Nonlinear Optimization (18799 B, PP)
IST-CMU PhD course
Spring 2013
Instructor: `jxavier@isr.ist.utl.pt`
TAs: `augustos@andrew.cmu.edu`, `ric.s.cabral@gmail.com`

Important: The homework is due April, 28.

Homework 5

Instructions: read sections 5.1 and 5.2 of [1].

Problem A. *(Distance between polyhedrons)* Let $P_1 = \{x : A_1 x \leq b_1\}$ and $P_2 = \{x : A_2 x \leq b_2\}$ be two given non-empty polyhedrons, with $A_i \in \mathbb{R}^{m \times n}$ and $b_i \in \mathbb{R}^m$ for $i = 1, 2$. The distance between the polyhedrons, denoted $d(P_1, P_2)$, is the optimal value of the optimization problem

$$\begin{array}{ll} \text{minimize} & \|x_1 - x_2\| \\ \text{subject to} & A_1 x_1 \leq b_1 \\ & A_2 x_2 \leq b_2, \end{array} \tag{1}$$

with (x_1, x_2) being the optimization variable. In this problem, we derive an alternative representation for $d(P_1, P_2)$ via duality.

(a) Let $s \in \mathbb{R}^n$ be given. Show that

$$\inf\left\{\|x\| - s^\top x \,:\, x \in \mathbb{R}^n\right\} = \left\{\begin{array}{ll} 0 & , \text{ if } \|s\| \leq 1 \\ -\infty & , \text{ if } \|s\| > 1. \end{array}\right.$$

Hint: the Cauchy-Schwartz inequality is useful.

(b) Obtaining the dual problem of (1) might be challenging (give it a try!). Here is an interesting trick: consider the equivalent problem

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & y = x_1 - x_2 \\ & A_1 x_1 \leq b_1 \\ & A_2 x_2 \leq b_2, \end{array} \tag{2}$$

with optimization variable (y, x_1, x_2). The point is that the variables x_1 and x_2 are no longer coupled in the objective; this will facilitate taking the dual. Obtain the dual of (2), simplify it, and show that it is equivalent to

$$\begin{array}{ll} \text{maximize} & -\mu_1^\top b_1 - \mu_2^\top b_2 \\ \text{subject to} & A_1^\top \mu_1 + A_2^\top \mu_2 = 0 \\ & \left\|A_1^\top \mu_1\right\| \leq 1, \left\|A_2^\top \mu_2\right\| \leq 1 \\ & \mu_1, \mu_2 \geq 0. \end{array} \tag{3}$$

Side remark: We could have dropped one of the inequalities $\left\|A_i^\top \mu_i\right\| \leq 1$ in (3), since the first constraint makes it redundant. But this would make the problem less symmetric and, therefore, not so beautiful.

(c) Invoke one of the strong duality theorems from the lecture slides (mention which one) to prove that the optimal value of (2) (hence, the optimal value of (1)) is equal to the optimal value of (3).

Side remark: The representation in (3) has many applications. For example, we see that $d(P_1, P_2)$ is a convex function of (b_1, b_2); indeed, the equality

$$d(P_1, P_2) = \sup\left\{-\mu_1^\top b_1 - \mu_2^\top b_2 \,:\, A_1^\top \mu_1 + A_2^\top \mu_2 = 0, \left\|A_i^\top \mu_i\right\| \leq 1, \mu_i \geq 0,\, i = 1, 2\right\}$$

expresses $d(P_1, P_2)$ has the pointwise supremum of convex functions (in fact, affine) of (b_1, b_2). Convexity of $d(P_1, P_2)$ with respect to (b_1, b_2) can also be established from (1), but (3) makes it really obvious. As another application, suppose you computed $d(P_1, P_2)$ by solving (3). Let $\mu_1^\star$ and $\mu_2^\star$ be the solution. Now, we change slightly the b_i's by adding a small perturbation, i.e., consider $\widehat{b}_i = b_i + \delta_i$, for small δ_i; we want to compute $\widehat{d}(P_1, P_2)$, the distance between the polyhedrons corresponding to the novel $\widehat{b}_i$'s. We may re-solve (3) to get the answer. However, there is a computationally cheap, approximate answer: pretend that $\mu_1^\star$ and $\mu_2^\star$ are also the solution of the perturbed problem. Plugging them into (3), yields the approximation $\widehat{d}(P_1, P_2) \simeq d(P_1, P_2) - (\mu_1^\star)^\top \delta_1 - (\mu_2^\star)^\top \delta_2$. (This heuristic argument is worth to be revisited once you learn about subgradients which, unfortunately, falls outside of the scope of this course.)

References

[1] S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge University Press, 2004.