Homework #1 Solutions

1. Consider the *Procrustes problem:*

$$\min_X \|AX - B\|_F \ : \ X^T X = I_n,$$

where $A \in \mathbf{R}^{m \times n}$, $B \in \mathbf{R}^{m \times n}$ are given, and the matrix $X \in \mathbf{R}^{n \times n}$ is the variable. (In the above, I_n stands for the identity matrix.)

(a) Is the problem, as given, convex?

(b) Show that the problem can be reduced to

$$\max_X \mathbf{Tr}\, B^T AX \ : \ X^T X = I_n.$$

(c) Show how to solve the problem using an SVD of the matrix $B^T A$.

Solution:

(a) The constraint involves a quadratic equality so the problem is not convex. In fact, in the scalar case $n = 1$, the constraint reduces to $X \in \{-1, 1\}$, which is not convex.

(b) We have
$$\|AX - B\|_F^2 = \|AX\|_F^2 - 2\,\mathbf{Tr}\, B^T AX + \|B\|_F^2.$$
Since $X^T X = I_n$ and X is square, it is invertible, and $X^{-1} = X^T$ (that is, X is a rotation in n-dimensional space). So $XX^T = I_n$ and $\|AX\|_F^2 = \mathbf{Tr}(A^T AXX^T) = \|A\|_F^2$ is constant. This proves the result.

(c) Let $M = B^T A \in \mathbf{R}^{n \times n}$ and assume it has the SVD $M = USV^T$ with S diagonal positive semi-definite. Using the new of variable $Y = V^T XU$, and noting that $X^T X = I_n$ if and only if $Y^T Y = I_n$, we obtain the equivalent formulation:

$$\max_Y \mathbf{Tr}\, SY \ : \ Y^T Y = I_n,$$

with X obtained from Y via $X = VYU^T$.

We observe that for any feasible Y, $|Y_{ij}| \leq 1$ for every i, j, hence, since S is diagonal, $\mathbf{Tr}\, SY \leq \mathbf{Tr}\, S$. This upper bound is attained for $Y = I$. Hence the optimal value of the problem is $\mathbf{Tr}\, S$ (the sum of the singular values of $M = B^T A$, also called the *nuclear norm* of M), and an optimizer is $X^* = VU^T$.

2. *Supporting hyperplane.* Let $f : \mathbf{R}^n \to \mathbf{R}$ be a convex, differentiable function, with domain $\mathbf{R}^n$. Assume that the set $\mathcal{F} := \{x \in \mathbf{R}^n \ : \ f(x) \leq 0\}$ is not empty, and let x_0 be a point on its boundary (so that $f(x_0) = 0$). Show that the set

$$\mathcal{H} := \left\{x \in \mathbf{R}^n \ : \ \nabla f(x_0)^T (x - x_0) = 0\right\}$$

defines a supporting hyperplane of $\mathcal{F}$ at x_0. (*Hint:* express the first-order convexity condition for f at x_0.)

Solution: The first-order convexity condition at x_0 is

$$\forall\, x \in \mathbf{R}^n \ : \ f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0).$$

With $f(x_0) = 0$, and $f(x) \leq 0$ for every $x \in \mathcal{F}$, we obtain

$$\forall\, x \in \mathcal{F} \ : \ 0 \geq \nabla f(x_0)^T (x - x_0).$$

This means that $\mathcal{H} := \{x \ : \ \nabla f(x_0)^T (x - x_0) = 0\}$ is a supporting hyperplane of $\mathcal{F}$ at x_0, with the gradient pointing orthogonally to it outwards.

3. *Maximum of a convex function over a polyhedron.* Show that the maximum of a convex function f over the polyhedron $\mathcal{P} = \mathbf{conv}\{v_1, \ldots, v_k\}$ is achieved at one of its vertices, i.e.,

$$\max_{x \in \mathcal{P}} f(x) = \max_{i=1,\ldots,k} f(v_i)$$

Hint: Assume the statement is false, and use Jensen's inequality.

Solution: The polytope $\mathcal{P}$ is the set of points of the form $V\lambda$ with $V = [v_1, \ldots, v_k]$ and $\lambda \in \mathbf{R}^k_+$, with $\lambda_1 + \ldots + \lambda_k = 1$. By convexity of f,

$$f(V\lambda) \leq \sum_{i=1}^{k} \lambda_i f(v_i) \leq \max_{i=1,\ldots,k} f(v_i),$$

so we see that the maximum of $f(x)$ over $x \in \mathcal{P}$ is bounded above by the maximum over the vertices. Since the vertices belong to $\mathcal{P}$, we have proved the result.

4. *A general vector composition rule:* suppose

$$f(x) = h(g_1(x), g_2(x), ..., g_k(x))$$

where $h : \mathbf{R}^k \to \mathbf{R}$ is convex, and $g_i : \mathbf{R}^n \to \mathbf{R}$. Suppose that for each i, one of the following holds:

(a) h is nondecreasing in the i-th argument, and g_i is convex;

(b) h is nonincreasing in the i-th argument, and g_i is concave;

(c) g_i is affine.

Show that f is convex. (This composition rule subsumes all the ones given in the book, and is the one used in software systems such as CVX.)

Solution: Let us use the epigraph criterion to prove convexity. Let us first assume that for every i, h is nondecreasing in the i-th argument, and g_i is convex.

The condition for $(x,t) \in \mathbf{R}^{n+1}$ $f(x) \le t$ is equivalent to the existence of $u \in \mathbf{R}^k$ such that

$$h(u) \le t, \quad g_i(x) \le u_i, \quad i = 1, \ldots, k. \tag{1}$$

Indeed, if the above holds for some $u \in \mathbf{R}^k$, then since h is nondecreasing with each argument, we have

$$f(x) = h(g_1(x), \ldots, g_k(x)) \le h(u_1, \ldots, u_k) \le t.$$

The converse is obvious, with the choice $u_i = g_i(x)$, $i = 1, \ldots, k$. The above proves our result, since it implies that the epigraph of f is the projection of a convex set (in the space of (x,t,u)-variables) on the space of (x,t)-variables.

To prove the general statement, we follow the same idea, replacing the statement (1) by

$$\begin{aligned} h(u) \le t, \quad & g_i(x) \le u_i, \quad i \in \mathcal{I}_{\text{convex}}, \\ & g_i(x) \ge u_i, \quad i \in \mathcal{I}_{\text{concave}}, \\ & g_i(x) = u_i, \quad i \in \mathcal{I}_{\text{affine}}, \end{aligned}$$

where the index sets $\mathcal{I}_{\text{convex}}$, $\mathcal{I}_{\text{concave}}$, and $\mathcal{I}_{\text{affine}}$ correspond to the indices where one of the three conditions (a), (b) or (c) as stated in the problem hold.

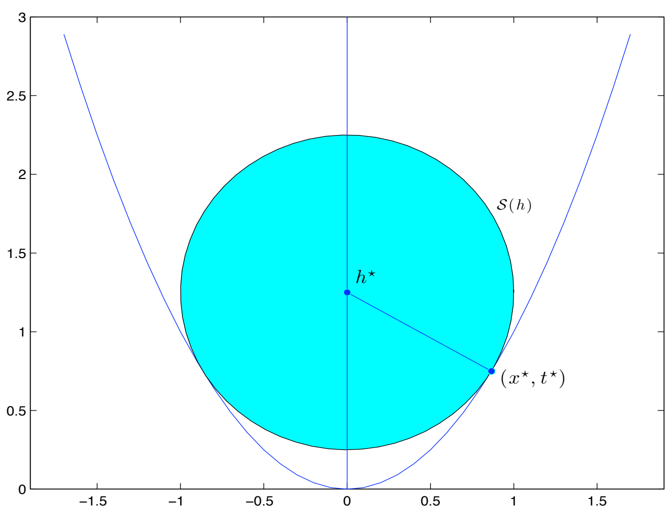

Figure 1: Geometric view of problem 5, with $q = 2$.

5. *Lowest altitude of sphere inside a paraboloid.* Consider the paraboloid $\mathcal{Q} = \{(x,t) : (1/2)qx^2 \le t\} \subseteq \mathbf{R}^2$, where $q > 0$ is given. We inscribe a sphere $\mathcal{S}(h)$, of radius 1 and centered on a point $(0,h)$, inside the paraboloid. We want to find the smallest value of h such that the paraboloid contains the sphere. (Fig. 1 shows the situation for $q = 2$.)

(a) Show that the paraboloid contains the sphere $\mathcal{S}(h)$ if and only if $h \geq 1$ and $f^* \geq 1$, with

$$f^*(h) := \min_{h-1 \leq t \leq h+1} f(t), \quad f(t) := \frac{2t}{q} + (t-h)^2.$$

(b) Solve the minimization problem above: show that

$$f^*(h) = \begin{cases} \frac{2h}{q} + 1 - \frac{2}{q} & \text{if } q < 1, \\ \frac{2h}{q} - \frac{1}{q^2} & \text{if } q \geq 1. \end{cases}$$

Find the corresponding optimal value for t, $t^*(h)$.

(c) Find the minimum altitude, h^*.

(d) Provide the optimal points of contact between the sphere and the paraboloid.

(e) You should have found that the value $q = 1$ plays a special role. Interpret this geometrically.

Solution:

(a) Let us fix h, with $h \geq 1$ (otherwise the sphere cuts the x-axis). The condition $\mathcal{S}(h) \subseteq \mathcal{Q}$ is equivalent to

$$\forall\,(x,t) \;:\; x^2 \leq \frac{2t}{q} \text{ whenever } x^2 + (t-h)^2 \leq 1.$$

Let us examine the above condition for fixed t, which means we are looking at horizontal slices of the sphere, and making sure those slices are inside the paraboloid. The condition is void if $|t-h| > 1$ or $t < 0$. Let us assume $|t-h| \leq 1$ and $t \geq 0$; since $h \geq 1$, this reduces to $t \in [h-1, h+1]$. The condition above writes

$$\forall\, t \in [h-1, h+1] \;:\; 1 - (t-h)^2 \leq \frac{2t}{q},$$

or:

$$1 \leq f^*(h) := \min_{t \in [h-1,h+1]} f(t), \quad f(t) := \frac{2t}{q} + (t-h)^2. \tag{2}$$

(b) The minimum of the quadratic function f (over all values of t) is obtained for $t_{\min} := h - 1/q$. There are three cases to consider, depending on the position of $t_{\min}$ with respect to the interval $[h-1, h+1]$.

If $t_{\min} \in [h-1, h+1]$, then the minimizer of f over $[h-1, h+1]$ is $t^*(h) = t_{\min}$, and the corresponding minimum value is $f^*(h) = 2h/q - 1/q^2$. The case $t_{\min} > h+1$ implies $q < -1$, so it is ruled out. Finally, in case $t_{\min} < h-1$, that is, $q < 1$, the minimizer is $t^*(h) = h-1$, and $f^*(h) = 2(h-1)/q + 1$. That is,

$$f^*(h) = \begin{cases} \frac{2h}{q} + 1 - \frac{2}{q} & \text{if } q < 1, \\ \frac{2h}{q} - \frac{1}{q^2} & \text{if } q \geq 1. \end{cases}, \quad t^*(h) = \begin{cases} h-1 & \text{if } q < 1, \\ h - \frac{1}{q} & \text{if } q \geq 1. \end{cases}$$

(c) The condition $f^*(h) \geq 1$ is equivalent to $h \geq h^*$, where

$$h^* := \begin{cases} 1 & \text{if } q < 1, \\ \frac{1}{2}(q + \frac{1}{q}) & \text{if } q \geq 1. \end{cases}$$

We conclude that the minimum altitude is h^* above.

(d) The optimal points of contact (x^*, t^*) are obtained by solving the equation

$$(1/2)q(x^*)^2 = t^*,$$

with t^* the minimizer of the function f as defined above, setting h to its optimal value h^*. That is:

$$t^* = \begin{cases} 0 & \text{if } q < 1, \\ \frac{1}{2}(q - \frac{1}{q}) & \text{if } q \geq 1. \end{cases}$$

Thus, the two points of contact are $(\pm x^*, t^*)$, with

$$x^* = \begin{cases} 0 & \text{if } q < 1, \\ \sqrt{1 - \frac{1}{q^2}} & \text{if } q \geq 1. \end{cases} = \sqrt{\max\left(1 - \frac{1}{q^2}, 0\right)}.$$

(e) When $q < 1$, the optimal sphere actually touches the origin, and the two points of contact coincide there. The paraboloid is then "flat enough" to allow the sphere to rest at the origin without touching it elsewhere. Otherwise, the points of contact are distinct, and above the x-axis.