EECS127 Course Notes

Anmol Parande

Spring 2021 - Professor Laurent El Ghaoui

Disclaimer: These notes reflect 127 when I took the course (Spring 2021). They may not accurately reflect current course content, so use at your own risk. If you find any typos, errors, etc, please raise an issue on the GitHub repository.

Contents

1 Linear Algebra

Definition 1 *An affine set is one of the form* $\mathcal{A} = \{\boldsymbol{x} \in \mathcal{X} : \ \boldsymbol{x} = \boldsymbol{v} + \boldsymbol{x_o}, \ \boldsymbol{v} \in \mathcal{V}\}$ *where* $\mathcal{V}$ *is a subspace of a vector space* $\mathcal{X}$ *and* x_0 *is a given point.*

Notice that by definition 1, a subspace is simply an affine set containing the origin. Also notice that the dimension of an affine set $\mathcal{A}$ is the same as the dimension of $\mathcal{V}$.

1.1 Norms

Definition 2 *A norm on the vector space* $\mathcal{X}$ *is a function* $\|\cdot\| : \mathcal{X} \to \mathbb{R}$ *which satisfies:*
1. $\|\boldsymbol{x}\| \geq 0$ *with equality if and only if* $\boldsymbol{x} = \boldsymbol{0}$

2. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

3. $\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$ *for any scalar* α*.*

Definition 3 *The* l_p *norms are defined by*

$$\|\boldsymbol{x}\|_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{\frac{1}{p}}, \ 1 \leq p \leq \infty$$

In the limit as $p \to \infty$,

$$\|\boldsymbol{x}\|_\infty = \max_k |x_k|.$$

Similar to vectors, matrices can also have norms.

Definition 4 *A function* $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ *is a matrix norm if*

$$f(A) \geq 0 \quad f(A) = 0 \Leftrightarrow A = 0 \quad f(\alpha A) = |\alpha| f(A) \quad f(A + B) \leq f(A) + f(B)$$

Definition 5 *The Froebenius norm is the* l_2 *norm applied to all elements of the matrix.*

$$\|A\|_F = \sqrt{\mathit{trace} AA^T} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$$

One useful way to characterize matrices is by measuring their "gain" relative to some l_p norm.

Definition 6 *The operator norms is defined as*

$$\|A\|_p = \max_{\boldsymbol{u} \neq 0} \frac{\|A\boldsymbol{u}\|_p}{\|u\|_p}$$

When $p = 2$, the norm is called the spectral norm because it relates to the largest eigenvalue of $A^T A$.

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

1.2 Inner Products

Definition 7 *An inner product on real vector space is a function that maps $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ to a non-negative scalar, is distributive, is commutative, and $\langle \boldsymbol{x}, \boldsymbol{x}, \rangle = 0 \Leftrightarrow \boldsymbol{x} = 0$.*

Inner products induce a norm $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$. In $\mathbb{R}^n$, the standard inner product is $\boldsymbol{x}^T \boldsymbol{y}$. The angle bewteen two vectors is given by

$$\cos\theta = \frac{\boldsymbol{x}^T \boldsymbol{y}}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2}.$$

In general, we can bound the absolute value of the standard inner product between two vectors.

Theorem 1 (Holder Inequality)

$$|\boldsymbol{x}^T \boldsymbol{y}| \leq \sum_{k=1}^{n} |x_k y_k| \leq \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q, \; p, q \geq 1 \; s.t \; p^{-1} + q^{-1} = 1.$$

Notice that for $p = q = 2$, theorem 1 turns into the Cauchy-Schwartz Inequality ($|\boldsymbol{x}^T \boldsymbol{y}| \leq \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2$).

1.3 Functions

We consider functions to be of the form $f : \mathbb{R}^n \to \mathbb{R}$. By contrast, a map is of the form $f : \mathbb{R}^n \to \mathbb{R}^m$. The components of the map f are the scalar valued functions f_i that produce each component of a map.

Definition 8 *The graph of a function f is the set of input-output pairs that f can attain.*

$$\left\{(x, f(x)) \in \mathbb{R}^{n+1} : \; x \in \mathbb{R}^n\right\}$$

Definition 9 *The epigraph of a function is the set of input-output pairs that f can achieve and anything above.*

$$\{(x,t) \in \mathbb{R}^{n+1} : \boldsymbol{x} \in \mathbb{R}^{n+1},\ t \geq f(x)\}$$

Definition 10 *The t-level set is the set of points that achieve exactly some value of f.*

$$\{\boldsymbol{x} \in \mathbb{R}^n :\ f(x) = t\}$$

Definition 11 *The t-sublevel set of f is the set of points achieving at most a value t.*

$$\{x \in \mathbb{R}^n :\ f(x) \leq t\}$$

Definition 12 *The half-spaces are the regions of space which a hyper-plane separates.*

$$H_{\text{-}} = \{x : \boldsymbol{a}^T\boldsymbol{x} \leq b\} \qquad H_+ = \{x : \boldsymbol{a}^T\boldsymbol{x} > b\}$$

Definition 13 *A polyhedron is the intersection of m half-spaces given by $\boldsymbol{a}_i^T\boldsymbol{x} \leq b_i$ for $i \in [1, m]$.*

When a polyhedron is bounded, it is called a polytope.

1.3.1 Types of Functions

Theorem 2 *A function is linear if and only if it can be expressed as $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$ for some unique pair $(\boldsymbol{a}, b)$.*

An affine function is linear when $b = 0$. A hyperplane is simply a level set of a linear function.

Theorem 3 *Any quadratic function can be written as the sum of a quadratic term involving a symmetric matrix and an affine term:*

$$q(x) = \frac{1}{2}\boldsymbol{x}^T H\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} + d.$$

Another special class of functions are polyhedral functions.

Definition 14 *A function $f : \mathbb{R}^n \to \mathbb{R}$ is polyhedral if its epigraph is a polyhedron.*

$$\text{epi } f = \left\{ (x, t) \in \mathbb{R}^{n+1} : \; C \begin{bmatrix} \boldsymbol{x} \\ t \end{bmatrix} \leq d \right\}$$

1.3.2 Vector Calculus

We can also do calculus with vector functions.

Definition 15 *The gradient of a function at a point x where f is differentiable is a column vector of first derivatives of f with respsect to the components of $\boldsymbol{x}$*

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

The gradient is perpendicular to the level sets of f and points from a point $\boldsymbol{x}_0$ to higher values of the function. In other words, it is the direction of steepest increase. It is akin to the derivative of a 1D function.

Definition 16 *The Hessian of a function f at point x is a matrix of second derivatives.*

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

The Hessian is akin to the second derivative in a 1D function. Note that the Hessian is a symmetric matrix.

1.4 Matrices

Matrices define a linear map between an input space and an output space. Any linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by a matrix.

Theorem 4 (Fundamental Theorem of Linear Algebra) *For any matrix $A \in \mathbb{R}^{m \times n}$,*

$$\mathcal{N}(A) \oplus \mathcal{R}(A^T) = \mathbb{R}^n \qquad \mathcal{R}(A) \oplus \mathcal{N}(A^T) = \mathbb{R}^m.$$

1.4.1 Symmetric Matrices

Recall that a symmetric matrix is one where $A = A^T$.

Theorem 5 (Spectral Theorem) *Any symmetric matrix is orthogonally similar to a real diagonal matrix.*

$$A = A^T \implies A = U\Lambda U^T = \sum_i \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T, \quad \|\boldsymbol{u}\| = 1, \quad \boldsymbol{u}_i^T \boldsymbol{u}_j = 0 \; (i \neq j)$$

Let $\lambda_{min}(A)$ be the smallest eigenvalue of symmetric matrix A and $\lambda_{max}(A)$ be the largest eigenvalue.

Definition 17 *The Rayleigh Quotient for $\boldsymbol{x} \neq \boldsymbol{0}$ is $\frac{\boldsymbol{x}^T A \boldsymbol{x}}{\|\boldsymbol{x}\|^2}$.*

Theorem 6 *For any $\boldsymbol{x} \neq \boldsymbol{0}$,*

$$\lambda_{min}(A) \leq \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\|\boldsymbol{x}\|^2} \leq \lambda_{max}(A).$$

Two special types of symmetric matrices are those with non-negative eigenvalues.

Definition 18 *A symmetric matrix is positive semi-definite if $\boldsymbol{x}^T A \boldsymbol{x} \geq 0 \implies \lambda_{min}(A) \geq 0$.*

Definition 19 *A symmetric matrix is poitive definite if $\boldsymbol{x}^T A \boldsymbol{x} > 0 \implies \lambda_{min}(A) > 0$.*

These matrices are important because they often have very clear geometric structures. For example, an ellipsoid in multi-dimensional space can be defined as the set of points

$$\mathcal{E} = \{x \in \mathbb{R}^m : \; \boldsymbol{x}^T P^{-1} \boldsymbol{x} \leq 1\}$$

where P is a positive definite matrix. The eigenvectors of P give the principle axes of this ellipse, and $\sqrt{\lambda}$ are the semi-axis lengths.

1.4.2 QR Factorization

Similar to how spectral theorem allows us to decompose symmetric matrices, QR factorization is another matrix decomposition technique that works for any general matrix.

Definition 20 *The QR factorization matrix are the orthogonal matrix Q and the upper triangular matrix R such that $A = QR$*

An easy way to find the QR factorization of a matrix is to apply Graham Schmidt to the columns of the matrix and express the result in matrix form. Suppose that our matrix A is full rank (i.e its columns $\boldsymbol{a}_i$ are linearly independent) and we have applied Graham-Schmidt to columns $\boldsymbol{a}_{i+1} \cdots \boldsymbol{a}_n$ to get orthogonal vectors $\boldsymbol{q}_{i+1} \cdots \boldsymbol{q}_n$. Continuing the procedure, the ith orthogonal vector $\boldsymbol{q}_i$ is

$$\tilde{\boldsymbol{q}}_i = \boldsymbol{a}_i - \sum_{k=i+1}^{n} (\boldsymbol{q}_k^T \boldsymbol{a}_k)\boldsymbol{q}_k \qquad \boldsymbol{q}_i = \frac{\tilde{\boldsymbol{q}}_i}{\|\tilde{\boldsymbol{q}}_i\|_2}.$$

If we re-arrange this, to solve for $\boldsymbol{a}_i$, we see that

$$\boldsymbol{a}_i = \|\tilde{\boldsymbol{q}}_i\|_2 \boldsymbol{q}_i + \sum_{k=i+1}^{n} (\boldsymbol{q}_k^T \boldsymbol{a}_k)\boldsymbol{q}_k.$$

Putting this in matrix form, we can see that

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix} \qquad r_{ij} = \boldsymbol{a}_i^T \boldsymbol{q}_{\boldsymbol{j}}, r_{ii} = \|\tilde{\boldsymbol{q}}_i\|_2.$$

1.4.3 Singular Value Decomposition

Definition 21 *A matrix $A \in \mathbb{R}^{m \times n}$ is a dyad if it can be written as $\boldsymbol{p}\boldsymbol{q}^T$.*

A dyad is a rank-one matrix. It turns out that all matrices can be decomposed into a sum of dyads.

Definition 22 *The Singular Value Decomposition of a matrix A is*

$$A = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

where σ_i are the singular values of A and $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are the left and right singular vectors.

Th singular values are ordered such that $\sigma_1 >= \sigma_2 >= \cdots$. The left singular values are the eigenvectors of AA^T and the right singular values are the eigenvectors of A^TA. The singular values are $\sqrt{\lambda_i}$ where λ_i are the eigenvalues of A^TA. Since AA^T and A^TA are symmetric, $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are orthogonal. The number of non-zero singular values is equal to the rank of the matrix. We can write the SVD in matrix form as

$$A = [U_r \quad U_{n-r}]\,\mathrm{diag}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0) \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix}$$

Writing the SVD tells us that

1. V_{n-r} forms a basis for $\mathcal{N}(A)$
2. U_r form a basis for $\mathcal{R}(A)$

The Frobenius norm and spectral norm are tightly related to the SVD.

$$\|A\|_F = \sum_i \sigma_i^2$$

$$\|A\|_2^2 = \sigma_1^2$$

2 Fundamentals of Optimization

Definition 23 *The standard form of optimization is*

$$p^\star = \min_{\boldsymbol{x}} f_0(\boldsymbol{x}) \textit{ such that } f_i(\boldsymbol{x}) \leq 0$$

- The vector $\boldsymbol{x} \in \mathbb{R}^n$ is known as the **decision variable**.
- The function $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the **objective**.
- The functions $f_i : \mathbb{R}^n \to \mathbb{R}$ is the **constraints**.
- $p^\star$ is the **optimal value**, and the $\boldsymbol{x}^\star$ which achieves the optimal value is called the **optimizer**.

Definition 24 *The feasible set of an optimization problem is*

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n : \ f_i(\boldsymbol{x}) \leq 0\}$$

Definition 25 *A point $\boldsymbol{x}$ is ϵ-suboptimal if it is feasible and satisfies*

$$p^\star \leq f_0(\boldsymbol{x}) \leq p^\star + \epsilon$$

Sometimes, optimizations in a particular formulation do not admit themselves to be solved easily. In many cases, we can introduce additional "slack" variable and constraints to massage the problem into a form which is easier to analyze. For example, if f_0 is a polyhedral function, then $\min_{\boldsymbol{x}} f_0(x)$ is equivalent to the problem

$$\min_{\boldsymbol{x},t} t \quad : \quad f_0(x) \leq t,$$

and the epigraphic constraint $f_0(x) \leq t$ can be further massaged into linear constraints since a polyhedron is the intersection of half-spaces. This works because by minimizing t, we are also minimizing how large $f_0(x)$ can get since $f_0(x) \leq t$, so at optimum, $f_0(x) = t$.

2.1 Robust Optimization

For a "nominal" problem

$$\min_{\boldsymbol{x}} f_0(\boldsymbol{x}) \quad : \quad \forall i \in [1, m],\ f_i(\boldsymbol{x}) \leq 0,$$

uncertainty can enter in the data used to create the f_0 and f_i. It can also enter during decision time where the $\boldsymbol{x}^\star$ which solves the optimization cannot be implemented exactly. These uncertainties can create unstable solutions or degraded performance. To make our optimization more robust to uncertainty, we add a new variable $\boldsymbol{u} \in \mathcal{U}$.

Definition 26 *For a nominal optimization problem $\min_{\boldsymbol{x}} f_0(\boldsymbol{x})$ subject to $f_i(\boldsymbol{x}) \leq 0$ for $i \in [1, m]$, the robust counterpart is*

$$\min_{\boldsymbol{x}} \max_{\boldsymbol{u} \in \mathcal{U}} f_0(\boldsymbol{x}, \boldsymbol{u}) \quad : \quad \forall i \in [1, m],\ f_i(\boldsymbol{x}, \boldsymbol{u}) \leq 0$$

3 Linear Algebraic Optimization

Many optimization problems can be solved using the machinery of Linear Algebra. These problems do not have inequality constraints or non-euclidean norms in the objective function.

3.1 Projection

The idea behind projection is to find the closest point in a set closest (with respect to particular norm) to a given point.

Definition 27 *Given a vector $\boldsymbol{x}$ in inner product space $\mathcal{X}$ and a subspace $S \subseteq \mathcal{X}$, the projection of $\boldsymbol{x}$ onto S is given by*

$$\Pi_S(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{y} \in S} \|\boldsymbol{y} - \boldsymbol{x}\|$$

where the norm is the one induced by the inner product.

Theorem 7 *There exists a unique vector $\boldsymbol{x}^* \in S$ which solves*

$$\min_{\boldsymbol{y} \in S} \|\boldsymbol{y} - \boldsymbol{x}\|.$$

It is necessary and sufficient for $\boldsymbol{x}^*$ to be optimal that $(\boldsymbol{x} - \boldsymbol{x}^*) \perp S$. The same condition applies when projecting onto an affine set.

3.1.1 Matrix Pseudo-inverses

Definition 28 *A pseudoinverse is a matrix $A^\dagger$ that satisfies:*

$$AA^\dagger A = A \quad A^\dagger AA^\dagger = A^\dagger \quad (AA^\dagger)^T = AA^\dagger \quad (A^\dagger A)^T = A^\dagger A$$

There are several special cases of pseudoinverses.

1. $A^\dagger = V_r \text{diag}\left(\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_r}\right) U_r^T$ is the Moore-Penrose Pseudo-inverse.
2. When A and non-singular, $A^\dagger = A^{-1}$.
3. When A is full column rank, $A^\dagger = (A^TA)^{-1}A^T$.
4. When A is full row rank, $A^\dagger = A^T(AA^T)^{-1}$

The pseudo-inverses are useful because they can easily compute the projection of a vector onto a related subspace of A.

1. $\text{argmin}_{z\in\mathcal{R}(A)} \|\boldsymbol{z}-\boldsymbol{y}\|_2 = AA^\dagger\boldsymbol{y}$
2. $\text{argmin}_{z\in\mathcal{R}(A)^\perp} \|\boldsymbol{z}-\boldsymbol{y}\|_2 = (I - AA^\dagger)\boldsymbol{y}$
3. $\text{argmin}_{z\in\mathcal{N}(A)} \|\boldsymbol{z}-\boldsymbol{y}\|_2 = (I - A^\dagger A)\boldsymbol{y}$
4. $\text{argmin}_{z\in\mathcal{N}(A)^\perp} \|\boldsymbol{z}-\boldsymbol{y}\|_2 = A^\dagger A\boldsymbol{y}$

3.2 Explained Variance

The Low Rank Approximation problem is to approximate a matrix A with a rank k matrix

$$\min_{A_k} \|A - A_k\|_F^2 \text{ such that } \text{rank}(A_k) = k.$$

The solution to the low rank approximation problem is simply the first k terms of the SVD:

$$A_K^\star = \sum_{i=1}^k \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T.$$

This is because the singular values give us a notion of how much of the Frobenius Norm (Total Variance) each dyad explains.

$$\eta = \frac{\|A_k\|_F^2}{\|A\|_F^2} = \frac{\sum_i^k \sigma_i^2}{\sum_i^r \sigma_i^2}$$

3.2.1 PCA

Suppose we had a matrix containing m data points in $\mathbb{R}^n$ (each data point is a column), and without loss of generality, assume this data is centered around 0 (i.e $\sum_i \boldsymbol{x}_i = 0$). The variance of this data along a particular direction $\boldsymbol{z}$ is given by $\boldsymbol{z}^T XX^T \boldsymbol{z}$. Principle Component Analysis is finding the directions $\boldsymbol{z}$ such that the variance is maximized.

$$\max_{z \in \mathbb{R}^n} \boldsymbol{z}^T XX^T \boldsymbol{z} \text{ such that } \|\boldsymbol{z}\|_2 = 1$$

The left singular vector corresponding to the largest singular value of the XX^T matrix is the optimizer of this problem, and the variance along this direction is σ_1^2. If we wanted to find subsequent directions of maximal variance, they are just the left singular vectors corresponding to the largest singular values.

3.3 Removing Constraints

Following from the Fundmental Theorem of Linear Algebra, if $A\boldsymbol{x} = \boldsymbol{y}$ has a solution, then the set of solutions can be expressed as

$$S = \{\bar{\boldsymbol{x}} + N\boldsymbol{z}\}$$

where $A\bar{\boldsymbol{x}} = \boldsymbol{y}$ and N is a basis for $\mathcal{N}(A)$. This means if we have a constrained optimization problem

$$\min_{\boldsymbol{x}} f_0(\boldsymbol{x}) \ : \ A\boldsymbol{x} = \boldsymbol{b},$$

we can write an equivalent unconstrained problem

$$\min_{\boldsymbol{z}} f_0(\boldsymbol{x}_0 + N\boldsymbol{z})$$

where $A\boldsymbol{x}_0 = \boldsymbol{b}$

4 Conic Programming

Conic programming is the set of optimization problems which deal with variables constrained to a second-order cone.

Definition 29 *A n-dimensional second-order cone is the set*

$$\mathcal{K}_n = \{(\boldsymbol{x}, t), \ \boldsymbol{x} \in \mathbb{R}^n, \ t \in \mathbb{R} : \ \|\boldsymbol{x}\|_2 \leq t\}$$

In spaces 3-dimensions and higher, we can rotate these cones.

Definition 30 *A rotated second order cone in $\mathbb{R}^{n+2}$ is the set*

$$\mathcal{K}_n^r = \{(\boldsymbol{x}, y, z), \boldsymbol{x} \in \mathbb{R}^n, y \in \mathbb{R}, z \in \mathbb{R} : \ \boldsymbol{x}^T \boldsymbol{x} \leq yz, y \geq 0, z \geq 0\}.$$

The rotated second-order cone can be interpreted as a rotation because the hyperbolic constraint $\|\boldsymbol{x}\|_2^2 \leq yz$ can be expressed equivalently as

$$\left\| \begin{bmatrix} 2\boldsymbol{x} \\ y - z \end{bmatrix} \right\|_2 \leq y + z.$$

Definition 31 *The standard Second Order Cone Constraint is*

$$\|A\boldsymbol{x} + \boldsymbol{b}\|_2 \leq \boldsymbol{c}^T \boldsymbol{x} + d.$$

A SOC constraint will confine $\boldsymbol{x}$ to a second order cone since if we let $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b} \in \mathbb{R}^m$ and $t = \boldsymbol{c}^T \boldsymbol{x} + d$, then $(\boldsymbol{y}, t) \in \mathcal{K}_m$.

Definition 32 *A second-order cone program in standard inequality form is given by*

$$\min \boldsymbol{c}^T \boldsymbol{x} \textit{ such that } \|A_i \boldsymbol{x} + \boldsymbol{b}_i\|_2 \leq \boldsymbol{c}_i^T \boldsymbol{x} + d_i.$$

4.1 Quadratic Programming

A special case of SOCPs are Quadratic Programs. These programs have constraints and an objective function which can be expressed as a quadratic function. In SOCP form, they look like

$$\begin{aligned} \min_{\boldsymbol{x}, t} \quad & \boldsymbol{a}_0^T \boldsymbol{x} + t \\ \text{s.t:} \quad & \left\| \begin{bmatrix} 2Q_0^{\frac{1}{2}} \boldsymbol{x} \\ t - 1 \end{bmatrix} \right\|_2 \leq t + 1 \\ & \left\| \begin{bmatrix} 2Q_i^{\frac{1}{2}} \boldsymbol{x} \\ b_i - \boldsymbol{a}_i^T \boldsymbol{x} - 1 \end{bmatrix} \right\|_2 \leq b_i - \boldsymbol{a}_i \boldsymbol{x} + 1 \end{aligned}$$

Definition 33 *The standard form of a quadratic constrained quadratic program is*

$$\min_{\boldsymbol{x}} \boldsymbol{x}^T Q_0 \boldsymbol{x} + \boldsymbol{a}_0^T \boldsymbol{x} \quad : \quad \forall i \in [1, m],\ \boldsymbol{x}^T Q_i \boldsymbol{x} + \boldsymbol{a}_i^T \boldsymbol{x} \leq b_i$$

To be a quadratic program, the matrix H must be positive semi-definite. If the $Q_i = 0$ in the constraints, then we get a normal quadratic program.

Definition 34 *The standard form of a quadratic program is given by*

$$\min_{\boldsymbol{x}} \frac{1}{2} \boldsymbol{x}^T H \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x} \quad : \quad \forall i \in [1, m],\ \boldsymbol{a}_i^T \boldsymbol{x} \leq b_i$$

Its SOCP form looks like

$$\begin{aligned} \min_{\boldsymbol{x}, y} \quad & \boldsymbol{c}^T \boldsymbol{x} + y \\ \text{s.t:} \quad & \left\| \begin{bmatrix} 2H^{\frac{1}{2}} \boldsymbol{x} \\ y - 1 \end{bmatrix} \right\|_2 \leq y + 1, \\ & \boldsymbol{a}_i \boldsymbol{x} \leq b_i \end{aligned}$$

In the special case where H is positive definite and we have no constraints, then

$$\frac{1}{2} \boldsymbol{x}^T H \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x} + d = \frac{1}{2} (\boldsymbol{x} + H^{-1} \boldsymbol{c})^T H (\boldsymbol{x} + H^{-1} \boldsymbol{c}) + d - (H^{-1} \boldsymbol{c})^T H (H^{-1} \boldsymbol{c})$$

Thus

$$\operatorname*{argmin}_{\boldsymbol{x}} \frac{1}{2} \boldsymbol{x}^T H \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x} + d = -H^{-1} \boldsymbol{c}$$

4.2 Linear Programming

If the matrix in the objective function of a quadratic program is 0 (and there are no quadratic constraints), then the resulting objective and constraints are affine functions. This is a linear program.

Definition 35 *The inequality form of a linear program is given by*

$$\min_{\boldsymbol{x}} \boldsymbol{c}^T \boldsymbol{x} + d \quad : \quad \forall i \in [1, m],\ \boldsymbol{a}_i^T \boldsymbol{x} \leq b_i$$

Since linear program is a special case of a quadratic program, it can also be expressed as an SOCP.

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t} \quad & \forall i \in [1, m],\ \|0 \boldsymbol{x} + 0\|_2 \leq b_i - \boldsymbol{a}_i^T \boldsymbol{x} \end{aligned}$$

Because of the constraints, the feasible set of a linear program is a polyhedron.