EE5138R: Problem Set 7
Assigned: 20/03/15
Due: 27/03/15

1. BV Problem 9.1

   Solution:

   (a) If P is not positive semidefinite, there exists a v such that $v^T Pv < 0$. With $x = tv$, we have

   $$f(x) = t^2(v^T Pv/2) + t(q^T v) + r$$

   which diverges to $-\infty$ as $t \to \infty$.

   (b) This means that q is not in the range of P. Express $q = \bar{q} + v$ where $\bar{q}$ is the Euclidean projection of q onto the range of P and take $v = q - \bar{q}$. This vector is nonzero and orthogonal to $\mathcal{R}(P)$, i.e., $v^T Pv = 0$. It follows that for $x = tv$, we have

   $$f(x) = tq^T v + r = t(\bar{q} + v)^T v + r = t(v^T v) + r$$

   which is unbounded below.

2. BV Problem 9.6

   Solution: For $k = 0$, we get the starting point $x^{(0)} = (\gamma, 1)$.

   The gradient at $x^{(k)}$ is $(x_1^{(k)}, x_2^{(k)})$ so we get

   $$x^{(k)} - t\nabla f(x^{(k)}) = \begin{bmatrix} (1-t)x_1^{(k)} \\ (1-\gamma t)x_2^{(k)} \end{bmatrix} = \left(\frac{\gamma - 1}{\gamma + 1}\right)^k \begin{bmatrix} (1-t)\gamma \\ (1-\gamma t)(-1)^k \end{bmatrix}$$

   and

   $$f(x^{(k)} - t\nabla f(x^{(k)})) = (\gamma^2(1-t)^2 + \gamma(1-\gamma t)^2) \left(\frac{\gamma - 1}{\gamma + 1}\right)^{2k}$$

   This is minimized with

   $$t = \frac{2}{1+\gamma}$$

   so

   $$\begin{aligned} x^{(k+1)} &= x^{(k)} - t\nabla f(x^{(k)}) \\ &= \begin{bmatrix} (1-t)x_1^{(k)} \\ (1-\gamma t)\gamma x_2^{(k)} \end{bmatrix} \\ &= \left(\frac{\gamma - 1}{\gamma + 1}\right) \begin{bmatrix} x_1^{(k)} \\ -x_2^{(k)} \end{bmatrix} \\ &= \left(\frac{\gamma - 1}{\gamma + 1}\right)^{k+1} \begin{bmatrix} \gamma \\ (-1)^k \end{bmatrix} \end{aligned}$$

3. (Optional) BV Problem 9.7
   Solutions provided in the next problem set.

4. Consider the problem

$$\min\ f(x) = 10x_1 + 3x_2, \qquad \text{s.t.} \qquad 5x_1 + x_2 \geq 4, x_1, x_2 = 0 \text{ or } 1.$$

   (a) Sketch the set of constraint-cost pairs

$$\{(4 - 5x_1 - x_2, 10x_1 + 3x_2) : x_1, x_2 = 0, 1\}$$

   Solution: This is obvious.

   (b) Sketch the dual function.
   Solution: The Lagrangian is

$$L(x, \mu) = 10x_1 + 3x_2 + \mu(4 - 5x_1 - x_2)$$

   and the dual function is

$$g(\mu) = \inf_{x_1, x_2 \in \{0,1\}} \{4\mu + (10 - 5\mu)x_1 + (3 - \mu)x_2\} = \begin{cases} 4\mu & \mu \in [0, 2] \\ 10 - \mu & \mu \in [2, 3] \\ 13 - 2\mu & \mu \in [3, \infty) \end{cases}$$

   (c) Solve the problem and its dual.
   Solution: By inspection, we see that $x^* = (1, 0)$ and $p^* = 10$. From the dual, we see that $d^* = 8$. Thus, there is a duality gap of $p^* - d^* = 2$.

5. **Solution:** A straightforward calculation yields the dual function

$$g(\lambda) = \min_{x \in \mathbf{R}^n} \left\{ \|z - x\|_2^2 + \lambda^T A x \right\} = -\frac{\|A^T \lambda\|_2^2}{4} + \lambda^T A z$$

   Thus the dual problem is equivalent to

$$\min_{\lambda \in \mathbf{R}^m} \left\{ \frac{\|A^T \lambda\|_2^2}{4} - \lambda^T A z + \|z\|_2^2 \right\}$$

   or

$$\min_{\lambda \in \mathbf{R}^m} \left\| z - \frac{A^T \lambda}{2} \right\|_2^2$$

   This is the problem of projecting z on the subspace spanned by the rows of A.

6. **Solutions**:

   (a) Lagrange optimality yields

$$\nabla f(x^*) + \nu^* \nabla h(x^*) = 0$$

   which is

$$2x^* + (\nu^*, \ldots, \nu^*)^T = 0$$

   Hence

$$x^* = -\frac{1}{2}(\nu^*, \ldots, \nu^*)^T$$

   But the sum of the vector x^* must be one by primal optimality so

$$x^* = (1/n, \ldots, 1/n)^T.$$

   Problem is convex so x^* is globally optimal.

(b) Lagrange optimality yields

$$(1, 1, \ldots, 1)^T + 2\nu^* x^* = 0$$

So

$$x^* = \left(-\frac{1}{2\nu^*}, \ldots, -\frac{1}{2\nu^*}\right)^T$$

Furthermore the norm of x^* must equal 1 so

$$x^* = \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)^T \quad \text{or} \quad x^* = -\left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)^T$$

Furthermore,

$$\nabla^2 f(x^*) + \nu^* \nabla^2 h(x^*) = 2\lambda^* I = -\sqrt{n} I \quad \text{or} \quad \sqrt{n} I$$

So the only local minimum is $x^* = -(1, 1, \ldots, 1)^T / \sqrt{n}$ so it is the global minimum.

7. **Solution**: We have

$$f(x) = \|x\|^{2+\beta} = (x_1^2 + \ldots + x_n^2)^{1+\beta/2}$$

so

$$\nabla f(x) = (1 + \beta/2)(x_1^2 + \ldots + x_n^2)^{\beta/2}(x_1, \ldots, x_n)^T \cdot 2 = (2 + \beta)\|x\|^\beta x.$$

We first check whether the Lipschitz condition is satisfied; i.e., whether for all $x, y \in \mathbf{R}^n$, there is some constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

or

$$(2 + \beta)\|\|x\|^\beta x - \|y\|^\beta y\| \leq L\|x - y\|$$

By letting $y = -x$, this yields $(2 + \beta)\|x\|^\beta \leq L$, so clearly the Lipschitz condition is not satisfied.

The behavior of the gradient descent method with constant stepsize s is described by the equation

$$x^{k+1} = x^k - s\nabla f(x^k) = x^k(1 - s(2 + \beta)\|x^k\|^\beta)$$

It is easy to show by induction that if $\|x^1\| < \|x^0\|$, then $\|x^{k+1}\| < \|x^k\|$ for all k, and that if $\|x^1\| \geq \|x^0\|$, then $\|x^{k+1}\| \geq \|x^k\|$ for all k. Thus in order for the method to converge, we must have

$$\|x^1\| = \|x^0(1 - s(2 + \beta)\|x^0\|^\beta)\| < \|x^0\|$$

or equivalently

$$|1 - s(2 + \beta)\|x^0\|^\beta| < 1$$

or equivalently

$$s(2 + \beta)\|x^0\|^\beta < 2$$

For the values of s, β, and x^0 satisfying the above inequality, the sequence $\{\|x^k\|\}$ is monotonically decreasing. We will show that for the same values, we have $x^k \to 0$. Indeed, let c be the limit of $\{\|x^k\|\}$. If $c = 0$, we have $x^k \to 0$ and we are done. If $c > 0$ then

$$\lim_{k \to \infty} \frac{\|x^{k+1}\|}{\|x^k\|} = 1$$

and from the fact that $c \leq \|x^0\|$ we have

$$|1 - s(2 + \beta)c^\beta| < 1$$

Combining this iteration with the iterations, we have

$$\lim_{k \to \infty} \frac{\|x^{k+1}\|}{\|x^k\|} = |1 - s(2 + \beta)c^\beta| < 1$$

a contradiction. Hence $c = 0$.