Chapter 1

Mathematical Foundations and Preliminaries

1.1 Introduction and Motivation

1.1.1 What is Optimization?

Optimization is everywhere around us. When you choose the shortest route to work, when Netflix recommends movies based on your preferences, when engineers design aircraft wings for maximum efficiency, or when investment managers construct portfolios to maximize returns while minimizing risk—all of these involve solving optimization problems.

Mathematically, an optimization problem seeks to find the "best" solution from a set of available alternatives. We express this as:

$$\begin{aligned} \text{minimize} \quad & f(x) && (1.1)\\ \text{subject to} \quad & x \in \mathcal{X} && (1.2) \end{aligned}$$

where:

- x is our **decision variable** (what we can control)
- $f(x)$ is our **objective function** (what we want to optimize)
- $\mathcal{X}$ is our **feasible set** (the constraints we must satisfy)

Example 1.1 (Portfolio Optimization)**.** Suppose you have \$10,000 to invest in three stocks. Let x_1, x_2, x_3 represent the amounts invested in each stock. Your objective might be to minimize risk (measured by portfolio variance) while achieving a target return of at least 8%.

$$\begin{aligned} \text{minimize} \quad & x^T \Sigma x \quad \text{(minimize risk)} && (1.3)\\ \text{subject to} \quad & \mu^T x \geq 0.08 \cdot 10000 \quad \text{(target return)} && (1.4)\\ & x_1 + x_2 + x_3 = 10000 \quad \text{(budget constraint)} && (1.5)\\ & x_1, x_2, x_3 \geq 0 \quad \text{(no short selling)} && (1.6) \end{aligned}$$

where Σ is the covariance matrix of returns and μ is the vector of expected returns.

1.1.2 Why Convex Optimization?

Not all optimization problems are created equal. Consider these two simple examples:

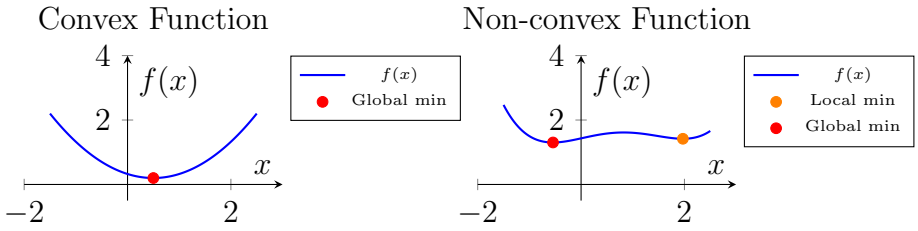


Figure 1.1: Convex vs non-convex optimization landscapes

In the convex case (left), every local minimum is a global minimum, so a local method such as gradient descent (with a suitable step size) converges to a global minimum. In the non-convex case (right), there are multiple local minima, and finding the global minimum is generally very difficult.

Key advantages of convex optimization:

1. **Global optimality**: Any local minimum is automatically a global minimum
2. **Efficient algorithms**: Many classes (LP, convex QP, SOCP, SDP) can be solved to any given accuracy in polynomial time
3. **Reliable software**: Robust, well-tested solvers are available
4. **Rich theory**: Beautiful mathematical framework
5. **Wide applicability**: Many practical problems are convex or can be reformulated as convex

1.1.3 Historical Perspective

The foundations of convex optimization were laid in the mid-20th century:

- **1940s-1950s**: Linear programming developed by Dantzig and others for military logistics
- **1960s**: Convex analysis formalized by Rockafellar and others
- **1980s**: Interior-point methods revolutionize the field (Karmarkar, 1984)
- **1990s-2000s**: Semidefinite programming and second-order cone programming emerge
- **2000s-present**: Applications explode in machine learning, signal processing, and data science

Today, convex optimization is an essential tool in engineering, economics, machine learning, and many other fields.

1.2 Linear Algebra Review

Since optimization deals with vectors and matrices extensively, let's review the essential linear algebra concepts.

1.2.1 Vectors and Vector Spaces

Definition 1.1 (Vector Space)**.** A vector space V over the real numbers $\mathbb{R}$ is a set equipped with two operations:

- Vector addition: $u + v \in V$ for all $u, v \in V$
- Scalar multiplication: $\alpha v \in V$ for all $\alpha \in \mathbb{R}, v \in V$

satisfying the usual axioms (associativity, commutativity, distributivity, etc.).

The most important vector space for us is $\mathbb{R}^n$, the set of all n-dimensional real vectors:

$$\mathbb{R}^n = \{x = (x_1, x_2, \ldots, x_n)^T : x_i \in \mathbb{R}\}$$

Definition 1.2 (Inner Product)**.** The standard inner product in $\mathbb{R}^n$ is:

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i$$

Definition 1.3 (Norm)**.** The Euclidean norm (or 2-norm) of a vector $x \in \mathbb{R}^n$ is:

$$\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Other important norms include:

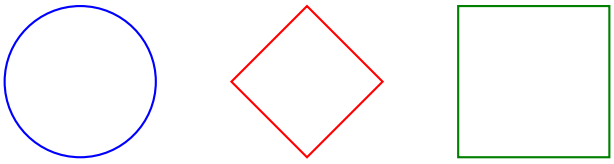
- ℓ_1 norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
- ℓ_∞ norm: $\|x\|_\infty = \max_i |x_i|$
- ℓ_p norm: $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$ for $p \geq 1$

1.2.2 Linear Independence and Bases

Definition 1.4 (Linear Independence)**.** Vectors $v_1, v_2, \ldots, v_k$ are linearly independent if the only solution to

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0$$

is $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.



ℓ_2 norm $\|x\|_2 \leq 1$ | ℓ_1 norm $\|x\|_1 \leq 1$ | ℓ_∞ norm $\|x\|_\infty \leq 1$

Figure 1.2: Unit balls for different norms in $\mathbb{R}^2$

Definition 1.5 (Span and Basis)**.** The span of vectors $v_1, \ldots, v_k$ is:

$$\text{span}\{v_1, \ldots, v_k\} = \left\{ \sum_{i=1}^{k} \alpha_i v_i : \alpha_i \in \mathbb{R} \right\}$$

A basis for a vector space V is a set of linearly independent vectors that span V.

Example 1.2 (Standard Basis)**.** The standard basis for $\mathbb{R}^3$ is:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Any vector $x \in \mathbb{R}^3$ can be written as $x = x_1 e_1 + x_2 e_2 + x_3 e_3$.

1.2.3 Matrices and Linear Transformations

Definition 1.6 (Matrix)**.** An $m \times n$ matrix A is a rectangular array of real numbers:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Every matrix $A \in \mathbb{R}^{m \times n}$ defines a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(x) = Ax$.
Important matrix properties:

- **Range (Column Space)**: $\text{range}(A) = \{Ax : x \in \mathbb{R}^n\}$
- **Null Space (Kernel)**: $\text{null}(A) = \{x : Ax = 0\}$
- **Rank**: $\text{rank}(A) = \dim(\text{range}(A))$

Theorem 1.1 (Fundamental Theorem of Linear Algebra)**.** For $A \in \mathbb{R}^{m \times n}$:

$$\text{rank}(A) + \dim(\text{null}(A)) = n$$

$$\text{rank}(A) = \text{rank}(A^T)$$

1.2.4 Eigenvalues and Eigenvectors

Definition 1.7 (Eigenvalue and Eigenvector)**.** For a square matrix $A \in \mathbb{R}^{n\times n}$, a scalar λ is an eigenvalue and $v \neq 0$ is the corresponding eigenvector if:

$$Av = \lambda v$$

The eigenvalues are the roots of the characteristic polynomial:

$$\det(A - \lambda I) = 0$$

Example 1.3 (2×2 Matrix Eigenvalues)**.** For $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$:

$$\det(A - \lambda I) = \det \begin{pmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix} = (3-\lambda)^2 - 1 = \lambda^2 - 6\lambda + 8$$

So $\lambda_1 = 4, \lambda_2 = 2$.

For $\lambda_1 = 4$: $(A - 4I)v = 0$ gives $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

For $\lambda_2 = 2$: $(A - 2I)v = 0$ gives $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

1.2.5 Positive Definite Matrices

Definition 1.8 (Positive Definite Matrix)**.** A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is:

- **Positive definite** ($A \succ 0$) if $x^T Ax > 0$ for all $x \neq 0$
- **Positive semidefinite** ($A \succeq 0$) if $x^T Ax \geq 0$ for all x
- **Negative definite** ($A \prec 0$) if $x^T Ax < 0$ for all $x \neq 0$
- **Negative semidefinite** ($A \preceq 0$) if $x^T Ax \leq 0$ for all x

Theorem 1.2 (Characterizations of Positive Definiteness)**.** For a symmetric matrix A, the following are equivalent:

1. $A \succ 0$ (positive definite)
2. All eigenvalues of A are positive
3. All leading principal minors of A are positive
4. There exists an invertible matrix B such that $A = B^T B$

Example 1.4 (Testing Positive Definiteness)**.** Is $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ positive definite?

Method 1 (Eigenvalues): $\det(A - \lambda I) = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = (\lambda - 1)(\lambda - 3)$ So $\lambda_1 = 3 > 0, \lambda_2 = 1 > 0$. Thus $A \succ 0$.

Method 2 (Principal Minors): $M_1 = 2 > 0$, $M_2 = \det(A) = 4 - 1 = 3 > 0$. Thus $A \succ 0$.

Method 3 (Definition): For any $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \neq 0$: $x^T Ax = 2x_1^2 + 2x_1x_2 + 2x_2^2 = 2(x_1 + \frac{x_2}{2})^2 + \frac{3x_2^2}{2} > 0$. Thus $A \succ 0$.

1.3 Multivariable Calculus Review

Optimization heavily relies on calculus, particularly for finding critical points and analyzing function behavior.

1.3.1 Partial Derivatives and Gradients

Definition 1.9 (Partial Derivative)**.** For a function $f : \mathbb{R}^n \to \mathbb{R}$, the partial derivative with respect to x_i at point x is:

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}$$

Definition 1.10 (Gradient)**.** The gradient of f at x is the vector of all partial derivatives:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

Geometric interpretation: The gradient $\nabla f(x)$ points in the direction of steepest increase of f at x.

Example 1.5 (Computing Gradients)**.** For $f(x_1, x_2) = x_1^2 + 3x_1x_2 + 2x_2^2$:

$$\frac{\partial f}{\partial x_1} = 2x_1 + 3x_2, \quad \frac{\partial f}{\partial x_2} = 3x_1 + 4x_2$$

So $\nabla f(x) = \begin{pmatrix} 2x_1 + 3x_2 \\ 3x_1 + 4x_2 \end{pmatrix}$

At the point $(1, 2)$: $\nabla f(1, 2) = \begin{pmatrix} 8 \\ 11 \end{pmatrix}$

1.3.2 The Hessian Matrix

Definition 1.11 (Hessian Matrix)**.** The Hessian of $f : \mathbb{R}^n \to \mathbb{R}$ at x is the $n \times n$ matrix of second partial derivatives:

$$\nabla^2 f(x) = H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

If f is twice continuously differentiable, then by Schwarz's theorem:

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

so the Hessian is symmetric.

Example 1.6 (Computing the Hessian)**.** For $f(x_1, x_2) = x_1^2 + 3x_1x_2 + 2x_2^2$:

$$\frac{\partial^2 f}{\partial x_1^2} = 2, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 3, \quad \frac{\partial^2 f}{\partial x_2^2} = 4$$

So $\nabla^2 f(x) = \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}$ (constant for quadratic functions)

1.3.3 Taylor's Theorem

Taylor's theorem provides local approximations to functions, which are crucial for optimization algorithms.

Theorem 1.3 (Taylor's Theorem (Multivariate))**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Then for any $x, h \in \mathbb{R}^n$:

First-order approximation:

$$f(x + h) = f(x) + \nabla f(x)^T h + o(\|h\|)$$

Second-order approximation:

$$f(x + h) = f(x) + \nabla f(x)^T h + \frac{1}{2} h^T \nabla^2 f(x) h + o(\|h\|^2)$$

where $o(\|h\|^k)$ denotes terms that go to zero faster than $\|h\|^k$.

Geometric interpretation:

- The first-order approximation is a hyperplane tangent to the function
- The second-order approximation is a quadratic function that matches the curvature

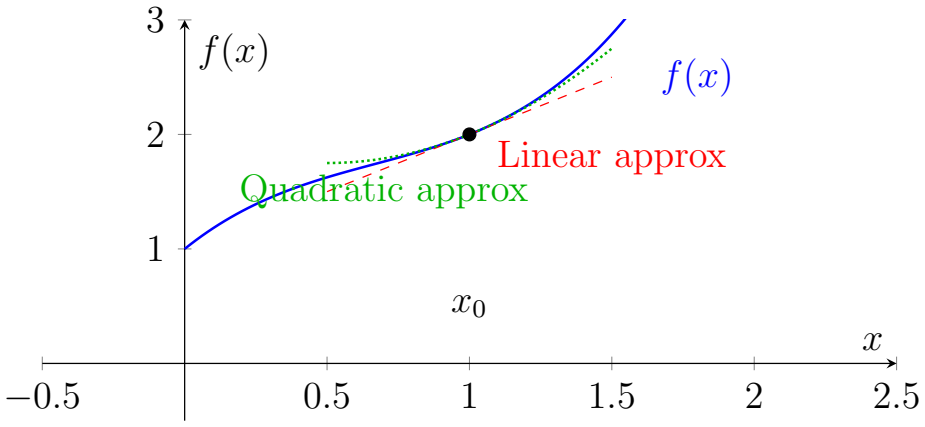


Figure 1.3: Taylor approximations: linear and quadratic

1.3.4 Critical Points and Optimality Conditions

Definition 1.12 (Critical Point)**.** A point x^* is a critical point of f if $\nabla f(x^*) = 0$.

Theorem 1.4 (Second Derivative Test)**.** Let x^* be a critical point of f (i.e., $\nabla f(x^*) = 0$). Then:

- If $\nabla^2 f(x^*) \succ 0$, then x^* is a strict local minimum
- If $\nabla^2 f(x^*) \prec 0$, then x^* is a strict local maximum
- If $\nabla^2 f(x^*)$ has both positive and negative eigenvalues, then x^* is a saddle point
- If $\nabla^2 f(x^*)$ is positive or negative semidefinite but singular, the test is inconclusive (e.g. x^4, $-x^4$ and x^3 all have $f''(0) = 0$)

Example 1.7 (Analyzing Critical Points)**.** Consider $f(x_1, x_2) = x_1^2 - x_2^2$:

$$\nabla f = \begin{pmatrix} 2x_1 \\ -2x_2 \end{pmatrix}, \quad \nabla^2 f = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

Critical point: $\nabla f = 0 \Rightarrow x^* = (0, 0)$
The Hessian has eigenvalues $\lambda_1 = 2 > 0$ and $\lambda_2 = -2 < 0$, so $(0, 0)$ is a saddle point.

1.4 Optimization Theory Basics

Now let's establish the fundamental concepts and terminology of optimization.

1.4.1 General Optimization Problem

Definition 1.13 (General Optimization Problem)**.** A general optimization problem has the form:

$$\begin{aligned} \text{minimize} \quad & f_0(x) && (1.7) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, 2, \ldots, m && (1.8) \\ & h_j(x) = 0, \quad j = 1, 2, \ldots, p && (1.9) \end{aligned}$$

where:

- $x \in \mathbb{R}^n$ is the **optimization variable** (or decision variable)
- $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the **objective function**
- $f_i : \mathbb{R}^n \to \mathbb{R}$ are **inequality constraint functions**
- $h_j : \mathbb{R}^n \to \mathbb{R}$ are **equality constraint functions**

Definition 1.14 (Feasible Set and Optimal Value)**.** The **feasible set** is:

$$\mathcal{X} = \{x : f_i(x) \leq 0, i = 1, \ldots, m; h_j(x) = 0, j = 1, \ldots, p\}$$

The **optimal value** is:

$$p^* = \inf\{f_0(x) : x \in \mathcal{X}\}$$

A point $x^* \in \mathcal{X}$ is **optimal** (or **globally optimal**) if $f_0(x^*) = p^*$.

1.4.2 Local vs Global Optimality

Definition 1.15 (Local Optimality)**.** A feasible point x^* is **locally optimal** if there exists $\delta > 0$ such that

$$f_0(x^*) \leq f_0(x) \text{ for all } x \in \mathcal{X} \cap \{x : \|x - x^*\| \leq \delta\}$$

Key insight: For general (non-convex) optimization problems, local optimality does not imply global optimality. However, for convex problems, any local optimum is also global—this is one of the main reasons convex optimization is so powerful!

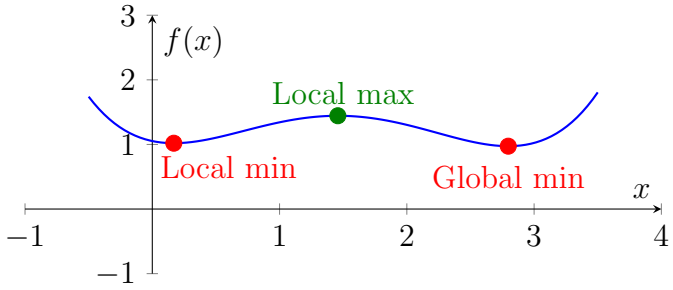


Figure 1.4: Non-convex function with multiple local optima

1.4.3 Special Classes of Optimization Problems

Linear Programming (LP)

Definition 1.16 (Linear Program)**.** A linear program has the form:

$$\begin{aligned} \text{minimize} \quad & c^T x && (1.10)\\ \text{subject to} \quad & Ax = b && (1.11)\\ & x \geq 0 && (1.12) \end{aligned}$$

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

Example 1.8 (Diet Problem)**.** A nutritionist wants to design a minimum-cost diet that meets certain nutritional requirements. Let x_j be the amount of food j to include.

$$\text{minimize} \quad \sum_{j=1}^{n} c_j x_j \quad \text{(total cost)} \tag{1.13}$$

$$\text{subject to} \quad \sum_{j=1}^{n} a_{ij} x_j \geq b_i, \quad i = 1, \ldots, m \quad \text{(nutritional requirements)} \tag{1.14}$$

$$x_j \geq 0, \quad j = 1, \ldots, n \quad \text{(non-negativity)} \tag{1.15}$$

where c_j is the cost per unit of food j, a_{ij} is the amount of nutrient i in food j, and b_i is the minimum requirement for nutrient i.

Quadratic Programming (QP)

Definition 1.17 (Quadratic Program)**.** A quadratic program has the form:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^T Px + q^T x + r && (1.16) \\ \text{subject to} \quad & Ax = b && (1.17) \\ & Gx \leq h && (1.18) \end{aligned}$$

where $P \in \mathbb{R}^{n \times n}$ is symmetric.

If $P \succeq 0$, the problem is convex. Quadratic programming includes least-squares problems as special cases.

1.4.4 Convexity Preview

We'll study convexity in detail in the next chapters, but here's a preview of why it's so important:

Definition 1.18 (Convex Problem (Informal))**.** An optimization problem is convex if:

- The objective function is convex
- The feasible set is convex

Theorem 1.5 (Fundamental Property of Convex Optimization)**.** For a convex optimization problem, any local optimum is a global optimum.

This means that if we can find any local optimum using an algorithm, we've actually found the best possible solution!

1.5 Exercises

Exercise 1.1 (Vector Norms)**.** Compute $\|x\|_1$, $\|x\|_2$, and $\|x\|_\infty$ for $x = (3, -4, 2)^T$.

Exercise 1.2 (Matrix Properties)**.** For $A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$:

(a) Find rank(A)

(b) Find null(A)

(c) Is A invertible?

Exercise 1.3 (Positive Definiteness)**.** Determine whether the following matrices are positive definite, positive semidefinite, negative definite, negative semidefinite, or indefinite:

(a) $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$

(b) $B = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

(c) $C = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$

Exercise 1.4 (Gradient and Hessian)**.** For $f(x, y) = x^3 + xy^2 - 3x$:

(a) Compute $\nabla f(x, y)$

(b) Compute $\nabla^2 f(x, y)$

(c) Find all critical points

(d) Classify each critical point as a local minimum, local maximum, or saddle point

Exercise 1.5 (Optimization Problem Formulation)**.** A company produces two products, A and B. Product A requires 2 hours of labor and 1 kg of raw material per item, with a profit of \$3 per item. Product B requires 1 hour of labor and 2 kg of raw material per item, with a profit of \$2 per item. The company has 100 hours of labor and 80 kg of raw material available.

Formulate this as an optimization problem to maximize profit.

Exercise 1.6 (Taylor Approximation)**.** For $f(x, y) = e^{x+y}$, write the second-order Taylor approximation around the point $(0, 0)$.

Exercise 1.7 (Linear Program Geometry)**.** Consider the linear program:

$$\begin{aligned} \text{minimize} \quad & x + 2y && (1.19) \\ \text{subject to} \quad & x + y \geq 1 && (1.20) \\ & 2x + y \geq 2 && (1.21) \\ & x, y \geq 0 && (1.22) \end{aligned}$$

(a) Sketch the feasible region

(b) Identify the vertices of the feasible region

(c) Solve graphically by finding the optimal vertex

Exercise 1.8 (Quadratic Function Analysis)**.** Consider $f(x) = \frac{1}{2}x^T Ax - b^T x$ where $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

(a) Show that f is convex

(b) Find the minimizer of f

(c) Compute the minimum value

1.6 Chapter Summary

In this chapter, we've established the mathematical foundations needed for studying convex optimization:

Key Concepts Covered:

- **Motivation**: Why optimization matters and why convex optimization is special
- **Linear Algebra**: Vectors, matrices, eigenvalues, positive definiteness
- **Calculus**: Gradients, Hessians, Taylor expansions, critical points
- **Optimization Basics**: Problem formulation, feasibility, optimality

Key Takeaways:

1. Optimization problems are everywhere in engineering, economics, and data science
2. Convex optimization is special because local optima are global optima
3. Linear algebra and calculus provide the tools for analyzing optimization problems
4. Different problem classes (LP, QP, etc.) have different computational properties

Looking Ahead: In the next chapter, we'll dive deep into convex sets—the geometric foundation of convex optimization. We'll learn what makes a set convex, study important examples, and explore operations that preserve convexity.

The mathematical tools we've developed here will be essential for understanding the theory and algorithms that follow. Make sure you're comfortable with vector and matrix operations, as well as computing gradients and Hessians, before moving on.