IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Introduction to Optimization

Departement Image et Traitement de l'Information

Mai Quyen Pham

Contents

Optimization problem

A mathematical optimization problem has the form

$$\begin{aligned}
\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \;\; & f(\mathbf{x}) \\
s.t. \;\; & h_i(\mathbf{x}) \leq 0 \quad \forall i = 0,\ldots,k-1 \\
& g_i(\mathbf{x}) = 0 \quad \forall i = 0,\ldots,m-1
\end{aligned} \tag{1}$$

- $\mathbf{x} = (x_1,\ldots,x_n)$: optimization variable of the problem
- $f : \mathbb{R}^n \to \mathbb{R}$: objective function
- $h_i : \mathbb{R}^n \to \mathbb{R}\ \forall i = 0,\ldots,k-1$: inequality constraint functions
- $g_i : \mathbb{R}^n \to \mathbb{R}\ \forall i = 0,\ldots,m-1$: equality constraint functions

$\mathbf{x}^*$ is called optimal of (1) if

- $h_i(\mathbf{x}^*) \leq 0\ \forall i = 0,\ldots,k-1$ and $g_i(\mathbf{x}^*) = 0\ \forall i = 0,\ldots,m-1$
- for any $\mathbf{z} \in \mathbb{R}^n$ such that $h_i(\mathbf{z}) \leq 0\ \forall i = 0,\ldots,k-1$ and $g_i(\mathbf{z}) = 0\ \forall i = 0,\ldots,m-1$, then $f(\mathbf{x}^*) \leq f(\mathbf{z})$

Examples

1. **Least-squares problems**

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \;\; f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$

where $\mathbf{A} \in \mathbb{R}^{k\times n}$ (with $k \geq n$) and $\mathbf{b} \in \mathbb{R}^k$
Solution: $\mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{A}^\top\mathbf{b}$
$\Longrightarrow$ **The least-squares problem can be solved in a time approximately proportional to** n^2k

2. **Linear programs**

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} &\;\; \mathbf{c}^\top\mathbf{x} \\ s.t. &\;\; \mathbf{a}_i^\top\mathbf{x} \leq b_i, i = 1, \ldots, m. \end{aligned}$$

where $\mathbf{c}, \mathbf{a}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for all $i = 1, \ldots, m$
There is no simple analytical formula for the solution of linear program, but there are a variety of very effective methods for solving them. However, we cannot give the exact number of arithmetic operations required to solve a linear program.

Families of optimization problems

- **Linear program**: for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$,

$$f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$$

- **Convex optimization**: for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$,

$$f(\alpha\mathbf{x} + \beta\mathbf{y}) \leq \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$$

- **Nonlinear problem**: when the objective or constraint functions are not linear
- **Nonconvex optimization**: when the objective or constraint functions are not convex
- **Local optimization:** minimizes the objective function among feasible points that are near it, but is not guaranteed to have a lower objective value than all other feasible points.
- **Global optimization:** The true global solution of the optimization problem is found and the compromise is efficiency

Convex set

- **line segment** between $\mathbf{x}$ and $\mathbf{y}$: all points $\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y},\ \alpha \in [0,1]$
- **convex set** contains line segment between any two points in the set

$$\mathbf{x}, \mathbf{y} \in C, \lambda \in [0,1] \Longrightarrow \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in C$$

- **Examples**

convex set　　non-convex set

- **Convex combination** of $\mathbf{x}_1, \ldots, \mathbf{x}_k$: any point $\mathbf{x}$ of the form $\mathbf{x} = \alpha_1\mathbf{x}_1 + \ldots + \alpha_k\mathbf{x}_k$ with $\alpha_1 + \ldots + \alpha_k = 1,\ \alpha_i \geq 0,\ \forall i = 1, \ldots, k.$
- **Convex hull conv** S: set of all convex combinations of points in S
- **Hyper-plane**: set of the form $\{\mathbf{x} : \mathbf{a}^\top\mathbf{x} = b\}$ $(\mathbf{a} \neq 0)$
- **Halfspace**: set of the form $\{\mathbf{x} : \mathbf{a}^\top\mathbf{x} \leq b\}$ $(\mathbf{a} \neq 0)$

Convex function

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\operatorname{dom} f$ is a convex set and if $\forall x, y \in \operatorname{dom} f, \lambda \in [0, 1]$,

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

- If $f \in C^1(\mathbb{R}^n)$ is convex if and only if $\forall x, y \in \mathbb{R}^n$

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x)$$

$$\iff (\nabla f(y) - \nabla f(x))^\top (y - x) \geq 0$$

- if $f \in C^2(\mathbb{R}^n)$ is convex if and only if $\forall x, y \in \mathbb{R}^n$,

$$(y - x)^\top \nabla^2 f(y)(y - x) \geq 0$$

Norms

Vector norm: $\|\mathbf{u}\|_p = (\sum_{i=1}^{n} |u_i|^p)^{1/p}$

- Minkowsky inequality: $\|\mathbf{u}+\mathbf{v}\|_p \leq \|\mathbf{u}\|_p + \|\mathbf{v}\|_p$
- Hölder inequality: $|\mathbf{u}^\top \mathbf{v}| \leq \|\mathbf{u}\|_p \|\mathbf{v}\|_q, \ (1/p + 1/q = 1)$

Matrix norms:

- $\|\mathbf{A}\| = \sup\limits_{\mathbf{x}\in\mathbb{C}^n} \dfrac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \sup\limits_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|$
- Frobenius norm: $\|\mathbf{A}\|_F = \sum_{i,j} |a_{i,j}|^2$
- $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$
- $\|\mathbf{A}\|_p = \sup\limits_{\|\mathbf{x}\|_p=1} \|\mathbf{A}\mathbf{x}\|_p$
- $\|\mathbf{A}\| = \|\mathbf{U}^H\mathbf{A}\mathbf{U}\|$ (if $\mathbf{U}^H\mathbf{U} = \mathbf{I}$)

Optimization: generalities

- Definitions : local, global and strict minimum/maximum.
- **Admissible direction**. Let $f : A \in \mathbb{R}^n \to \mathbb{R}$ and $x \in A$. d an admissible direction at x if

$$\exists \alpha > 0, \forall \lambda \in [0, \alpha], \; x + \lambda d \in A$$

- Descent direction : an admissible direction d is a descent direction at x if

$$\exists \beta > 0, \forall \lambda \in [0, \alpha], \; f(x + \lambda d) \leq f(x)$$

- A set in the form $\mathcal{V}(x) = \{x \in \mathbb{R}^n : \; g_i(x) = 0, \; \forall i = 1, \ldots, m\}$ is a **variety set** of $\mathbb{R}^n$
- **x** is a **regular point** of the variety $\mathcal{V}$ if rank $\nabla \mathbf{g}(x) = m$
- **The tangent space** $\mathcal{T}(x)$ of $\mathcal{V}$ at a regular point x is

$$\mathcal{T}(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n : \nabla \mathbf{g}(\mathbf{x})^\top \mathbf{y} = 0\}$$

Duality (I)

Given a minimization problem

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \;\; & f(\mathbf{x}) \\ s.t. \quad & h_i(\mathbf{x}) \leq 0 \quad \forall i = 0, \ldots, k-1 \\ & g_i(\mathbf{x}) = 0 \quad \forall i = 0, \ldots, m-1 \end{aligned} \tag{2}$$

we defined the **Lagrangian**:

$$L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}^\top \mathbf{h}(\mathbf{x}) + \mathbf{v}^\top \mathbf{g}(\mathbf{x})$$

and **Lagrange dual function:**

$$\ell(\mathbf{u}, \mathbf{v}) = \min_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x}, \mathbf{u}, \mathbf{v})$$

Duality (II)

The subsequent **dual problem** is

$$\max_{\mathbf{u},\mathbf{v}} \ell(\mathbf{u},\mathbf{v}) \tag{3}$$
$$s.t. \quad u_i \geq 0 \quad \forall i = 0,\dots,k-1$$

Important properties:

- Dual problem is always convex, i.e., ℓ is always concave (even if primal problem is not convex)
- The primal and dual optimal values, f^* and ℓ^*, always satisfy weak duality: $f^* \geq \ell^*$
- Slater's condition: for convex primal, if there is an $\mathbf{x}$ such that $h_1(\mathbf{x}) < 0,\dots,h_k(\mathbf{x}) < 0$ and $g_1(\mathbf{x}) = 0,\dots,g_m(\mathbf{x}) = 0$ then **strong duality** holds: $f^* = \ell^*$. (Can be further refined to strict inequalities over nonaffine $h_i, i = 1,\dots,k$)

Karush-Kuhn-Tucker conditions

Given a minimization problem

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \;\; & f(\mathbf{x}) \\ s.t. \quad & h_i(\mathbf{x}) \leq 0 \quad \forall i = 0, \dots, k-1 \\ & g_i(\mathbf{x}) = 0 \quad \forall i = 0, \dots, m-1 \end{aligned} \tag{4}$$

Theorem (NC1-Khun and Tucker conditions)

If x is a regular point of $\mathcal{V}$ and f has a minimum at x, then there exists $\mathbf{u} \in \mathbb{R}^k$ and $\mathbf{v} \in \mathbb{R}^m$ such that

$$\begin{aligned} & \nabla f(x) + \mathbf{u}\nabla\mathbf{h}(x) + \mathbf{v}\nabla\mathbf{g}(x) = 0 && \textit{(stationarity)} \\ & u_i h_i(x) = 0, \;\forall i && \textit{(complementary slackness)} \\ & h_i(x) \leq 0, \; g_j(\mathbf{x}) = 0 \;\forall i, j && \textit{(primal feasiblility)} \\ & u_i \geq 0, \;\forall i && \textit{(dual feasiblility)} \end{aligned}$$

For convex problems, these necessary conditions are sufficient

Second-order sufficient conditions

Theorem (NC2)

If x is a regular point of $\mathcal{V}$ and f has a minimum at x, then there exists $\mathbf{u} \in \mathbb{R}^k$ and $\mathbf{v} \in \mathbb{R}^m$ such that

$$\nabla f(x) + \mathbf{u}\nabla \mathbf{h}(x) + \mathbf{v}\nabla \mathbf{g}(x) = 0 \qquad \textit{(stationarity)}$$

$$u_i h_i(x) = 0,\ \forall i \qquad \textit{(complementary slackness)}$$

$$h_i(x) \leq 0,\ g_j(\mathbf{x}) = 0\ \forall i, j \qquad \textit{(primal feasiblility)}$$

$$u_i \geq 0,\ \forall i \qquad \textit{(dual feasiblility)}$$

$$\nabla^2 L(\mathbf{x}, \mathbf{u}, \mathbf{v}) \textit{is } \textbf{\textit{positive semi-definite}} \textit{ over } \mathcal{T}$$

Theorem (SC2)

If $\mathbf{x}$ is a regular point of $\mathcal{V}$ there exists $\mathbf{u} \in \mathbb{R}^k_+$ and $\mathbf{v} \in \mathbb{R}^m$ such that (NC1) is satisfied and $\nabla^2 L(\mathbf{x}, \mathbf{u}, \mathbf{v})$ is ***positive definite*** *over $\mathcal{T}(x)$, then f has a strict local minimum at $\mathbf{x}$.*

Unconstrained optimization

Problem: $\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\, f(\mathbf{x})$

Optimization via relaxation:

- Iterative minimization of $f(x)$ is successive directions:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \gamma_k \mathbf{d}_k$$

- Cyclic optimization along coordinate axes ($\mathbf{d}_k = \mathbf{e}_k$) yields **relaxation method**

$$f(\mathbf{x}_k) = \min_{\gamma} f(\mathbf{x}_{k-1} + \gamma \mathbf{e}_k)$$

- **Convergence of relaxation method**: The relaxation method converges for C^1 strongly convex functions
- Note : although not involved in the algorithm the derivability assumption is necessary to state convergence in the theorem above.

Gradient descend algorithm

- **Relaxation algorithm:** $\mathbf{x}_{k+1} = \mathbf{x}_k + \gamma_k \mathbf{d}_k$
- **Gradient descend algorithm:** $\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$
- Gradient descend algorithm with optimal step-size:

$$\gamma_k = \text{argmin}\ _{\gamma \in \mathbb{R}_+} f(\mathbf{x}_k - \gamma \nabla f(\mathbf{x}_k))$$

Theorem (Convergence of the Gradient descend algorithm with optimal step-siez)

If f is strongly convex, the optimal step-size gradient algorithm converges. Letting $S = \{\mathbf{x} : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$, $p^* = \min_{\mathbf{x}} f(\mathbf{x})$ and $f \in C^2$, then there exist c and $C \in \mathbb{R}_+^*$, $c\mathbf{I} \leq \|\nabla^2 f(\mathbf{x})\| \leq C\mathbf{I}$ for $\mathbf{x} \in S$ and

$$f(\mathbf{x}_k) - p^* \leq \left(1 - \frac{c}{C}\right)^k (f(\mathbf{x}_0) - p^*)$$

Backtracking algorithm

We want to find a simple way to choose the step-size which minimize the following problem

$$\gamma_k = \text{argmin }_\gamma f(\mathbf{x}_k - \gamma\nabla f(\mathbf{x}_k))$$

Backtracking algorithm

Input: Set $0 < \alpha < 1/2$, $0 < \beta < 1$, and $\gamma = 1$
Iterations:
While $f(\mathbf{x} + \gamma\mathbf{d}) > f(\mathbf{x}) + \alpha\gamma\nabla f(\mathbf{x})^\top \mathbf{d}$
$\gamma = \beta\gamma$
Output: γ

Theorem (Convergence of gradient algorithm with backtracking)

If f is strongly convex, the optimal step-size gradient algorithm converges if $c\mathbf{I} \leq \|\nabla^2 f(\mathbf{x})\| \leq C\mathbf{I}$ for $\mathbf{x} \in S = \{\mathbf{x} : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$,

$$f(\mathbf{x}_k) - p^* \leq (1 - 2c\alpha \min\{1, \beta/C\})^k (f(\mathbf{x}_0 - p^*))$$

Newton algorithm

- $\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k[\nabla^2 f(\mathbf{x}_k)]^{-1}\nabla f(\mathbf{x}_k)$
- γ_k can be chosen vie Backtracking algorithm
- Problem: compute of $\nabla^2 f(\mathbf{x}_k) > 0$
- Solution: occasional update of $\nabla^2 f$, rank perturbation, diagonal approximation, $\ldots$

Theorem (Convergence Newton algorithm with backtracking)

If f is strongly convex, the optimal step-size gradient algorithm converges. For Backtracking with parameters α, β, if $c\mathbf{I} \leq \|\nabla^2 f(\mathbf{x})\| \leq C\mathbf{I}$ for $\mathbf{x} \in S = \{\mathbf{x} : \ f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ and $\nabla^2 f(\mathbf{x})$ is L-lipschitz, a number k of iterations that ensures better than machine precision accuracy $(\sim 10^{-20})$ for $f(\mathbf{x}_k) - p^*$ is given by

$$6 + \frac{C^2L^2}{c^5\alpha^\beta \min\{1, 9(1-2\alpha)^2\}}(f(\mathbf{x}_0) - p^*)$$

Constrained optimization

Solve the constrained optimization problem:

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \;\; & f(\mathbf{x}) \\ s.t. \quad & h_i(\mathbf{x}) \leq 0 \quad \forall i = 0, \ldots, k-1 \\ & g_i(\mathbf{x}) = 0 \quad \forall i = 0, \ldots, m-1 \end{aligned}$$

Let $C_1 = \{\mathbf{x} \in \mathbb{R}^n : h_i(\mathbf{x}) \leq 0, \forall i = 0 \ldots, k-1\}$
and $C_2 = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) = 0, \forall i = 0 \ldots, m-1\}$

The problem becomes:

$$\underset{\mathbf{x}\in C = C_1 \cap C_2}{\text{minimize}} \;\; f(\mathbf{x})$$

Forward-Backward algorithm

$$\underset{\mathbf{x}\in C=C_1\cap C_2}{\text{minimize}} \; f(\mathbf{x})$$

If C is a convex set and the projection of a point $\bar{\mathbf{x}}$ onto this set can be computed i.e.

$$P_C(\bar{\mathbf{x}}) = \underset{\mathbf{x}\in C}{\text{minimize}} \; \frac{1}{2}\|\mathbf{x}-\bar{\mathbf{x}}\|^2$$

Forward-Backward algorithm

Input: Set $\gamma_k > 0$ and $\mathbf{x}_0 \in C$
Iterations:
For $k = 0, 1, \ldots$
$\tilde{\mathbf{x}}_k = \mathbf{x}_k - \gamma_k \nabla f(\mathbf{x}_k)$ Forward step
$\mathbf{x}_{k+1} = P_C(\tilde{\mathbf{x}}_k)$ Backward step
Output: $\hat{\mathbf{x}} = \mathbf{x}_{k+1}$

Interior-point method

- If the computation of the projection onto convex set C is difficult
- Let $\varphi(\mathbf{x}) = -\sum_{i=0}^{k-1} \log[-h_i(\mathbf{x})]$: **logarithmic barrier function**
- If f and h_i are convex and the projection onto the convex set C_2 is computable, we then consider the problem:

$$\underset{\mathbf{x}\in C_2}{\text{minimize}}\, f(\mathbf{x}) + \lambda\varphi(\mathbf{x})$$

The log-barrier method

If $C_2 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\}$

Log-barrier algorithm

Input: Set $\lambda > 0$, $\alpha > 1$ and $\mathbf{x}_0 \in C_2$

Iterations:

For $k = 0, 1, \ldots$

Solve

$$\min_{\mathbf{y} \in C_2} \mathbf{y}^\top (\nabla f(\mathbf{x}_k) + \lambda \nabla \varphi(\mathbf{x}_k)) + \frac{1}{2} \mathbf{y}^\top (\nabla^2 f(\mathbf{x}_k) + \lambda \nabla^2 \varphi(\mathbf{x}_k))\mathbf{y}$$

set $\mathbf{x}_{k+1} = \mathbf{x}_k + \gamma \mathbf{y}$, where $\gamma = \text{argmin}\,_t [f + \lambda\varphi](\mathbf{x}_k + t\mathbf{y})$

$\lambda = \lambda / \alpha$

if $m\lambda < \epsilon$

break

Output: $\hat{\mathbf{x}} = \mathbf{x}_{k+1}$