Optimization

Homework 2 Solutions

1.

(a). The value of x^* (in terms of a, b and c) that minimizes f is $x^* = b/a$.

(b). We have $f'(x) = ax - b$. Therefore, the recursive equation for the DDS algorithm is $x^{(k+1)} = x^{(k)} - \alpha(a\, x^{(k)} - b) = (1 - \alpha a)\, x^{(k)} + \alpha b$.

(c). Let $\bar{x} = \lim_{k\to\infty} x^{(k)}$. Taking limits of both sides of $x^{(k+1)} = x^{(k)} - \alpha(a\, x^{(k)} - b)$

(from part b), we get $\bar{x} = \bar{x} - \alpha(a\, x^{(k)} - b)$. Hence, we get $\bar{x} = b/a = x^*$.

(d). To find the order of convergence, we compute

$$\frac{|\mathbf{x}^{(\mathbf{k+1})} - \mathbf{b/a}|}{|\mathbf{x}^{(\mathbf{k})} - \mathbf{b/a}|^{\mathbf{p}}} = \frac{|(1-\alpha a)\mathbf{x}^{(\mathbf{k})} + \alpha b - \mathbf{b/a}|}{|\mathbf{x}^{(\mathbf{k})} - \mathbf{b/a}|^{\mathbf{p}}}$$

$$= \frac{|(1-\alpha a)\mathbf{x}^{(\mathbf{k})} - (1-\alpha a)\mathbf{b/a}|}{|\mathbf{x}^{(\mathbf{k})} - \mathbf{b/a}|^{\mathbf{p}}} = |1 - \alpha a\ ||x^{(k)} - b/a|^{1-p} \ .$$

Let $z^{(k)} = |1 - \alpha a\ ||x^{(k)} - b/a|^{1-p}$. Note that $z^{(k)}$ converges to a finite nonzero number if and only if $p=1$ (if $p<1$, then $z^{(k)} \to 0$, and if $p>1$, then $z^{(k} \to \infty$). Therefore, the order of convergence of $\{ x^{(k)} \}$ is 1.

(e). Let $y^{(k)} = |x^{(k)} - b/a|$. From part d, after some manipulation we obtain $y^{(k+1)} = |1 - \alpha a|\ y^{(k)} = |1 - \alpha a|^{k+1}\, y^{(0)}$. The sequence $\{x^{(k)}\}$ converges (to b/a) if and only if $y^{(k)} \to 0$. This holds if and only if $|1 - \alpha a| < 1$, which is equivalent to $0 < \alpha < 2/a$.

2.

(a). We have $f(x) = ||Ax - b||^2 = (Ax - b)^T(Ax - b)$

$$= (x^T A^T - b^T)(Ax - b) = x^T (A^T A)\, x - 2(A^T b)\, x + b^T b\ .$$

which is a quadratic function. The gradient is given by $\nabla f(x) = 2(A^T A)\, x - 2(A^T b)$ and the Hessian is given by $F(x) = 2(A^T A)$.

(b). The fixed step size gradient algorithm for solving the above optimization problem is given by $x^{(k)} = x^{(k)} - \alpha(2(A^T A)\, x^{(k)} - 2A^T b)$

$$= x^{(k)} - 2\alpha A^T (Ax^{(k)} - b)\ .$$

(c). The largest range of values for α such that the algorithm in part b converges to the solution of the problem is given by $0 < \alpha < \frac{2}{\lambda\max(2A^T A)} = \frac{1}{4}$.

3. The steepest descent algorithm applied to the quadratic function f has the form

$$x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)} = x^{(k)} - \frac{g^{(k)T} g^{(k)}}{g^{(k)T} Q g^{(k)}}\, g^{(k)}\ .$$

⇒**:** if $x^{(1)} = Q^{-1}b$, then $Q^{-1}b = x^{(0)} - \alpha_0 g^{(0)}$.

Rearranging the above yields $Qx^{(0)} - b = \alpha_0 Q g^{(0)}$.

Since $g^{(0)} = Qx^{(0)} - b \neq 0$, we have $Qg^{(0)} = \frac{1}{\alpha_0} g^{(0)}$.

Which means that $g^{(0)}$ is an eigenvector of Q (with corresponding eigenvalue $\frac{1}{\alpha_0}$).

⇐**:** By assumption, $Qg^{(0)} = \lambda g^{(0)}$. Where $\lambda \in R$. We want to show that $Qx^{(1)} = b$. We have $Qx^{(1)} = Q(x^{(0)} - \frac{g^{(0)T} g^{(0)}}{g^{(0)T} Q g^{(0)}} g^{(0)})$.

$$= Q x^{(0)} - \frac{1}{\lambda}\frac{g^{(0)T} g^{(0)}}{g^{(0)T} g^{(0)}} Q g^{(0)} .$$

$$= Q x^{(0)} - g^{(0)} = b.$$

4.

For the given algorithm we have $\gamma_k = \beta(2-\beta)\left(\frac{(g^{(k)T} g^{(k)})^2}{(g^{(k)T} Q g^{(k)})(g^{(k)T} Q^{-1} g^{(k)})}\right)$.

If $0 < \beta < 2$, then $\beta(2-\beta) > 0$, and by Lemma 8.2, $\gamma_k \geq \beta(2-\beta)\left(\frac{\lambda_{\min}(Q)}{\lambda_{\max}(Q)}\right) > 0$.

which implies that $\sum_{k=0}^{\infty} \gamma_k = \infty$. Hence, by Theorem 8.1, $x^{(k)} \to x^*$ for any $x^{(0)}$.

If $\beta \leq 0$ or $\beta \geq 0$, then $\beta(2-\beta) \leq 0$, and by Lemma 8.2,

$\gamma_k \leq \beta(2-\beta)\left(\frac{\lambda_{\min}(Q)}{\lambda_{\max}(Q)}\right) < 0$.

By Lemma 8.1, $V(x^{(k)}) \geq V(x^{(0)})$. Hence, if $x^{(0)} \neq x^*$, then $\{V(x^{(k)})\}$ does not converge to 0, and consequently $x^{(k)}$ does not converge to x^* .

5.

By Lemma 8.1, $V(x^{(k+1)}) = (1-\gamma_k)V(x^{(k)})$ for all k. Note that the algorithm has a descent property if an only if $V(x^{(k+1)}) < V(x^{(k)})$ whenever $g^{(k)} \neq 0$. Clearly, whenever $g^{(k)} \neq 0$, $V(x^{(k+1)}) < V(x^{(k)})$ if and only if $1-\gamma_k < 1$. The desired result follows immediately.

6.

(a). $x^{(k+1)} = x^{(k+1)} - F(x^{(k)})^{-1} g^{(k)}$. $\nabla f(x) = \begin{bmatrix} -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$.

$F(x) = \begin{bmatrix} -400x_1(x_2 - x_1^2) + 800x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$.

$k=0$, $x^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $g^{(0)} = \nabla f(x^{(0)}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $F(x^{(0)}) > 0$.

$x^{(1)} = x^{(0)} - F(x^{(0)})^{-1}g^{(0)} = \begin{bmatrix}1\\1\end{bmatrix} = x^{(0)}$, $g^{(1)} = \nabla f(x^{(1)}) = \begin{bmatrix}0\\0\end{bmatrix}$.

Hence, $x^{(0)}$ is a global minimizer of f.

(b). $x^{(0)} = \begin{bmatrix}0\\0\end{bmatrix}$, $\nabla f(x) = \begin{bmatrix}-400x_1(x_2 - x_1^2) - 2(1 - x_1)\\200(x_2 - x_1^2)\end{bmatrix}$.

$$F(x) = \begin{bmatrix}-400x_1(x_2 - x_1^2) + 800x_1^2 + 2 & -400x_1\\-400x_1 & 200\end{bmatrix}.$$

Iteration 1:

$g^{(0)} = \nabla f(x^{(0)}) = \begin{bmatrix}-2\\0\end{bmatrix}$, $F(x^{(0)}) = \begin{bmatrix}2 & 0\\0 & 200\end{bmatrix}$, $F(x^{(0)})^{-1} = \begin{bmatrix}\frac{1}{2} & 0\\0 & \frac{1}{200}\end{bmatrix}$.

$F(x^{(0)})^{-1} g^{(0)} = \begin{bmatrix}\frac{1}{2} & 0\\0 & \frac{1}{200}\end{bmatrix}\begin{bmatrix}-2\\0\end{bmatrix} = \begin{bmatrix}-1\\0\end{bmatrix}$. $x^{(1)} = x^{(0)} - F(x^{(0)})^{-1}g^{(0)} = \begin{bmatrix}1\\0\end{bmatrix}$.

Iteration 2:

$g^{(1)} = \nabla f(x^{(1)}) = \begin{bmatrix}400\\-200\end{bmatrix}$, $F(x^{(1)}) = \begin{bmatrix}402 & -400\\-400 & 200\end{bmatrix}$,

$F(x^{(1)})^{-1} = \frac{-1}{79600}\begin{bmatrix}200 & 400\\400 & 402\end{bmatrix}$.

$F(x^{(1)})^{-1} g^{(1)} = \frac{-1}{79600}\begin{bmatrix}200 & 400\\400 & 402\end{bmatrix}\begin{bmatrix}400\\-200\end{bmatrix} = \begin{bmatrix}0\\-1\end{bmatrix}$.

$x^{(2)} = x^{(1)} - F(x^{(1)})^{-1}g^{(1)} = \begin{bmatrix}1\\1\end{bmatrix}$.

(c). Use the gradient algorithm $x^{(k+1)} = x^{(k)} - \alpha^{(k)} g^{(k)}$.

$\alpha^{(0)} = \alpha^{(1)} = 0.05$

$x^{(0)} = \begin{bmatrix}0\\0\end{bmatrix}$., $\nabla f(x) = \begin{bmatrix}-400x_1(x_2 - x_1^2) - 2(1 - x_1)\\200(x_2 - x_1^2)\end{bmatrix}$.

Iteration 1:

$g^{(k)} = \nabla f(x^{(0)}) = \begin{bmatrix}-2\\0\end{bmatrix}$.

$x^{(1)} = x^{(0)} - \alpha^{(0)} g^{(0)} = \begin{bmatrix}0.1\\0\end{bmatrix}$..

Iteration 2:

$g^{(1)} = \nabla f(x^{(1)}) = \begin{bmatrix}-1.4\\2\end{bmatrix}$.

$x^{(2)} = x^{(1)} - \alpha^{(1)} g^{(1)} = \begin{bmatrix}0.17\\0.1\end{bmatrix}$..

7. If $x^{(0)} = x^*$, we are done. So, assume $x^{(0)} \neq x^*$. Since the standard Newton's method reaches the point x^* in one step, we have

$f(x^*) = f(x^{(0)}+Q^{-1}g^{(0)})= \min f(x^*) \leq f(x^{(0)}+\alpha Q^{-1}g^{(0)})$.

For any $\alpha \geq 0$, Hence $\alpha_0=\arg\min f(x^{(0)}+\alpha Q^{-1}g^{(0)}) =1$.

Hence, in the case, the modified Newton's algorithm is equivalent to the standard Newton's algorithm and thus $x^{(1)} = x^*$,