Nonlinear Optimization	Homework 1
MATH 408 Spring 2019	

Reading in the supplementary text (Nonlinear Optimization): Chapter 1, Chapter 2: page 9 to top of page 10, section 2.2, Sections 3-5,8; Chapter 4: Sections 1-4.

Exercises: Do the following exercises, justifying all steps.

1. Consider the system

$$\begin{array}{rcrcrcl} 4x_1 & & & - & x_3 & = & 200 \\ 9x_1 & + & x_2 & - & x_3 & = & 200 \\ 7x_1 & - & x_2 & + & 2x_3 & = & 200. \end{array}$$

   (a) (1 point) Write the augmented matrix corresponding to this system.

$$\begin{pmatrix} 4 & 0 & -1 & 200 \\ 9 & 1 & -1 & 200 \\ 7 & -1 & 2 & 200 \end{pmatrix}$$

   (b) (3 point) Reduce the augmented system in part (a) to echelon form.

$$\begin{pmatrix} 1 & 0 & 0 & 30 \\ 0 & 1 & 0 & -150 \\ 0 & 0 & 1 & -80 \end{pmatrix}$$

   (c) (2 point) Describe the set of solutions to the given system.
   The unique solution is $(x_1, x_2, x_3) = (30, -150, -80)$.

2. (6 points) Solve the following system of linear equations

$$\begin{array}{rcrcrcl} x_1 & + & 2x_2 & & & = & 1 \\ -x_1 & - & 4x_2 & + & x_3 & = & 2 \\ & & 2x_2 & + & x_3 & = & 0. \end{array}$$

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ -1 & -4 & 1 & 2 \\ 0 & 2 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 5/2 \\ 0 & 1 & 0 & -3/4 \\ 0 & 0 & 1 & 3/2 \end{pmatrix}$$

so the unique solution is $(x_1, x_2, x_3) = (5/2, -3/4, 3/2)$.

3. (3 points) Represent the linear span of the four vectors as the range space of some matrix:

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 4 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 5 \end{bmatrix}, \quad x_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -2 \end{bmatrix}, \quad x_4 = \begin{bmatrix} 7 \\ 2 \\ 1 \\ 1 \end{bmatrix}.$$

$$\text{Span}\{x_1, x_2, x_3, x_4\} = \mathit{Range}(A) \text{ where } A = \begin{pmatrix} 1 & 1 & -1 & 7 \\ 2 & 2 & 1 & 2 \\ 2 & 2 & 1 & 1 \\ 4 & 5 & -2 & 1 \end{pmatrix}$$

4. (3 points) Compute a basis for $Nul(A^T)^{\perp}$ where A is the matrix

$$A = \begin{bmatrix} 1 & 1 & -1 & 7 \\ 2 & 2 & 1 & 2 \\ 2 & 2 & 1 & 1 \\ 4 & 5 & -2 & 1 \end{bmatrix}.$$

By the Fundamental Theorem of the Alternative, $Nul(A^T)^{\perp} = \mathit{Range}(A)$. A straightforward computation confirms the columns of A are linearly independent, so $Nul(A^T)^{\perp} = \mathit{Range}(A) = \mathbb{R}^4$.

5. Consider the matrix

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

(a) (3 points) Find the eigenvalues of the matrix A. Is any eigenvalue repeated?

$$\det(A - \lambda I) = -\lambda^3 + 6\lambda^2 - 9\lambda + 4 = -(\lambda - 1)^2(\lambda - 4)$$

so eigenvalues of A are $\lambda_1 = 4$, $\lambda_2 = \lambda_3 = 1$.

(b) (4 points) Find three eigenvectors u_1, u_2, u_3 of A that are orthonormal.

Find eigenvectors u_i by determining $Nul(A - \lambda_i I)$ and use Gram-Schmidt to orthonormalize. There are many possibilities for u_2, u_3 below.

$$u_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \quad u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad u_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -1 \\ -2 \end{pmatrix}$$

(c) (1 point) State a spectral (eigenvalue) decomposition of A.

$$A = U\Lambda U^T \text{ where } U = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} \text{ from above and } \Lambda = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

6. Recall that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is said to be differentiable at a point $x \in \mathbb{R}^n$ if there is a vector $g \in \mathbb{R}^n$ such that

$$f(y) = f(x) + g^T(y - x) + o(\|y - x\|) .$$

The vector g is called the gradient of f at x and is denoted $g = \nabla f(x)$. Note that, when defined, the relation $x \mapsto \nabla f(x)$ is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$, i.e. $\nabla f\colon \mathbb{R}^n \to \mathbb{R}^n$. We say that f is continuously differentiable at $x \in \mathbb{R}^n$ if the mapping ∇f is continuous at x. When f is continuously differentiable at $x \in \mathbb{R}^n$, then $\nabla f(x)$ is easily computed as the vector of partial derivatives of f at x, i.e.

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

Compute the gradient of the following functions.

(a) (2 points) $f(x) = x_1^3 + x_2^3 - 3x_1 - 15x_2 + 25$ on $\mathbb{R}^2$

$$\nabla f(x) = \begin{pmatrix} 3x_1^2 - 3 \\ 3x_2^2 - 15 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} 6x_1 & 0 \\ 0 & 6x_2 \end{pmatrix}$$

(b) (2 points) $f(x) = x_1^2 + x_2^2 - \sin(x_1 x_2)$ on $\mathbb{R}^2$

$$\nabla f(x) = \begin{pmatrix} 2x_1 - x_2 \cos(x_1 x_2) \\ 2x_2 - x_1 \cos(x_1 x_2) \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} 2 + x_2^2 \sin(x_1 x_2) & -\cos(x_1 x_2) + x_1 x_2 \sin(x_1 x_2) \\ -\cos(x_1 x_2) + x_1 x_2 \sin(x_1 x_2) & 2 + x_1^2 \sin(x_1 x_2) \end{pmatrix}$$

(c) (2 points) $f(x) = \|x\|^2 = \sum_{j=1}^n x_j^2$ on $\mathbb{R}^n$

For each of the below, the gradient and Hessian is specified component-wise.

$$(\nabla f(x))_i = 2x_i$$

$$(\nabla^2 f(x))_{ij} = \begin{cases} 2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

(d) (2 points) $f(x) = e^{\|x\|^2}$ on $\mathbb{R}^n$

$$(\nabla f(x))_i = 2x_i e^{\|x\|^2}$$

$$(\nabla^2 f(x))_{ij} = \begin{cases} 2e^{\|x\|^2} + 4x_i^2 e^{\|x\|^2} & \text{if } i = j \\ 4x_i x_j e^{\|x\|^2} & \text{if } i \neq j \end{cases}$$

(e) (2 points) $f(x) = x_1 x_2 x_3 \cdots x_n$ on $\mathbb{R}^n$

$$(\nabla f(x))_i = \prod_{j \neq i} x_j$$

$$(\nabla^2 f(x))_{ij} = \begin{cases} 0 & \text{if } i = j \\ \prod_{k \neq i,j} x_k & \text{if } i \neq j \end{cases}$$

(f) (2 points) $f(x) = -\log(x_1 x_2 x_3 \cdots x_n)$ on the set $\{x \in \mathbb{R}^n : x_i > 0 \text{ for all } i = 1, \ldots, n\}$

$$(\nabla f(x))_i = -\frac{\prod_{j \neq i} x_j}{\prod_{i=1}^{n} x_j} = -\frac{1}{x_i}$$

$$(\nabla^2 f(x))_{ij} = \begin{cases} \frac{1}{x_i^2} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

7. (12 points) Let $\mathbb{R}^{n \times n}$ denote the set of real $n \times n$ square matrices. A function $f \colon \mathbb{R}^n \to \mathbb{R}$ is said to be twice differentiable at a point $x \in \mathbb{R}^n$ if is differentiable at x and there is a matrix $H \in \mathbb{R}^{n \times n}$ such that

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \tfrac{1}{2}(y - x)^T H (y - x) + o(\|y - x\|^2).$$

The matrix H is called the Hessian of f at x and is denoted $\nabla^2 f(x)$. Note that, when defined, the relation $x \mapsto \nabla^2 f(x)$ is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^{n \times n}$, i.e. $\nabla^2 f \colon \mathbb{R}^n \to \mathbb{R}^{n \times n}$. We say that f is twice continuously differentiable at $x \in \mathbb{R}^n$ if the mapping $\nabla^2 f$ is continuous at x. It can be shown that if f is twice continuously differentiable at a point $x \in \mathbb{R}^n$, then the matrix $\nabla^2 f(x)$ is symmetric, i.e. $\nabla^2 f(x) = \nabla^2 f(x)^T$, in which case $\nabla^2 f(x)$ is the matrix of second partial derivatives of f at x:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{bmatrix}.$$

Compute the Hessian of the functions given in problem 6 above.

See above.