Optimization Problems and Convexity

Andrew Nobel

March, 2021

General Optimization Problem

Problem: Minimize a function $f : \mathbb{R}^d \to \mathbb{R}$ over a set $A \subseteq \mathbb{R}^d$ of interest.
Often expressed in the form of a *mathematical program*:

$$\min f(x) \text{ subject to } x \in A$$

- Function f called *objective function*
- Set A represents *constraints* on the arguments x of interest
- Points $x \in A$ called *feasible*
- Usually interested in $\min_A f(x)$ and $\operatorname{argmin}_A f(x)$

General Optimization Problem, cont.

Global and local minima

- Feasible $x \in A$ is a *global minimum* of f if $f(x) \leq f(y)$ for all $y \in A$
- Feasible $x \in A$ is a *local minimum* of f if there exists an $r > 0$ such that $f(x) \leq f(y)$ for all $y \in A$ with $||x - y|| \leq r$

Notes: A global minimum is a local minimum. Other issues

- Is there a global min? Is it unique?
- Is there a closed form solution for the global min?
- Are there good iterative or approximate solutions?
- Does f have many local minima?

Review of Convex Sets and Functions

1. A set $C \subseteq \mathbb{R}^d$ is convex if for every pair $x, y \in C$ and every $\alpha \in [0,1]$

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

2. An intersection of convex sets is convex

3. A function $f : C \to \mathbb{R}$ is convex if for every pair $x, y \in C$ and $\alpha \in [0,1]$

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

and is *strictly convex* if $\leq$ replaced by $<$ whenever $x \neq y$ and $\alpha \in (0,1)$

4. The maximum of convex functions is convex

Convexity and Optimization

Fact: If $C \subseteq \mathbb{R}^d$ is convex and $f : C \to \mathbb{R}$ is convex then

1. Any local minimum is a global minimum
2. If f is strictly convex any global minimum is unique

In general: If $C \subseteq \mathbb{R}^d$ and $f : C \to \mathbb{R}$ are convex then there are efficient iterative methods to find the global minimum of f when it exists

Optimization Examples

1. *Largest eigenvalue.* Given $A \in \mathbb{R}^{n \times n}$ symmetric, $\lambda_{\text{max}}(A)$ is the solution of

$$\max\ v^t A v \ \text{ s.t. } \ ||v|| \leq 1$$

2. *Sample variance.* Given $x_1, \ldots, x_n \in \mathbb{R}$

$$\min \sum_{i=1}^{n} (x_i - a)^2 \ \text{ s.t. } \ a \in \mathbb{R}$$

Optimization Examples

3. *PCA.* Given $x_1, \ldots, x_n \in \mathbb{R}^p$ with $\sum_{i=1}^n x_i = 0$

$$\min \sum_{i=1}^{n} \|x_i - \text{proj}_V(x_i)\|^2 \ \text{ s.t. } \ V \text{ a } k\text{-dim subspace of } \mathbb{R}^p$$

4. *K-means clustering.* Given $x_1, \ldots, x_n \in \mathbb{R}^p$

$$\min \sum_{i=1}^{n} \min_{1 \leq j \leq k} ||x_i - c_j||^2 \ \text{ s.t. } \ c_1, \ldots, c_k \in \mathbb{R}^p$$

Optimization Examples

5. *Best linear classification rule.* Given $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^p \times \{0, 1\}$, find hyperplane $H(\beta) = \{x : x^t\beta = \beta_0\}$ that best separates 0s and 1s

$$\min \sum_{i=1}^{n} \mathbb{I}\{(2y_i - 1)(x_i^t\beta - \beta_0) \geq 0\} \text{ s.t. } \beta \in \mathbb{R}^{p+1}$$

6. *Least squares regression.* Given $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^p \times \mathbb{R}$, find linear function of x minimizing sum of squared errors

$$\min \sum_{i=1}^{n} (y_i - x_i^t\beta - \beta_0)^2 \text{ s.t. } \beta \in \mathbb{R}^{p+1}$$

Optimization Examples

8. *Maximum Likelihood Estimation.* Given data $x_1, \ldots, x_n \in \mathcal{X}$ and family $\mathcal{P} = \{f(x : \theta) : \theta \in \Theta\}$ of densities on $\mathcal{X}$, maximum likelihood estimate is

$$\hat{\theta}^{\mathsf{MLE}} = \operatorname*{argmin}_{\theta \in \Theta} \sum_{i=1}^{n} \log f_\theta(x_i)$$