Jesse Cai
MATH 164 Homework 2
304634445

1. **9.3 Consider minimizing** $f(x) = x^{\frac{4}{3}}$

   (a) We can find $\nabla f(x) = g(x) = \frac{4}{3}\sqrt[3]{x}$ and $f"(x) = \frac{4}{9}x^{-\frac{2}{3}}$

   $$x^{k+1} = x^k - F(x^k)^{-1}g^k = x^k - (\frac{4}{9}x^{-\frac{2}{3}})^{-1}\frac{4}{3}x^{\frac{1}{3}} = x^k - 3x^k = -2x^k$$

   (b) We can expand the fomula to get $x^k = -2^k x^0$ which does not converge to 0 except when $x^0 = 0$.

2. **9.4 Consider minimizing** $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$

   (a) Note $f(x) \geq 0$, so the minima occurs when $f(x) = 0$. For this to be true, we can see that both $(1 - x_1)^2$ and $(100(x_2 - x_1^2)^2$ are 0. Solving the first equation yields $x_1 = 1$ and plugging this into the second part and solving yields $x_2 = 1$. So the unique minima is $[1, 1]$.

   (b) $\nabla f(x) = [400x_1^3 - 400x_1x_2 + 2x_1 - 2, 200(x_2 - x_1^2)]^T$

   $$F(x) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

   $$F(x)^{-1} = \frac{1}{80000(x_1^2 - x_2) + 400}\begin{bmatrix} 200 & 400x_1 \\ 400x_1 & 1200x_1^2 - 400x_2 + 2 \end{bmatrix}$$

   $x_0 = [0, 0]^T$, $g_0 = [-2, 0]$, $x_1 = x_0 - F(x^0)^{-1}g^0 = -\frac{1}{400}\begin{bmatrix} 200 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

   $x_1 = [1, 0]^T$, $g_1 = [400, -200]$, $x_2 = x_1 - F(x^1)^{-1}g^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \frac{1}{80400}\begin{bmatrix} 200 & 400 \\ 400 & 1202 \end{bmatrix}\begin{bmatrix} 400 \\ -200 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

   (c) $x_0 = [0, 0]^T$, $g_0 = [-2, 0]$, $x_1 = x_0 - 0.05g_0 = [0.1, 0]^T$
   $x_1 = [0.1, 0]^T$, $g_1 = [-1.4, 2]$, $x_2 = x_1 - 0.05g_1 = [0.17, 0.1]^T$

3. **10.1 Show that** $d^0 \dots d^{n-1}$ **are** Q **conjugate**

   We can do this via induction. Let $P(n) = \forall i < n : d_i^T Q d_n = 0$.

   Base Case: $d_0 = p_0$ and $d_1 = p_1 - \frac{p_1^T Q d_0}{d_0^T Q d_0} d_0$

   $$d_0^T Q d_1 = p_0^T Q(p_1 - \frac{p_1^T Q p_0}{p_0^T Q p_0} p_0) = p_0^T Q p_1 - \frac{p_1^T Q p_0}{p_0^T Q p_0} p_0^T Q p_0 = (p_1^T Q p_0)(1 - 1) = 0$$

   Inductive Step: Assume $P(n)$ holds, then we want to show $P(n) \implies P(n+1)$.

   $$d_{k+1} = p_{k+1} - \sum_{i=0}^{k} \frac{p_{k+1}^T Q d_i}{d_i^T Q d_i} d_i \implies d_j^T Q d_{k+1} = d_j^T Q p_{k+1} - \sum_{i=0}^{k} \frac{p_{k+1}^T Q d_i}{d_i^T Q d_i} d_i$$

   But we know by $P(n)$ that $i \neq j \implies d_j^T d_i = 0$ and then by symmetry of Q.

   $$d_j^T Q d_{k+1} = d_j^T Q p_{k+1} - \frac{p_{k+1}^T Q d_j}{d_j^T Q d_j} d_j^T Q d_j = d_j^T Q p_{k+1} - p_{k+1}^T Q d_j = 0$$

4. **10.2 Show that if** $g_k^T d_k \neq 0$ **then** $\{d_0 \dots d_{n-1}\}$ **are Q-conjugate.**

   $$f(x_k + \alpha d_k) = \frac{1}{2}(x_k + \alpha d_k)^T Q(x_k + \alpha d_k) - (x_k + \alpha d_k)^T b = \frac{1}{2} d_k^T Q d_k \alpha^2 + g_k^T d_k \alpha + C$$

   We can take the derivitave of ϕ with respect to α, which is 0 due to Lemma 10.2.
   $\phi'(\alpha_k) = \nabla f(x_k + \alpha_k d_k)^T d_k = g_{k+1}^t d_k = 0$
   But then since $g_k^T d_i = 0$:

   $$d_k^T Q d_i = \frac{1}{\alpha_k}(x_{k+1} - x_k)^T Q d_i = \frac{1}{\alpha_k}(g_{k+1} - g_k)^T Q d_i = 0$$

5. **10.3 Show that in the conjugate gradient method for a standard quadratic $d_k^T Q d_k = -d_k^T Q g_k$**
   Again we can do this via induction. Let $P(k) = d_k^T Q d_k = -d_k^T Q g_k$
   Base Case: By definition $g_0 = -d_0 \implies d_0^T Q d_0 = -d_0^T g_0$ so $P(0)$ holds.
   Inductive Step: Assume $d_k^T Q d_k = -d_k^T Q g_k$. By definition $d_{k+1} = -g_{k+1} - \beta_k d_k$
   Then since d_k, d_{k+1} are Q-conjugate:
   $$d_{k+1}^T Q d_{k+1} = d_{k+1}^T Q(-g_{k+1} - \beta_k d_k) = -d_{k+1}^T Q g_{k+1} - \beta_k d_{k+1}^T Q d_k = -d_{k+1}^T Q g_{k+1}$$

6. **10.4 Let Q be a real symmetric matrix**
   (a) Since Q is a real symmetric matrix, there exists an orthogonal basis of eigenvectors $\{v_1 \dots v_n\}$. This basis is $Q - conjugate$.
   Pick i, j s.t. $i \neq j$. Then because v_j is an eigenvector and since v_i, v_j are orthogonal.
   $$\implies v_i^T Q v_j = v^i \lambda_j v_j = \lambda_j v_i^T v_j = 0$$
   (b) If $\{d_1 \dots d_n\}$ is Q-conjugate and also orthogonal then.
   $$d_i^T Q d_j = 0 = d_i^T d_j = \lambda_j d_i^T d_j = d_i^T \lambda_j d_j \implies Q d_j = \lambda_j d_j$$
   So d_j must be an eigenvector.

7. **10.5 For a standard quadratic with $d_{k+1} = \gamma_k g_{k+1} + d_k$ find an expression for γ_k**
   Recall that our original definition of the conjugate gradient method gives $d_{k+1} = -g_{k+1} + \beta_k d_k$.
   So we can plug in and solve:
   $$\gamma_k g_{k+1} + d_k = -g_{k+1} + \beta_k d_k \implies \gamma_k = \frac{-1}{\beta_k} = \frac{d_k^T Q d_k}{g_{k+1}^T Q d_k}$$

8. **10.6 Suppose we are minimizing a standard quadratic with update rule $d_{k+1} = \alpha_k g_{k+1} + \beta_k d, \alpha_k, \beta_k \in \mathbb{R}$.**
   (a) Show that $d_k \in V_{k+1}$ and $x_k \in V_k$. We can prove this by induction. Let $P(k) = d_k \in V_{k+1} \wedge x_k \in V_k$.
   Base Case $P(0)$:
   $d_0 = a_0 g_0 = -a_0 b \in V_1$ since $b \in V_1$ $x_0 = 0 \in V_0$ as 0 is in every subspace.
   Inductive Step: Assume $d_k \in V_{k+1} \wedge x_k \in V_k$.
   Note $V_k \subset V_{k+1} \implies x_k \in V_{k+1}$ so
   $x_{k+1} = x_k + \alpha_k d_k \implies x_{k+1} \in V_{k+1}$
   $d_{k+1} = a_k g_{k+1} + b_k d_k$. But $Q x_{k+1} - b \in V_{k+2}$ by definition so $d_{k+1} \in V_{k+2}$
   (b) The conjugate gradient algorithm finds the min of each subspace V_k along each step.

9. **10.7 Consider a standard quadratic function $f(x)$ and $\phi(a) = f(x_0 + Da)$ where D is a matrix of rank r. Show ϕ is a standard quadratic**
   Note $f(x) = \frac{1}{2} x^T Q x - x^T b + c$
   $$\phi(a) = \frac{1}{2}(x_0 + Da)^T Q(x_0 + Da) - (x_0 + Da)^T b = x_0^T Q x_0 + (aD)^T Q Da - x_0^T b - (Da)^T b$$
   Using the fact that x_0 and b are constants, and then $(Da)^T = a^T D^T$ we get:
   $$\phi(a) = (Da)^T Q Da - (Da)^T b + c = a D^T Q D a = a^T D b \implies Q' = D^T Q D, b' = Db$$
   So $\phi(a)$ can be written in the standard quadratic form.

10. **10.8 Consider a conjugate gradient algorithm applied to a quadratic function**

(a) WTS $\forall 0 \leq k \leq n-1 \wedge 0 \leq i \leq k : g_{k+1}^T g_i = 0$.
Recall that $d_{k+1} = -g_{k+1} + \beta_k d_k \implies g_{k+1} = \beta_k d_k - d_{k+1}$.

$$g_{k+1}^T g_i = (\beta_k d_k - d_{k+1})^T g_i$$

and by Lemma 10.2 $d_k^T g_i$ and $d_{k+1}^T g_i = 0$

(b) $g_{k+1}^T Q g_i = (\beta_k d_k - d_{k+1})^T Q(\beta_{i-1} d_{i-1} - d_i)$

$$= \beta_k \beta_{i-1} d_k^T Q d_{i-1} \beta_k d_k^T Q d_i - \beta_{i-1} d_{k+1}^T Q d_{i-1} + d_{k+1}^T Q d_i$$

Then by Q-conjuacy this is 0 so g_{k+1} and g_i are Q-conjugate. $rank(D) = r \implies Da = 0 \iff a = 0 \implies Q'$ is positive definite.

11. **10.9 Represent $f(x_1, x_2) = \frac{5}{2}x_1^2 + x_2^2 - 3x_1x_2 - x_2 - 7$ in standard form and find d_1.**

$$f(x) = \frac{1}{2}x^T \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} x - [0,1]^T x - 7$$

Recall $d_0 = g_0 = Qx_0 - b = Q(0) - b = [0,1]^T$

$x_1 = x_0 - \alpha_0 d_0 = -\frac{g_0^T d_0}{d_0^T Q d_0} d_0 = [0, \frac{1}{2}]^T$

$g_1 = \nabla f(x_1) = Q[0, \frac{1}{2}] - b = [-\frac{3}{2}, 0]^T$

$d_1 = -[-\frac{3}{2}, 0]^T + \frac{[-\frac{3}{2},0]^T[0,1]}{2}[0,1]^T = [\frac{3}{2}, \frac{9}{4}]^T$

12. **10.10 Consider minimizing $f(x_1, x_2) = \frac{5}{2}x_1^2 + \frac{1}{2}x_2^2 + 2x_1x_2 - 3x_1 - x_2$.**

(a)

$$f(x) = \frac{1}{2}x^T \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} x - [3,1]^T x$$

(b) $x_0 = 0, -d_0 = g_0 = [3,1]^T, x_1 = -\frac{g_0^T g_0}{g_0^T Q g_0} d_0 = \frac{5}{29}[3,1]^T$
$g_1 =, d_1 = -g_1 + \beta_0 d_0$
$x_2 = [1, -1]^T$.

(c) Solving this analyticaly, we get the same answeras above.

$$x^* = Q^{-1}b = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

13. **11.1**

(a) $\phi(\alpha) = f(x_k + \alpha d_k) \implies \phi'(\alpha) = d_k^T g_k$. But ϕ' is continuous so if $d_k^T g_k < 0 \implies \exists \alpha : \forall \alpha \in (0, \alpha) \phi(\alpha) > \phi(0)$

(b) We know that $alpha_k = \arg\min_{\alpha \geq 0} \phi(\alpha)$ so either $alpha_k > 0$ or $\alpha_k = 0$, but $\alpha_k = 0$ leads to a contradiction with a) so $alpha_k > 0$.

(c) $g_{k+1} = f'(x_k + \alpha d_k) = \phi'(alpha_k) = 0$

(d) i. $d_k = -g_k \implies d_k^T g_k = -g_k^T g_k = -\|g_k\|^2 < 0$
 ii. $d_k = -F(x_k)^{-1} g_k$. But Q is positive definite $\implies Q^{-1}$ is positive definite.
 So $d_k^T g_k = -g_k^T F(x_k^{-1]}) g_k < 0$
 iii. $d_k = -g_k + \beta_{k-1} d_{k-1}$ So using Lemma 10.2 we get $d_k^T g_k = -g_k^T g_k + \beta_{k-1} d_{k-1}^T g_k = -g_k^T g_k < 0$
 iv. $d_k = -H_k g_k$ If $H_k > 0 \implies d_k^T g_k = -g_k^T H g_k < 0$

(e) $\alpha_k = -\frac{d_k^T g_k}{d_k^T Q d_k}$

14. **11.3 Consider minimizing $\phi(\alpha) = f(x_k + \alpha d_k)$ where f is a standard quadratic.**

(a) $\phi'(\alpha) = 0 = d_k^T f'(x + \alpha d_k)$

$$= d_k^T (Q(x + \alpha d_k) - b) = d_k^T (Qx - b) + \alpha d_k^T Q d_k$$

Solving for α yields $\alpha = \frac{d_k^T g_k}{d_k^T Q d_k}$

(b) we can expand $d_k^T Q d_k$ as $g_k^T H_k^T Q H g_k$ so we know that H_k must be positive definite to ensure Q is as well.

15. **11.5 Minimize $f(x) = \frac{1}{2}x^T \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} x + x^T \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.**

$g_0 = Qx_0 - b = -b = [-1, 1]^T$ and $d_0 = -H_0 g_0 = -g_0 = [1, -1]^T$

$\alpha_0 = \frac{g_0^T d_0}{d_0^T Q d_0} = \frac{2}{3}$

$x_1 = x_0 + \frac{2}{3}[1, -1]^T = [\frac{2}{3}, -\frac{2}{3}]^T$

$g_1 = Qx_1 - b = [-\frac{1}{3}, -\frac{1}{3}]^T$

$\Delta x_0 - H_0 \Delta g_0 = \begin{bmatrix} 0 - \frac{2}{3} \\ 0 + \frac{2}{3} \end{bmatrix} - \begin{bmatrix} 1 + \frac{1}{3} \\ -1 + \frac{1}{3} \end{bmatrix} = [0, \frac{2}{3}]$

$\Delta g_0^T (\Delta x_0 - H_0 \Delta g_0) = -\frac{8}{9}$

$$H_1 = H_0 + \frac{(\Delta x_0 - H_0 \Delta g_0)(\Delta x_0 - H_0 \Delta g_0)^T}{\Delta g_0^T (\Delta x_0 - H_0 \Delta g_0)} = I + -\frac{9}{8} \begin{bmatrix} 0 & 0 \\ 0 & \frac{9}{4} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$d_1 = -H_1 g_1 = -\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{6} \end{bmatrix}$

$\alpha_1 = \frac{g_1^T d_1}{d_1^T Q d_1} = 1$

$x_2 = x_* = \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix}$

16. **11.7 Show that if $H_k > 0$ and $\Delta g_k^T (\Delta x_k - H_k \Delta g_k) > 0$ then $H_{k+1} > 0$.**

Recall that

$$H_{k+1} = H_k + \frac{(\Delta x_k - H_k \Delta g_k)(\Delta x_k - H_k \Delta g_k)^T}{g_k^T (\Delta x_k - H_k \Delta g_k)}$$

.

$$x^T H_{k+1} x = x^T H_k x + \frac{x^T (\Delta x_k - H_k \Delta g_k)(\Delta x_k - H_k \Delta g_k)^T x}{g_k^T (\Delta x_k - H_k \Delta g_k)}$$

$$= x^T H_k x + \frac{[x^T (\Delta x_k - H_k \Delta g_k)]^2}{g_k^T (\Delta x_k - H_k \Delta g_k)}$$

$$=> 0 + > 0$$

since there's a square and we know the denominator is > 0 by assumption.

17. **11.9 Use rank-one correction method to generate two Q-conjugate directions.**

This is the exact same equation as 11.5, so we can take the two vectors generated there:

$$d_0 = -H_0 g_0 = -g_0 = [1, -1]^T$$

$$d_1 = -H_1 g_1 = -\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{6} \end{bmatrix}$$

18. **11.10 Apply the rank-1 algorithm to $f(x) = \frac{1}{2}x^T \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} x - x^T \begin{bmatrix} -1 \\ 1 \end{bmatrix}$**

$g_0 = Qx_0 - b = -b = [-1, 1]^T$ and $d_0 = -H_0 g_0 = -g_0 = [1, -1]^T$

$\alpha_0 = \frac{g_0^T d_0}{d_0^T Q d_0} = \frac{2}{3}$

$x_1 = x_0 + \frac{2}{3}[1, -1]^T = [\frac{2}{3}, -\frac{2}{3}]^T$

$g_1 = Qx_1 - b = [\frac{4}{3}, 0]^T$

$\Delta x_0 - H_0 \Delta g_0 = \begin{bmatrix} \frac{2}{3} \\ -\frac{2}{3} \end{bmatrix} - \begin{bmatrix} \frac{4}{3} - 1 \\ 0 + 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{-5}{3} \end{bmatrix}$

$\Delta g_0^T(\Delta x_0 - H_0\Delta g_0) = [\frac{1}{3}, 1]^T \begin{bmatrix} \frac{1}{3} \\ -\frac{5}{3} \end{bmatrix} = -\frac{14}{9}$

$$H_1 = H_0 + \frac{(\Delta x_0 - H_0\Delta g_0)(\Delta x_0 - H_0\Delta g_0)^T}{\Delta g_0^T(\Delta x_0 - H_0\Delta g_0)} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$d_1 = -H_1 g_1 = -\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} = 0$

So this algorithm fails for this problem.