Math 171B: Numerical Optimization: Nonlinear Problems

Instructor: Michael Holst

Spring Quarter 2015

Solutions for Homework Assignment #1

Exercise 1.1. *If x is an eigenvector of A, show that βx is also an eigenvector for any $\beta \neq 0$. What is the associated eigenvalue? Use this result to show that the unit vector $x/\|x\|$ formed from an eigenvector x is also an eigenvector of A corresponding to the same eigenvalue as that of x.*

If $Ax = \lambda x$, then clearly $A(\beta x) = \beta(Ax) = \beta(\lambda x) = \lambda(\beta x)$ holds, simply by the properties of scalar-vector and scalar-matrix multiplication, and by the definition of an eigenpair. Therefore, if x is an eigenvector then so is βx, and both have the same eigenvalue. Moreover, if x is an eigenvector, we know that $x \neq 0$, so that $\|x\| \neq 0$ (property of the norm), so that taking $\beta = 1/\|x\|$ is well-defined. Therefore, if x is an eigenvector, so is $x/\|x\|$, for the same eigevalue.

Exercise 1.2. *Let $(x, y) : V \mapsto \mathbb{R}$ be an inner-product on a vector space V with associated scalar field $\mathbb{R}$. We know that (x, y) must satisfy the three properties of an inner-product:*

1. $(x, x) \geq 0, \quad \forall x \in V, \quad (x, x) = 0 \text{iff} x = 0.$
2. $(x, y) = (y, x), \quad \forall x, y \in V.$
3. $(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z), \quad \forall \alpha, \beta \in \mathbb{R}, \quad \forall x, y, z \in V.$

Use these three properties to show that the induced norm $\|x\| = (x, x)^{1/2}$ satisfies the three properties of a norm:

1. $\|\alpha x\| = |\alpha| \; \|x\|, \quad \forall \alpha \in \mathbb{R}, \quad \forall x \in V.$
2. $\|x\| \geq 0, \quad \forall x \in V, \quad \|x\| = 0 \text{iff} x = 0.$
3. $\|x + y\| \leq \|x\| + \|y\|, \quad \forall x, y \in V.$

Hint: Showing the first two properties is very easy; to show the last property (triangle inequality), assume the Cauchy-Schwarz inequality holds: $|(x, y)| \leq \|x\| \; \|y\|$.

Property 1 By properties 2 and 3 of the inner-product we have $\|\alpha x\| = (\alpha x, \alpha x)^{1/2} = \left\{\alpha^2 (x, x)\right\}^{1/2} = |\alpha| \; \|x\|$.

Property 2 By property 1 of the inner-product $\|x\| = (x, x)^{1/2} \geq 0 \;\; \forall x \in V$, and $\|x\| = (x, x)^{1/2} = 0 \text{iff} x = 0$.

Property 3 By properties 2 and 3 of the inner-product, and by the Cauchy-Schwarz inequality, it holds that:

$$\begin{aligned}
\|x + y\|^2 &= (x + y, x + y) \\
&= (x, x) + 2(x, y) + (y, y) \\
&\leq (x, x) + 2|(x, y)| + (y, y) \\
&\leq (x, x) + 2 \; \|x\| \; \|y\| + (y, y) \\
&= \|x\|^2 + 2 \; \|x\| \; \|y\| + \|y\|^2 \\
&= (\|x\| + \|y\|)^2,
\end{aligned}$$

which gives $\|x + y\| \leq \|x\| + \|y\|$.

Exercise 1.3. *Let $F(x)$ denote a twice-differentiable function of one variable. Assuming only the mean-value theorem of integral calculus: $F(b) = F(a) + \int_a^b F'(t)\, dt$, derive the following variants of the Taylor-series expansion with integral remainder:*

(a) $F(x + h) = F(x) + \int_x^{x+h} F'(t)\, dt.$

This follows from the mean-value theorem on substituting $a = x$ and $b = x + h$.

(b) $F(x+h) = F(x) + h\int_0^1 F'(x+\xi h)\,d\xi$.

The idea is to transform the variable of integration in part (a). Consider the new variable ξ such that $t = x + h\xi$. As t varies between the limits of integration x and $x+h$, we find that ξ varies between 0 and 1. Substituting for t in part (a) and using $dt = h\,d\xi$ gives the result.

(c) $F(x+h) = F(x) + hF'(x) + h\int_0^1 [F'(x+\xi h) - F'(x)]\,d\xi$.

From part (b) we have

$$\begin{aligned} F(x+h) &= F(x) + h\int_0^1 F'(x+\xi h)\,d\xi \\ &= F(x) + hF'(x) - hF'(x) + h\int_0^1 F'(x+\xi h)\,d\xi. \end{aligned}$$

Since $hF'(x)$ is independent of ξ, the variable of integration, we may take it under the integral sign to give

$$F(x+h) = F(x) + hF'(x) + h\int_0^1 [F'(x+\xi h) - F'(x)]\,d\xi$$

as required.

(d) $F(x+h) = F(x) + hF'(x) + h^2\int_0^1 F''(x+\xi h)(1-\xi)\,d\xi$. *(Hint: Try expanding $F'(x+h)$ using a formula analogous to part (b) and differentiate with respect to h using the chain rule.)*

Since $F'(x)$ is differentiable, we can apply the formula of part (b) to the function F', giving

$$F'(x+h) = F'(x) + h\int_0^1 F''(x+\xi h)\,d\xi. \tag{1.1}$$

Another expression for $F'(x+h)$ can be found by considering the result of part (b):

$$F(x+h) = F(x) + h\int_0^1 F'(x+\xi h)\,d\xi. \tag{1.2}$$

For fixed x we can differentiate $F(x+h)$ with respect to h. Applying the chain-rule for differentiation gives

$$\frac{dF(x+h)}{dh} = F'(x+h) = \int_0^1 F'(x+\xi h)\,d\xi + h\int_0^1 F''(x+\xi h)\xi\,d\xi. \tag{1.3}$$

Combining (1.1), (1.2) and (1.3), so that the terms $F'(x+h)$ and $\int_0^1 F'(x+\xi h)\,d\xi$ are eliminated, we obtain the required identity. In particular, taking $2\times(1.2) - h\times(1.3) + h\times(1.1)$ gives the result.

Exercise 1.4. *Find the gradient vector $F(x) = \nabla f(x)$ of the following functions, and then find the Jacobian matrix of $F(x)$. (The Jacobian matrix of $F(x) = \nabla f(x)$ is the same as the Hessian matrix $\nabla^2 f(x)$ of $f(x)$).*

(a) $f(x) = 2\big(x_2 - x_1^2\big)^2 + \big(x_1 - 3\big)^2$.

First we form the row vector $f'(x)$ by forming the first partials with respect to each variable. This gives

$$f'(x) = \begin{pmatrix} -8x_1(x_2 - x_1^2) + 2(x_1 - 3) & 4(x_2 - x_1^2) \end{pmatrix}.$$

The gradient of f is just the transpose of this vector, giving

$$g(x) = (f'(x))^T = \begin{pmatrix} -8x_1(x_2 - x_1^2) + 2(x_1 - 3) \\ 4(x_2 - x_1^2) \end{pmatrix}.$$

The Hessian matrix is defined as the derivative of the gradient g, so that

$$H(x) = \begin{pmatrix} -8(x_2 - x_1^2) + 16x_1^2 + 2 & -8x_1 \\ -8x_1 & 4 \end{pmatrix}.$$

Since the Hessian is always a *symmetric matrix* (if the second partials are continuous), you can use this as a check on your work: if you work out the off-diagonal partial derivatives separately, they should come out to be the same value.

(b) $f(x) = (2x_1 + x_2)^2 + 4(x_2 - x_3)^4$.

Forming the gradient vector as in part (a), we get

$$g(x) = \begin{pmatrix} 8x_1 + 4x_2 \\ 4x_1 + 2x_2 + 16(x_2 - x_3)^3 \\ -16(x_2 - x_3)^3 \end{pmatrix}.$$

The Hessian matrix is just the "Jacobian" of this column vector. Plugging away at it gives the Hessian matrix $H(x)$ as

$$\begin{pmatrix} 8 & 4 & 0 \\ 4 & 2 + 48(x_2 - x_3)^2 & -48(x_2 - x_3)^2 \\ 0 & -48(x_2 - x_3)^2 & 48(x_2 - x_3)^2 \end{pmatrix}.$$

Exercise 1.5. *Find $f'(x)$, $\nabla f(x)$ and $\nabla^2 f(x)$ for the following functions of n variables.*

(a) $f(x) = \frac{1}{2}x^T Hx$, *where H is an $n \times n$ constant matrix.*

The trick is to and work out the derivatives without getting too buried in lots of indices. First, we write H in terms of its columns $\{h_j\}$,

$$H = \begin{pmatrix} h_1 & h_2 & \cdots & h_n \end{pmatrix}, \text{where} h_j = \begin{pmatrix} h_{1j} \\ h_{2j} \\ \vdots \\ h_{nj} \end{pmatrix}.$$

This partition of H allows us to write $f(x)$ as

$$f(x) = \tfrac{1}{2}x^T Hx = \tfrac{1}{2}x^T \left(\sum_{j=1}^{n} h_j x_j \right) = \tfrac{1}{2} \sum_{j=1}^{n} (x^T h_j) x_j.$$

Taking the partial derivative with respect to x_i, and using standard rules about differentiating products, we obtain

$$\frac{\partial f(x)}{\partial x_i} = \tfrac{1}{2} \sum_{j=1}^{n} \frac{\partial (x^T h_j)}{\partial x_i} x_j + \tfrac{1}{2} \sum_{j=1}^{n} (x^T h_j) \frac{\partial x_j}{\partial x_i} = \tfrac{1}{2} \sum_{j=1}^{n} \frac{\partial (x^T h_j)}{\partial x_i} x_j + \tfrac{1}{2} x^T h_i.$$

The first term under the summation can be written as

$$\frac{\partial (x^T h_j)}{\partial x_i} = \frac{\partial}{\partial x_i} \sum_{k=1}^{n} x_k h_{kj} = h_{ij}.$$

Since $x^T h_i = \sum_{j=1}^{n} h_{ji} x_j$, it follows that

$$\frac{\partial f(x)}{\partial x_i} = \tfrac{1}{2} \sum_{j=1}^{n} h_{ij} x_j + \tfrac{1}{2} \sum_{j=1}^{n} h_{ji} x_j,$$

and $\partial f(x)/\partial x_i$ is half the sum of the ith elements of Hx and $H^T x$, with

$$f'(x) = \tfrac{1}{2} x^T (H + H^T).$$

The gradient $\nabla f(x)$ is the transpose of $f'(x)$, giving

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

The Hessian matrix $\nabla^2 f(x)$ is defined as the first derivative (i.e., Jacobian) of the gradient vector. Hence

$$\nabla^2 f(x) = (\nabla f(x))' = \tfrac{1}{2}((H^T x)' + (Hx)').$$

Given any constant matrix A, the Jacobian of Ax is A, which implies that

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

Notice that the Hessian is a symmetric matrix, even though H is not symmetric.

(b) $f(x) = b^T Ax - \frac{1}{2}x^T A^T Ax$, *where A is an $m \times n$ constant matrix and b is a constant m-vector.*

If we use $H = A^T A$ in part (a) we have

$$f'(x) = b^T A - \tfrac{1}{2}x^T(A^T A + (A^T A)^T) = b^T A - x^T A^T A.$$

the last equality following from the fact that $A^T A$ is symmetric. Forming the transpose for $g(x)$ gives

$$\nabla f(x) = f'(x)^T = A^T b - A^T Ax = A^T(b - Ax).$$

Finally, we have from part (a)

$$\nabla^2 f(x) = (\nabla f(x))' = \tfrac{1}{2}(A^T A + (A^T A)^T) = A^T A.$$

(c) $f(x) = \|x\| = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$.

First we compute $\partial f/\partial x_j$:

$$\frac{\partial f(x)}{\partial x_j} = \tfrac{1}{2}\Big(\sum_{i=1}^n x_i^2\Big)^{-1/2} 2x_j = \frac{1}{\|x\|}x_j.$$

This implies that

$$f'(x) = \begin{pmatrix} \frac{1}{\|x\|}x_1 & \frac{1}{\|x\|}x_2 & \cdots & \frac{1}{\|x\|}x_n \end{pmatrix} = \frac{1}{\|x\|}x^T,$$

and $\nabla f(x) = x/\|x\|$.
Differentiating with respect to x_i for $i \neq j$, we obtain

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f(x)}{\partial x_j}\right) = \frac{\partial}{\partial x_i}\left(\frac{x_j}{\|x\|}\right) = -\frac{x_i x_j}{\|x\|^3}.$$

Similarly, for $i = j$, we have

$$\frac{\partial^2 f(x)}{\partial x_j^2} = \frac{\partial}{\partial x_j}\left(\frac{\partial f(x)}{\partial x_j}\right) = \frac{\partial}{\partial x_j}\left(\frac{x_j}{\|x\|}\right) = \frac{\|x\|^2 - x_j^2}{\|x\|^3}.$$

Therefore, the Hessian of $f(x)$ may be written:

$$\begin{aligned}
\nabla^2 f(x) &= \frac{1}{\|x\|^3}\begin{pmatrix} \|x\|^2 - x_1^2 & -x_1x_2 & \cdots & -x_1x_n \\ -x_2x_1 & \ddots & & \vdots \\ \vdots & & & \\ -x_nx_1 & \cdots & & \|x\|^2 - x_n^2 \end{pmatrix} \\
&= \frac{1}{\|x\|^3}(\|x\|^2 I - xx^T) = \frac{1}{\|x\|}\left(I - \hat{x}\hat{x}^T\right),
\end{aligned}$$

where $\hat{x}$ is the unit vector $x/\|x\|$.

Exercise 1.6.* *Create a* MATLAB *m-file of the form:*

```
function [F,J] = D(x)
  F = [ a ; b ];
  J = [ c d ; e f ];
```

where the expressions for a, b, c, d, e, f are chosen so that the function returns the 2×1-vector-valued function $F(x)$ and the 2×2 Jacobian matrix $J(x)$ for the function $F(x)$ from part (a) of Problem 1.4. Use this m-file to compute F and J at $x = (1,\, 0)^T$; and $x = (1,\, 1)^T$. Capture the output from the computation and turn it in with the homework.

See the TA for the solution to this problem.