Stochastic Optimization Methods

Radoslav Harman

FMFI UK Bratislava

General Introduction to Optimization I

Two "components" of an optimization problem:

- X ... set of feasible (permissible) solutions
- $f : X \to \mathbb{R}$ ... objective (loss, error; utility, fitness) function, sometimes also "criterion"

A typical **minimization problem** is to find one element x^*, called **global optimum**, of:

$$\operatorname{argmin}_{x \in X} f(x) := \{x^* \in X : f(x^*) \leq f(x) \text{ for all } x \in X\}.$$

$$\text{Notation: } \min f(x) \text{ subject to (s.t.) } x \in X.$$

Variants: • Sometimes, we need to find more than one, possibly *all* elements of $\operatorname{argmin}_{x \in X} f(x)$. • More often, it is enough to find one or more $\tilde{x}$'s from X that are *"close"* to $\operatorname{argmin}_{x \in X} f(x)$.

Analogously we define a **maximization problem**.

General Introduction to Optimization II

Sometimes the objective function is $f : X \to \mathbb{R}^m$ for $m \geq 2$, in which case we get the so-called **multicriterial optimization** problem. This is a separate topic with specific principles.

Exercise: Assume that for each $x \in X$ we have a system of **neighbourhoods** $N_\epsilon(x) \subseteq X$; $\epsilon > 0$. How would you define the intuitive notion of a **"local" optimum**?

Exercise: Suggest optimization problems that may appear in practice, and the corresponding X and f. (Optimization of the movement of a robot, "training" in machine learning, ...)

Important questions:

- **Is there any solution?** ($\operatorname{argmin}_{x \in X} f(x) \neq \emptyset$?)
- **Is there a unique solution?** (Is $\operatorname{argmin}_{x \in X} f(x)$ a singleton?)
- **Can we verify the optimality** of a given $x^? \in X$?

General Introduction to Optimization III

Exercise: Give simple examples of optimization problems for which the solution does not exist, for which the solution exists and is unique and for which the solution does exist but is not unique.

Basic mathematical (analytic) conditions which guarantee the existence of the solution of an optimization problem:

- X is finite
- $X \subset \mathbb{R}^n$ **compact** & f **continuous** on X (in Eucl. metric).

Note: $X \subset \mathbb{R}^n$ is compact if it is closed and bounded.

Basic mathematical (analytic) condition that guarantees the uniqueness of the solution of a minimization problem:

- $X \subset \mathbb{R}^n$ is compact, **convex**, and f is **strictly convex** on X.

Note: $X \subset \mathbb{R}^n$ is convex and $f : X \in \mathbb{R}$ is strictly convex on X if for all $x_1, x_2 \in X$ and all $\alpha \in (0, 1)$ we have $\alpha x_1 + (1 - \alpha)x_2 \in X$ and

$$f(\alpha x_1 + (1 - \alpha)x_2) < \alpha f(x_1) + (1 - \alpha)f(x_2).$$

General Introduction to Optimization IV

Some **classes of optimization problems** (loose definitions):

- X is finite or countable ... discrete optimization (combinatorial optimization, integer programming with $X \subseteq \mathbb{Z}^n$). Examples of applications: Optimum scheduling, Experimental design etc.
- $X \subseteq \mathbb{R}^n$ is not a discrete set ... continuous optimization. Examples of applications: Parameter estimation, "Training" of ANNs, etc.

- $X \subseteq \mathbb{R}$ ... univariate optimization. Example of application: Line step search in multivariate optimization.
- $X \subseteq \mathbb{R}^n$, $n \geq 2$ ... multivariate optimization; almost all optimization problems in data science are multivariate.

- $X = \mathbb{R}^n$ for some $n \geq 1$ ... unconstrained optimization
- $X \neq \mathbb{R}^n$ ... constrained optimization

General Introduction to Optimization V

- $X \subseteq \mathbb{R}^n$, f are convex ... convex optimization Applications: SVM, Some MLE, optimal approximate experimental design...
- Either X or f non-convex ... non-convex optimization; much of optimization in data science is non-convex

---

- $X \subsetneq \mathbb{R}^n$ is polytopic, f is linear ... linear optimization (linear programming). Example of application: L1 regression.
- $X \subseteq \mathbb{R}^n$ is not polytopic or f is non-linear ... non-linear optimization (non-linear programming); most optimization problems in data science are non-linear.

- f is differentiable on $X \subseteq \mathbb{R}^n$ ... differentiable optimization
- f is not differentiable in some point of X (in particular in the optimum) ... non-differentiable optimization

General Introduction to Optimization VI

We will focus on heuristic multivariate non-convex optimization without derivatives, typically including stochastic elements.

- Non-population methods: Local search with modifications to overcome local optima (Simulated annealing, Multistart search, Iterated local search, Tabu search, ...)
- Population methods for discrete optimization: Discrete genetic algorithms, Estimation-of-distribution algorithms, Ant-colony optimization, ...
- Population methods for continuous optimization: Nelder-Mead algorithm, Continuous genetic algorithms, Evolution strategies, Particle-swarm optimization, Differential Evolution,...