Introduction to Convex Optimization

Xuezhi Wang

Computer Science Department
Carnegie Mellon University

10701-recitation, Jan 29

Outline

Outline

Convex Sets

- Definition
  For $x, x' \in X$ it follows that $\lambda x + (1-\lambda)x' \in X$ for $\lambda \in [0,1]$
- Examples
  - Empty set $\emptyset$, single point $\{x_0\}$, the whole space $\mathbb{R}^n$
  - Hyperplane: $\{x \mid a^\top x = b\}$, halfspaces $\{x \mid a^\top x \leq b\}$
  - Euclidean balls: $\{x \mid ||x - x_c||_2 \leq r\}$
  - Positive semidefinite matrices: $\mathbf{S}^n_+ = \{A \in \mathbf{S}^n | A \succeq 0\}$ ($\mathbf{S}^n$ is the set of symmetric $n \times n$ matrices)

Convexity Preserving Set Operations

Convex Set C, D

- Translation $\{x + b \mid x \in C\}$
- Scaling $\{\lambda x \mid x \in C\}$
- Affine function $\{Ax + b \mid x \in C\}$
- Intersection $C \cap D$
- Set sum $C + D = \{x + y \mid x \in C, y \in D\}$

Outline

Convex Functions

dom f is convex, $\lambda \in [0, 1]$
$\lambda f(x) + (1 - \lambda) f(y) \geq f(\lambda x + (1 - \lambda) y)$

- **First-order condition**: if f is differentiable,

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

- **Second-order condition**: if f is twice differentiable,

$$\nabla^2 f(x) \succeq 0$$

- **Strictly convex**: $\nabla^2 f(x) \succ 0$
  Strongly convex: $\nabla^2 f(x) \succeq dI$ with $d > 0$

Convex Functions

- **Below-set of a convex function** is convex:
  $f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y)$
  hence $\lambda x + (1 - \lambda) y \in X$ for $x, y \in X$
- **Convex functions don't have local minima**:
  Proof by contradiction:
  linear interpolation breaks local minimum condition
- **Convex Hull**:
  $Conv(X) = \{\bar{x} \mid \bar{x} = \sum \alpha_i x_i \textit{ where } \alpha_i \geq 0 \textit{ and } \sum \alpha_i = 1\}$
  Convex hull of a set is always a convex set

Convex Functions examples

- Exponential. e^{ax} convex on $\mathbb{R}$, any $a \in \mathbb{R}$
- Powers. x^a convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$.
- Powers of absolute value. $|x|^p$ for $p \geq 1$, convex on $\mathbb{R}$.
- Logarithm. $\log x$ concave on $\mathbb{R}_{++}$.
- Norms. Every norm on $\mathbb{R}^n$ is convex.
- $f(x) = \max\{x_1, ..., x_n\}$ convex on $\mathbb{R}^n$
- Log-sum-exp. $f(x) = \log(e^{x_1} + ... + e^{x_n})$ convex on $\mathbb{R}^n$.

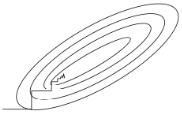
Convexity Preserving Function Operations

Convex function $f(x), g(x)$

- Nonnegative weighted sum: $af(x) + bg(x)$
- Pointwise Maximum: $f(x) = \max\{f_1(x), ..., f_m(x)\}$
- Composition with affine function: $f(Ax + b)$
- Composition with nondecreasing convex g: $g(f(x))$

Outline

Gradient Descent



given a starting point $x \in \mathbf{dom} f$.
repeat
1. $\Delta x := -\nabla f(x)$
2. Choose step size t via exact or backtracking line search.
3. update. $x := x + t\Delta x$.

Until stopping criterion is satisfied.

- Key idea
  - Gradient points into descent direction
  - Locally gradient is good approximation of objective function
- Gradient Descent with line search
  - Get descent direction
  - Unconstrained line search
  - Exponential convergence for strongly convex objective

Convergence Analysis

- Assume ∇f is L-Lipschitz continuous, then gradient descent with fixed step size $t \leq 1/L$ has convergence rate $O(1/k)$
  i.e., to get $f(x^{(k)}) - f(x^*) \leq \epsilon$, need $O(1/\epsilon)$ iterations
- Assume strong convexity holds for f, i.e., $\nabla^2 f(x) \succeq dI$ and ∇f is L-Lipschitz continuous, then gradient descent with fixed step size $t \leq 2/(d+L)$ has convergence rate $O(c^k)$, where $c \in (0, 1)$,
  i.e., to get $f(x^{(k)}) - f(x^*) \leq \epsilon$, need $O(\log(1/\epsilon))$ iterations

Outline

Newton's method

- Convex objective function f
- Nonnegative second derivative

$$\partial_x^2 f(x) \succeq 0$$

- Taylor expansion

$$f(x+\delta) = f(x) + \delta^\top \partial_x f(x) + \frac{1}{2}\delta^\top \partial_x^2 f(x)\delta + O(\delta^3)$$

- Minimize approximation & iterate til converged

$$x \leftarrow x - [\partial_x^2 f(x)]^{-1} \partial_x f(x)$$

Convergence Analysis

- Two Convergence regimes
  - As slow as gradient descent outside the region where Taylor expansion is good

$$||\partial_x f(x^*) - \partial_x f(x) - \langle x^* - x, \partial_x^2 f(x)\rangle|| \leq \gamma ||x^* - x||^2$$

  - Quadratic convergence once the bound holds

$$||x_{n+1} - x^*|| \leq \gamma ||[\partial_x^2 f(x_n)]^{-1}|| \; ||x_n - x^*||^2$$

Outline

Constrained Optimization

Primal problem:

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{subject to } h_i(x) \leq 0, i = 1, \ldots, m$$
$$l_j(x) = 0, j = 1, \ldots, r$$

Lagrangian:

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j l_j(x)$$

where $u \in \mathbb{R}^m$, $v \in \mathbb{R}^r$, and $u \geq 0$.

Lagrange dual function:

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

Constrained Optimization

Dual problem:

$$\max_{u,v} g(u, v)$$
$$\text{subject to } u \geq 0$$

- Dual problem is a convex optimization problem, since g is always concave (even if primal problem is not convex)
- The primal and dual optimal values always satisfy weak duality: $f^* \geq g^*$
- **Slater's condition**: for convex primal, if there is an x such that $h_1(x) < 0, ..., h_m(x) < 0$ and $l_1(x) = 0, ..., l_r(x) = 0$ then strong duality holds: $f^* = g^*$.

Outline

1. Convexity
   - Convex Sets
   - Convex Functions
2. Unconstrained Convex Optimization
   - First-order Methods
   - Newton's Method
3. Constrained Optimization
   - Primal and dual problems
   - KKT conditions

KKT conditions

If x^*, u^*, v^* are primal and dual solutions, with zero duality gap (strong duality holds), then x^*, u^*, v^* satisfy the KKT conditions:

- stationarity: $0 \in \partial f(x) + \sum u_i \partial h_i(x) + \sum v_j \partial l_j(x)$
- complementary slackness: $u_i h_i(x) = 0$ for all i
- primal feasibility: $h_i(x) \leq 0, l_j(x) = 0$ for all i, j
- dual feasibility: $u_i \geq 0$ for all i

Proof:
$$\begin{aligned} f(x^*) &= g(u^*, v^*) \\ &= \min_{x \in \mathbb{R}^n} f(x) + \sum u_i^* h_i(x) + \sum v_j^* l_J(x) \\ &\leq f(x^*) + \sum u_i^* h_i(x^*) + \sum v_j^* l_j(x^*) \\ &\leq f(x^*) \end{aligned}$$

Hence all these inequalities are actually equalities

For Further Reading I

Boyd and Vandenberghe
Convex Optimization.