Numerical optimization : theory and applications

Ammar Mian
Associate professor, LISTIC, Université Savoie Mont Blanc

Outline

Context and Motivation

- **Unconstrained optimization:** We could freely minimize $f(\mathbf{x})$ over $\mathbb{R}^n$
- **Real-world problems:** Often have restrictions on variables
- **Examples:** Resource limits, physical constraints, design specifications

Goal: Characterize solutions when constraints are present, extending our knowledge from unconstrained optimization.

Problem Formulation

General Constrained Optimization Problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \quad \text{subject to} \quad \begin{cases} c_i(\mathbf{x}) = 0, & i \in \mathcal{E} \\ c_i(\mathbf{x}) \geq 0, & i \in \mathcal{I} \end{cases}$$

where:

- f: objective function
- $c_i, i \in \mathcal{E}$: equality constraints
- $c_i, i \in \mathcal{I}$: inequality constraints
- $\Omega = \{\mathbf{x} \mid c_i(\mathbf{x}) = 0, i \in \mathcal{E}; c_i(\mathbf{x}) \geq 0, i \in \mathcal{I}\}$: feasible set

Impact of Constraints on Solutions

Constraints can:

- **Simplify:** Exclude many local minima $\Rightarrow$ easier to find global minimum
- **Complicate:** Create infinitely many solutions

Example 1: $\min \|\mathbf{x}\|_2^2$ subject to $\|\mathbf{x}\|_2^2 \geq 1$

- Unconstrained: unique solution $\mathbf{x} = \mathbf{0}$
- Constrained: any $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$ solves the problem

Solution Definitions

Definition (Local solution)

$\mathbf{x}^\star$ is a **local solution** if $\mathbf{x}^\star \in \Omega$ and there exists neighborhood $\mathcal{N}$ of $\mathbf{x}^\star$ such that

$$f(\mathbf{x}) \geq f(\mathbf{x}^\star) \text{ for all } \mathbf{x} \in \mathcal{N} \cap \Omega$$

Definition (Strict local solution)

$\mathbf{x}^\star$ is a **strict local solution** if $\mathbf{x}^\star \in \Omega$ and there exists neighborhood $\mathcal{N}$ of $\mathbf{x}^\star$ such that

$$f(\mathbf{x}) > f(\mathbf{x}^\star) \text{ for all } \mathbf{x} \in \mathcal{N} \cap \Omega, \mathbf{x} \neq \mathbf{x}^\star$$

Smoothness and Constraint Representation

Key insight: Nonsmooth boundaries can often be described by smooth constraint functions.

Diamond example: $\|\mathbf{x}\|_1 = |x_1| + |x_2| \leq 1$

- **Nonsmooth:** Single constraint with absolute values
- **Smooth equivalent:** Four linear constraints:

$$x_1 + x_2 \leq 1, \quad x_1 - x_2 \leq 1, \quad -x_1 + x_2 \leq 1, \quad -x_1 - x_2 \leq 1$$

[Visualization: Diamond constraint representation]

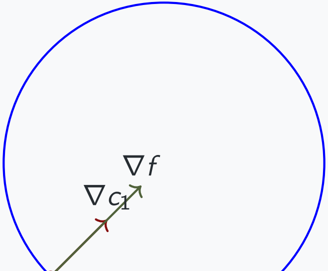
Example 1: Single Equality Constraint

Problem

$$\min x_1 + x_2 \quad \text{subject to} \quad x_1^2 + x_2^2 - 2 = 0$$

- **Feasible set:** Circle of radius $\sqrt{2}$
- **Solution:** $\mathbf{x}^\star = (-1, -1)^T$ (by inspection)
- **Key observation:** At solution, $\nabla f(\mathbf{x}^\star)$ and $\nabla c_1(\mathbf{x}^\star)$ are parallel

$x_1^2 + x_2^2 = 2$



Optimality Condition for Equality Constraints

Necessary Condition

At solution $\mathbf{x}^\star$, there exists $\lambda_1^\star$ such that:

$$\nabla f(\mathbf{x}^\star) = \lambda_1^\star \nabla c_1(\mathbf{x}^\star)$$

Intuition:

- For feasible descent direction $\mathbf{d}$: $\nabla c_1(\mathbf{x})^T \mathbf{d} = 0$ (stay on constraint)
- For improvement: $\nabla f(\mathbf{x})^T \mathbf{d} < 0$
- No such $\mathbf{d}$ exists when gradients are parallel

Lagrangian formulation:

$$\mathcal{L}(\mathbf{x}, \lambda_1) = f(\mathbf{x}) - \lambda_1 c_1(\mathbf{x})$$

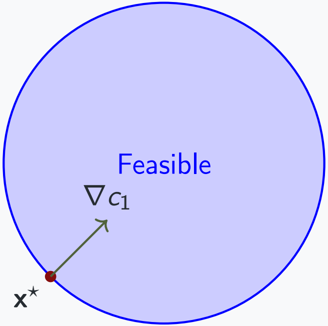
$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^\star, \lambda_1^\star) = \mathbf{0}$$

Example 2: Single Inequality Constraint

Problem

$$\min x_1 + x_2 \quad \text{subject to} \quad 2 - x_1^2 - x_2^2 \geq 0$$

- **Feasible set:** Disk of radius $\sqrt{2}$ (circle + interior)
- **Solution:** $\mathbf{x}^\star = (-1, -1)^T$ (same as before)
- **Key difference:** Sign of Lagrange multiplier matters



Two Cases for Inequality Constraints

Case I: Interior point $(c_1(\mathbf{x}) > 0)$

- Constraint not restrictive
- Necessary condition: $\nabla f(\mathbf{x}) = \mathbf{0}$
- Lagrange multiplier: $\lambda_1 = 0$

Case II: Boundary point $(c_1(\mathbf{x}) = 0)$

- Constraint is active
- Feasible descent direction **d**: $\nabla c_1(\mathbf{x})^T \mathbf{d} \geq 0$
- No such direction when: $\nabla f(\mathbf{x}) = \lambda_1 \nabla c_1(\mathbf{x})$ with $\lambda_1 \geq 0$

Complementarity Condition

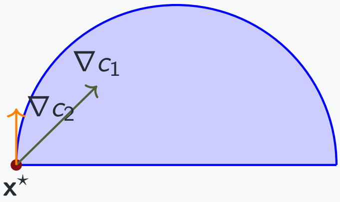
$$\lambda_1 c_1(\mathbf{x}) = 0$$

Example 3: Two Inequality Constraints

Problem

$$\min x_1 + x_2 \quad \text{subject to} \quad 2 - x_1^2 - x_2^2 \geq 0, \quad x_2 \geq 0$$

- **Feasible set:** Half-disk
- **Solution:** $\mathbf{x}^\star = (-\sqrt{2}, 0)^T$
- **Both constraints active** at solution



Multiple Constraints: KKT Conditions Preview

Lagrangian:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \lambda_1 c_1(\mathbf{x}) - \lambda_2 c_2(\mathbf{x})$$

Optimality conditions:

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) = \mathbf{0} \tag{1}$$

$$\lambda_i^\star \geq 0 \quad \text{for all } i \in \mathcal{I} \tag{2}$$

$$\lambda_i^\star c_i(\mathbf{x}^\star) = 0 \quad \text{for all } i \tag{3}$$

For Example 3: $\boldsymbol{\lambda}^\star = (1/(2\sqrt{2}), 1)^T$

Active Set and Constraint Qualification

Definition (Active Set)

$$\mathcal{A}(\mathbf{x}) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(\mathbf{x}) = 0\}$$

Definition (Linear Independence Constraint Qualification (LICQ))

At point $\mathbf{x}^\star$, LICQ holds if the set of active constraint gradients $\{\nabla c_i(\mathbf{x}^\star), i \in \mathcal{A}(\mathbf{x}^\star)\}$ is linearly independent.

Purpose: Ensures constraint gradients are well-behaved and don't vanish inappropriately.

Karush-Kuhn-Tucker (KKT) Conditions

Theorem (First-Order Necessary Conditions)

If $\mathbf{x}^\star$ *is a local solution and LICQ holds at* $\mathbf{x}^\star$, *then there exists* $\boldsymbol{\lambda}^\star$ *such that:*

$$\begin{aligned}
\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star) &= \mathbf{0} && \text{(Stationarity)}\\
c_i(\mathbf{x}^\star) &= 0, \quad i \in \mathcal{E} && \text{(Equality feasibility)}\\
c_i(\mathbf{x}^\star) &\geq 0, \quad i \in \mathcal{I} && \text{(Inequality feasibility)}\\
\lambda_i^\star &\geq 0, \quad i \in \mathcal{I} && \text{(Dual feasibility)}\\
\lambda_i^\star c_i(\mathbf{x}^\star) &= 0, \quad i \in \mathcal{E} \cup \mathcal{I} && \text{(Complementarity)}
\end{aligned}$$

General Lagrangian

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(\mathbf{x})$$

KKT Conditions: Interpretation

Stationarity: $\nabla f(\mathbf{x}^\star) = \sum_{i \in \mathcal{A}(\mathbf{x}^\star)} \lambda_i^\star \nabla c_i(\mathbf{x}^\star)$

- Objective gradient is linear combination of active constraint gradients

Complementarity: $\lambda_i^\star c_i(\mathbf{x}^\star) = 0$

- Either constraint is active ($c_i = 0$) or multiplier is zero ($\lambda_i = 0$)
- Cannot have both $c_i > 0$ and $\lambda_i > 0$

Dual feasibility: $\lambda_i^\star \geq 0$ for inequality constraints

- Sign restriction crucial for inequality constraints
- No sign restriction for equality constraint multipliers

Economic Interpretation of Lagrange Multipliers

Sensitivity analysis: How does optimal value change when constraints are perturbed?

Consider perturbed constraint: $c_i(\mathbf{x}) \geq -\epsilon \|\nabla c_i(\mathbf{x}^\star)\|$

Key Result

$$\frac{df(\mathbf{x}^\star(\epsilon))}{d\epsilon} = -\lambda_i^\star \|\nabla c_i(\mathbf{x}^\star)\|$$

Interpretation:

- $\lambda_i^\star$ measures sensitivity of optimal value to constraint i
- Large $\lambda_i^\star \Rightarrow$ constraint i is "tight" or "binding"
- $\lambda_i^\star = 0 \Rightarrow$ constraint i has little impact on optimal value

Strongly vs. Weakly Active Constraints

Definition (Strongly Active Constraints)

Inequality constraint c_i is **strongly active** if $i \in \mathcal{A}(\mathbf{x}^\star)$ and $\lambda_i^\star > 0$.

Definition (Weakly Active Constraints)

Inequality constraint c_i is **weakly active** if $i \in \mathcal{A}(\mathbf{x}^\star)$ and $\lambda_i^\star = 0$.

Economic interpretation:

- **Strongly active:** Relaxing constraint would improve objective
- **Weakly active:** Small constraint relaxation has no first-order effect

Feasible Sequences Approach

Definition (Feasible Sequence)

Given feasible point $\mathbf{x}^\star$, sequence $\{\mathbf{z}_k\}$ is feasible if:

1. $\mathbf{z}_k \neq \mathbf{x}^\star$ for all k
2. $\lim_{k\to\infty} \mathbf{z}_k = \mathbf{x}^\star$
3. $\mathbf{z}_k$ is feasible for all k sufficiently large

Definition (Limiting Direction)

Vector $\mathbf{d}$ is a limiting direction if:

$$\lim_{k\to\infty} \frac{\mathbf{z}_k - \mathbf{x}^\star}{\|\mathbf{z}_k - \mathbf{x}^\star\|} = \mathbf{d}$$

for some feasible sequence $\{\mathbf{z}_k\}$.

First-Order Necessary Condition via Feasible Sequences

Theorem (Feasible Sequence Necessary Condition)

If $\mathbf{x}^\star$ is a local solution, then for all feasible sequences $\{\mathbf{z}_k\}$ and their limiting directions $\mathbf{d}$:

$$\nabla f(\mathbf{x}^\star)^T \mathbf{d} \geq 0$$

Proof idea:

- If $\nabla f(\mathbf{x}^\star)^T \mathbf{d} < 0$, then by Taylor expansion:

$$f(\mathbf{z}_k) = f(\mathbf{x}^\star) + \|\mathbf{z}_k - \mathbf{x}^\star\| \mathbf{d}^T \nabla f(\mathbf{x}^\star) + o(\|\mathbf{z}_k - \mathbf{x}^\star\|)$$

- For large k: $f(\mathbf{z}_k) < f(\mathbf{x}^\star)$ contradicting optimality

Linearized Feasible Directions

Definition (Linearized Feasible Directions)

$$F_1 = \left\{ \alpha \mathbf{d} \mid \alpha > 0, \begin{array}{ll} \mathbf{d}^T \nabla c_i(\mathbf{x}^\star) = 0, & i \in \mathcal{E} \\ \mathbf{d}^T \nabla c_i(\mathbf{x}^\star) \geq 0, & i \in \mathcal{A}(\mathbf{x}^\star) \cap \mathcal{I} \end{array} \right\}$$

Lemma (Characterization of Limiting Directions)

When LICQ holds:

1. *Every limiting direction satisfies the conditions defining* F_1
2. *Every direction in* F_1 *is a limiting direction of some feasible sequence*

Consequence: Under LICQ, optimality requires $\nabla f(\mathbf{x}^\star)^T \mathbf{d} \geq 0$ for all $\mathbf{d} \in F_1$.

From Geometry to Algebra

Lemma (Lagrange Multiplier Characterization)

There is no direction $\mathbf{d} \in F_1$ *with* $\mathbf{d}^T \nabla f(\mathbf{x}^\star) < 0$ *if and only if there exists* $\boldsymbol{\lambda}$ *such that:*

$$\nabla f(\mathbf{x}^\star) = \sum_{i \in \mathcal{A}(\mathbf{x}^\star)} \lambda_i \nabla c_i(\mathbf{x}^\star)$$

with $\lambda_i \geq 0$ *for* $i \in \mathcal{A}(\mathbf{x}^\star) \cap \mathcal{I}$.

Geometric intuition:

- Objective gradient must lie in cone generated by active constraint gradients
- Farkas' lemma: Either system has solution or alternative system has solution

Need for Second-Order Analysis

First-order conditions are not sufficient!

Consider directions $\mathbf{w}$ where first-order information is inconclusive:

$$\mathbf{w}^T \nabla f(\mathbf{x}^\star) = 0$$

Question: Does moving along $\mathbf{w}$ increase or decrease f?

Definition (Critical Cone)

$$F_2(\boldsymbol{\lambda}^\star) = \left\{ \mathbf{w} \in F_1 \mid \nabla c_i(\mathbf{x}^\star)^T \mathbf{w} = 0, \text{ all } i \in \mathcal{A}(\mathbf{x}^\star) \cap \mathcal{I} \text{ with } \lambda_i^\star > 0 \right\}$$

Key property: For $\mathbf{w} \in F_2(\boldsymbol{\lambda}^\star)$: $\mathbf{w}^T \nabla f(\mathbf{x}^\star) = 0$

Second-Order Necessary Conditions

Theorem (Second-Order Necessary Conditions)

If $\mathbf{x}^\star$ is a local solution, LICQ holds, and $\boldsymbol{\lambda}^\star$ satisfies KKT conditions, then:

$$\mathbf{w}^T \nabla_{\mathbf{xx}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\mathbf{w} \geq 0 \quad \text{for all } \mathbf{w} \in F_2(\boldsymbol{\lambda}^\star)$$

Theorem (Second-Order Sufficient Conditions)

If $\mathbf{x}^\star$ is feasible, KKT conditions hold, and:

$$\mathbf{w}^T \nabla_{\mathbf{xx}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\mathbf{w} > 0 \quad \text{for all } \mathbf{w} \in F_2(\boldsymbol{\lambda}^\star), \mathbf{w} \neq \mathbf{0}$$

then $\mathbf{x}^\star$ is a strict local solution.

Projected Hessian Matrices

When strict complementarity holds: $F_2(\boldsymbol{\lambda}^\star) = \text{Null}(\mathbf{A})$
where $\mathbf{A} = [\nabla c_i(\mathbf{x}^\star)]^T_{i \in \mathcal{A}(\mathbf{x}^\star)}$

Let $\mathbf{Z}$ be matrix whose columns span $\text{Null}(\mathbf{A})$.

Projected Hessian Conditions

Necessary: $\mathbf{Z}^T \nabla_{\mathbf{xx}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\mathbf{Z} \succeq 0$
Sufficient: $\mathbf{Z}^T \nabla_{\mathbf{xx}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\mathbf{Z} \succ 0$

Computational approach: Use QR factorization of $\mathbf{A}^T$ to find $\mathbf{Z}$.

Summary: Characterizing Optimal Solutions

Complete Characterization

Point $\mathbf{x}^\star$ is a local solution if:

1. **First-order:** KKT conditions hold
2. **Second-order:** $\mathbf{w}^T \nabla_{\mathbf{xx}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)\mathbf{w} \geq 0$ for $\mathbf{w} \in F_2(\boldsymbol{\lambda}^\star)$

Practical verification:

- Check LICQ (linear independence of active constraint gradients)
- Solve KKT system for $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$
- Verify projected Hessian conditions

Next: Algorithms to find points satisfying these conditions!