Intro to optimization

Data Science 2: AC 209b

Javier Zazo Pavlos Protopapas

February 20, 2019

Abstract

We present the basic concepts of unconstrained and constrained optimization. This will allow you to understand the derivations to obtain the dual problem of the optimal transport formulation.

1 Intro to optimization

We say an optimization problem is unconstrained when we minimize in the whole Euclidean space, i.e., $x \in \mathbb{R}^n$:

$$\min_{x \in \mathbb{R}^n} \quad f(x). \tag{1}$$

We have a constrained optimization problem when the minimization is with respect to $X \subset \mathbb{R}^n$:

$$\min_{x \in X} \quad f(x). \tag{2}$$

A set $X \subseteq \mathbb{R}^n$ is convex if every point between two points belonging to the set, also belongs to the same set. Examples of convex sets include the whole Euclidean space, half-spaces (subspaces divided by hyperplanes), hyperplanes, polytopes (the intersection of multiple halfspaces), etc. See also Figure 1.

A function $f(x)$ is convex in an open set X, if for every two points x_1 and $x_2 \in X$, the points connecting $f(x_1)$ and $f(x_2)$ are greater than or equal to the function f evaluated at those points. If the function $f(x)$ is doubly differentiable, the function is convex if its Hessian is positive semidefinite on every point $x \in X$. An example is given in Figure 2.

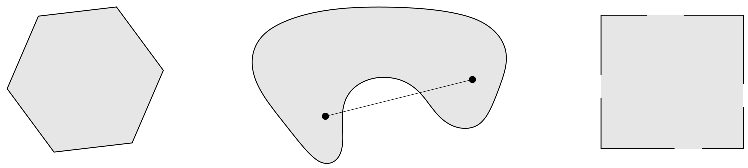

Figure 1: Three sets. The hexagon on the left is convex, the kidney shaped set is non-convex, the squared set excluding part of the boundary is also non-convex.

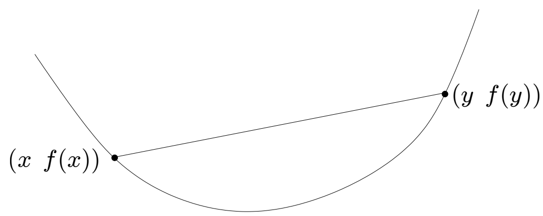

Figure 2: Example of a convex function.

2 Unconstrained optimization

We want to solve problem (1). If the function is differentiable, a necessary condition for optimality on point x^* is that its gradient is null evaluated on that point, i.e.,

$$\nabla_x f(x^*) = 0. \tag{3}$$

If $f(x)$ is additionally a convex function, then the condition is both necessary and sufficient.

An example is to minimize the convex parabola $f_1(x) = ax^2 + bx + c$ with $a > 0$. Its derivate is $\frac{d}{dx} f(x) = 2ax + b$, and its minimum becomes $x^* = \frac{-b}{2a}$. We can generalize to the multivariate case:

$$f_2(x) = x^T Ax + 2b^T x + c, \tag{4}$$

with A being a symmetric positive definite matrix. The gradient is

$$\nabla_x f_2(x) = 2Ax + 2b, \tag{5}$$

and finding its root we obtain $x^* = -A^{-1}b$.

3 Constrained optimization

We want to solve problem (2). We can assume that X is represented in analytical with equality and inequality equations as follows:

$$X = \{\, x \in \mathbb{R}^n \mid g_i(x) \leq 0 \wedge h_j(x) = 0, \quad i \in \{1, \ldots, m\}, j \in \{1, \ldots, p\}\,\}. \tag{6}$$

This allows us to rewrite (2) in standard form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0 \quad i \in \{1, \ldots, m\} \\ & h_j(x) = 0 \quad j \in \{1, \ldots, p\}\,. \end{aligned} \tag{7}$$

We say that problem (7) is convex if $f(x)$ is convex, every $g_i(x)$ is convex, and every $h_j(x)$ are affine functions. Otherwise, the problem is non-convex. The SVM problem that we introduced in the course is convex.

If we have a constrained convex problem, and it satisfies a special constraint qualification, then we can use duality theory to solve it. The motivation to derive the dual is threefold: it allows to check specific conditions for optimality; it introduces other optimization tools to solve the original problem, hopefully more efficient; it may give some theoretical insights about the problem, such as pricing of a certain resource in an economic model.

Regarding the constraint qualification we mentioned, we need to verify if the problem satisfies Slater's condition:

$$\exists \hat{x} \,|\, g_i(\hat{x}) < 0 \quad \forall i \text{ and } h_j(\hat{x}) = 0 \quad \forall j. \tag{8}$$

The previous expression can be relaxed to a simple feasibility requirement as $g_i(\hat{x}) \leq 0$, if g_i is an affine expression.

We call (7) the primal problem, because we optimize in the primal variable x. We will derive now the dual problem. First we form the Lagrangian:

$$L(x, \lambda, \nu) = f(x) + \sum_i \lambda_i g_i(x) + \sum_j \nu_j h_j(x). \tag{9}$$

The dual function is the minimum of the Lagrangian over variable x, and it is a function over λ_i and ν_j:

$$q(\lambda, \nu) = \min_x L(x, \lambda, \nu). \tag{10}$$

And finally, the dual problem consists on the maximization of the dual function over $\lambda_i \geq 0$:

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^m, \nu \mathbb{R}^p} \quad & q(\lambda, \nu) \\ \text{s.t.} \quad & \lambda_i \geq 0 \quad \forall i. \end{aligned} \tag{11}$$

The motivation behind using duality theory to solve problem (7) is that sometimes it is easy to solve the minimum over x in the Lagrangian, and the dual problem has an amenable form. Notice that the minimization over x of the Lagrangian is an unconstrained problem, and therefore it is necessary that

$$\nabla_x L(x^*, \lambda, \nu) = 0 \tag{12}$$

for any candidate solution x^*. This is the first necessary condition of the Karush-Kuhn-Tucker (KKT) conditions. The rest of them refer to feasibility:

$$g_i(x^*) \leq 0 \quad \forall i \tag{13a}$$
$$h_j(x^*) = 0 \quad \forall j \tag{13b}$$
$$\lambda_i^* \geq 0 \quad \forall i \tag{13c}$$
$$\nu_j^* \in \mathbb{R} \quad \forall j, \tag{13d}$$

and complementarity slackness:

$$\sum_i \lambda_i^* g_i(x^*) = 0 \tag{14a}$$
$$\sum_j \nu_j^* h_j(x^*) = 0. \tag{14b}$$

The reason of imposing (14) is to have the following relation:

$$\begin{aligned} \max_{\lambda,\nu} \min_x \quad & L(x, \lambda, \nu) \\ = \quad & f(x^*) + \sum_i \underbrace{\lambda_i^* g_i(x^*)}_{=0} + \sum_j \underbrace{\nu_j^* h_j(x^*)}_{=0} = f(x^*). \end{aligned}$$

We see then that when the KKT conditions are satisfied for points x^*, λ_i^*, ν_j^*, and the problem is convex, then we achieve optimality of the primal problem. The KKT conditions (provided that Slater condition holds) are then necessary and sufficient.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2003.

Acknowledgments

Figures 1 and 2 are borrowed from [1].