Homework 4

陈远洋

Problem 1

Consider the function $f:\mathbb{R}^n \longmapsto \mathbb{R}$ given by

$$f(x) = \|x\|^{3/2},$$

and the method of steepest descent with a constant stepsize. Show that for this function, the Lipschitz condition $\|f(x) - f(y)\| \leq L\|x-y\|$ all x and y is not satisfied for any L. Furthermore, for any value of constant stepsize, the method either converges in a finite number of iterations to the minimizing point $x^* = 0$ or else it does not converge to x^*.

证明. 题目实际包含两个命题的证明：

- （1）不存在 L，使得对任意的 $x, y \in \mathbb{R}^n$,$\|f(x) - f(y)\| \leq L\|x-y\|$。
- （2）对于任意固定步长 α，要么在有限步内点列收敛至 $\vec{0}$, 要么点列不会收敛至 $\vec{0}$。

我们首先来看（1）。容易知道 $\nabla f(x) = \dfrac{\mathrm{d}f(x)}{\mathrm{d}\|x\|}\nabla\|x\|(x) = \dfrac{1.5}{\sqrt{\|x\|}}x$。令 $\Delta x = y - x$, 则有 $f(y) = f(x) + \nabla f(x)^T\Delta x + o(\|\Delta x\|)$。可知：

$$\begin{aligned}\|f(y) - f(x)\| &= \|1.5\sqrt{\|x\|}\frac{x^T\Delta x}{\|x\|} + o(\|\Delta x\|)\| \\ &\geq 1.5\sqrt{\|x\|}\|\Delta x\|\frac{< x, \Delta x >}{\|x\|\|\Delta x\|} - o(\|\Delta x\|)\end{aligned}$$

故有：$\dfrac{\|f(y) - f(x)\|}{\|y-x\|} \geq 1.5\sqrt{\|x\|}\cos\theta - o(1)$，其中 θ 是 x 和 Δx 之间的夹角。即 $\forall x\,, \forall \epsilon > 0, \exists t > 0$ 当 $\|\Delta x\| \leq t$ 时 $\dfrac{\|f(y) - f(x)\|}{\|y-x\|} \geq 1.5\sqrt{\|x\|}\cos\theta - \epsilon$，进一步，$\forall M > 0$ 取 $\|x\| > \dfrac{4}{9}M^2\,, \Delta x = tx$ 此时 $\dfrac{\|f(y) - f(x)\|}{\|y-x\|} \geq M - \epsilon$. 若存在题设的 L，取 $M > L + \epsilon$ 即可导出矛盾！故（1）成立！

再来看（2）：由（1）的推导过程 $x_{k+1} = x_k - \dfrac{1.5\alpha}{\sqrt{\|x_k\|}} x_k$。所以序列 $\{x^k\}$ 共线, 令 $t_k = \|x_k\|$, 有 $t_{k+1} = |t_k - c\sqrt{t_k}| = h(t_k)$ 其中 $c = 1.5\alpha$。可以看到当 $\alpha \geq 0$ 时，$\{t_k\}$ 是单调递增的永不收敛至 0. 故我们只需考虑 $\alpha < 0$ 的情况。考虑迭代的第 i 步，若此时未停止迭代（$h(t_i) \neq 0$，否则已经在有限步内收敛至 0），作以下分类讨论：

- $t_i > c^2$。由于在 $t_i > c^2$ 时有：$t_{i+1} = t_i - c\sqrt{t_i} < t_i - c^2$。取 $j = \lceil \dfrac{t_i - c^2}{c^2} \rceil$，容易知道在 j 步之内，必有 $k < i + j$, 使得 $t_k \in [0, c^2]$。当 $t_k = 0 \, or \, t_k = c^2$ 时，$h(t_k) = 0$，此时已经收敛至 0。（即有限步内收敛）。否则转步第二种情况。

- $t_i \in (0, c^2)$。容易得到在 $(0, c^2)$ 内，$h(t) \leq h(\dfrac{c^2}{4}) = \dfrac{c^2}{4}$. 故对以后的每一步，$h(t_m) \leq \dfrac{c^2}{4}$. 则由迭代格式 $\forall m > i+1$, $\dfrac{\sqrt{t_{m+1}} - \frac{c}{2}}{\sqrt{t_m} - \frac{c}{2}} = \dfrac{-\sqrt{t_m} + \frac{c}{2}}{\sqrt{t_{m+1}} + \frac{c}{2}} < 1$，可知此后 $\{t^k\}$ 开始递增，且以 $\dfrac{c^2}{4}$ 为上界。故有 $\dfrac{\sqrt{t_{m+1}} - \frac{c}{2}}{\sqrt{t_m} - \frac{c}{2}} \leq \dfrac{\frac{c}{2} - \sqrt{t_{i+1}}}{\frac{c}{2} + \sqrt{t_{i+1}}}$。可知序列 $\{t^k\}$ 至少线性收敛至 $\dfrac{c^2}{4}$，对应 $\{x^k\}$ 最后趋近于在 $\dfrac{c^2}{4}\dfrac{x_0}{\|x_0\|}$ 与 $-\dfrac{c^2}{4}\dfrac{x_0}{\|x_0\|}$ 之间正负横跳。不收敛至 0.

综上所述，（2）成立！ □

Problem 2

Let $f(x) = \frac{1}{2}x^T Q x$, where Q is symmetric, invertible, and has at least one negative eigenvalue. Consider the steepest descent method with constant stepsize and show that unless the starting point x^0 belongs to the subspace spanned by the eigenvectors of Q corresponding to the non-negative eigenvalues, the generated sequence $\{x^k\}$ diverges.

证明. 容易知道 $\nabla f(x) = Qx$，令 $x_k = \sum_{i=1}^n \beta_{i,k} y_i$，其中 $y_1 \cdots y_n$ 是 Q 的互相正交的单位特征向量，对应特征值 $\lambda_1 \leq \cdots \leq \lambda_i < 0 < \cdots \leq \lambda_n$。设固定的步长为 α. 则有：$x_{k+1} = x_k - \alpha Q x_k \Rightarrow \beta_{s,k+1} = (1 - \alpha\lambda_s)\beta_{s,k}$。若 $\exists s \leq i$, 且 $\beta_{s,0} \neq 0$ 此时：

- $\alpha > 0$ 由上面推导过程 $\beta_{s,n} = (1 - \alpha\lambda_s)^n \beta_{s,0}$，且 $1 - \alpha\lambda_s > 1$。故 $\lim_{n\to\infty} \|x_n\|^2 \geq \lim_{n\to\infty} |\beta_{s,n}|^2 = \lim_{n\to\infty}(1 - \alpha\lambda_s)^n |\beta_{s,0}|^2 = \infty$。序列发散！

- $\alpha < 0$ 若存在 $\exists \bar{s} > i$, 且 $\beta_{\bar{s},0} \neq 0$，则情况类似于上述过程，一定发散！否则 x_0 落在 Q 的负特征空间。而 $x_k^T Q x_k = \sum_{j=1}^i \lambda_j \beta_{j,k}^2$。如果 $\forall j < i$ 且 $\beta_{j,0} \neq 0$ 有：$1 - \alpha\lambda_j \geq -1$（否则类似于上面的情况，必然发散），此时 $|\beta_{j,k}|$ 是关于 k 线性衰减的（当全部的 $1 - \alpha\lambda_j = -1$ 时函数值不变），所以此时函数值不会下降，不符合函数值下降的要求，不成立！

- $\alpha = 0$ 此时相当于 x_k 没变（$x_k = x_0$），无意义！

综上所述，当 $x_0 \notin \text{span}\{y_{i+1}, \cdots, y_n\}$ 时，$\{x^k\}$ 必然发散！ $\square$

Problem 3

Consider the steepest descent method $x^{k+1} = x^k - \alpha^k \left(\nabla f(x^k) + e^k\right)$, where e^k is an error satisfying $\|e^k\| \le \delta$ for all k. Assume that ∇f is Lipschitz continuous. Show that for any $\delta' > \delta$, there exists a range of positive stepsizes $[\underline{\alpha}, \bar{\alpha}]$ such that if $\alpha^k \in [\underline{\alpha}, \bar{\alpha}]$ for all sufficiently large k, then either $f(x^k) \to -\infty$ or $\|\nabla f(x^k)\| < \delta'$ for infinitely many values of k. (Hint: Using the reasoning of Prop.1.2.2 in the *Nonlinear Programming* textbook.)

证明. 先证几个辅助命题：

- （1）若 $\|\nabla f(x_k)\| \ge \delta'$，且 $\|e_k\| \le \delta$，则有：$(\nabla f(x_k)+e_k)^T \nabla f(x_k) = \|\nabla f(x_k)\|^2 + e_k^T \nabla f(x_k) \ge \|\nabla f(x_k)\|^2 - \|\nabla f(x_k)\|\|e_k\|$。最右端表达式在 $\|\nabla f(x_k)\| > \|e_k\|$ 时单调递增，所右端项在 $\|\nabla f(x_k)\| = \delta'$ 时取得最小值。同时表达式关于 $\|e_k\|$ 单调递减，故在 $\|e_k\| = \delta$ 取得最小值。故 $(\nabla f(x_k) + e_k)^T \nabla f(x_k) \ge \delta'(\delta' - \delta) > 0$。

- （2）若 $\|\nabla f(x_k)\| \ge \delta'$，且 $\|e_k\| \le \delta$，并令 θ 为 $\nabla f(x_k)$ 与 e_k 夹角。则有：

$$\frac{(\nabla f(x_k) + e_k)^T \nabla f(x_k)}{\|\nabla f(x_k) + e_k\|^2} = \frac{\|\nabla f(x_k)\|^2 + \|\nabla f(x_k)\|\|e_k\| \cos\theta}{\|\nabla f(x_k)\|^2 + \|e_k\|^2 + 2\|\nabla f(x_k)\|\|e_k\| \cos\theta}$$
$$\xlongequal{t=\|\nabla f(x_k)\|(\|\nabla f(x_k)\|+\|e_k\|\cos\theta)} \frac{1}{2 + \dfrac{\|e_k\|^2 - \|\nabla f(x_k)\|^2}{t}} = g(t)$$

  显然 $g(t)$ 关于 t 单调递减。所以固定梯度与误差的模长时 $g(t)$ 在 $\cos\theta = 1$ 时取得最小值 $\dfrac{\|\nabla f(x_k)\|}{\|\nabla f(x_k)\| + \|e_k\|} = h(\|e_k\|, \|\nabla f(x_k)\|)$。又显然 $h(\|e_k\|, \|\nabla f(x_k)\|)$ 关于 $\|e_k\|$ 单调递减，关于 $\|\nabla f(x_k)\|$ 单调递增。故有 $h(\|e_k\|, \|\nabla f(x_k)\|) \ge \dfrac{\delta'}{\delta' + \delta}$。即 $\dfrac{(\nabla f(x_k) + e_k)^T \nabla f(x_k)}{\|\nabla f(x_k) + e_k\|^2} \ge \dfrac{\delta'}{\delta' + \delta}$。

- （3）若有 $x_{k+1} = x_k + \alpha d$ 且 ∇f 是 Lipschitz 连续的，则有：

$$\begin{aligned}
f(x_{k+1}) - f(x_k) &= \int_0^{\alpha} d^T \nabla f(x_k + td)\,\mathrm{d}t \\
&= \int_0^{\alpha} d^T \nabla f(x_k) + d^T(\nabla f(x_k + td) - \nabla f(x_k))\,\mathrm{d}t \\
&\leq \alpha d^T \nabla f(x_k) + \int_0^{\alpha} tL\|d\|^2\,\mathrm{d}t \\
&= \alpha d^T \nabla f(x_k) + \frac{\alpha^2 L}{2}\|d\|^2
\end{aligned}$$

下面开始证明。采用反证法。取足够小的 $\epsilon > 0$，使得 $0 < \underline{\alpha} = \epsilon < \dfrac{(2-\epsilon)\delta'}{L(\delta+\delta')} = \bar{\alpha}$。若对于所有足够大的 k，$\alpha^k \in [\underline{\alpha}, \bar{\alpha}]$，并且不满足题设条件。即此时只有有限多个 k 满足 $\|\nabla f(x^k)\| < \delta'$（可以推出有无限多个大于等于）并且 $f(x^k)$ 并不趋于负无穷。对于每一个足够大，满足 $\nabla f(x_k) \geq \delta'$ 且 $\alpha^k \in [\underline{\alpha}, \bar{\alpha}]$ 的 k 则有：

$$\begin{aligned}
f(x_{k+1}) - f(x_k) &\leq -\alpha_k(\nabla f(x_k) + e_k)^T \nabla f(x_k) + \frac{\alpha_k^2 L}{2}\|\nabla f(x_k) + e_k\|^2 \\
&\leq \alpha_k(\bar{\alpha}\frac{L\|\nabla f(x_k) + e_k\|^2}{2} - (\nabla f(x_k) + e_k)^T \nabla f(x_k)) \\
&\leq \alpha_k((1 - \frac{\epsilon}{2})(\nabla f(x_k) + e_k)^T \nabla f(x_k) - (\nabla f(x_k) + e_k)^T \nabla f(x_k)) \\
&\leq \frac{-\epsilon^2}{2}(\nabla f(x_k) + e_k)^T \nabla f(x_k) \\
&\leq \frac{-\epsilon^2}{2}(\delta'(\delta' - \delta))
\end{aligned}$$

对于那些有限个的满足 $\|\nabla f(x^k)\| < \delta'$ 的 k，假设其每步使得函数值变化的绝对值的总和为 M。而对于那些无穷多个满足 $\nabla f(x_k) \geq \delta'$ 条件的 k。不妨记对应的 x_k 构成的子列为 $x_{k_i} \Rightarrow$

$$\lim_{n\to\infty} f(x_{k_n}) < \lim_{n\to\infty} M + f(x_{k_0}) - n(\frac{\epsilon^2}{2}(\delta'(\delta' - \delta))) = -\infty$$

与 $f(x^k)$ 并不趋于负无穷互相矛盾。所以对于所有足够大的 k，若 $\alpha^k \in [\underline{\alpha}, \bar{\alpha}]$，此时“只有有限多个 k 满足 $\|\nabla f(x^k)\| < \delta'$ 并且 $f(x^k)$ 并不趋于负无穷”这一条件并不成立。换而言之，此时要么 $f(x_k) \to -\infty$，要么有无穷多个 $\|\nabla f(x_k)\| < \delta'$。 □