

Homework Assignment 6

1. While searching for the minimum of

$$f(x) = [x_1^2 + (x_2 + 1)^2][x_1^2 + (x_2 - 1)^2]$$

the algorithm terminates at the following points:

- a) $x^{(1)} = [0, 0]^T$
- b) $x^{(2)} = [0, 1]^T$
- c) $x^{(3)} = [0, -1]^T$
- d) $x^{(4)} = [1, 1]^T$

Classify each point.

2. Consider the quadratic function with the parameter M:

$$f(x) = 3x_1 + x_2 + 2x_3 + 4x_1^2 + 3x_2^2 + 2x_3^2 + (M-2)x_1x_2 + 2x_2x_3$$

For $M = 0$ find all stationary points. Are they optimal? Find the path of optimal solutions as M increases from zero.

3. In Powell damping, the BFGS update is modified if $s^T y$ is not sufficiently positive by defining $\bar{y} = \theta y + (1 - \theta)B^k s$ and substituting for y in the BFGS formula.

a) Show that θ can be found by solving the one-dimensional linear program:

$$\max \theta \text{ s. t. } \theta s^T y + (1 - \theta)s^T B^k s \geq 0.2 s^T B^k s, \theta \in [0, 1]$$

b) If $s^T y \geq 0.2 s^T B^k s$ show that Powell damping corresponds to a normal BFGS update.

c) If $s^T y \rightarrow -\infty$, show that Powell damping corresponds to skipping the BFGS update.

4. Show that if B^k is positive definite $\cos \theta^k > 1/\kappa(B^k)$ where $\kappa(B^k)$ is the condition number of B^k , based on the 2-norm.

5. Derive a stepsize rule for α for the Armijo line search that minimizes the quadratic interpolant from the Armijo inequality.

6. Consider the convex problem:

$$\min x_1 \text{ s. t. } x_2 \leq 0, x_2 - x_1^2 \geq 0$$

Show that this problem does not satisfy LICQ and does not satisfy the KKT conditions at its optimum solution.

7. Consider the convex problem

$$\min f(x) \text{ s.t. } g(x) \leq 0$$

and the equivalent problem

$$\min f(x) \text{ s.t. } g(x) + s = 0, s \geq 0.$$

- a) Show that the KKT conditions of the two problems are equivalent.
- b) If the second problem has a local solution. Show that this is also a global solution.

Solution Hw 6

1. Minimize

$$\begin{aligned}
 f(x) &= (x_1^2 + (x_2 + 1)^2)(x_1^2 + (x_2 - 1)^2) \\
 &= x_1^4 + x_1^2(x_2^2 + 2x_2 + 1) + x_1^2(x_2^2 - 2x_2 + 1) \\
 &\quad + (x_2 + 1)^2(x_2 - 1)^2 \\
 &= x_1^4 + x_1^2(2x_2^2 + 2) + (x_2^2 - 1)^2 \\
 &= x_1^4 + 2x_1^2x_2^2 + 2x_1^2 + x_2^4 - 2x_2^2 + 1
 \end{aligned}$$

$$\nabla f(x) = \begin{bmatrix} 4x_1^3 + 4x_1x_2^2 + 4x_1 \\ 4x_1^2x_2 + 4x_2^3 - 4x_2 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} 12x_1^2 + 4x_2^2 + 4 & 8x_1x_2 \\ 8x_1x_2 & 4x_1^2 + 12x_2^2 - 4 \end{bmatrix}$$

a) $x_a^T = [0 \ 0]$, $\nabla f(x_a) = 0$, $\nabla^2 f(x_a) = \begin{bmatrix} 4 & 0 \\ 0 & -4 \end{bmatrix}$

saddle point

b) $x_b^T = [0 \ 1]$, $\nabla f(x_b) = 0$, $\nabla^2 f(x_b) = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}$

local minimum - SSOC

c) $x_c^T = [0 \ -1]$, $\nabla f(x_c) = 0$, $\nabla^2 f(x_c) = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}$

local min - SSOC

d) $x_d = [1, 1]^T$, $\nabla f(x_d) = \begin{bmatrix} 12 \\ 4 \end{bmatrix}$
not a stationary p.

2.
$$f(x) = 3x_1 + x_2 + 2x_3 + 4x_1^2 + 3x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_2x_3$$

$$= \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}^T x + x^T \begin{bmatrix} 4 & \frac{M-1}{2} & \\ \frac{M-1}{2} & 3 & 1 \\ & 1 & 2 \end{bmatrix} x = a^T x + x^T A x$$

for $M=0$

$$f(x) = a^T x + x^T A x \quad \text{where } A = \begin{bmatrix} 4 & -1 & \\ -1 & 3 & 1 \\ & 1 & 2 \end{bmatrix}$$

$$2Ax + a = 0$$

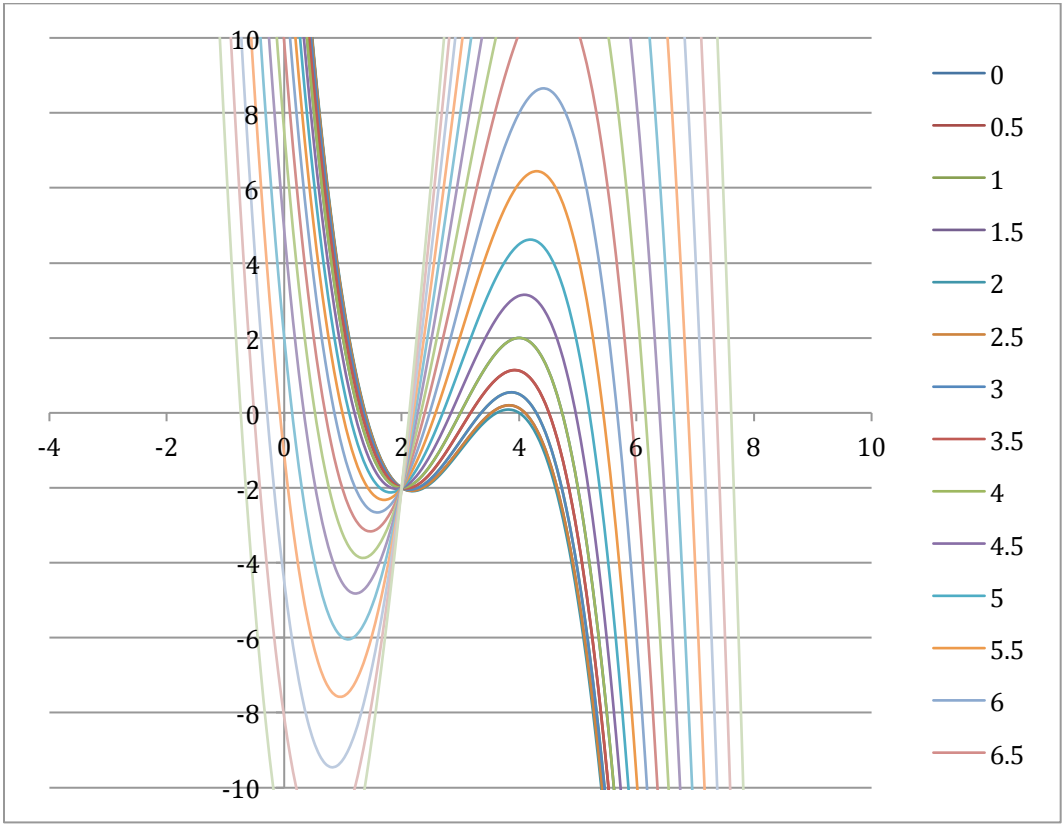
$x^* = \frac{1}{2} A^{-1} a$ gives the stationary points ($\nabla f(x^*) = 0$)

Eigenvalues of A determine whether $\nabla^2 f(x^*) = 2A$ is positive semi-definite or not.

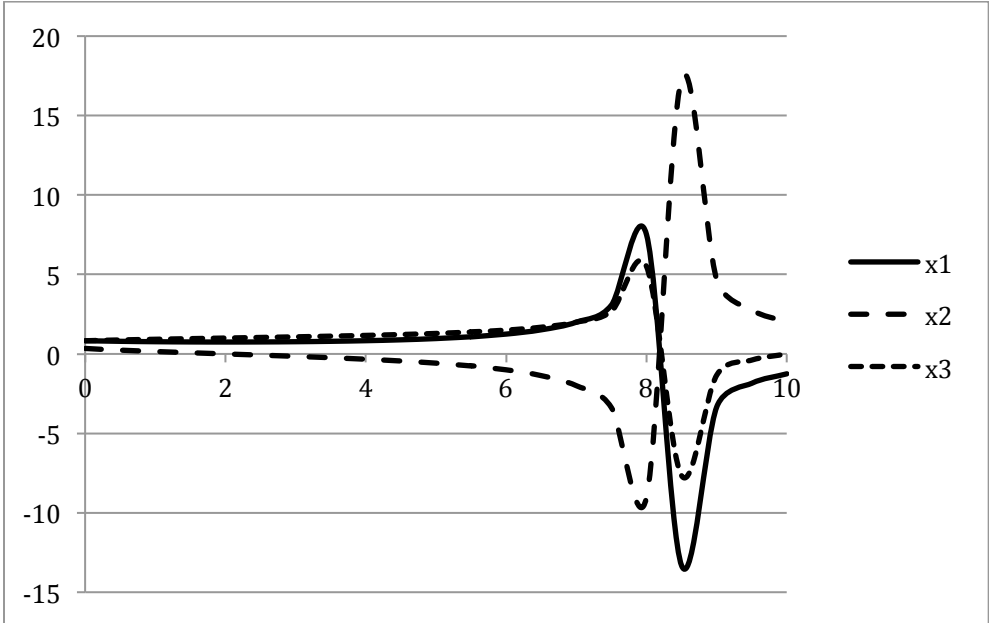
The attached curves show how x_1^*, x_2^*, x_3^* change with $M > 0$.

For $M > 8$, A is no longer positive definite and x^* is a saddle point. This is shown in the attached chart.

Roots of: $\det(A - \lambda I) = (4-\lambda)(3-\lambda)(2-\lambda) = (4-\lambda)(2-\lambda)(M/2 - 1)^2 = 0$
 for $M \in [0, 10]$.



Solution of $a^T x + x^T A x = 0$ as a function of M.



3. Powell damping:

$$\bar{y} = \theta y + (1-\theta) B^k s$$

so that $\bar{y}^T s \geq 0.2 s^T B^k s$ (new)

Using

$$B^{k+1} = B^k - \frac{B^k s s^T B^k}{s^T B^k s} + \frac{y y^T}{s^T y}$$

$$\begin{aligned} s^T B^{k+1} s &= s^T B^k s - \frac{(s^T B^k s)(s^T B^k s)}{s^T B^k s} \\ &\quad + \frac{s^T y y^T s}{s^T y} = s^T y \\ &= \theta (s^T y - s^T B^k s) + s^T B^k s \end{aligned}$$

$$\begin{aligned} \text{Now } \bar{y}^T s &= \theta (s^T y - s^T B^k s) + s^T B^k s \\ &\geq 0.2 s^T B^k s \end{aligned}$$

$$\text{so } \theta \geq \frac{0.8 s^T B^k s}{s^T B^k s - s^T y}$$

$$\text{and } s^T B^{k+1} s \geq 0.2 s^T B^k s$$

and B^{k+1} is p.d. along s .

a)

Max θ

$$\begin{aligned} \text{s.t. } \theta (y^T s - s^T B^k s) &\geq -0.8 s^T B^k s \\ 0 \leq \theta &\leq 1 \end{aligned}$$

$$-\theta - (y^T s - s^T B^k s) u + u_1 = 0$$

$$0 \leq \theta \perp u_1 \geq 0$$

$$0 \leq u \perp \theta (y^T s - s^T B^k s) + 0.8 s^T B^k s \geq 0$$

b) $\forall y^T s \geq 0.2 s^T B s$ then
 $\theta(y^T s - s^T B s) > -0.8 s^T B s$
 for any $\theta \in [0, 1]$ and $u = 0$
 with $\theta = 1, u_1 = 1$.

$\forall y^T s < 0.2 s^T B s$ then
 $\theta = \frac{0.8 s^T B s}{s^T B s - s^T y}, u_1 = 0$
 if $\theta > 0$.
 and $\bar{y}^T s = 0.2 s^T B s$

c) $\forall s^T y \rightarrow -\infty$ then $\theta = \frac{0.8 s^T B s}{s^T B s - s^T y} \rightarrow 0$

for $\theta = 0$ we have

$$s^T \bar{y} = s^T B s, \bar{y} = B s$$

$$B^{k+1} = B^k - \frac{B^k s s^T B^k}{s^T B^k} + \frac{\bar{y} \bar{y}^T}{s^T \bar{y}}$$

$$= B^k$$

and no update is taken

4. B^k is positive definite

$$p = -(B^k)^{-1} \nabla f(x^k)$$

$$\begin{aligned} \cos \theta^k &= \frac{-p^T \nabla f(x^k)}{\|p\| \|\nabla f(x^k)\|} = \frac{\nabla f(x^k)^T (B^k)^{-1} \nabla f(x^k)}{\|(B^k)^{-1} \nabla f(x^k)\| \|\nabla f(x^k)\|} \\ &\geq \frac{\nabla f(x^k)^T (B^k)^{-1} \nabla f(x^k)}{\|\nabla f(x^k)\|^2 \|(B^k)^{-1}\|} \quad (1) \end{aligned}$$

also with $(B^k)^{-1} = (L^T)^{-1} L^{-1}$

$$\begin{aligned} \nabla f(x^k)^T B^k \nabla f(x^k) &= \|\nabla f(x^k) L^{-T}\|^2 \\ \text{and } \|\nabla f(x^k)\|^2 &= \|\nabla f(x^k) L^{-T} B^k L^{-1} \nabla f(x^k)\| \\ &\leq \|\nabla f(x^k) L^{-T}\|^2 \|B^k\|^2 \end{aligned}$$

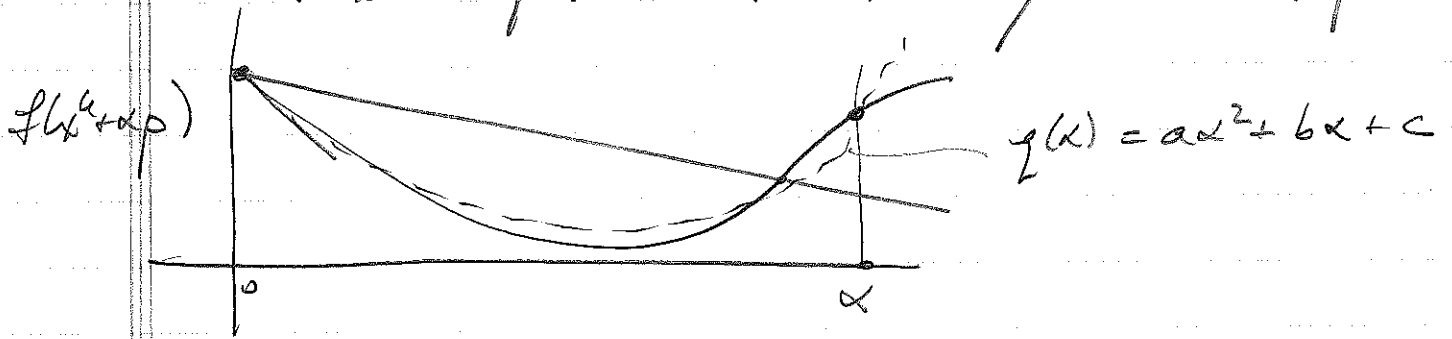
$$\text{so } \|\nabla f(x^k) L^{-T}\|^2 \geq \frac{\|\nabla f(x^k)\|^2}{\|B^k\|} \quad (2)$$

Substituting (2) into (1) leads to:

$$\begin{aligned} \cos \theta^k &\geq \frac{\|\nabla f(x^k)\|^2}{\|B^k\| \|\nabla f(x^k)\|^2 \|(B^k)^{-1}\|} \\ &= \frac{1}{\|B^k\| \|(B^k)^{-1}\|} = \frac{1}{\kappa(B^k)} \end{aligned}$$

5. Armijo line search

$$f(x^k + \alpha p) \leq f(x^k) + \gamma \alpha \nabla f(x^k)^T p$$



$$f(x^k) = q(0) = c$$

$$f(x^k + \alpha p) = q(\alpha) = a\alpha^2 + b\alpha + c$$

$$\nabla f(x^k)^T p = q'(0) = b$$

$$f(x^k + \alpha p) = a\alpha^2 + (\nabla f(x^k)^T p)\alpha + f(x^k)$$

$$\rightarrow (f(x^k + \alpha p) - f(x^k) - (\nabla f(x^k)^T p)\alpha) / \alpha^2 = a$$

quadratic min is $q'(\bar{\alpha}) = 0 = 2a\bar{\alpha} + b$

$$\bar{\alpha} = -b / 2a = \tau \alpha$$

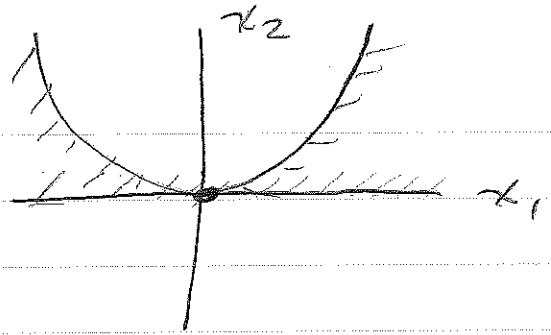
$$= \frac{\nabla f(x^k)^T p \alpha^2}{2(f(x^k + \alpha p) - f(x^k) - \alpha \nabla f(x^k)^T p)}$$

and

$$\tau = \frac{\nabla f(x^k)^T p \alpha}{2(f(x^k + \alpha p) - f(x^k) - \alpha \nabla f(x^k)^T p)}$$

6.

$$\begin{aligned} \text{Min } & x_1 \\ \text{s.t. } & x_2 \geq 0 \\ & x_2 \geq x_1^2 \end{aligned}$$



$$L = x_1 - u_1 x_2 - u_2 (x_2 - x_1^2)$$

KKT conditions

$$1 + 2x_1 u_2 = 0$$

$$x^* = 0$$

$$-u_1 - u_2 = 0$$

so KKT conditions fail.

$$A^\top = \begin{bmatrix} 0 & -1 \\ 2x_1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & -1 \end{bmatrix} \text{ @ } x^* \text{ and LICQ fails.}$$

7.

$$\text{Min } f(x) \text{ s.t. } g(x) \leq 0 \quad (1)$$

(convex problem)

$$\text{Min } f(x) \text{ s.t. } g(x) + s = 0, s \geq 0 \quad (2)$$

KKT conditions of (1)

$$\nabla f(x) + \nabla g(x) u = 0$$

$$0 \leq u \perp g(x) \leq 0$$

KKT conditions of (2)

$$\nabla f(x) + \nabla g(x) u = 0$$

$$u - \sigma = 0$$

$$0 \leq \sigma \perp s \geq 0$$

since $\sigma = u$, $s = -g(x)$ at solution and KKT conditions of (1) + (2) are identical since every solution of (2) is a solution of (1) & (1) has only global solutions, then (2) has only global solutions.