Problem Set 2 Solutions

February 20, 2005

1.1.6

(a) The cost function is convex so the necessary and sufficient condition for optimality of x^* is

$$\sum_{i=1}^{m} w_i \frac{x^* - y_i}{\|x^* - y_i\|} = 0,$$

which is the same as the condition for the equilibrium of forces in the Varignon frame mechanical model.

(b) The solution is not always unique. Consider the case where there are two weights and $w_1 = w_2$. Then any position x where x is between y_1 and y_2 minimizes the sum of weighted distances.

(c) Let H be the height of the board, measured from the reference level, and let l_i be the length of the string from the knot to the ith weight. Then, when the position of the knot is x, the height of the ith weight is $H - (l_i - \|x - y_i\|)$, and the potential energy of the system is

$$E(x) = \sum_{i=1}^{m} w_i(H - (l_i - \|x - y_i\|)) = \sum_{i=1}^{m} w_i(H - l_i) + \sum_{i=1}^{m} w_i\|x - y_i\|.$$

Therefore, minimizing $E(x)$ over x is equivalent to minimizing $\sum_{i=1}^{m} w_i\|x - y_i\|$, which the problem of part (a).

1.2.2

We have

$$f(x) = \|x\|^{2+\beta} = (x_1^2 + \cdots + x_n^2)^{1+\frac{\beta}{2}},$$

so

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

To invoke Prop. 1.2.3, we need to check whether the Lipschitz condition is satisfied; i.e., whether for all $x, y \in \Re^n$, there is some constant $L > 0$ such that

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

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 steepest descent method with constant stepsize s is described by the equation

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

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

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

or, equivalently,

$$|1 - s(2+\beta)\|x^0\|^\beta| < 1, \tag{2}$$

or equivalently

$$s(2+\beta)\|x^0\|^\beta < 2. \tag{3}$$

For the values of s, β, and x^0 satisfying Eq. (3), 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 Eqs. (2) and (3), and the fact $c \leq \|x^0\|$, we have

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

Combining the above relation and Eq. (1), we obtain

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

a contradiction. Hence, we must have $c = 0$.

1.2.3

Consider the Lipschitz condition when y is taken to be $-x$:

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

or

$$\|2\nabla f(x)\| \leq L\|2x\|$$

or

$$3\|x\|^{\frac{1}{2}} \leq 2L\|x\|.$$

This is not true, however, for $\|x\|^{1/2} < \frac{3}{2L}$. Thus the Lipschitz condition does not hold for all x and y. In fact, it does not hold on any set containing the optimal point $x^* = 0$ in its interior.

Now consider the behavior of the algorithm if started at a point $x^0 \neq 0$, for any value of the constant stepsize α. We have at every iteration

$$x^{k+1} = x^k \left(1 - \frac{3\alpha}{2\|x^k\|^{1/2}}\right).$$

If, for some k, we have $||x^k|| = 9\alpha^2/4$, then $x^{k+1} = 0$, and the method will have converged to the minimizing point $x^* = 0$ in a finite number of iterations. Now if this does not occur, note that

$$\|x^{k+1}\| \begin{cases} > \|x^k\| & \text{if } 0 < \|x^k\| < 9\alpha^2/16, \\ < \|x^k\| & \text{if } \|x^k\| > 9\alpha^2/16, \\ = \|x^k\| & \text{if } \|x^k\| = 9\alpha^2/16. \end{cases}$$

Thus, the values of $||x^k||$ will oscillate around $9\alpha^2/16$ unless $||x^k|| = 0$ or $||x^k|| = 9\alpha^2/16$ for some k. In the former case, $\{x^k\}$ never converges to the minimizing value $x^* = 0$. If $||x^{k_0}|| = 9\alpha^2/16$ for some k_0, then $x^{k_0+1} = -x^{k_0}$ and x^k will oscillate between those two points for all $k \geq k_0$. Finally, if $||x^k|| = 0$ for some k then $\{x^k\}$ converges in a finite number of iterations.

More formally, let's suppose that $\{x^k\}$ converges to $x^* = 0$ in an infinite number of iterations. Then for all $\epsilon > 0$, there exists $\bar{k}$ such that $\|x^k\| < \epsilon, \forall\, k > \bar{k}$. Now suppose $\epsilon = \frac{9\alpha^2}{16}$ and $\bar{k}$ is such that $\|x^k\| < \epsilon, \forall\, k > \bar{k}$. (Note that $\|x^k\| > 0$ since we assume that $\{x^k\}$ does not converge finitely to zero.) Then from above, we must have $\|x^k\| \geq \|x^{\bar{k}}\|, \forall\, k > \bar{k}$. But then, for $\epsilon = \|x^{\bar{k}}\|$, there cannot exist a $\hat{k}$ such that $\|x^k\| < \|x^{\hat{k}}\|, \forall\, k > \hat{k}$, yielding a contradiction. Thus $\{x^k\}$ cannot converge to $x^* = 0$ in an infinite number of iterations.

1.2.6

(a) We have

$$\|\nabla f(x) - \nabla f(y)\|^2 = \|Q(x-y)\|^2 = (x-y)'Q^2(x-y) \leq \lambda_{max}(Q^2)\|x-y\|^2.$$

Since Q is positive definite and symmetric, $\lambda_{max}(Q^2)$, the maximum eigenvalue of Q^2, equals the square of the maximum eigenvalue of Q, proving the desired relation.

(b) The iteration $x^{k+1} = x^k - sDf(x^k)$ is written as

$$x^{k+1} - x^* = (I - sDQ)(x^k - x^*).$$

Define $y^k = D^{1/2}x^k, \ \ y^* = D^{1/2}x^*$, and substitute in the above equation. We have

$$y^{k+1} - y^* = (I - sD^{1/2}QD^{1/2})(y^k - y^*).$$

This iteration converges if and only if all the eigenvalues of $(I - sD^{1/2}QD^{1/2})$ are in the interval $(-1, 1)$. (Note that these eigenvalues are real because $D^{1/2}QD^{1/2}$ is symmetric. Note also that DQ need not be symmetric, which is the reason we introduced the y-coordinate system.)

Thus the iteration of y^k (and equivalently the iteration of x^k) converges if and only if the eigenvalues of $sD^{1/2}QD^{1/2}$ are in the interval $(0, 2)$. Equivalently s must lie between 0 and $2/\lambda_{max}(D^{1/2}QD^{1/2})$.

1.2.15

Since

$$f(x) = 2x^2\text{sign}(x) + x,$$

the steepest descent iteration has the following form

$$x^{k+1} = x^k\Big(1 - \frac{(2|x^k| + 1)}{k+1}\Big). \tag{1}$$

(a) Let = 1. If $x^k \geq k+1$, then from Eq. (1) we have

$$x^{k+1} = x^k(1 - \frac{2x^k+1}{k+1}) = \frac{x^k(k-2x^k)}{k+1} \leq -(k+2).$$

If $x^k \leq -(k+1)$, then from Eq. (1) we similarly obtain $x^{k+1} \geq k+2$. Hence, for all k,

$$|x^k| \geq k+1 \qquad \Rightarrow \qquad |x^{k+1}| \geq k+2.$$

Since $|x^0| \geq 1$, recursively we obtain $|x^k| \geq k+1$ for all k.

(b) Define $y^k = |x^k|$ and let

$$\gamma(2y^0+1) < 2.$$

Then from Eq. (1) we have

$$y^{k+1} = y^k|1 - \frac{(2y^k+1)}{k+1}|. \tag{2}$$

Based on this relation and induction, we show that for all k

$$\gamma(2y^k+1) < 2. \tag{3}$$

Note that the equation (3) holds for $k=0$ by the choice of y^0. Now, assume that Eq. (3) holds for some $k>0$ and let us prove that it holds for $k+1$. By using induction hypothesis, we obtain

$$1 - \frac{\gamma(2y^k+1)}{k+1} > 1 - \frac{2}{k+1} \geq 0, \tag{4}$$

which in view of Eq. (2) implies that

$$y^{k+1} \leq y^k(1 - \frac{\gamma(2y^k+1)}{k+1}) < y^k,$$

so that

$$\gamma(2y^{k+1}+1) < \gamma(2y^k+1) < 2,$$

where we again use induction hypothesis. Hence Eq. (3) holds for all k.

Next we prove that $y^k \to 0$. Let c be the limit of the monotonically decreasing sequence $\{y^k\}$. If $c=0$, we are done, so assume that $c>0$. Let $\bar{k}$ be sufficiently large, so that

$$0 < 1 - \frac{\gamma(2y^k+1)}{k+1}, \qquad \forall\, k \geq \bar{k}.$$

Since $y^k > c$, we have from Eq. (2), and all $k \geq \bar{k}$,

$$y^{k+1} = y^k - \frac{\gamma y^k(2y^k+1)}{k+1} < y^k - \frac{\gamma c(2c+1)}{k+1}.$$

Adding over all $k \geq \bar{k}$, we obtain

$$y^{m+1} < y^{\bar{k}} - \sum_{k=\bar{k}}^{m} \frac{\gamma c(2c+1)}{k+1}, \qquad \forall\, m \geq \bar{k}.$$

Since $\sum_{k=\bar{k}}^{\infty} \frac{1}{k+1} = \infty$, we obtain a contradiction. Hence $c=0$, $y^k \to 0$, and $x^k \to 0$.

(c) Proposition 1.2.4 is not applicable because f does not satisfy the Lipschitz condition.