)}, S^{(i)})$ is close to the central path at μ (but we won't specify this further).

2. Perform one step of Newton's Method w.r.t. $F_{\sigma\mu}^{X^{(i)}}$, where

$$\sigma = 1 - \frac{\delta}{\sqrt{n}}$$

is the *centrality parameter*, for some constant δ. This means, compute $\Delta X, \Delta \mathbf{y}, \Delta S$ by solving (4.17), (4.18) and (4.21), and set

$$\begin{aligned} X^{(i+1)} &:= X^{(i)} + \Delta X, \\ \mathbf{y}^{(i+1)} &:= \mathbf{y}^{(i)} + \Delta \mathbf{y}, \\ S^{(i+1)} &:= S^{(i)} + \Delta S. \end{aligned}$$

It can be shown that $X^{(i+1)}, S^{(i+1)} \succ 0$, and that $(X^{(i+1)}, \mathbf{y}^{(i+1)}, S^{(i+1)})$ is again close to the central path [12]. Moreover, we made progress:

$$\mathrm{Tr}(S^{(i+1)} X^{(i+1)}) = \sigma \, \mathrm{Tr}(S^{(i)} X^{(i)}).$$

A bound on the runtime directly follows from the geometric decrease of the sequence $(\mathrm{Tr}(S^{(i)}X^{(i)}))_{i\in\mathbb{N}}$.

4.5.1 Theorem. *Fix $\varepsilon > 0$ and suppose that we have $(\tilde{X}, \tilde{\mathbf{y}}, \tilde{S})$ such that*

$$A(\tilde{X}) = \mathbf{b},\ \tilde{X} \succ 0 \quad \textit{(strict primal feasibility)},$$

$$A^T(\tilde{\mathbf{y}}) - \tilde{S} = C,\ \tilde{S} \succ 0 \quad \textit{(strict dual feasibility)},$$

and $(\tilde{X}, \tilde{\mathbf{y}}, \tilde{S})$ is close to the central path at $\mu = \mathrm{Tr}(\tilde{S}\tilde{X})/n$. With $X^{(0)} = \tilde{X}, \mathbf{y}^{(0)} = \tilde{\mathbf{y}}, S^{(0)} = \tilde{S}$, the above algorithm computes an iterate $X^{(k)}, \mathbf{y}^{(k)}, S^{(k)}$ such that

(i) $A(X^{(k)}) = \mathbf{b},\ X^{(k)} \succ 0,$

(ii) $A^T(\mathbf{y}^{(k)}) - S^{(k)} = C,\ S^{(k)} \succ 0,$

(iii) $\mathrm{Tr}(S^{(k)}X^{(k)}) \le \varepsilon$

in at most

$$\log_{1/\sigma} \frac{\mathrm{Tr}(\tilde{S}\tilde{X})}{\varepsilon} = O\left(\sqrt{n}\log \frac{\mathrm{Tr}(\tilde{S}\tilde{X})}{\varepsilon}\right)$$

iterations, where $\sigma = 1 - \Theta(1/\sqrt{n})$.

Recall that the condition $\mathrm{Tr}(S^{(k)}X^{(k)}) \le \varepsilon$ leads to small duality gap and therefore a primal solution that is optimal up to an additive error of ε (see the proof of Corollary 4.3.3).

We note that on top of requring strictly feasible primal and dual points, we also need to require that the initial solution is close to the central path. Given a semidefinite program, we may not have this; it is possible, though, to achieve this property, at the cost of solving an auxiliary semidefinite program first for which these conditions are fulfilled. We won't go into the details.

4.6 Complexity: Theory Versus Practice

The whole interior-point approach has one severe drawback: it only leads to polynomial-time bounds in the RAM model, since there are no good bounds for the size (in bits) of the numbers that appear in intermediate computations, even if we assume integral input. We actually get a polynomial-time bound in the RAM model from Theorem 4.5.1 if $\log \mathrm{Tr}(\tilde{S}\tilde{X})$ is polynomially bounded in n for the initial $\tilde{X}, \tilde{S}$.

But even this cannot always be guaranteed: there are semidefinite programs with bounded integer coordinates, such that *every* feasible solution has exponential bit size (exponential in n). In other words, we already need exponential time just to write out $\tilde{X}$ (we give an example below). This is in contrast with linear programming, where such pathologies cannot occur.

If you really want to (approximately) solve a semidefinite program in polynomial time (in the bit model), your only option is the *ellipsoid method* [8]. And even here, you need to require the existence of a feasible solution of bounded size in order to get started (usually this not a problem). But in the pathological situation that we describe below, there is no such solution.

The main virtue of interior-point methods for semidefinite programming is that they are easy to implement and work well in practice. On top of that, they do in principle come with theoretical guarantees, but in practice, the analyzeable algorithms are rather slow and get replaced by faster heuristic variants.

In fact, the focus of many interior-point papers is not on worst-case runtimes but on efficient implementations of the primitives that arise (like solving systems of linear equations, taking matrix square roots, etc.). For example, the algorithm of Helmberg [9] has been shown to converge, but without any bounds on the convergence rate. Still, Helmberg is offering an efficient semidefinite programming solver based on his algorithm; the convergence is fast in practice, and the primitive operations are very simple (no matrix square roots, for example).

In a way, the situation is similar to that of linear programming: there are proven polynomial-time methods (interior-point methods are polynomial even in the bit model), but the fastest methods in practice are based on the simplex method for which no polynomial-time bounds are known.

From a theoretical point of view, it may be somewhat unsatisfactory, but the main message is this: **Semidefinite programs can efficiently be solved in practice, using interior-point methods**, and we have outlined how this works. This fact makes methods based on semidefinite programming attractive for problems that are hard in practice, like MAX-CUT and a lot of other NP-hard optimization problems.

4.6.1 A Semidefinite Program with Only Huge Feasible Solutions

Let us consider a semidefinite program with the following constraints:

$$\begin{pmatrix} 1 & 2 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 2 & x_1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & x_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & x_1 & x_2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_2 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & x_2 & x_3 & \cdots & 0 & 0 \\ & & & \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & x_{n-1} \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & x_{n-1} & x_n \end{pmatrix} \succeq 0.$$

This is in fact a constraint of the form $X \succeq 0$, along with various equalities involving entries of X. Due to the block structure, we have $X \succeq 0$ if and only if

$$\begin{pmatrix} 1 & x_{i-1} \\ x_{i-1} & x_i \end{pmatrix} \succeq 0, \quad i = 1, \ldots, n,$$

where $x_0 := 2$. But this implies

$$\det \begin{pmatrix} 1 & x_{i-1} \\ x_{i-1} & x_i \end{pmatrix} = x_i - x_{i-1}^2 \geq 0, \quad i = 1, \ldots, n,$$

equivalently $x_i \geq x_{i-1}^2, i = 1, \ldots, n$. It follows that

$$x_n \geq 2^{2^n}$$

for every feasible solution, which is doubly-exponential in n. Hence, the encoding size of x_n (when written as a rational number, say) is exponential in n and also in the number of variables.

4.7 Exercises

4.7.1 Exercise. *Consider the problem of minimizing a function $f(\mathbf{x})$ (with continuous partial derivatives) subject to linear constraints $g_i(\mathbf{x}) = \mathbf{a}_i\mathbf{x} - b_i = 0, i = 1, \ldots, m$ (here, the $\mathbf{a}_i$'s are row vectors). Show that if $\tilde{\mathbf{x}}$ is a maximizer of f subject to the constraints $g_i(\mathbf{x}) = 0$, then there exists $\mathbf{y} \in \mathbb{R}^m$ such that*

$$\nabla f(\tilde{\mathbf{x}}) = \sum_{i=1}^{m} y_i \mathbf{a}_i.$$

In particular, we do not need the requirement that $\tilde{\mathbf{x}}$ is a regular point.

Hint: *You may assume correctness of the general method of Lagrange multipliers as outlined in Section 4.3.1.*

4.7.2 Exercise. *Prove that for any real number M, the set*

$$\{X \in \mathcal{S}_n : A(X) = \mathbf{b}, X \succ 0, f_\mu(X) \geq M\}$$

is closed.

4.7.3 Exercise. *Let $Q \subseteq \mathcal{S}_n^+$ be a set of matrices for which all eigenvalues are bounded by some global constant c. Prove that Q is bounded as well (interpreted as a subset of R^{n^2}).*

4.7.4 Exercise. *Prove that if $\mathbf{x}$ and $y_1, \ldots, y_m$ satisfy (4.3), where the g_i are linear functions $g_i(\mathbf{x}) = \mathbf{a}_i\mathbf{x} - b_i$, and f is concave in a small neighborhood U of $\mathbf{x}$, then $\mathbf{x}$ is a local maximum of f subject to $g_i(\mathbf{x}) = 0, i = 1, \ldots, m$.*

4.7.5 Exercise. *Prove that the function*

$$\mu \mapsto (X^*(\mu), \mathbf{y}^*(\mu), S^*(\mu))$$

that maps μ to the unique solution of (4.8) is continuous.

4.7.6 Exercise. *Let $F : \mathbb{R}^k \to \mathbb{R}^k$ be a linear function, and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$. Prove that*

$$DF(\mathbf{x})\mathbf{y} = F(\mathbf{y}).$$

4.7.7 Exercise. *Let $\tilde{S}, \tilde{X} \succ 0$, and let $A : \mathcal{S}_n \to \mathbb{R}^m$ be a linear operator such that $A(X)_i = \mathrm{Tr}(A_i^T X)$, with the matrices A_i being linearly independent. Prove that there is a matrix $M \succ 0$ such that*

$$A(\tilde{S}^{-1} A^T(\mathbf{y}) \tilde{X}) = M\mathbf{y}, \quad \forall \mathbf{y} \in \mathbb{R}^m.$$

4.7.8 Exercise. *Let P be an invertible matrix and consider the function*

$$S_P(M) = \frac{1}{2}\left(PMP^{-1} + (PMP^{-1})^T\right).$$

Prove that $S_P(M) = Q$ implies $\mathrm{Tr}(M) = \mathrm{Tr}(Q)$.

Bibliography

[1] Richard J. Duffin. Infinite programs. In Harold W. Kuhn and Albert W. Tucker, editors, *Linear Inequalities and Related Systems*, volume 38 of *Annals of Mathematical Studies*, pages 157–170. 1956.

[2] Yuli Eidelman, Vitali Milman, and Antonis Tsolomitis. *Functional Analysis: An Introduction*, volume 66 of *Graduate Studies in Mathematics*. American Mathematical Society, 2004.

[3] M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified NP-complete graph problems. *Theoretical Computer Science*, 1(3):237–267, February 1976.

[4] Michael R. Garey and David S. Johnson. *Computers and Intractability : A Guide to the Theory of NP-Completeness (Series of Books in the Mathematical Sciences)*. W. H. Freeman & Co Ltd, January 1979.

[5] Bernd Gärtner and Jiří Matoušek. *Understanding and Using Linear Programming*. Universitext. Springer-Verlag, Berlin Heidelberg, 2007.

[6] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.

[7] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, 3rd edition, 1996.

[8] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin Heidelberg, 1988.

[9] C. Helmberg, R. Rendl, R. Vanderbei, and GH. Wolkowicz. An interior-point method fpr semidefinite programming. *SIAM J. Optim.*, 6(2):342–461, 1996.

[10] Nicholas J. Higham. Analysis of the Cholesky decomposition of a semidefinite matrix. In M. G. Cox and S. Hammarling, editors, *Reliable Numerical Computation*, pages 161–185. Oxford University Press.

[11] Lázló Lovász. On the Shannon capcity of a graph. *IEEE Transactions on Information Theory*, IT-25:1–7, 1979.

[12] R. Monteiro. Primal-dual path-following algorithms for semidefinite programming. *Siam J. Opt.*, 7(3):663–678, 1997.

[13] R. Monteiro and M. Todd. Path-following methods. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*. Kluwer Academic Publishers, 2000.

[14] A. L. Peressini, F. E. Sullivan, and J. J. Uhl. *The Mathematics of Nonlinear Programming*. Undergraduate Texts in Mathematics. Springer-Verlag, 1988.

[15] C. Shannon. The zero-error capacity of a noisy channel. *IRE Transactions on Information Theory*, 2:9–19, 1956.