Lecture #18: November 15, 2018

- Recall our geometric condition for boundary optimization: ~~use~~

(★) $\operatorname{argmin} f(x)$, s.t. $h(x) \geq 0$, where $\{x \mid h(x) \geq 0\} = R \subset \mathbb{R}^d$ is the feasible set. Suppose $h = (h_1, \dots, h_m)$ is continuous and that R is a bounded set.

- Then we know there is a solution x^* to (★), and moreover

$$\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*), \quad \text{some } \lambda_i \geq 0.$$

- Idea: If we knew $\{\lambda_i\}_{i=1}^m$, then we could solve

$$\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*)$$

to get x^*; this may be hard if $\nabla f, \nabla h_i$ are complicated, but in principle it can be done.

- We will see that in the context of an SVM, the constraints are simple, so that the resulting system is linear, hence more easily solvable.

- Consider $f_\lambda(x) = f(x) - \sum_{i=1}^{m} \lambda_i h_i(x)$, for $\{\lambda_i\}_{i=1}^m$ fixed. Then

$$\nabla f_\lambda(x^*) = 0,$$ so that x^* is a critical point of f_λ.

• Moreover, since $h(x) \geq 0$ by hypothesis on the feasible set,

$$f_\lambda(x) \leq f(x), \quad \forall x.$$

$$\Rightarrow \quad \min_{x \in \mathbb{R}^d} f_\lambda(x) \leq \min_{x \in R} f_\lambda(x) \leq \min_{x \in R} f(x).$$

• Thinking in terms of $\lambda = (\lambda_1, \dots, \lambda_m)$ varying, set $g(\lambda) = \min_{x \in \mathbb{R}^d} f_\lambda(x)$

<u>Exercise</u>: $\max_{\substack{\lambda \in \mathbb{R}^m \\ \lambda_i \geq 0}} g(\lambda) \leq \min_{x \in R} f(x).$

<u>Dual Formulation</u>

$$\operatorname{argmax}_{\lambda_i \geq 0} g(\lambda)$$

$$\shortparallel$$

$$\operatorname{argmax}_{\lambda_i \geq 0} \min_{x \in \mathbb{R}^d} f_\lambda(x)$$

<u>Primal (Original) Formulation</u>

$$\operatorname{argmin}_{h(x) \geq 0} f(x) = \operatorname{argmin}_{x \in R} f(x)$$

• Solving the dual problem gives a lower bound on the primal problem, though this may not be tight... the gap is $\min_{x \in R} f(x) - \max_{\substack{\lambda \in \mathbb{R}^m \\ \lambda_i \geq 0}} g(\lambda)$

is the duality gap ... would be great if $=0$.

Benefit of Dual Problem: $g(\lambda) = \min_{x \in \mathbb{R}^n} f(x) - \sum_{i=1}^{m} \lambda_i h_i(x)$

is always concave! (i.e. $-g(\lambda)$ is convex). This is independent of the convexity of the primal problem.

. It would be nice to handle the minimization needed to compute $g(\lambda)$, since then we can easily (hopefully) do the outer maximization $\max_{\lambda \geq 0} g(\lambda)$, which allows us to estimate the solution to the original problem.

. In fact, if f is a convex quadratic and each h_i is linear, we can directly minimize $f(x) - \sum_{i=1}^{m} \lambda_i h_i(x)$!

Let $f(x) = x^T A x + b^T x + c$, $A \in \mathbb{R}^{n \times n}$ positive definite, $b \in \mathbb{R}^{n \times 1}$, $c \in \mathbb{R}$

WLOG, we suppose $b = c = 0$ and consider $h_i(x) = z_i^T x + r_i$, $z_i \in \mathbb{R}^n$, $r_i \in \mathbb{R}$.

This yields: $\arg\min x^T A x$ s.t. $z_i^T x + r_i \geq 0$, $i = 1, \dots, n$.

- Since A is symmetric positive definite, $A = U^T \Lambda U$ for U an orthogonal matrix, Λ diagonal with positive entries.

- Setting $\sqrt{A} = U^T \sqrt{\Lambda} U$, we see (exercise)

$$\arg\min \; x^T A x \quad \text{s.t.} \quad z_i x + r_i \geq 0, \; i = 1, \dots, m$$

$$\Updownarrow \quad y = A^{\frac{1}{2}} x$$

$$\arg\min \; \|y\|^2 \quad \text{s.t.} \quad (A^{-\frac{1}{2}} z_i)^T y + r_i \geq 0, \; i = 1, \dots, m$$

Since A is invertible, this is equivalent to

$$(\star\star) \quad \arg\min \; \|x\|^2 \quad \text{s.t.} \quad z_i^T x + r_i \geq 0, \; i = 1, \dots, m$$

- Let's think about $(\star\star)$ in the context of KKT: $\nabla g_i = z_i$, so need $x^* = \frac{1}{2} \sum_{i \in J} \lambda_i^* z_i$, where J is the set of indices s.t. $z_i^T x^* + r_i = 0$

$$\Updownarrow$$

$$x^* = \frac{1}{2} \sum_{i=1}^{n} \lambda_i^* z_i, \quad \lambda_i^* \geq 0 \quad \text{and} \quad \lambda_i^* (z_i^T x^* + r_i) = 0$$