Optimization Theory Summary

Mathematical Notes

October 19, 2025

Contents

1 Linear Programming

1.1 Standard Form

Definition 1.1. A linear programming problem in standard form:

$$\begin{aligned} \min \quad & c^T x && (1) \\ \text{s.t.} \quad & Ax = b && (2) \\ & x \geq 0 && (3) \end{aligned}$$

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

1.2 Basic Solutions

Definition 1.2. A **basic solution** is obtained by setting $n - m$ variables to zero and solving the resulting system. If all variables are nonnegative, it's a **basic feasible solution**.

1.3 Simplex Method

Definition 1.3. The simplex method:

1. Start with a basic feasible solution
2. Choose entering variable (most negative reduced cost)
3. Choose leaving variable (minimum ratio test)
4. Pivot to new basic feasible solution
5. Repeat until optimal

1.4 Duality

Definition 1.4. The dual of the primal problem $\min\{c^T x : Ax = b, x \geq 0\}$ is:

$$\max \quad b^T y \quad \text{s.t.} \quad A^T y \leq c$$

Theorem 1.1 (Strong Duality)**.** If the primal has an optimal solution, then so does the dual, and their optimal values are equal.

2 Convex Optimization

2.1 Convex Sets

Definition 2.1. A set $C \subseteq \mathbb{R}^n$ is **convex** if for all $x, y \in C$ and $\lambda \in [0, 1]$:

$$\lambda x + (1 - \lambda) y \in C$$

2.2 Convex Functions

Definition 2.2. A function $f : C \to \mathbb{R}$ is **convex** if for all $x, y \in C$ and $\lambda \in [0, 1]$:

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

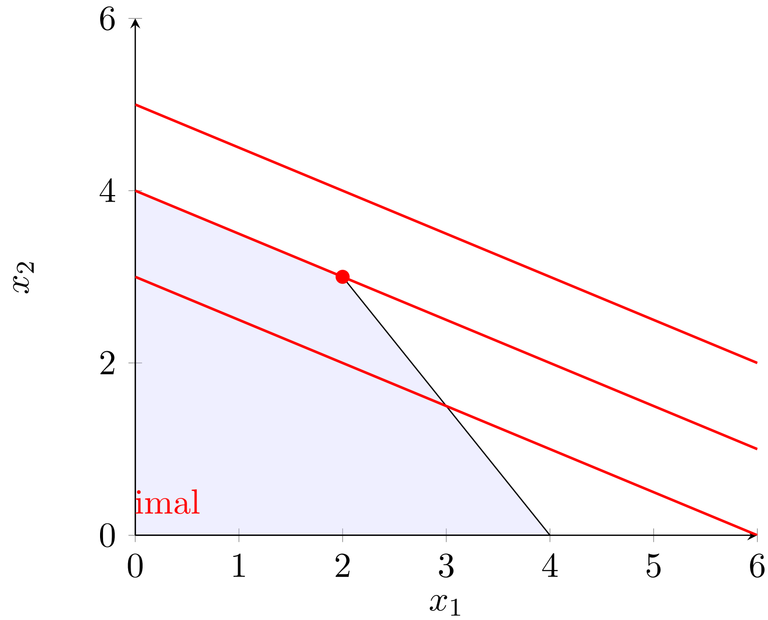

Figure 1: Linear programming feasible region and objective contours

2.3 Convex Optimization Problem

Definition 2.3. A convex optimization problem:

$$\begin{aligned} \min \quad & f(x) && (4) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m && (5) \\ & h_j(x) = 0, \quad j = 1, \ldots, p && (6) \end{aligned}$$

where f and g_i are convex, and h_j are affine.

2.4 Optimality Conditions

Theorem 2.1 (KKT Conditions)**.** For a convex optimization problem, x^* is optimal if and only if there exist Lagrange multipliers $\lambda_i \geq 0$ and ν_j such that:

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{p} \nu_j \nabla h_j(x^*) = 0 \tag{7}$$

$$g_i(x^*) \leq 0, \quad i = 1, \ldots, m \tag{8}$$

$$h_j(x^*) = 0, \quad j = 1, \ldots, p \tag{9}$$

$$\lambda_i g_i(x^*) = 0, \quad i = 1, \ldots, m \tag{10}$$

3 Unconstrained Optimization

3.1 Gradient Descent

Definition 3.1. Gradient descent for minimizing $f(x)$:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

where α_k is the step size.

3.2 Newton's Method

Definition 3.2. Newton's method for optimization:

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

3.3 Convergence Analysis

Theorem 3.1. *If f is strongly convex with Lipschitz gradient, gradient descent with constant step size converges linearly:*

$$f(x_k) - f(x^*) \leq \rho^k (f(x_0) - f(x^*))$$

for some $\rho < 1$.

4 Constrained Optimization

4.1 Lagrange Multipliers

Definition 4.1. For the problem $\min f(x)$ subject to $h(x) = 0$, the Lagrangian is:

$$L(x, \lambda) = f(x) + \lambda^T h(x)$$

4.2 Penalty Methods

Definition 4.2. The penalty method approximates constrained problems by:

$$\min f(x) + \mu \sum_{i=1}^{m} [g_i(x)]_+^2 + \mu \sum_{j=1}^{p} h_j(x)^2$$

where $[z]_+ = \max(0, z)$.

4.3 Barrier Methods

Definition 4.3. The barrier method uses:

$$\min f(x) - \mu \sum_{i=1}^{m} \log(-g_i(x))$$

for inequality constraints.

5 Integer Programming

5.1 Branch and Bound

Definition 5.1. Branch and bound for integer programming:

1. Solve LP relaxation
2. If solution is integer, stop

3. Branch on fractional variable
4. Bound using LP relaxation
5. Prune infeasible or suboptimal nodes

5.2 Cutting Planes

Definition 5.2. Cutting plane methods add valid inequalities to tighten the LP relaxation:

1. Solve LP relaxation
2. If solution is fractional, find cutting plane
3. Add cut and resolve
4. Repeat until integer solution

6 Nonlinear Programming

6.1 Sequential Quadratic Programming

Definition 6.1. SQP solves the subproblem:

$$\begin{aligned} \min \quad & \frac{1}{2}d^T B_k d + \nabla f(x_k)^T d && (11) \\ \text{s.t.} \quad & \nabla g_i(x_k)^T d + g_i(x_k) \leq 0 && (12) \\ & \nabla h_j(x_k)^T d + h_j(x_k) = 0 && (13) \end{aligned}$$

where B_k approximates the Hessian.

6.2 Trust Region Methods

Definition 6.2. Trust region methods solve:

$$\min_{d:\|d\|\leq\Delta_k} m_k(d)$$

where $m_k(d)$ is a model of $f(x_k + d)$.

7 Stochastic Optimization

7.1 Stochastic Gradient Descent

Definition 7.1. SGD for minimizing $E[F(x, \xi)]$:

$$x_{k+1} = x_k - \alpha_k \nabla F(x_k, \xi_k)$$

where ξ_k is a random sample.

7.2 Robust Optimization

Definition 7.2. Robust optimization considers uncertainty in parameters:

$$\min_x \max_{\xi \in \mathcal{U}} f(x, \xi)$$

where $\mathcal{U}$ is the uncertainty set.

8 Applications

8.1 Operations Research

Optimization is used in:

- Supply chain management
- Scheduling problems
- Resource allocation
- Network design

8.2 Machine Learning

Applications include:

- Training neural networks
- Support vector machines
- Regularized regression
- Clustering algorithms

8.3 Finance

Used for:

- Portfolio optimization
- Risk management
- Option pricing
- Algorithmic trading

9 Important Theorems

9.1 Farkas' Lemma

Theorem 9.1. Exactly one of the following holds:

1. There exists $x \geq 0$ such that $Ax = b$
2. There exists y such that $A^T y \leq 0$ and $b^T y > 0$

9.2 Complementary Slackness

Theorem 9.2. For optimal solutions x^* and y^* of primal and dual:

$$x_i^* > 0 \Rightarrow (A^T y^*)_i = c_i$$

$$(A^T y^*)_i < c_i \Rightarrow x_i^* = 0$$

9.3 Minimax Theorem

Theorem 9.3. For a convex-concave function $f(x, y)$:

$$\min_x \max_y f(x, y) = \max_y \min_x f(x, y)$$

10 Complexity

10.1 Computational Complexity

- Linear programming: Polynomial time (interior point methods)
- Convex optimization: Polynomial time
- Integer programming: NP-hard
- General nonlinear programming: NP-hard

10.2 Approximation Algorithms

Definition 10.1. An α-approximation algorithm produces a solution within factor α of optimal in polynomial time.