Introduction to Machine Learning - CS725
Instructor: Prof. Ganesh Ramakrishnan
Lecture 11 - Constrained Optimization, KKT Conditions, Duality, SVM Dual

Convex Optimization Problem

- Formally, a convex optimization problem is an optimization problem of the form

$$\begin{aligned} \textit{minimize } & f(\mathbf{x}) && (1) \\ \textit{subject to } c & \in C && (2) \end{aligned}$$

  where f is a convex function, C is a convex set, and **x** is the optimization variable.

- A specific form of the above would be

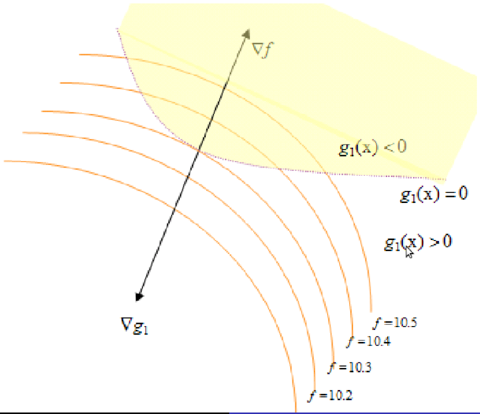
$$\begin{aligned} \textit{minimize } f(\mathbf{x}) & && (3) \\ \textit{subject to } g_i(\mathbf{x}) & \leq 0, \; i = 1, ..., m && (4) \\ h_i(\mathbf{x}) & = 0, i = 1, ..., p && (5) \end{aligned}$$

  where f is a convex function, g_i are convex functions, and h_i are affine (linear) functions, and **x** is the vector of optimization variables.

Constrained convex problems

Q. *How to solve constrained problems of the above-mentioned type?*
A. Canonical example:

$$\textit{Minimize } f(\mathbf{w}) \ s.t. \ g_1(\mathbf{w}) \leq 0 \tag{6}$$



- If $\mathbf{w}^*$ is on the boundary of g_1, *i.e.*, if $g_1(\mathbf{w}^*) = 0$,

$$\nabla f(\mathbf{w}^*) = -\lambda \nabla g_1(\mathbf{w}^*) \textit{ for some } \lambda \geq 0$$

- **Intuition:**

[1] $\nabla_\perp g_1(\mathbf{w}^*)$ is the direction orthogonal to $\nabla g_1(\mathbf{w}^*)$

[2] `Section 4.4, pg-72: cs725/notes/BasicsOfConvexOptimization.pdf`

Constrained Convex Problems

- If $\mathbf{w}^*$ is on the boundary of g_1, *i.e.*, if $g_1(\mathbf{w}^*) = 0$,

$$\nabla f(\mathbf{w}^*) = -\lambda \nabla g_1(\mathbf{w}^*) \text{ for some } \lambda \geq 0$$

- **Intuition:** If the above didn't hold, then we would have $\nabla f(\mathbf{w}^*) = \lambda_1 \nabla g_1(\mathbf{w}^*) + \lambda_2 \nabla_{\perp} g_1(\mathbf{w}^*)$, where, by moving in direction[1] $\pm \nabla_{\perp} g_1(\mathbf{w}^*)$ (or $-\nabla g_1(\mathbf{w}^*)$), we remain on boundary $g_1(\mathbf{w}^*) = 0$, (or within $g_1(\mathbf{w}^*) \leq 0$) while decreasing the value of f, which is not possible at the point of optimality.
- Thus, at the point of optimality[2],

[1] $\nabla_{\perp} g_1(\mathbf{w}^*)$ is the direction orthogonal to $\nabla g_1(\mathbf{w}^*)$

[2] Section 4.4, pg-72: cs725/notes/BasicsOfConvexOptimization.pdf

Constrained Convex Problems

- If $\mathbf{w}^*$ is on the boundary of g_1, *i.e.*, if $g_1(\mathbf{w}^*) = 0$,

$$\nabla f(\mathbf{w}^*) = -\lambda \nabla g_1(\mathbf{w}^*) \textit{ for some } \lambda \geq 0$$

- **Intuition:** If the above didn't hold, then we would have $\nabla f(\mathbf{w}^*) = \lambda_1 \nabla g_1(\mathbf{w}^*) + \lambda_2 \nabla_\perp g_1(\mathbf{w}^*)$, where, by moving in direction[1] $\pm \nabla_\perp g_1(\mathbf{w}^*)$ (or $-\nabla g_1(\mathbf{w}^*)$), we remain on boundary $g_1(\mathbf{w}^*) = 0$, (or within $g_1(\mathbf{w}^*) \leq 0$) while decreasing the value of f, which is not possible at the point of optimality.
- Thus, at the point of optimality[2], for some $\lambda \geq 0$,

$$\textit{Either } g_1(\mathbf{w}^*) < 0 \quad \& \quad \nabla f(\mathbf{w}^*) = 0 \qquad (7)$$

$$\textit{Or } g_1(\mathbf{w}^*) = 0 \quad \& \quad \nabla f(\mathbf{w}^*) = -\lambda \nabla g_1(\mathbf{w}^*) \qquad (8)$$

---

[1] $\nabla_\perp g_1(\mathbf{w}^*)$ is the direction orthogonal to $\nabla g_1(\mathbf{w}^*)$

[2] `Section 4.4, pg-72: cs725/notes/BasicsOfConvexOptimization.pdf`

Explaining the Figure

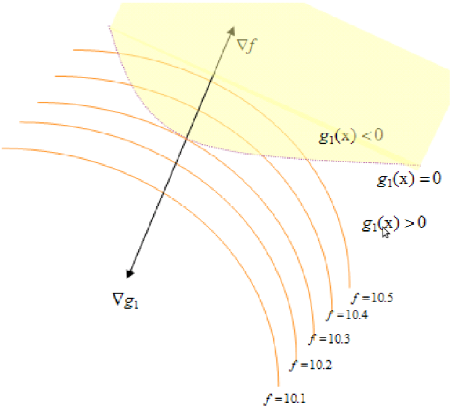


Figure 2: Two conditions under which a minimum can occur: a) When the minimum is on the constraint function boundary, in which case the gradients are in opposite directions; b) When point of minimum is inside the constraint space (shown in yellow shade), in which case $\nabla f(\mathbf{w}^*) = \mathbf{0}$.

- The first condition occurs when minima lies on the boundary of function g. In this case, gradient vectors corresponding to the functions f and g, at $\mathbf{w}^*$, point in opposite directions barring multiplication by a real constant.
- Second condition represents the case that point of minimum lies inside the constraint space. This space is shown shaded in Figure 1. Clearly, for this case, $\nabla f(\mathbf{w}) = \mathbf{0}$.
- An Alternative Representation: $\nabla L(\mathbf{w}, \lambda) = 0$ for some $\lambda \geq 0$ where

$$L(\mathbf{w}, \lambda) = f(\mathbf{w}) + \lambda \mathbf{g}(\mathbf{w}); \lambda \in \mathbb{R}$$

  is called the lagrange function which has objective function augmented by weighted sum of constraint functions

Duality and KKT conditions

For a convex objective and constraint function, the minima, $\mathbf{w}^*$, can satisfy one of the following two conditions:

1. $g(\mathbf{w}^*) = \mathbf{0}$ and $\nabla f(\mathbf{w}^*) = -\lambda \nabla \mathbf{g}(\mathbf{w}^*)$
2. $g(\mathbf{w}^*) < \mathbf{0}$ and $\nabla f(\mathbf{w}^*) = \mathbf{0}$

Duality and KKT conditions

- Here, we wish to penalize higher magnitude coefficients, hence, we wish $g(\mathbf{w})$ to be negative while minimizing the lagrangian. In order to maintain such direction, we must have $\lambda \geq 0$. Also, for solution $\mathbf{w}$ to be feasible, $\nabla g(\mathbf{w}) \leq \mathbf{0}$.
- Due to complementary slackness condition, we further have $\lambda g(\mathbf{w}) = \mathbf{0}$, which roughly suggests that the lagrange multiplier is zero unless constraint is active at the minimum point. As $\mathbf{w}$ minimizes the lagrangian $L(\mathbf{w}, \lambda)$, gradient must vanish at this point and hence we have $f(\mathbf{w}) + \lambda \nabla \mathbf{g}(\mathbf{w}) = \mathbf{0}$