Math 6640 : Introduction to Optimisation

Rebecca Hardenbrook[1], Chee Han Tan[1], and Nathan Willis[1]

[1]*Department of Mathematics, University of Utah*

Last modified : October 16, 2018

Contents

Preface

These notes are largely based on **Math 6640: Introduction to Optimisation** course, taught by Braxton Osting in Fall 2018, at the University of Utah. The main textbook is [Bec14], but additional examples or remarks or results from other sources are added as we see fit, mainly to facilitate our understanding. These notes are by no means accurate or applicable, and any mistakes here are of course our own. Please report any typographical errors or mathematical fallacy to us by email rebeccah@math.utah.edu or tan@math.utah.edu or willis@math.utah.edu.

1. C1: Chee Han
2. C2: Chee Han
3. C3: Nathan
4. C4: Rebecca
5. C5: Nathan
6. C6: Chee Han

Chapter 1

Linear Algebra

We are interested in the following optimisation problem: minimise an objective functional $f(x)$ subject to the constraints

$$\begin{aligned} g_i(x) &= 0 \quad \text{for } i = 1, \dots, m \\ h_j(x) &= 0 \quad \text{for } j = 1, \dots, p \end{aligned}$$

We denote the feasible (admissible or constraint) set S as the set of $x \in \mathbb{R}^n$ that satisfy the inequality and equality constraint functions. A vector $x^* \in \mathbb{R}^n$ is a **global minimiser** if $x^* \in S$ and $f(x^*) \leq f(x)$ for all $x \in S$. The **optimal value** is $f^* := f(x^*)$. A vector $x^* \in \mathbb{R}^n$ is a **local** minimiser if $x^* \in S$ and $f(x^*) \leq f(x)$ for all x in a neighbourhood of x^*.

This problem collapses to the standard single-variable calculus in the case of $n = 1$, but what about higher dimensional? We are also interested in optimising functionals that are non-differentiable. Equality constraints correspond to the method of Lagrange multipliers, the difficult bits to handle are the inequality constraints. Some of the important questions that we care about:

1. Well-posedness of the problem, *i.e.* existence and uniqueness of a (global) minimiser. This depends on the properties of the functional and the feasible set.

2. Characterisation of solution set.

3. How to numerically approximate solutions.

We point out that the following problems are equivalent

$$\max_{x \in S} f(x) \iff -\min_{x \in S}(-f(x)).$$

We will mainly focus on continuous optimisation problem, *i.e.* the admissible set varies continuously. We can always rewrite the equality constraint $h(x) = 0$ as two inequality constraints $h(x) \leq 0$ and $h(x) \geq 0$. Below are some real-life applications:

1. Pattern recognition/classifications (binary)

2. Medical imaging

3. Shape optimisation

4. Portfolios optimisation

5. Scheduling problems
6. Advertisement optimisation

Problems that have special structures:

1. Least squares problems (medical imaging, regression problem). This has the form
$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2.$$
The problem actually has analytical solution $x=(A^TA)^{-1}A^Tb$.
2. Linear programming (LP) problems. This has the form
$$\min_{a_i^Tx\leq b_i, i=1,\ldots,m} \quad c^Tx$$
[Constraints could be affine.]
3. Convex problems, where the objective functional f and inequality constraint functions g_i are convex functions. No analytical solution, but there are very efficient numerical methods.
4. Nonconvex problems. These are generally very difficult.

1. Nonnegative orthant:
$$\mathbb{R}^n_+ = \{x\in\mathbb{R}^n \colon x\geq 0 \text{ pointwise}\}.$$
2. Positive orthant:
$$\mathbb{R}^n_{++} = \{x\in\mathbb{R}^n \colon x > 0 \text{ pointwise}\}.$$
3. Closed line segments: For any $x,y\in\mathbb{R}^n$,
$$[x,y] = \{x+\alpha(y-x)\in\mathbb{R}^n \colon \alpha\in[0,1]\}.$$
4. Unit simplex Δ_n:
$$\Delta_n = \{x\in\mathbb{R}^n \colon x\geq 0, x^Te=1\}.$$
Here, $e=(1,\ldots,1)^T\in\mathbb{R}^n$ is the 1-vector. (Intersection of planes passing through certain points)

1.1 Inner Products and Norms

Definition 1.1.1. An **inner product** on $\mathbb{R}^n$ is a map $\langle\cdot,\cdot\rangle\colon \mathbb{R}^n\times\mathbb{R}^n\mapsto\mathbb{R}$ satisfying

(a) $\langle x,y\rangle = \langle y,x\rangle$ for all $x,y\in\mathbb{R}^n$

(b) $\langle x,y+z\rangle = \langle x,y\rangle + \langle x,z\rangle$

(c) $\langle \lambda x,y\rangle = \lambda\langle x,y\rangle$ for all $\lambda\in\mathbb{R}, x,y\in\mathbb{R}^n$

(d) $\langle x,x\rangle\geq 0$ and $\langle x,x\rangle = 0 \iff x=\mathbf{0}$

A standard example of inner product is the vector dot product. We can also construct a weighted inner product as follows: for any $w \in \mathbb{R}^n_{++}$,

$$\langle x, y \rangle_w = \sum_{i=1}^{n} w_i x_i y_i.$$

Definition 1.1.2. A norm on $\mathbb{R}^n$ is a function $\|\cdot\|\colon \mathbb{R}^n \mapsto \mathbb{R}$ satisfying

(a) $\|x\| \geq 0$ for all $x \in \mathbb{R}^n$ and $\|x\| = 0 \iff x = \mathbf{0}$.

(b) $\|\lambda x\| = |\lambda| \|x\|$ for all $\lambda \in \mathbb{R}, x \in \mathbb{R}^n$

(c) $\|x + y\| \leq \|x\| + \|y\|$

Any inner product induces a norm $\|x\| = \sqrt{\langle x, x \rangle}$. The Euclidean norm is

$$\|x\|_2 = x^T x = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

The generalisation of Euclidean norm is the ℓ^p-norm for $1 \leq p \leq \infty$:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

One can show that the ℓ^∞-norm is the limit of the ℓ^p-norm as $p \longrightarrow \infty$:

$$\|x\|_\infty = \max_{i=1,\dots,n} |x_i|.$$

Lemma 1.1.3 (Cauchy-Schwarz)**.** *The following holds for all $x, y \in \mathbb{R}^n$*

$$|x^T y| \leq \|x\| \|y\|$$

with equality holds for $x = \alpha y$ for all $\alpha \in \mathbb{R}$.

Let's look at the induced matrix norm. More precisely, given two norms $\|\cdot\|_a$ and $\|\cdot\|_b$,

$$\|A\|_{a,b} = \max_{x \in \mathbb{R}^n} \Big\{ \|Ax\|_b \colon \|x\|_a \leq 1 \Big\}.$$

An immediate consequence of the definition is the following:

$$\|Ax\|_{a,b} \leq \|A\|_{a,b} \|x\|_a.$$

We can also consider the spectral norm:

$$\|A\|_2 = \|A\|_{2,2} = \sqrt{\lambda_{\max}(A^T A)} = \sigma_{\max}(A).$$

We can also look at the induced matrix 1-norm:

$$\|A\|_1 = \|A\|_{1,1} = \max_{j=1,\dots,n} \left(\sum_{i=1}^{m} |a_{ij}| \right) = \text{maximum column sum}$$

$$\|A\|_\infty = \|A\|_{\infty,\infty} = \max_{i=1,\dots,m} \left(\sum_{j=1}^{n} |a_{ij}| \right) = \text{maximum row sum}$$

Finally, we define the Frobenius norm (this cannot be induced!) as

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}.$$

1.2 Eigenvalues and Eigenvectors

Definition 1.2.1. Let $A \in \mathbb{R}^{n\times n}$. Then $v \in \mathbb{R}^n, v \neq \mathbf{0}$ is an eigenvector of A with eigenvalue $\lambda \in \mathbb{C}$ if $Av = \lambda v$. Note that eigenvectors are scalar-invariant, so typically we choose v such that $\|v\|_2 = 1$.

A well-known fact is the following: the eigenvalues of a symmetric matrix exist and are real. We order the eigenvalues from the largest to the smallest:

$$\lambda_{\max}(A) = \lambda_1(A) \geq \lambda_2(A) \geq \ldots \cdots \geq \lambda_n(A) = \lambda_{\min}(A).$$

Theorem 1.2.2 (Spectral Decomposition)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then there exists an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ (i.e. $U^TU = I_n = UU^T$) and a diagonal matrix $D = diag(\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A))$ such that $A = UDU^T$. Moreover, $\lambda_j(A), j = 1, 2, \ldots, n$ are the eigenvalues of A and columns u_j of U are the corresponding eigenvector of λ_j.*

Corollary 1.2.3. *The following identity holds for symmetric (normal) matrices:*

$$Tr(A) = \sum_{i=1}^{n} \lambda_i(A)$$
$$\det(A) = \prod_{i=1}^{n} \lambda_i(A)$$

Definition 1.2.4. For $x \neq \mathbf{0}$, the **Rayleigh Quotient** $R_A \colon \mathbb{R}^n \mapsto \mathbb{R}$ is defined by

$$R_A(x) = \frac{x^TAx}{x^Tx}.$$

Lemma 1.2.5.

$$\lambda_{min}(A) \leq R_A(x) \leq \lambda_{max}(A) \quad \textit{for all } x \in \mathbb{R}^n, x \neq \mathbf{0}.$$

Moreover,

$$\lambda_{max}(A) = \max_{x\neq\mathbf{0}} R_A(x)$$
$$\lambda_{min}(A) = \min_{x\neq\mathbf{0}} R_A(x)$$

1.3 Point-Set Topology

Definition 1.3.1. The **open ball** with center $c \in \mathbb{R}^n$ and radius $r > 0$ is

$$B(c, r) = \{x \in \mathbb{R}^n \colon \|x - c\| < r\}.$$

The **closed ball** with center $c \in \mathbb{R}^n$ and radius $r > 0$ is

$$B[c, r] = \{x \in \mathbb{R}^n \colon \|x - c\| \leq r\}.$$

Definition 1.3.2. Given $U \subset \mathbb{R}^n$, a point $x \in U$ is an **interior point** of U if there exists $r > 0$ for which $B(x, r) \subset U$. The interior of U is the set of all interior points and denoted $\text{int}(U)$.

Definition 1.3.3. An open set is a set that contains only interior points.

Definition 1.3.4. A set U is closed if its complement $U^c \equiv \mathbb{R}^n \setminus U$ is open. Equivalently, it contains all limits of convergent sequences in U. For example, the closed ball $B[c, r]$, unit simplex Δ_n and line segment $[a, b]$ are closed sets in $\mathbb{R}^n$.

Proposition 1.3.5. *Let f be a continuous function defined on a closed set $S \subset \mathbb{R}^n$. For any $\alpha \in \mathbb{R}$, the sets*

$$\begin{aligned} L(f, \alpha) &= \{x \in S \colon f(x) \leq \alpha\} = \alpha\text{-sublevel set of } f \\ Con(f, \alpha) &= \{x \in S \colon f(x) = \alpha\} = \alpha\text{-contour of } f \end{aligned}$$

are closed.

Definition 1.3.6. Given a set $U \subset \mathbb{R}^n$, a point $x \in U$ is called a **boundary point** if for all $r > 0$, the open ball $B(x, r)$ contains at least one point in U and one point in U^c.

For example, $\text{bd}(B[c, r]) = \{x \in \mathbb{R}^n \colon \|x - c\| = r\}$. Also, $\text{bd}(\Delta_n) = \Delta_n$.

Definition 1.3.7. The closure of $U \subset \mathbb{R}^n$ is the smallest closed set that contains U, *i.e.*

$$\text{cl}(U) = \cap\{T \colon U \subset T, T \text{ closed}\}.$$

One can show that $\text{cl}(U) = U \cup \text{bd}(U)$.

Definition 1.3.8. A set $U \subset \mathbb{R}^n$ is bounded if there exists an $M > 0$ such that $U \subset B(0, M)$.

Definition 1.3.9 (Heine-Borel)**.** A set $U \subset \mathbb{R}^n$ is **compact** if and only if it is closed and bounded.

1.4 Differentiability

Let $f \colon S \mapsto \mathbb{R}$ be a real-valued function on $S \subset \mathbb{R}^n$, $x \in \text{int}(S)$ and $d \in \mathbb{R}^n$, $\|d\| = 1$. If

$$f'(x; d) := \lim_{\varepsilon \to 0+} \frac{f(x + \varepsilon d) - f(x)}{\varepsilon}$$

exists, it is called the **directional derivative** (unclear).

Consider the standard canonical basis $\{e_1, \ldots, e_n\} \subset \mathbb{R}^n$. The partial derivative of f at x is given by

$$\frac{\partial f}{\partial x_i} = f'(x; e_i) = \lim_{\varepsilon \to 0^+} \frac{f(x + \varepsilon e_i) - f(x)}{\varepsilon}.$$

If all partial derivatives exist, we define the gradient ∇f as

$$\nabla f(x) = (\partial_{x_1} f, \partial_{x_2} f, \ldots, \partial_{x_n} f)^T \in \mathbb{R}^n.$$

We say that a function $f \colon \mathbb{R}^n \mapsto \mathbb{R}$ is continuously differentiable on $U \subset \mathbb{R}^n$ and written $f \in C^1(U)$ if all partial derivatives exist and are continuous. In this case,

$$f'(x; d) = \nabla f(x) \cdot d.$$

If $f \colon S \subset \mathbb{R}^n \mapsto \mathbb{R}^m$ is a vector-valued function, we compute the **Jacobian** by computing the gradient of components

$$\nabla f(x) = [\nabla f_1 | \nabla f_2 | \ldots | \nabla f_n] \in \mathbb{R}^{n \times m}.$$

Example 1.4.1. Take $w \in \mathbb{R}^n$ and consider the function $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by $f(x) = w^T x$. Then

$$\nabla f(x) = (w_1, w_2, \ldots, w_n)^T = w.$$

Example 1.4.2. Take $B \in \mathbb{R}^{m\times n}$ and consider the function $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ defined by $f(x) = Bx$. Then $\nabla f(x) = B^T$.

Example 1.4.3. Take $A \in \mathbb{R}^{n\times n}$ and consider the function $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by $f(x) = x^T Ax$. Show that $\nabla f = (A + A^T)x$. In particular, $\nabla f(x) = 2Ax$.

Definition 1.4.4. If all partial derivatives of $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ are continuously differentiable, we can compute the second partial derivatives $\dfrac{\partial f(x)}{\partial x_i \partial x_j}$. The matrix of all second partial derivatives is the **Hessian** of f at x:

$$[\nabla^2 f]_{i,j} = \frac{\partial f}{\partial x_i \partial x_j} = \nabla^T f \in \mathbb{R}^{n\times n}.$$

We say that f is **twice-differentiable** on $U \subset \mathbb{R}^n$ and write $f \in C^2(U)$ if the Hessian of f is continuous on U. In this case, the Hessian is symmetric, *i.e.* $\nabla^2 f(x) = [\nabla^2 f(x)]^T$.

Theorem 1.4.5 (Chain Rule)**.** *Let $f\colon \mathbb{R}^m \longrightarrow \mathbb{R}$ and $g\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be differentiable functions. The gradient of the function $h\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by $h(x) = f(g(x))$ is given by*

$$\nabla h(x) = \sum_{i=1}^{n} \frac{\partial f(g(x))}{\partial x_i} \nabla g_i(x) = \nabla g \cdot \nabla f(g(x)).$$

As an example, consider the function $g\colon \mathbb{R}^p \longrightarrow \mathbb{R}^m$ defined by $g(x) = f(Ax + b)$, where $A \in \mathbb{R}^{n\times p}$ and $b \in \mathbb{R}^n$. Then

$$\nabla g(x) = A^T \nabla f(Ax + b).$$

For $m = 1$, one can show that the Hessian of g is $\nabla^2 g(x) = A^T \nabla^2 f(Ax + b)A \in \mathbb{R}^{p\times p}$.

Theorem 1.4.6. *Let $U \subset \mathbb{R}^n$ be open, $f \in C^2(U)$ and $x \in U, r > 0$ such that $B(x, r) \subset U$. Then for all $y \in B(x, r)$, there exists $z_y \in [x, y]$ (typically unknown) such that the following linear approximation holds:*

$$f(y) = f(x) + (y - x)^T \nabla f(x) + \underbrace{\frac{1}{2}(y - x)^T \nabla^2 f(z_y)(y - x)}_{\textit{remainder term}}.$$

The remainder term is $o(\|y - x\|)$ as $\|y - x\| \longrightarrow 0$. Furthermore, for any $y \in B(x, r)$

$$f(y) = f(x) + (y - x)^T \nabla f(x) + \frac{1}{2}(y - x)^T \nabla^2 f(y)(y - x) + \underbrace{o\left(\|y - x\|^2\right)}_{\textit{as } \|y - x\| \longrightarrow 0}.$$

For example, the quadratic approximation of the exponential function is

$$e^x = 1 + x + \frac{x^2}{2} + o(x^2) \quad \text{as } x \longrightarrow 0.$$

Chapter 2

Optimality Conditions for Unconstrained Optimisation

2.1 Global and Local Optima

Definition 2.1.1. Let $f\colon S \longrightarrow \mathbb{R}$ be defined on a constrained set $S \subset \mathbb{R}^n$. Then

1. $x^* \in S$ is a **global minimiser** of f over S if $f(x) \geq f(x^*)$ for all $x \in S$
2. $x^* \in S$ is a **strict global minimiser** of f over S if $f(x) > f(x^*)$ for all $x \in S$.

Example 2.1.2. A global minimiser may not exist. For example, $\min_{x\in\mathbb{R}} e^{-x} = 0$ and $\min_{x\in(0,1)} x = 0$ but they are never actually attained. Another example would be a quadratic function over $\mathbb{R}$, but with a hole at the vertex.

Definition 2.1.3. The minimal value of f over S is the infimum of f over S:

$$f^* = \min\{f(x), x \in S\} = \inf\{f(x), x \in S\}.$$

The set of global minimisers of f over S is denoted by

$$\operatorname{argmin}\{f(x)\colon x \in S\}.$$

Note that this set might be empty.

Example 2.1.4. Consider maximising the function $f(x_1, x_2) = x_1 + x_2$ over the closed unit ball $S = B[0, 1]$. From Cauchy-Schwarz,

$$f(x) = e^T x \leq\leq \|e\|\|x\| \leq \sqrt{2}.$$

On the other hand, choosing $(x_1, x_2) = (1/\sqrt{2}, 1/\sqrt{2}) \in B[0, 1]$ gives

$$f(1/\sqrt{2}, 1/\sqrt{2}) = \frac{2}{\sqrt{2}} = \sqrt{2}.$$

Thus $f^* = \sqrt{2}$ and $(1/\sqrt{2}, 1/\sqrt{2})$ is a global maximiser.

Definition 2.1.5. Let $f\colon S \longrightarrow \mathbb{R}$ be defined on $S \in \mathbb{R}^n$.

1. $x^* \in S$ is a **local minima** of f over S if there exists an $r > 0$ such that
$$f(x) \geq f(x^*) \quad \text{for all } x \in S \cap B(x^*, r)$$

2. $x^* \in S$ is a **strict local minima** of f over S if there exists an $r > 0$ such that
$$f(x) > f(x^*) \quad \text{for all } x \in S \cap B(x^*, r)$$

An example would be the double well potential function.

Theorem 2.1.6 (Necessary condition for local optima)**.** *Let $f\colon U \longrightarrow \mathbb{R}$ be a function defined on a set $U \subset \mathbb{R}^n$. Suppose $x^* \in int(U)$ is a local optimum point and all partial derivatives of f exist at x^*. Then $\nabla f(x^*) = 0$.*

Definition 2.1.7. Let $f\colon U \longrightarrow \mathbb{R}$ be a function defined on a set $U \subset \mathbb{R}^n$. $x^* \in \operatorname{int}(U)$ is a **stationary point** of f if f is differentiable in a neighbourhood of x^* and $\nabla f(x^*) = 0$.

In 1D, Taylor's theorem asserts that
$$f(x^* + d) = f(x^*) + f'(x^*)d + \frac{1}{2}\underbrace{f''(x^*)}_{\text{tells us something about the curvature}} d^2 + o(d^2).$$

with x^* a stationary point. We have a local minima if $f''(x^*) > 0$, local maxima if $f''(x^*) < 0$, inconclusive if $f''(x^*) = 0$. In higher dimension, we need to look at the Hessian $\nabla^2 f(x)$. (look at principal eigenvalues?)

2.2 Classification of Matrices

Definition 2.2.1. Let $A \in \mathbb{R}^{n \times n}$ be symmetric.

1. A is **positive semidefinite**, denoted $A \succeq 0$ if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$.
2. A is **positive-definite**, denoted $A \succ 0$ if $x^T A x > 0$ for all $x \in \mathbb{R}^n, x \neq \mathbf{0}$.
3. A is **negative (semi)definite** if $-A$ is positive (semi)definite, denoted $A \preceq 0$ ($A \prec 0$).
4. A is **indefinite** if it is neither positive or negative semi-definite.

Example 2.2.2. Note that positive definiteness does not require positivity of entries. Let $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$. Then
$$x^T A x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 - 2x_1x_2 + 2x_2^2 = x_1^2 + x_2^2 + (x_1 - x_2)^2 \geq 0.$$

Moreover $x^T A x = 0 \iff x = \mathbf{0}$ since $x^T A x$ is a sum of squares. We conclude that A is positive-definite.

Example 2.2.3. Consider the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Choosing $x = (1, -1)$ gives $x^T A x = -2$. But choosing $x = (1, 0)$ gives $x^T A x = 1$. So the matrix A is indefinite.

Theorem 2.2.4. *If A is positive-definite, then the diagonal elements of A are positive. Similarly if A is positive semidefinite, then the diagonal elements of A are nonnegative.*

Example 2.2.5. Suppose $A \in \mathbb{R}^{n\times n}$ is a diagonal matrix, $A = \text{diag}(a)$. Then $A \succ 0 \iff a > 0$ and $A \succeq 0 iff a \geq 0$. Easy proof.

Theorem 2.2.6 (Eigenvalue Characterisation Theorem)**.** *Let $A \in \mathbb{R}^{n\times n}$ be symmetric.*

1. *A is positive (negative) definite if and only if all its eigenvalues are positive (negative).*

2. *A is positive (negative) semidefinite if and only if all its eigenvalues are nonnegative (nonpositive).*

3. *A is indefinite if and only if it has at least one positive and one negative eigenvalue.*

Proof. It follows from the spectral decomposition theorem that A is diagonalisable, *i.e.* there exists an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ such that $A = UDU^T$. Then

$$x^T Ax = x^T UDU^T x = (U^T x)^T D(U^T x) = y^T Dy = \sum_{i=1}^{n} \lambda_i y_i^2.$$

■

How to check if the matrix is positive/negative definite in general?

1. Look at the diagonal entries. (note this doesn't imply that it is positive definite, rather it gives you a candidate)

2. Compute the eigenvalues

3. Use a rule that guarantees that eigenvalues are positive/negative: 2×2 matrix and diagonally dominant matrix.

Consider a symmetric 2×2 matrix $A = \begin{bmatrix} a & b \\ b & \\ d & \end{bmatrix}$. Then

$$\det(A - \lambda I_2) = \lambda^2 - \text{Tr}(A)\lambda + \det(A)$$

and so the eigenvalue λ is given by

$$\lambda = \frac{\text{Tr}(A) \pm \sqrt{(\text{Tr}(A))^2 - 4\det(A)}}{2}.$$

First, $\lambda_1, \lambda_2 \geq 0$ if and only if $\text{Tr}(A) \geq 0$ and $\det(A) \geq 0$.

2.2.1 Diagonally dominant matrices

Definition 2.2.7. Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then

1. A is **diagonally dominant** if $|A_{ii}| \geq \sum_{j \neq i} |A_{ij}|$ for all $i = 1, 2, \ldots, n$.
2. A is **strictly diagonally dominant** if $|A_{ii}| > \sum_{j \neq i} |A_{ij}|$ for all $i = 1, 2, \ldots, n$.

Theorem 2.2.8. *1. Let A be a diagonally dominant matrix, where the diagonal elements are nonnegative. Then A is positive semidefinite.*

2. Let A be a strictly diagonally dominant matrix, where the diagonal elements are positive. Then A is positive definite.

Proof. Uses Gershgorin's circle theorem. Check. ■

Example 2.2.9. Consider the second-order finite-difference matrix

$$D = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 \\ -1 & 2 & 1 & 0 & \ldots \end{bmatrix}.$$

This is positive semidefinite since D is diagonally dominant and its diagonal elements are nonnegative.

2.3 Second Order Optimality Conditions

2.4 Quadratic Functions

Definition 2.4.1. A **quadratic function** on $\mathbb{R}^n$ is a function of the form

$$f(x) = x^T A x + 2b^T x + c,$$

where $A \in \mathbb{R}^{n \times n}$ is symmetric, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

Lemma 2.4.2. *Let f be a quadratic function.*

(a) x is a stationary point of f if and only if $Ax = -b$.

(b) If $A \succeq 0$, then x is a global minimum point of f if and only if $Ax = -b$.

(c) If $A \succ 0$, then $x = -A^{-1}b$ is a strict global minimum point of f with value $f(x) = c - b^T A^{-1} b$.

Chapter 3

Least Square Problems

Suppose we want to solve a linear system $Ax = b$, where $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ and $\text{rank}(A) = n \leq m$, *i.e.* A has full column rank (A is tall and skinny). Note that if $m = n$, then the solution is simply $A^{-1}b$. If $m > n$, we say that the system is overdetermined and typically inconsistent. Let's look for a solution in the sense that the residual vector $r = Ax - b$ is as small as possible in the ℓ^2-norm. This gives rise to the least square problem:

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|^2.$$

We choose the ℓ^2-norm precisely because then the objective function becomes a quadratic function:

$$f(x) = (Ax - b)^T(Ax - b) = x^T A^T Ax - 2b^T Ax + b^T b.$$

By Lemma 2.41, a strict global minimum point is guaranteed if $A^T A \succ 0$, and this holds if A has full column rank. Simplifying $\nabla f(x^*) = \mathbf{0}$ yields the **normal equation**:

$$A^T Ax^* = A^T b$$

which then says that $x^* = (A^T A)^{-1}A^T b$, if the least square solution exists. Note that we recover the expected solution if $m = n$:

$$x^* = (A^T A)^{-1}A^T b = A^{-1}(A^T)^{-1}A^T b = A^{-1}b.$$

3.1 Regularised Least Squares and Denoising

Here the goal is to denoise a noisy signal b by solving the **regularised least squares problem**:

$$\min_{x\in\mathbb{R}^n} \|x - b\|^2 + \lambda R(x).$$

Here, notice that $A = I$ and we have a second term, called a **regularisation term**. The constant $\lambda > 0$ is called the regularisation parameter. The regularisation function, $R(x)$, is chosen to "regularise" the solution. In this example, we will choose $R(x) = \|Dx\|^2$, where $D \in \mathbb{R}^{p\times n}$ is a given matrix chosen so that the solution is "smoother" than the origina ldata. Stationary points of this regularised least squares problem satisfy

$$(I + \lambda D^T D)x^* = b.$$

This follows from expanding the objective function and notice that it is a quadratic function:

$$\begin{aligned}\|x-b\|^2+\lambda\|Dx\|^2 &= x^Tx-2b^Tx+b^TB+\lambda x^TD^TDx \\ &= x^T(I+\lambda D^TD)x-2b^Tx+b^Tb\end{aligned}$$

The question is how should we choose the matrix D?

3.2 Nonlinear Least Squares

We are given a system of nonlinear equations

$$f_i(x)=b_i,\quad i=1,\dots,n.$$

The nonlinear least squares problem is the following:

$$\min_{\boldsymbol{x}\in\mathbb{R}_n}\sum_{i=1}^{n}\left(f_i(x)-b_i\right)^2.$$

Good algorithms for solving nonlinear least squares problems, *e.g.* Gauss-Newton (later); note that this method doesn't guarantee that it will converge to an optimal point.

Example 3.2.1 (Circle fitting)**.** Given m points $a_1,\dots a_m\in\mathbb{R}^2$, the circle fitting problem is to find the circle

$$C(x,r)=\{y\in\mathbb{R}^2\colon\|y-x\|<r\}$$

that "best" fits the m points. The associated nonlinear equations is $f_i=\|x-a_i\|\approx r$, which gives the following problem:

$$\min_{x\in\mathbb{R}^2,r>0}\sum_{i=1}^{m}\left(\|x-a_i\|^2-r^2\right)^2.$$

Let us expand the objective functional:

$$\begin{aligned}\sum_{i=1}^{m}\left(\|x-a_i\|^2-r^2\right)^2 &= \sum_{i=1}^{m}\left(\|x\|^2-2a_i^Tx+\|a_i\|^2-r^2\right)^2 \\ &= \sum_{i=1}^{m}\left(R-2a_i^Tx+\|a_i\|^2\right)^2\end{aligned}$$

where $R=\|x\|^2-r^2$; the condition $r>0$ implies that we need to impose $R\le\|x\|^2$. The problem reduces to

$$\min_{x\in\mathbb{R}^2,R\le\|x\|^2}f(x,r)=\min_{x\in\mathbb{R}^2,R\le\|x\|^2}\left(\sum_{i=1}^{m}R-2a_i^Tx+\|a_i\|^2\right)^2.$$

We claim the constraint $R\le\|x\|^2$ is not necessary. Let us check if $f(x,R)$ is a quadratic function:

$$f(x,R)=\sum_{i=1}^{m}\left(\begin{bmatrix}2a_i & -1\end{bmatrix}\begin{bmatrix}x\\R\end{bmatrix}-\|a_i\|^2\right)^2$$

$$= \|\tilde{A}x - b\|^2$$

for $\tilde{A} \in \mathbb{R}^{m\times 3}$, $y \in \mathbb{R}^3$ and $b \in \mathbb{R}^m$. Finally, the problem transforms to a linear least squares problem in the new variable:

$$\min_{y\in\mathbb{R}^3} \|\tilde{A}x - b\|^2.$$

Let us prove the claim now. Suppse $(\hat{x}, \hat{R})$ is optimal with $\|x\|^2 < R$. Then

$$-2a_i^T\hat{x} + \hat{R} + \|a_i\|^2 > -2a_i^T\hat{x} + \|\hat{x}\|^2 + \|a_i\|^2 = \|\hat{x} - a_i\|^2 \geq 0.$$

Squaring each side and sum over $i = 1, \ldots, m$, we get

$$f(\hat{x}, \hat{R}) > f(\hat{x}, \|\hat{x}\|^2).$$

Since $f(\hat{x}, \|\hat{x}\|^2) < f(\hat{x}, \hat{R})$, this contradicts the optimality of $(\hat{x}, \hat{R})$.

Chapter 4

The Gradient Method

4.1 sdf

4.2 sdf

4.3 The Condition Number

Consider minimising the quadratic function $f(x) = x^T Ax$.

Lemma 4.3.1. *Let* $\{x_k\}_{k=0}^{\infty}$ *be the sequence generated by the gradient descent method with exact line search. Then for any* $k = 0, 1, 2, \dots$

$$f(x_{k+1}) \leq \left(\frac{M-m}{M+m}\right)^2 f(x_k),$$

where $M = \lambda_{max}(A)$ *and* $m = \lambda_{min}(A)$.

Proof.

$$\begin{aligned}
f(x_{k+1}) &= x_{k+1}^T A x_{k+1} \\
&= (x_k - t_k d_k)^T A (x_k - t_k d_k) \\
&= x_k^T A x_k - 2 t_k d_k^T A x_k + t_k^2 d_k^T A d_k \\
&= x_k^T A x_k - t_k d_k^T d_k + t_k^2 d_k^T A d_k \\
&= x_k^T A x_k - \left(\frac{d_k^T d_k}{2 d_k^T A d_k}\right) d_k^T d_k + \left(\frac{d_k^T d_k}{2 d_k^T A d_k}\right)^2 d_k^T A d_k \\
&= x_k^T A x_k - \frac{1}{4}\left(\frac{\left(d_k^T d_k\right)^2}{d_k^T A d_k}\right) \\
&= \left(1 - \frac{1}{4}\frac{\left(d_k^T d_k\right)^2}{\left(d_k^T A d_k\right)\left(x_k^T A A^{-1} A x_k\right)}\right) f(x_k) \\
&= \left(1 - \frac{1}{4}\frac{\left(d_k^T d_k\right)^2}{\left(d_k^T A d_k\right)\left(d_k^T A^{-1} d_k\right)}\right) f(x_k)
\end{aligned}$$

Recall the Kantorovich inequality: for any positive definite matrix $A \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n, x \neq \mathbf{0}$ we have that

$$\frac{\left(x^T x\right)^2}{\left(x^T A x\right)\left(x^T A^{-1} x\right)} \geq \frac{4\lambda_{\max}(A)\lambda_{\min}(A)}{\left(\lambda_{\max}(A)+\lambda_{\min}(A)\right)^2}.$$

Thus

$$f(x_{k+1}) \leq \left(1 - \frac{4Mm}{(M+m)^2}\right) f(x_k) = \left(\frac{M-m}{M+m}\right)^2 f(x_k).$$

■

Note that $f(x_k)$ is a sequence bounded above by a decreasing geometric sequence. Indeed,

$$f(x_k) \leq \left(\frac{M-m}{M+m}\right)^2 f(x_{k-1}) \leq \cdots \leq \left(\frac{M-m}{M+m}\right)^{2k} f(x_0).$$

We say that the sequence $f(x_k)$ converges linearly.

4.4 Convergence Analysis of Steepest Descent Method

Consider a general unconstrained problem

$$\min_{x\in\mathbb{R}^n} f(x)$$

where we assume that $f \in C^1(\mathbb{R}^n)$ and the gradient is globally Lipschitz over $\mathbb{R}^n$. The class of functions with Lipschitz gradient with constant L is denoted by $C_L^{1,1}(\mathbb{R}^n)$. As an example, the quadratic function $f(x) = x^T A x$ is in $C_{2\|A\|_2}^{1,1}(\mathbb{R}^n)$.

Theorem 4.4.1. *Let $f \in C^2(\mathbb{R}^n)$. The following are equivalent:*

1. *$f \in C_L^{1,1}(\mathbb{R}^n)$.*
2. *$\|\nabla^2 f(x)\|_2 \leq L$ for any $x \in \mathbb{R}^n$.*

Theorem 4.4.2. *Let $f \in C_L^{1,1}(\mathbb{R}^n)$ and $\{x_k\}_{k\geq 0}$ be the sequence generated by the steepest descent method for solving*

$$\min_{x\in\mathbb{R}^n} f(x)$$

with one of the following stepsize strategies:

1. *constant stepsize $\tilde{t} \in \left(0, \frac{2}{L}\right)$,*
2. *exact line search,*
3. *backtracking line search with parameters $s \in \mathbb{R}_{++}, \alpha \in (0,1)$ and $\beta \in (0,1)$.*

Assume f is bounded below over $\mathbb{R}^n$. Then

(a) The sequence $\{f(x_k)\}_{k\geq 0}$ is nonincreasing with $f(x_k) > f(x_{k+1})$ unless $\nabla f(x_k) = 0$.

(b) $\nabla f(x_k) \longrightarrow 0$ as $k \longrightarrow \infty$.

(c) $f(x_k) - f(x_{k+1}) \geq M\|\nabla f(x_k)\|^2$, *where*

$$M = \begin{cases} \tilde{t}\left(1 - \dfrac{\tilde{t}L}{2}\right) & \textit{for constant stepsize,} \\ \dfrac{1}{2L} & \textit{for exact line search,} \\ \alpha \min\left\{s, \dfrac{2\beta(1-\alpha)}{L}\right\} & \textit{for backtracking.} \end{cases}$$

(d) Let f^* *be the limit of the convergent sequence* $\{f(x_k)\}_{k\geq 0}$*. Then for any* $n = 0, 1, 2, \ldots$ *we have*

$$\min_{k=0,1,\ldots,n} \|\nabla f(x_k)\| \leq \sqrt{\frac{f(x_0) - f^*}{M(n+1)}}.$$

Statement (4) says that the gradient of one of the previous step is small, *i.e.* the gradient at the present step might not be small.

Lemma 4.4.3 (Descent Lemma)**.** *Let* $f \in C_L^{1,1}(\mathbb{R}^n)$ *for some* $L > 0$*. Then for any* $x, y \in \mathbb{R}^n$ *we have*

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|x - y\|^2.$$

Proof. By the fundamental theorem of calculus,

$$\begin{aligned} f(y) - f(x) &= \int_0^1 \langle \nabla f(x + t(y-x)), y_x \rangle \, dt \\ &= \langle \nabla f(x), y - x \rangle + \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), y - x \rangle \, dt \end{aligned}$$

Thus

$$\begin{aligned} |f(y) - f(x) - \langle \nabla f(x), y - x \rangle| &\leq \int_0^1 |\langle \nabla f(x + t(y-x)) - \nabla f(x), y_x \rangle| \, dt \\ &\leq \int_0^1 \|\langle \nabla f(x + t(y-x)) - \nabla f(x)\| \cdot \|y - x\| \, dt \\ &\leq \int_0^1 tL\|y - x\|^2 \, dt \\ &= \frac{L}{2}\|y - x\|^2. \end{aligned}$$

■

Lemma 4.4.4 (Sufficient Decrease Lemma)**.** *Suppose that* $f \in C_L^{1,1}(\mathbb{R}^n)$*. Then for any* $x \in \mathbb{R}^n$ *and* $t > 0$ *we have*

$$f(x) - f(x - t\nabla f(x)) \geq t\left(1 - \frac{Lt}{2}\right)\|\nabla f(x)\|^2.$$

Proof. This follows directly from the descent lemma:

$$f(x - t\nabla f(x)) \leq f(x) - t\|\nabla f(x)\|^2 + \frac{Lt^2}{2}\|\nabla f(x)\|^2$$

$$= f(x) - t\left(1 - \frac{Lt}{2}\right)\|\nabla f(x)\|^2.$$

■

Proof of giant theorem. For part (c), the sufficient decrease lemma gives

$$f(x_k) - f(x_{k+1}) \geq t\left(1 - \frac{Lt}{2}\right)\|\nabla f(x_k)\|^2$$

If we choose $\tilde{t} \in \left(0, \frac{2}{L}\right)$, then the difference greater than 0. For the exact line search, we want to maximise the function $t\left(1 - \frac{Lt}{2}\right)$ over the interval $\left(0, \frac{2}{L}\right)$. The maximum is attained at $t = 1/L$.

For part (a),

$$f(x_k) - f(x_{k+1}) \geq M\|\nabla f(x_k)\|^2 \geq 0$$

for some constant $M > 0$, and hence $f(x_k) > f(x_{k+1})$ unless $\nabla f(x_k) = 0$.

For part (b), note that since the sequence $\{f(x_k)\}_{k\geq 0}$ is nonincreasing and bounded below, it converges. In particular,

$$0 \leq \|\nabla f(x_k)\|^2 \leq \frac{1}{M}\left[f(x_k) - f(x_{k+1})\right] \longrightarrow 0 \quad \text{as } k \longrightarrow \infty.$$

For part (d), summing the inequality

$$f(x_k) - f(x_{k+1}) \geq M\|\nabla f(x_k)\|^2$$

over $k = 0, 1, \ldots, n$, the LHS is a telescopic sum and so

$$f(x_0) - f(x_{n+1}) \geq M\sum_{k=0}^{n}\|\nabla f(x_k)\|^2$$

Since $f(x_{n+1}) \geq f^*$, we obtain

$$f(x_0) - f^* \geq M\sum_{k=0}^{n}\|\nabla f(x_k)\|^2 \geq M(n+1)\min_{k=0,1,\ldots,n}\|\nabla f(x_k)\|^2.$$

■

Chapter 5

Newton's Method

5.1 xxx

Chapter 6

Convex Optimisation

CHT: Motivation, Introduction

6.1 Convex Sets

A set $C \subseteq \mathbb{R}^n$ is called **convex** if for any $x, y \in C$ and $\lambda \in [0,1]$, the point $\lambda x + (1-\lambda)y$ belongs to C. Equivalently, we say that C is a convex set if the closed line segment $[x, y]$ connecting any two points $x, y \in C$ is in C.

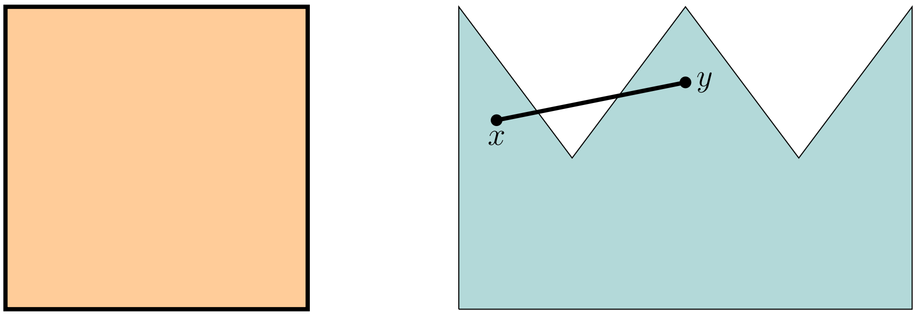


Figure 6.1: Examples of convex and nonconvex sets. *Left:* A square is convex. *Right:* The sawtooth-shaped set is not convex.

6.1.1 Important examples

The empty set is vacuously convex. The simplest examples of nonempty convex sets are singletons and the Euclidean space $\mathbb{R}^n$. It is geometrically obvious that hyperplanes, half-spaces, norm balls and ellipsoids are convex sets, but we include their proofs to illustrate how one can establish convexity of sets using the definition of a convex set.

Lemma 6.1.1 (Convexity of hyperplanes and half-spaces)**.** *Let $a \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $b \in \mathbb{R}$. The following sets are convex:*

*(a) the **hyperplane** $H = \{x \in \mathbb{R}^n \colon a^T x = b\}$;*

*(b) the **(closed) half-space** $H^- = \{x \in \mathbb{R}^n \colon a^T x \leq b\}$;*

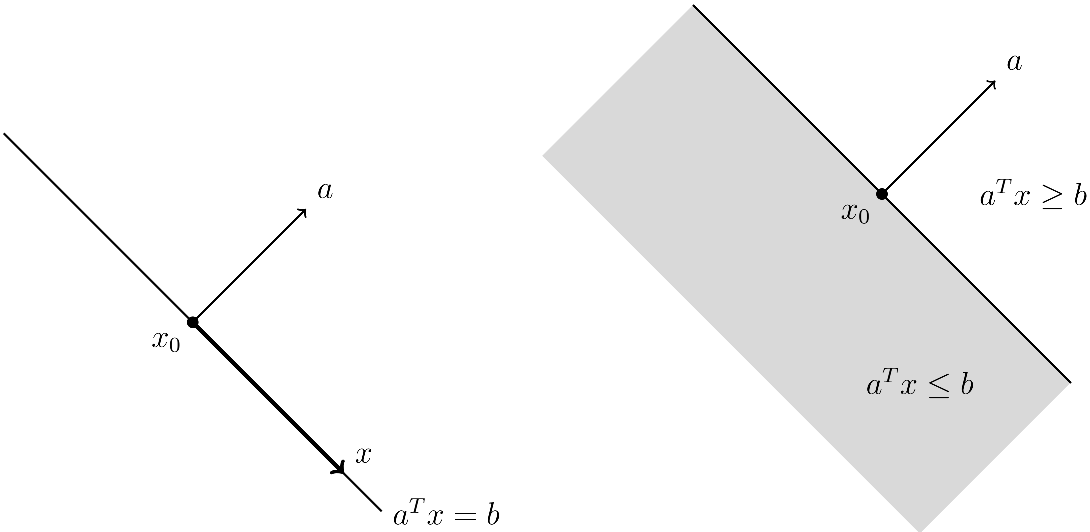
(c) the open half-space $\{x \in \mathbb{R}^n : a^T x < b\}$.

Proof. Let $x, y \in H^-$ and $\lambda \in [0,1]$. Then

$$a^T[\lambda x + (1-\lambda y)] = \lambda a^T x + (1-\lambda)a^T y \leq \lambda b + (1-\lambda)b = b$$

where we crucially use the fact that $\lambda \in [0,1]$. This shows that $\lambda x + (1-\lambda y) \in H^-$ and so H^- is convex. An identical argument replacing inequality with equality sign shows that hyperplanes are convex.

■



(a) Hyperplane in $\mathbb{R}^2$ (b) Half-space in $\mathbb{R}^2$

Figure 6.2: Hyperplane and half-space in $\mathbb{R}^2$, with outward normal vector $a \in \mathbb{R}^2$ and a point x_0 in the hyperplane. Observe that a hyperplane defines two halfspaces (Adopted from [BV04]).

Lemma 6.1.2 (Convexity of balls)**.** *Let* $c \in \mathbb{R}^n$ *and* $r > 0$*. For any arbitrary norm* $\|\cdot\|$ *defined on* $\mathbb{R}^n$*, the open ball* $B(c,r)$ *and closed ball* $B[c,r]$ *are convex.*

Proof. We will show the convexity of $B[c,r]$ as the proof of the convexity of $B(c,r)$ is almost identical. Let $x, y \in B[c,r]$ and $\lambda \in [0,1]$. It follows from the triangle inequality of $\|\cdot\|$ that

$$\begin{aligned} \|\lambda x + (1-\lambda)y - c\| &= \|\lambda(x-c) + (1-\lambda)(y-c)\| \\ &= |\lambda|\|x-c\| + |1-\lambda|\|y-c\| \\ &\leq \lambda r + (1-\lambda)r = r \end{aligned}$$

where we again crucially use the fact that $\lambda \in [0,1]$. This shows that $\lambda x + (1-\lambda y) \in B[c,r]$ and so $B[c,r]$ is convex.

■

Lemma 6.1.3 (Convexity of ellipsoids)**.** *Let* $Q \in \mathbb{R}^{n\times n}$ *be positive semidefinite,* $b \in \mathbb{R}^n$ *and* $c \in \mathbb{R}$*. Then the ellipsoid*

$$E = \{x \in \mathbb{R}^n : x^T Q x + 2b^T x + c \leq 0\}$$

is convex.

Proof. Let $x, y \in E$, $\lambda \in [0,1]$ and define $z = \lambda x + (1-\lambda)y$. We need to show that

$$z^T Q z + 2b^T z + c \leq 0.$$

Expanding the first term, we obtain

$$\begin{aligned} z^T Q z &= [\lambda x + (1-\lambda)y]^T Q [\lambda x + (1-\lambda)y] \\ &= \lambda^2 x^T Q x + (1-\lambda)^2 y^T Q y + 2\lambda(1-\lambda) x^T Q y. \end{aligned} \tag{6.1.1}$$

Since Q is positive semidefinite, $Q^{1/2}$ exists and it follows from the Cauchy-Schwarz inequality that

$$\begin{aligned} 2x^T Q y = x^T Q^{1/2} Q^{1/2} y = 2\left(Q^{1/2}x\right)^T \left(Q^{1/2}y\right) & \qquad \left[Q^{1/2} \text{ is symmetric.}\right] \\ \leq 2\|Q^{1/2}x\|_2 \|Q^{1/2}y\|_2 & \qquad \left[\text{Cauchy-Schwarz inequality.}\right] \\ = 2\sqrt{x^T Q x}\sqrt{y^T Q y} & \\ \leq x^T Q x + y^T Q y. & \qquad \left[\text{Young's inequality for product.}\right] \end{aligned}$$

Substituting this into (6.1.1), we obtain

$$\begin{aligned} z^T Q z &\leq \lambda^2 x^T Q x + (1-\lambda)^2 y^T Q y + \lambda(1-\lambda)\left[x^T Q x + y^T Q y\right] \\ &= \left[\lambda^2 + \lambda(1-\lambda)\right] x^T Q x + \left[(1-\lambda)^2 + \lambda(1-\lambda)\right] y^T Q y \\ &= \lambda x^T Q x + (1-\lambda) y^T Q y \end{aligned}$$

and hence

$$\begin{aligned} z^T Q z + 2b^T z + c &\leq \lambda x^T Q x + (1-\lambda) y^T Q y + 2b^T\left(\lambda x + (1-\lambda)y\right) + c \\ &= \lambda\left(x^T Q x + 2b^T x + c\right) + (1-\lambda)\left(y^T Q y + 2b^T y + c\right) \leq 0, \end{aligned}$$

since $x, y \in E$. This shows that $z \in E$ and the desired result follows.

■

6.1.2 Algebraic operations with convex sets

Establishing convexity of sets directly from the definition of a convex set can be tedious and often requires a cunning observation, as seen in Lemma 6.1.3. We will describe some operations that preserve convexity of sets, and these operations allow us to prove that a set is convex by constructing it from simple sets for which convexity is known. Two standard set operations that yield convex sets are intersection and Cartesian product.

Lemma 6.1.4. *Let $C_i \subset \mathbb{R}^{k_i}$ be a convex set for any $i = 1, \dots, m$. Then the Cartesian product*

$$C_1 \times C_2 \times \cdots \times C_m = \left\{(x_1, x_2, \dots, x_m) \in \prod_{i=1}^m \mathbb{R}^{k_i} : x_i \in C_i,\ i = 1, 2, \dots, m\right\}$$

is convex.

Proof. Suppose $x, y \in C_1 \times C_2 \times \cdots \times C_m$ and $\lambda \in [0,1]$. Then

$$\begin{aligned}\lambda x + (1-\lambda)y &= \lambda\,(x_1, x_2, \ldots, x_m) + (1-\lambda)\,(y_1, y_2, \ldots, y_m) \\ &= \sum_{i=1}^{m} (\lambda x_i + (1-\lambda) y_i)\, \boldsymbol{e}_i \in C_1 \times C_2 \times \cdots \times C_m,\end{aligned}$$

since C_i is convex and $x_i, y_i \in C_i$ for each $i = 1, \ldots, m$. The desired statement follows. ■

Lemma 6.1.5 (Closed under arbitrary intersections)**.** *Let $C_i \subseteq \mathbb{R}^n$ be a convex set for any $i \in I$, where I is an index set (possibly infinite). Then the set $C = \bigcap_{i \in I} C_i$ is convex.*

Proof. Suppose $x, y \in C$ and $\lambda \in [0,1]$. Then $x, y \in C_i$ for any $i \in I$ and since C_i is convex, it follows that $\lambda x + (1-\lambda)y \in C_i$ for any $i \in I$, *i.e.* $\lambda x + (1-\lambda) y \in \bigcap_{i \in I} C_i = C$. ■

Example 6.1.6. Consider the (convex) polytope P defined by

$$P = \{x \in \mathbb{R}^n \colon Ax \leq b\},$$

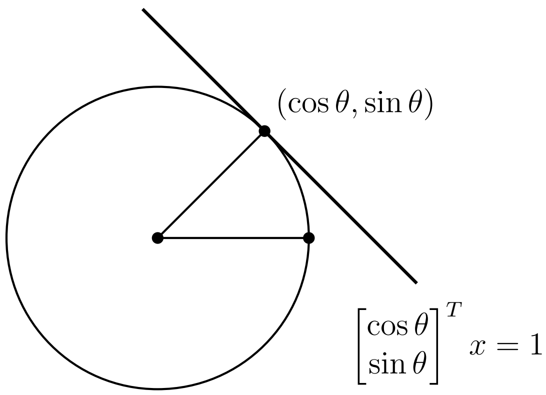
where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, *i.e.* P is the solution set of a finite number of linear inequalities. Actually, we can write P as an intersection of half-spaces:

$$P = \bigcap_{i=1}^{m} \left\{x \in \mathbb{R}^n \colon a_i^T x \leq b_i\right\},$$

where a_i^T is the ith-row of A. Since half-spaces are convex (see Lemma 6.1.1), it follows from Lemma 6.1.5 that P is also convex.

Example 6.1.7. It was proven in Lemma 6.1.2 that the unit ball $B[0,1]$ in $\mathbb{R}^2$ is convex. We present an alternate proof of this fact using Lemma 6.1.5. Indeed, the unit ball in $\mathbb{R}^2$ can be represented as the intersection of infinitely many half-spaces:

$$B[0,1] = \bigcap_{\theta \in [0, 2\pi]} \left\{ x \in \mathbb{R}^2 \colon \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}^T x \leq 1 \right\}.$$



Let us verify this expression for certain special angles:

$$\begin{aligned}
\theta = 0 &\implies x_1 \leq 1 \\
\theta = \frac{\pi}{4} &\implies \frac{1}{\sqrt{2}}(x_1 + x_2) \leq 1 \\
&\implies x_2 \leq \sqrt{2} - x_1 \\
\theta = \frac{\pi}{2} &\implies x_2 \leq 1 \\
\theta = \pi &\implies -x_1 \leq 1 \implies x_1 \geq -1
\end{aligned}$$

Lemma 6.1.8 (Closed under affine transformations)**.** *Let $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be an affine function.*

(a) If $C \subseteq \mathbb{R}^n$ is a convex set, then the image of C under f,

$$f(C) = \{f(x) \in \mathbb{R}^m \colon x \in C\},$$

is convex.

(b) If $D \subseteq \mathbb{R}^m$ is a convex set, then the inverse image of D under f,

$$f^{-1}(D) = \{x \in \mathbb{R}^n \colon f(x) \in D\},$$

is convex.

Proof. Since f is an affine function, there exists $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ such that

$$f(x) = Ax + b \quad \text{for all } x \in \mathbb{R}^n.$$

Let $y_1, y_2 \in f(C) \subseteq \mathbb{R}^m$ and $\lambda \in [0,1]$, then there exists corresponding vectors $x_1, x_2 \in C \subseteq \mathbb{R}^n$ such that $f(x_1) = y_1$ and $f(x_2) = y_2$. We will show that the point $z = \lambda y_1 + (1-\lambda)y_2 \in f(C)$. Indeed,

$$\begin{aligned} z &= \lambda f(x_1) + (1-\lambda) f(x_2) \\ &= \lambda(Ax_1 + b) + (1-\lambda)(Ax_2 + b) \\ &= A\left[\lambda x_1 + (1-\lambda)x_2\right] + b \in f(C), \end{aligned}$$

since C is convex and $x_1, x_2 \in C$. This proves part (a). On the other hand, let $x_1, x_2 \in f^{-1}(D) \subseteq \mathbb{R}^n$ and $\lambda \in [0,1]$, then there exists corresponding vectors $y_1, y_2 \in D \subseteq \mathbb{R}^m$ such that $f(x_1) = y_1$ and $f(x_2) = y_2$. We will show that the point $z = \lambda x_1 + (1-\lambda)x_2 \in f^{-1}(D)$, or equivalently, $f(z) \in D$. Indeed,

$$\begin{aligned} f(z) = f\left(\lambda x_1 + (1-\lambda)x_2\right) &= A\left[\lambda x_1 + (1-\lambda)x_2\right] + b \\ &= \lambda\left[Ax_1 + b\right] + (1-\lambda)\left[Ax_2 + b\right] \\ &= \lambda f(x_1) + (1-\lambda) f(x_2) \\ &= \lambda y_1 + (1-\lambda) y_2 \in D, \end{aligned}$$

since D is convex. This proves part (b).

■

Corollary 6.1.9. *The following set operations preserve convexity.*

(a) Scaling and translation: If $C \subseteq \mathbb{R}^n$ is a convex set, $\alpha \in \mathbb{R}$ and $b \in \mathbb{R}^n$, then the sets αC and $C + b$ are convex, where

$$\alpha C = \{\alpha x \in \mathbb{R}^n \colon x \in C\} \quad \textit{and} \quad C + b = \{x + b \colon x \in C\}.$$

(b) Coordinate projection: If $C \coloneqq \prod_{i=1}^m C_i \subseteq \prod_{i=1}^m \mathbb{R}^{k_i}$ is a convex set, then the projection of C onto C_i, $proj_i(C)$, is convex for any $i = 1, \dots, m$, where

$$proj_i(C) = \left\{x_i \in C_i \colon x = (x_1, \dots, x_m) \in \prod_{i=1}^m C_i\right\}.$$

(c) *Minkowski sum: If $C_1, \dots, C_k \subseteq \mathbb{R}^n$ are convex sets and $\mu_1, \dots, \mu_k \in \mathbb{R}$, then the Minkowski sum $\mu_1 C_1 + \cdots + \mu_k C_k$ is convex, where*

$$\mu_1 C_1 + \cdots + \mu_k C_k = \left\{ \sum_{i=1}^{k} \mu_i x_i \colon x_i \in C_i,\ i = 1, \dots, k \right\}.$$

Proof. These results are consequences of Lemma 6.1.8 with suitably chosen affine functions. Part (a) follows by choosing $f(x) = \alpha x$ for scaling and $f(x) = x + b$ for translation. Part (b) follows by realising the projection of C_i as image of C under the affine function $f\colon C \longrightarrow C_i$ defined by

$$f(x) = \boldsymbol{e}_i^T x, \quad x = (x_1, \dots, x_m)^T \in \prod_{i=1}^{m} \mathbb{R}^{k_i}.$$

Part (c) follows by realising the Minkowski sum as the image of the Cartesian product $\prod_{i=1}^{k} C_k$ under the linear function $f(x_1, \dots, x_k) = \mu_1 x_1 + \cdots + \mu_k x_k$. CHT: Rewrite part (b). ■

6.1.3 Convex hull

Definition 6.1.10. Given $x_1, \dots, x_k \in \mathbb{R}^n$, a **convex combination** of these vectors is a vector of the form

$$\lambda_1 x_1 + \cdots + \lambda_k \lambda_k$$

where $\lambda = (\lambda_1, \dots, \lambda_k) \in \Delta_k$.

For instance, points on a line segment are convex combinations of the endpoints. Given $x_1, \dots, x_k \in \mathbb{R}^n$, the mean $\frac{1}{k} \sum_{i=1}^{k} x_k$ is a convex combination of WHAT?

Theorem 6.1.11. *Let $C \subset \mathbb{R}^n$ be a convex set and $x_1, \dots, x_m \in C$. Then for any $\lambda \in \Delta_m$, the convex combination $\sum_{i=1}^{m} \lambda_i x_i \in C$.*

Definition 6.1.12. Let $S \in \mathbb{R}^n$. The **convex hull** of S, denoted $\mathrm{conv}(S)$ is the set of all convex combinations of vectors in S. That is,

$$\mathrm{conv}(S) = \left\{ \sum_{i=1}^{k} \lambda_i x_i \colon x_1, \dots, x_k \in S,\ \lambda \in \Delta_k,\ k \in \mathbb{N} \right\}.$$

Lemma 6.1.13. *Let $S \subset \mathbb{R}^n$. If $S \subset T$ for some convex set T, then $conv(S) \subset T$. This implies that $conv(S)$ is the smallest convex set containing S.*

Proof. Suppose $S \subset T$ for a convex set T. Let $z \in \mathrm{conv}(S)$, then there exists $x_1, \dots, x_k \in S \subset T$ and $\lambda \in \Delta_k$ such that

$$z = \sum_{i=1}^{k} \lambda_i x_i.$$

By Theorem 6.10, $z \in T$ since T is convex and this shows that $\mathrm{conv}(S) \subset T$. ■

Theorem 6.1.14 (Caratheodory Theorem)**.** *Let $S \subset \mathbb{R}^n$ and $x \in conv(S)$. Then there exists $x_1, \ldots, x_{n+1} \in S$ such that $x \in conv(\{x_1, \ldots, x_{n+1}\})$*

Proof. Let $x \in \text{conv}(S)$. There exists $x_1, \ldots, x_k \in S$ and $\lambda \in \Delta_k$ such that

$$x = \sum_{i=1}^{k} \lambda_i x_i.$$

The question now is how small can we choose k? WLOG, we can take $\lambda_i > 0$ for all $i = 1, \ldots, k$. If $k \geq n+2$, consider the vectors $x_2 - x_1, \ldots, x_k - x_1$ which are more than n vectors in $\mathbb{R}^n$. They form a linearly dependent set of vectors in $\mathbb{R}^n$ and so there exists $\mu_2, \ldots, \mu_k$ such that

$$\sum_{i=2}^{k} \mu_i (x_i - x_1) = 0.$$

Define $\mu_1 = -\sum_{i=2}^{k} \mu_i$, then $\sum_{i=1}^{k} \mu_i = 0$ and in particular $\sum_{i=1}^{k} \mu_i x_i = \mathbf{0}$. We claim that at least one of the μ_i is negative. Let $\alpha \geq 0$. Then

$$x = \sum_{i=1}^{k} \lambda_i x_i = \sum_{i=1}^{k} \lambda_i + \alpha \sum_{i=1}^{k} \mu_i x_i = \sum_{i=1}^{k} (\lambda_i + \alpha \mu_i) x_i.$$

Note that we have $\sum_{i=1}^{k} \lambda_i + \alpha \mu_i = 1$ and so the above is a convex combination representation if and only if

$$\lambda_i + \alpha \mu_i \geq 0 \quad \text{for all } i = 1, \ldots, k.$$

Since $\lambda_i \geq 0$ for all i, the above inequalities are satisfied if we choose

$$\alpha = \min_{i \colon \mu_i < 0} -\frac{\lambda_i}{\mu_i}.$$

But this implies that $\lambda_i + \alpha \mu_i = 0$ for some i, which means that x is a convex combination of $k-1$ vectors. ■

6.1.4 Convex cone

Definition 6.1.15. A set $S \subset \mathbb{R}^n$ is a **cone** if for any $x \in S$ and $\lambda \geq 0$, we have $\lambda x \in S$.

Lemma 6.1.16. *A set S is a convex cone if and only if*

(a) $x, y \in S \implies x + y \in S$

(b) $x \in S, \lambda \geq 0 \implies \lambda x \in S$

Example 6.1.17. The nonnegative orthant $\mathbb{R}^n_+$ is a convex cone. Let's verify this using the lemma above. Let $x, y \in \mathbb{R}^n_+$. Clearly $x + y \in \mathbb{R}^n_+$ and $\lambda x \in S$ for any $\lambda \geq 0$.

Example 6.1.18. Another example would be the **Lorentz cone**, given by

$$L^n = \left\{ (x, t) \in \mathbb{R}^{n+1} \colon \|x\| \leq t,\, x \in \mathbb{R}^n,\, t \in \mathbb{R} \right\}.$$

It can be shown that the Lorentz cone is actually a convex cone.

6.2 Convex Functions

Definition 6.2.1. A function $f\colon C \longrightarrow \mathbb{R}$ defined on a convex set $C \subset \mathbb{R}^n$ is convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \quad \text{for all } x, y \in C \text{ and } \lambda \in [0,1]$$

and strictly convex if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y) \quad \text{for all } x, y \in C \text{ and } \lambda \in (0,1)$$

Examples of convex functions include affine functions and norms.

Theorem 6.2.2 (Jensen's inequality)**.** *Suppose f is convex, $x_1, \ldots, x_n \in C$ and $\lambda \in \Delta_k$. Then*

$$f\left(\sum_{i=1}^{k} \lambda_i x_i\right) \leq \sum_{i=1}^{k} \lambda_i f(x_i).$$

Theorem 6.2.3 (Theorem 7.6)**.** *Suppose $f \in C^1(C)$. Then f is convex if and only if*

$$f(x) + \nabla f(x)^T (x-y) \leq f(y) \quad \textit{for any } x, y \in C.$$

Proposition 7.8.

Theorem 6.2.4. *Suppose $f \in C^2(C)$. Then f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in C$.*

6.2.1 Operations preserving convexity

lalallalalalaallala

Theorem 6.2.5. *Let $f_1, \ldots, f_p\colon C \longrightarrow \mathbb{R}$ be p convex functions over a convex set $C \subseteq \mathbb{R}^n$. Then the maximum function*

$$f(x) = \max_{i=1,\ldots,p} f_i(x)$$

is convex over C.

Proof. Let $x, y \in C$ and $\lambda \in [0,1]$. Then

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\leq \max_{i=1,\ldots,p} f_i\left(\lambda x + (1-\lambda)y\right) \\ &\leq \max_{i=1,\ldots,p} \left[\lambda f_i(x) + (1-\lambda)f_i(y)\right] \\ &\leq \lambda \max_{i=1,\ldots,p} f_i(x) + (1-\lambda) \max_{i=1,\ldots,p} f_i(y) \\ &= \lambda f(x) + (1-\lambda)f(y). \end{aligned}$$

∎

Example 6.2.6. Example 7.27

Theorem 6.2.7 (Theorem 7.2.8)**.** *Let $f\colon C \times D \longrightarrow \mathbb{R}$ be a convex function where $C \subseteq \mathbb{R}^m$ and $D \subseteq \mathbb{R}^n$ are convex. Then*

$$g(x) = \min_{y \in D} f(x, y)$$

is convex over C.

Example 6.2.8. Let $C \subseteq \mathbb{R}^n$ be a convex set. Then

$$f(x) = d(x, C) = \min_{y \in C} \|x - y\|$$

is convex.

6.2.2 Sublevel sets of convex functions

Definition 6.2.9. Let $f\colon S \longrightarrow \mathbb{R}$ be defined on $S \subseteq \mathbb{R}^n$. Then a **α-sublevel set of f** is

$$\operatorname{Lev}(f,\alpha) = \{x \in S \colon f(x) \leq \alpha\}.$$

Theorem 6.2.10. *Let $f\colon C \longrightarrow \mathbb{R}$ be a convex function on a convex set $C \subseteq \mathbb{R}^n$. For any $\alpha \in \mathbb{R}$, the α-sublevel set of f is a convex set.*

Proof. Let $x, y \in \operatorname{Lev}(f,\alpha)$ and $\lambda \in [0,1]$. Then

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\leq \lambda f(x) + (1-\lambda) f(y) \\ &\leq \lambda\alpha + (1-\lambda)\alpha = \alpha. \end{aligned}$$

This shows that $\lambda x + (1-\lambda)y \in \operatorname{Lev}(f,\alpha)$. ■

Example 6.2.11. This can be used to show that $B[0,1] \subset \mathbb{R}^n$ is a convex set. Indeed,

$$B[0,1] = \operatorname{Lev}(f,1), \quad \text{where } f(x) = \|x\|.$$

Example 6.2.12. Consider the following subset of $\mathbb{R}^n$:

$$D = \left\{ x \in \mathbb{R}^n \colon \left(x^T Q x + 1\right)^2 + \ln\left(\sum_{i=1}^n e^{x_i}\right) \leq 10 \right\},$$

where $Q \in \mathbb{R}^{n\times n}, Q \succeq \mathbf{0}$. The set D is convex since $D = \operatorname{Lev}(f,10)$, where

$$f(x) = \left(x^T Q x + 1\right)^2 + \ln\left(\sum_{i=1}^n e^{x_i}\right).$$

Sublevel sets do not characterise convex functions, take a function with a cusp for example (or Heaviside function). Convex implies quasiconvex, not the other way.

Definition 6.2.13. A function $f\colon C \longrightarrow \mathbb{R}$ defined on a convex st $C \subseteq \mathbb{R}^n$ is quasiconvex if $\operatorname{Lev}(f,\alpha)$ are convex for all $\alpha \in \mathbb{R}$.

6.2.3 Maxima of convex functions

Theorem 6.2.14. *Let $f\colon C \longrightarrow \mathbb{R}$ be a convex function which is not constant over a convex set $C \subseteq \mathbb{R}^n$. Then f does not attain its maximum at a point in the interior of C. That is, if f attains its maxima, then it must be at the boundary.*

Theorem 6.2.15. *Let $f\colon C \longrightarrow \mathbb{R}$ be a convex, continuous function over a convex, compact set $C \subseteq \mathbb{R}^n$. Then there exists at least one maximiser of f over C that is an extreme point of C.*

Proof. Let x^* be a maximiser of f over C. Krein-Milman asserts that

$$C = \operatorname{conv}(\operatorname{ext}(C)),$$

i.e. there exists $x_1, \ldots x_k \in \text{ext}(C)$ and $\lambda \in \Delta_k$ such that

$$x^* = \sum_{i=1}^{k} \lambda_i x_i.$$

By Jensen's inequality,

$$f(x^*) \leq \sum_{i=1}^{k} \lambda_i f(x_i).$$

In particular,

$$\sum_{i=1}^{k} \lambda_i \left[f(x_i) - f(x^*)\right] \geq 0.$$

Since x^* is a maximiser of f over C, we have that $f(x_i) \leq f(x^*)$ for each $i = 1, \ldots, k$. This means that $f(x_i) = f(x^*)$ for each $i = 1, \ldots, k$ since $\lambda_i \geq 0$ and the LHS of the inequality is a sum of nonpositive terms. Hence these extreme points are maximisers of f over C. ∎

Example 6.2.16. Let $Q \succeq \mathbf{0}$ and consider the optimisation problem

$$\max_{x \in \mathbb{R}^n, \|x\|_\infty = 1} x^T Q x.$$

Since the objective function is convex and the admissible set is compact and convex, it follows from the theorem that we only have to check extreme points $\{-1, 1\}^n$

Example 6.2.17. Linear programming: $\max c^T c$ such that $Ax \leq b$. The admissible set is a polyhedron which is a compact, convex set. So we "only" have to consider the vertices of this polyhedron to find the maximiser. Unfortunately, the number of vertices of a polyhedron grows exponentially with the dimension.

Bibliography

[Bec14] A. Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. Vol. 19. MOS-SIAM Series on Optimization. SIAM, 2014. DOI: 10.1137/1.9781611973655.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.