M 507 — Mathematical Optimization — Fall 2021

Elliott Pryor

Homework 3 — due October 5, on gradescope

1 MATLAB (completed in class — do not turn in)

Exercise 8.25 (slightly modified) Write a simple MATLAB program for implementing the steepest descent algorithm using built-in `fminsearch` for the line search. For the stopping criterion, use the condition $\|\mathbf{g}^k\| \leq \varepsilon$, where $\varepsilon = 10^{-6}$. Test your program using Example 8.1:

$$f(x_1,x_2,x_3) = (x_1-4)^4 + (x_2-3)^2 + 4(x_3+5)^4,$$

with initial point $\mathbf{x}^0 = \begin{bmatrix} 4 & 2 & -1 \end{bmatrix}^\top$. Try also starting from $\mathbf{x}^0 = \begin{bmatrix} -4 & 5 & 1 \end{bmatrix}^\top$, and compare the number of iterations required to satisfy the stopping criterion. Both times, evaluate the objective function f at the final point $\mathbf{x}^k$ to see how close it is to 0.

Exercise 9.4 (strongly modified)

1. Implement Newton's method in MATLAB, using anonymous functions for f, its gradient, and its Hessian.
2. Consider *Rosenbrock's Function*: $f(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1-x_1)^2$, where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top$.
   (a) Prove that $\begin{bmatrix} 1 & 1 \end{bmatrix}^\top$ is the unique global minimizer of f over $\mathbb{R}^2$.
   (b) With a starting point of $\mathbf{x}^0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$, apply Newton's method (numerically).

2 Paper

Exercise 8.1

1. Perform two iterations of steepest descent (including line search for α_k) to minimize the function

$$f(x_1,x_2) = x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_1^2 + x_2^2 + 3$$

with the starting point $\mathbf{x}^0 = 0$. (by hand) $\nabla f = \begin{matrix} 1 + x_1 \\ \frac{1}{2} + 2x_2 \end{matrix}$

$x^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$g^0 = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$

$\alpha_0 = \underset{\alpha}{\operatorname{argmin}} \left(f(x^0 - \alpha g^0) \right)$

$= \operatorname{argmin} \left(-\alpha - \frac{1}{4}\alpha + \frac{1}{2}\alpha^2 + \alpha^2 + 3 \right)$

$= \operatorname{argmin} \left(\frac{3}{2}\alpha^2 - \frac{5}{4}\alpha + 3 \right)$

Newton's, Quadratic => 1step

$\alpha_0 = 0 \quad g' = 3\alpha - \frac{5}{4}$
$\quad\quad\quad\;\; g'' = 3$

$\alpha_1 = 0 + \frac{5}{4} * \frac{1}{3} = \frac{5}{12}$

$\alpha_0 = \frac{5}{12}$

$X_1 = \begin{matrix} 0 \\ 0 \end{matrix} - \frac{5}{12} \begin{pmatrix} 1 \\ 1/2 \end{pmatrix} = \begin{pmatrix} -5/12 \\ -5/24 \end{pmatrix}$

$X_1 = \begin{pmatrix} -5/12 \\ -5/24 \end{pmatrix}$

$g^1 = \begin{pmatrix} 7/12 \\ 1/12 \end{pmatrix}$

$\alpha_1 = \operatorname{argmin} \left(f(x^1 - \alpha g^1) \right)$

$= \operatorname{argmin} \left((-5/12 - \alpha\, 7/12) + \frac{1}{2}(-5/24 - \alpha \frac{2}{24}) + \frac{1}{2}(-5/12 - \alpha\, 7/12)^2 + 3 \right)$

$= \operatorname{argmin} \left(-\frac{20}{48} - \alpha\frac{14}{24} - \frac{5}{48} - \alpha\frac{1}{24} + \frac{1}{2}\left(\alpha^2 \frac{49}{144} + 2\alpha\frac{35}{144} + \frac{25}{144}\right) + 3 \right)$

$= \operatorname{argmin} \left(-\alpha\frac{15}{24} + \frac{119}{48} + \alpha^2\frac{49}{288} + \alpha\frac{35}{144} + \frac{25}{288} \right)$

$= \operatorname{argmin} \left(\alpha^2 \frac{49}{288} - \alpha\frac{55}{144} + ? \right)$

Newton

$\alpha_0 = 0 \quad g' = \frac{49}{144}\alpha - \frac{55}{144}$
$\quad\quad\quad\;\; g'' = \frac{49}{144}$

$\alpha_1 = 0 + \frac{55}{\cancel{144}} \cdot \frac{\cancel{144}}{49} = \frac{55}{49}$

$X_2 = \begin{pmatrix} -5/12 \\ -5/24 \end{pmatrix} - \frac{55}{49}\begin{pmatrix} 7/12 \\ 1/12 \end{pmatrix} = \begin{pmatrix} -15/14 \\ -355/1176 \end{pmatrix}$

2. Determine an optimal solution (minimizer of f), analytically.

$\nabla f = \begin{matrix} 1+x_1 \\ \frac{1}{2}+2x_2 \end{matrix}$

$F = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \succ 0$ always

$0 = \begin{matrix} 1+x_1 \\ \frac{1}{2}+x_2 \end{matrix}$ $\quad x^* = \begin{matrix} -1 \\ -\frac{1}{2} \end{matrix}$

Exercise 8.8 Consider the function

$$f(\mathbf{x}) = 3(x_1^2 + x_2^2) + 4x_1x_2 + 5x_1 + 6x_2 + 7,$$

where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top \in \mathbb{R}^2$. Suppose that we use a fixed-step-size gradient algorithm to find the minimizer of f:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha \nabla f(\mathbf{x}^k).$$

$Q = \begin{matrix} a & b \\ c & d \end{matrix}$

Find the largest range of values of α for which the algorithm is globally convergent.

$x^T(Qx) = (x_1 \; x_2)\begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix}$

$= ax_1^2 + bx_1x_2 + cx_1x_2 + dx_2^2$

$a, d = 3$, $b, c = 2$

Converges iff $0 < \alpha < \frac{2}{\lambda_{max}}$

$\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$ $\quad (3-\lambda)^2 - 4$

$\lambda^2 - 6\lambda + 9 - 4 = 0$

$\lambda^2 - 6\lambda + 5 = 0$

$(\lambda - 5)(\lambda - 1)$

$\lambda = 5, 1$

$\boxed{0 < \alpha < \frac{2}{5}}$

Exercise 8.16 Consider the linear least squares optimization

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2,$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m \geq n$, and $\mathbf{b} \in \mathbb{R}^m$.

1. Write the objective function for this problem as a quadratic function, and write down its gradient and Hessian.

from M441 if x satisfies $A^TAx = A^Tb$ then it is least squares soln

$Q = A^TA \qquad d = A^Tb$

$f = \frac{1}{2}x^TQx - d^Tx = \frac{1}{2}x^TA^TAx - A^Tbx$

from notes

$\nabla f = (x^TQ - d^T)^T = (x^TA^TA - A^Tb)^T$

$F = Q = A^TA$

2. Write down the fixed-step-size gradient algorithm for solving this optimization problem.

$$x^{k+1} = x^k - \alpha\left(x^T A^T A - A^T b\right)^T$$

Can we solve for α or bound it at all?

3. Suppose that

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Find the largest range of values for α such that the algorithm above converges to the solution of the problem.

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \qquad \lambda_{max} = 4$$

$$0 < \alpha < \frac{1}{2}$$

Exercise 8.18 Let $f: \mathbb{R}^n \to \mathbb{R}$ be given by $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x} - \mathbf{x}^\top\mathbf{b}$, where $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{Q}$ is a real, symmetric, positive definite $n \times n$ matrix. Suppose that we apply the steepest descent method to this function, with $\mathbf{x}^0 \neq \mathbf{Q}^{-1}\mathbf{b}$. Show that the method converges to $\mathbf{x}^1 = \mathbf{Q}^{-1}\mathbf{b} = \mathbf{x}^*$ in one step if $\mathbf{x}^0$ is chosen such that $\mathbf{g}^0 = \nabla f(\mathbf{x}^0) = \mathbf{Q}\mathbf{x}^0 - \mathbf{b}$ is an eigenvector of $\mathbf{Q}$.

if g^0 is eigenvector of Q then $Qg^0 = \lambda g^0$

Thus our α becomes $\frac{g^{0T} g^0}{g^{0T}(Qg^0)} = \frac{g^0 \cdot g^0}{g^0 \cdot \lambda g^0} = \frac{1}{\lambda}$

then $x^1 = x^0 - \frac{1}{\lambda} g^0$

FONC

$$Q(x^0 - \tfrac{1}{\lambda} g^0) - b \overset{?}{=} 0$$

$$= Qx^0 - \tfrac{1}{\lambda} Qg^0 - b$$

$$= Qx^0 - \tfrac{1}{\cancel{\lambda}}\cancel{\lambda} g^0 - b$$

$$= Qx^0 - g^0 - b$$

$$Qx^0 - g^0 - b \overset{?}{=} 0$$

$$Qx^0 - (Qx^0 - b) - b \overset{?}{=} 0$$

$$Qx^0 - Qx^0 + b - b \overset{✓}{=} 0$$

x^1 satisfies FONC. There is only one point satisfying FONC with this $f(x)$. Therefore $x^1 = x^* = Q^{-1}b$