

6.3

If x^* is a local minimizer of f over the feasible set (F), then we know by the definition local minimizers that there exists an open neighborhood $U \subset \mathcal{F}$ around x^* such that $f(x^*) \leq f(x) \forall x \in U$. Now, since $U \subset \mathcal{F} \subset \mathcal{F}'$, we have that $U \subset \mathcal{F}'$ and therefore, x^* is also a minimizer of f on \mathcal{F}'

For the counterexample, take $f(x) = x^{\frac{1}{3}}$. Let $\mathcal{F} = [0, 1]$. In this case, 0 is a local minimizer. However, let $\mathcal{F}' = [-1, 1]$, which satisfies $\mathcal{F} \subset \mathcal{F}'$, but 0 is not a local minimizer for \mathcal{F}' .

6.5

We conduct this proof by contradiction. Let there be some sequence of solutions $x_k \in \mathcal{F}$ and a particular solution $x \notin \mathcal{F}$ such that $x_k \rightarrow x$. Since G is continuous we have that

$$\begin{aligned} G(x_k) \in G(\mathcal{F}) &\implies G(x_k) \preceq b \\ G(x) \notin G(\mathcal{F}) &\implies G(x) \succ b \\ G(x_k) &\rightarrow G(x) \end{aligned}$$

Now let $\varepsilon = \frac{G(x)-b}{2}$. We have a contradiction because we can't choose a k large enough such that $G(x_k) - G(x) \preceq \varepsilon$. We have, then, that $G(x) \in G(\mathcal{F})$ and, therefore, that $G(\mathcal{F})$ is closed, which implies that \mathcal{F} is closed. We therefore know that the feasible set of any such constraint function is closed, and since the feasible set of the function is simply the intersection of all the feasible sets of the constraint function, and the intersection of a finite number of closed sets is closed, we know that \mathcal{F} is closed.

An example is x^{-1} .

6.6 (iii)

$$\text{Min}_{m,b,\delta} \quad \|\delta\|_2^2 + \|\mathbf{y} - m\delta + m\mathbf{x} + b\mathbf{e}\|_2^2$$

6.9

We know that the *MLVUE* is given by

$$\hat{\mathbf{x}} = (A^T Q^{-1} A)^{-1} A^T Q^{-1} \mathbf{b}$$

Since we know that Q is the variance-covariance matrix and that we have homoskedasticity and no serial correlation, we know that Q is simply a diagonal matrix with a constant variance σ^2 all along the diagonal. Substituting, then, we have that

$$\begin{aligned}
\hat{\mathbf{x}} &= (A^T \frac{1}{\sigma^2} A)^{-1} A^T \frac{1}{\sigma^2} \mathbf{b} \\
&= \sigma^2 (A^T A)^{-1} A^T \frac{1}{\sigma^2} \mathbf{b} \\
&= (A^T A)^{-1} A^T \mathbf{b}
\end{aligned}$$

6.10

Setting the general odds ratio equal to the student odds ratio, we have that

$$r_i = \frac{p}{1-p} \implies p = \frac{r_i}{1+r_i}$$

And we have the optimization problem

$$\text{Max}_{\beta_0, \beta_1} \prod_{i=1}^n \frac{r_i}{1+r_i}^{y_i} (1 - \frac{r_i}{1+r_i})^{1-y_i}$$

6.11

$$\begin{aligned}
&\text{Max}_{k,m} \quad .07m + .05k \\
&\text{s.t.} \quad 240000 \geq 4m + 3k \\
&\quad \quad 6000 \geq 2m + 1k
\end{aligned}$$

6.12

Let Σ , the variance-covariance matrix, \mathbf{w} , the weights, \mathbf{r} , the expected returns, and \bar{r} , the return to the portfolio, be expressed as follows:

$$\mathbf{w} = [w_1, w_2]^T \quad \mathbf{r} = [\bar{r}_1, \bar{r}_2]^T$$

The minimization problem is as follows:

$$\begin{aligned}
&\text{Min}_w \quad \mathbf{w}' \Sigma \mathbf{w} \\
&\text{s.t.} \quad \mathbf{w}' \mathbf{1} = 1 \\
&\quad \quad \mathbf{w}' \mathbf{r} = \bar{r}
\end{aligned}$$

We have the following first-order conditions:

$$2 \Sigma \mathbf{w} + \lambda_1 \mathbf{1} + \lambda_2 \mathbf{r} = 0$$

$$\mathbf{1}'\mathbf{w} - 1 = 0$$

$$\mathbf{w}'\mathbf{r} - \bar{r} = 0$$

Considering this equation in matrix form, we have

$$\begin{bmatrix} 2\sigma_1^2 & 2\sigma_{12} & 1 & \bar{r}_1 \\ 2\sigma_{21} & 2\sigma_2^2 & 1 & \bar{r}_2 \\ 1 & 1 & 0 & 0 \\ \bar{r}_1 & \bar{r}_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \bar{r} \end{bmatrix}$$

Since A , the first matrix in the above equation, is full-rank, there is a unique solution to the system of equations, which is the minimized solution. Since the variance $\mathbf{w}'\sum\mathbf{w}$ is quadratic in \mathbf{w} , and \mathbf{w} is linear in \bar{r} , the variance is quadratic in \bar{r} . The bottom half is useless because it has the same variance but lower expected return, so the upper half is preferred.

6.13 (i)

The optimization problem is as follows:

$$\max_{C_0} u(C_0) + \beta u(m - C_0)$$

We have that

$$\max_{C_0} \log(C_0) + \beta \log(m - C_0)$$

First order conditions

$$\implies \frac{1}{C_0} + \beta \frac{-1}{m - C_0} = 0$$

And we have that

$$\implies C_0 = \frac{m}{1 + \beta} \quad C_1 = \frac{\beta m}{1 + \beta}$$

6.13 (ii)

We have that

$$\max_{C_0} (1 - e^{-\alpha(C_0)}) + \beta(1 - e^{-\alpha(C_0 - m)})$$

First order conditions

$$\begin{aligned} \alpha e^{-\alpha C_0} - \alpha \beta e^{\alpha(C_0 - m)} &= 0 \\ \implies e^{-\alpha C_0} &= \beta e^{\alpha(C_0 - m)} \end{aligned}$$

upon which taking the log yields

$$\begin{aligned} -\alpha C_0 &= \ln \beta + \alpha C_0 - \alpha m \\ \implies C_0 &= \frac{\alpha m - \ln \beta}{2\alpha} \quad C_1 = \frac{\alpha m + \ln \beta}{2\alpha} \end{aligned}$$

7.3 (i)

Note that

$$\begin{aligned}
f(x) &= \frac{1}{2} \|Ax - b\|_2^2 \\
&= \frac{1}{2} (Ax - b)^T (Ax - b) \\
&= \frac{1}{2} (x^T A^T Ax - x^T A^T b - b^T Ax + b^T b)
\end{aligned}$$

$$\implies \frac{df}{dx} = \frac{1}{2} (2A^T Ax - 2A^T b) = 0$$

which yields

$$A^T Ax^* = A^T b$$

7.3 (ii)

Note that

$$\frac{df}{dx} = \frac{1}{2} (2A^T Ax - 2A^T b) = 0$$

we also know that

$$D^2 f(x) = A^T A$$

Now, consider an arbitrary $n \times 1$ vector $v \in \mathbb{R}_n$. We need that

$$v^T A^T A v \geq 0$$

Note, though, that

$$v^T A^T A v = \langle vA, vA \rangle \geq 0$$

and we have that

$$A^T A \geq 0$$

7.4 (i)

By a previous exercise, we know that if A has full rank, then $A^T A$ does as well, implying that its eigenvalues are all positive, which implies that $A^T A$ is positive definite.

7.4 (ii)

The full rank of $A^T A$ implies that it is invertible, and from 7.3 we know that

$$(A^T A)x^* = A^T b$$

Therefore the unique solution of this equation can be written as

$$x^* = (A^T A)^{-1} A^T b$$

7.7 (i)

Using $f'(x) = \frac{4}{3}x^{\frac{1}{3}}$ and $f''(x) = \frac{4}{9}x^{-\frac{2}{3}}$, we use Newton's formula to find that

$$x_{k+1} = x_k - \frac{\frac{4}{3}x_k^{\frac{1}{3}}}{\frac{4}{9}x_k^{-\frac{2}{3}}} = x_k - 3x_k = -2x_k$$

7.7 (ii)

We know that

$$x_{k+1} = -2x_k$$

so the sequence doubles in value and switches sign every iteration, which implies divergence. Generalizing we get

$$x_k = (-2)^k x_0$$

which is non-zero. Thm 7.2.2 doesn't apply as $f''(x)$ is not defined at 0.

7.9

```
from __future__ import division
import numpy as np

def newton(initialguess, fderiv, f2deriv, maxiter = 250, tolerance=1e-10):
    xmin = 0
    err = 100
    iterations = 0
    try:
        fderiv(initialguess)
    except ZeroDivisionError:
        string = "fderiv zero division"

        return string

    try:
        f2deriv(initialguess)
    except ZeroDivisionError:
        string = "f2deriv zero division"
        return string

    try:
        fderiv(0)/f2deriv(0)
```

```

except ZeroDivisionError:
    string = "Zero division"
    return string

epsilon = 1e-5

xold = initialguess

while err > tolerance and iterations < maxiter:
    xnew = xold - fderiv(xold)/f2deriv(xold)
    err = abs(xnew - xold)*abs(xold)*epsilon
    xold = xnew
    iterations = iterations + 1

if iterations == maxiter:
    string = "Doesn't converge"
    return string

xmin = xnew
return xmin

```

7.12

Note that

$$(A + BCD)[A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}] = I$$

Then we have that

$$\begin{aligned}
 & (A + BCD)[A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}] \\
 &= I + BCDA^{-1} - (B + BCDA^{-1}B)(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\
 &= I + BCDA^{-1}BC(C^{-1} + DA^{-1}B)(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\
 &= I + BCDA^{-1} - BCDA^{-1} \\
 &= I
 \end{aligned}$$

7.13

Let $\Delta x_n = x_n - x_{n-1}$ and $\Delta g_n = g(x_n) - g(x_{n-1})$. Then substituting in we can express A_n as follows:

$$A_n = A_{n-1} + \frac{\Delta g_n - A_{n-1}(\Delta x_n)}{\Delta x_n^T \Delta x_n} (\Delta x_n^T)$$

Remember that

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

Now let

$$A = A_{n-1} \quad B = \frac{\Delta g_n - A_{n-1}(\Delta x_n)}{\Delta x_n^T \Delta x_n} \quad C = 1 \quad D = \Delta x_n^T$$

Now substituting in and applying the inverse to A_n , we have that

$$\begin{aligned} A_n^{-1} &= A_{n-1}^{-1} - \frac{A_{n-1}^{-1} \frac{\Delta g_n - A_{n-1}(\Delta x_n)}{\Delta x_n^T \Delta x_n} \Delta x_n^T A_{n-1}^{-1}}{1 + \Delta x_n^T A_{n-1}^{-1} \frac{\Delta g_n - A_{n-1}(\Delta x_n)}{\Delta x_n^T \Delta x_n}} \\ &= A_n^{-1} - \frac{(A_{n-1}^{-1} \Delta g_n - \Delta x_n) \Delta x_n^T A_{n-1}^{-1}}{\Delta x_n^T \Delta x_n + \Delta x_n^T A_{n-1}^{-1} \Delta g_n - \Delta x_n^T \Delta x_n} \\ &= A_{n-1}^{-1} + \frac{((x_n - x_{n-1}) - A_{n-1}^{-1}(g(x_n) - g(x_{n-1}))) (x_n - x_{n-1})^T A_{n-1}^{-1}}{(x_n - x_{n-1})^T A_{n-1}^{-1} (g(x_n) - g(x_{n-1}))} \end{aligned}$$

7.14

Initializing with an initial point \mathbf{x}_0 and compute its derivative given by

$$A_0 = Dg(x_0) \approx \frac{g(x_1) - g(x_0)}{x_1 - x_0}$$

And with $Dg(x_1) = A_1$, we have that

$$A_1 = A_0 + \frac{g(x_1) - g(x_0) - A_0(x_1 - x_0)}{\|x_1 - x_0\|} (x_1 - x_0)^T$$

To show that A_1 minimizes $\|A_n - A_0\|$ note that

$$\|A_n - A_0\| = \|A_{n-1} + \frac{g(x_n) - g(x_{n-1}) - A_{n-1}(x_n - x_{n-1})}{\|x_n - x_{n-1}\|} * (x_n - x_{n-1})^T - A_0\| \geq 0$$

and that

$$\|A_1 - A_0\| = \|A_0 + \frac{g(x_1) - g(x_0) - A_0(x_1 - x_0)}{\|x_1 - x_0\|} * (x_1 - x_0)^T - A_0\|$$

By constraint, we have that

$$A_0(x_1 - x_0) = g(x_1) - g(x_0)$$

Since $\|A_1 - A_0\| \geq 0$, we know that

$$A_1 \leq \|A_n - A_0\|$$

which is non-negative, so we have the minimizer.

7.15

We start with

$$g_k = Df(x_k)$$

$$y_{k+1} = g_{k+1} - g_k$$

$$s_{k+1} = x_{k+1} - x_k = \alpha_k d_k \quad \alpha \in \mathbb{R}$$

Let d_k represent the direction vector.

$$\begin{aligned} A_{k+1} &= A_k + \frac{y_{k+1}y_{k+1}^T}{y_{k+1}^T s_{k+1}} - \frac{A_k s_{k+1} s_{k+1}^T A_k}{s_{k+1}^T A_k s_{k+1}} \\ &= A_k + \frac{y_{k+1}y_{k+1}^T}{y_{k+1}^T s_{k+1}} + \frac{A_k s_{k+1} s_{k+1}^T A_k}{-(s_{k+1}^T A_k) s_{k+1}} \end{aligned}$$

Yielding

$$A_k + \frac{y_{k+1}y_{k+1}^T}{y_{k+1}^T s_{k+1}} + \frac{\alpha_k A_k d_k \alpha_k d_k^T A_k}{\alpha_k g_k^T \alpha_k d_k} = A_k + \frac{y_{k+1}y_{k+1}^T}{y_{k+1}^T s_{k+1}} + \frac{g_k g_k^T}{g_k^T d_k}$$

Then by the Sherman-Morrison-Woodbury formula, we have that

$$(A - uv^T)^{-1} = A^{-1} + \alpha A^{-1} uv^T A^{-1}$$

Which applying twice yields the following:

$$A_{k+1}^{-1} = A_k^{-1} - \frac{s_{k+1}y_{k+1}^T A_k^{-1} + A_k^{-1} y_{k+1} s_{k+1}^T}{s_{k+1}^T y_{k+1}} + \left(1 + \frac{y_{k+1}^T A_k^{-1} y_{k+1}}{s_{k+1}^T y_{k+1}}\right) \frac{s_{k+1} s_{k+1}^T}{s_{k+1}^T y_{k+1}}$$

Multiplying through and rearranging we get:

$$\begin{aligned} A_{k+1}^{-1} &= A_k^{-1} - \frac{s_{k+1}y_{k+1}^T A_k^{-1} + A_k^{-1} y_{k+1} s_{k+1}^T}{s_{k+1}^T y_{k+1}} + \frac{s_{k+1} s_{k+1}^T}{s_{k+1}^T y_{k+1}} + \frac{y_{k+1}^T A_k^{-1} y_{k+1}}{s_{k+1}^T y_{k+1}} \frac{s_{k+1} s_{k+1}^T}{s_{k+1}^T y_{k+1}} \\ &= A_k^{-1} - \frac{A_k^{-1} y_{k+1} s_{k+1}^T + s_{k+1} y_{k+1}^T A_k^{-1}}{s_{k+1}^T y_{k+1}} + \frac{s_{k+1} s_{k+1}^T}{s_{k+1}^T y_{k+1}} + \frac{y_{k+1}^T A_k^{-1} y_{k+1}}{s_{k+1}^T y_{k+1}} \frac{s_{k+1} s_{k+1}^T}{s_{k+1}^T y_{k+1}} \\ &= A_k^{-1} - \frac{A_k^{-1} y_{k+1} s_{k+1}^T + s_{k+1} y_{k+1}^T A_k^{-1}}{s_{k+1}^T y_{k+1}} + \frac{(s_{k+1}^T y_{k+1}) s_{k+1} s_{k+1}^T}{(s_{k+1}^T y_{k+1})^2} + \frac{y_{k+1}^T A_k^{-1} y_{k+1} s_{k+1} s_{k+1}^T}{(s_{k+1}^T y_{k+1})^2} \\ &= A_k^{-1} + \frac{(s_{k+1}^T y_{k+1} + y_{k+1}^T A_k^{-1} y_{k+1}) s_{k+1} s_{k+1}^T}{(s_{k+1}^T y_{k+1})^2} - \frac{A_k^{-1} y_{k+1} s_{k+1}^T A_k^{-1}}{s_{k+1}^T y_{k+1}} \end{aligned}$$

7.16

For the first part, as A is invertible, we have that

$$Q^T = \left(\frac{1}{2}(A^T + A)\right)^T = \frac{1}{2}((A^T)^T + A^T) = \frac{1}{2}(A + A^T) = \frac{1}{2}(A^T + A) = Q$$

so Q is symmetric.

For the second part, let $b \in \mathbb{R}^n$.

We show x^* is a minimizer of $f(x) = x^T A x - b^T x$ if and only if it is a minimizer of $g(x) = x^T Q x - b^T x$.

Taking the derivative of the first equation and setting it equal to 0 gives:

$$Df(x) = (x^*)^T (A + A^T) - b = 0$$

$$\implies (x^*)^T \left(\frac{1}{2}(A^T + A) + \frac{1}{2}(A^T + A)\right) = b$$

$$\implies (x^*)^T (A^T + A) = b$$

along with the second equation

$$Df(x) = (x^*)^T (Q + Q^T) - b = 0$$

$$\implies (x^*)^T (Q + Q) = b$$

$$\implies (x^*)^T \left(\frac{1}{2}(A^T + A) + \frac{1}{2}(A^T + A)\right) = b$$

$$\implies (x^*)^T (A^T + A) = b$$

and so x^* is a minimizer for $f(x) = x^T A x - b^T x$

7.19

Minimizing Φ_k , we have that

$$\frac{\partial \left(\frac{1}{2}(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T Q (\mathbf{x}_k + \alpha_k \mathbf{d}_k) - \mathbf{b}^T (\mathbf{x}_k + \alpha_k \mathbf{d}_k) \right)}{\partial \alpha}$$

$$\implies 0 = \mathbf{d}_k^T Q (\mathbf{x}_k + \alpha_k \mathbf{d}_k) + \mathbf{b}^T \mathbf{d}_k$$

$$\alpha_k = \frac{\mathbf{b}^T \mathbf{d}_k - \mathbf{d}_k^T Q \mathbf{x}_k}{\mathbf{d}_k^T Q \mathbf{d}_k}$$

$$= \frac{\mathbf{b}^T \mathbf{d}_k - (Q \mathbf{x}_k)^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k}$$

$$= \frac{\mathbf{r}_k^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k}$$

and we have that

$$\mathbf{r}_k = \mathbf{b} - Q\mathbf{x}_k$$

7.20(i)

Plugging in for ϵ_0 , we have that

$$\begin{aligned}\mathbf{d}_k^T Q \epsilon_0 &= \mathbf{d}_k^T Q \sum_{j=0}^{n-1} \delta_j \mathbf{d}_j \\ &= \mathbf{d}_k^T Q (\delta_0 \mathbf{d}_0 + \delta_1 \mathbf{d}_1 + \dots + \delta_{n-1} \mathbf{d}_{n-1}) \\ &= \delta_k \mathbf{d}_k^T Q \mathbf{d}_k\end{aligned}$$

and we have that

$$\Rightarrow \quad \delta_k = \frac{\mathbf{d}_k^T Q \epsilon_0}{\mathbf{d}_k^T Q \mathbf{d}_k}$$

7.20 (ii)

By the definitions,

$$\begin{aligned}\epsilon_k &= \mathbf{x}_k - \mathbf{x}^* \\ \mathbf{x}_k &= \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 + \alpha_1 \mathbf{d}_1 + \dots + \alpha_{k-1} \mathbf{d}_{k-1} = \mathbf{x}_0 + \sum_{j=0}^{k-1} \alpha_j \mathbf{d}_j \\ \Rightarrow \quad \epsilon_k &= \mathbf{x}_0 + \sum_{j=0}^{k-1} \alpha_j \mathbf{d}_j - \mathbf{x}^* \\ &= (\mathbf{x}_0 - \mathbf{x}^*) + \sum_{j=0}^{k-1} \alpha_j \mathbf{d}_j \\ &= \epsilon_0 + \sum_{j=0}^{k-1} \alpha_j \mathbf{d}_j\end{aligned}$$

Now note that

$$\begin{aligned}\frac{\mathbf{d}_k^T Q \epsilon_k}{\mathbf{d}_k^T Q \mathbf{d}_k} &= \frac{\mathbf{d}_k^T Q (\epsilon_0 + \sum_{j=0}^{k-1} \alpha_j \mathbf{d}_j)}{\mathbf{d}_k^T Q \mathbf{d}_k} \\ &= \frac{\mathbf{d}_k^T Q \epsilon_0 + (\mathbf{d}_k^T Q \sum_{j=0}^{k-1} \alpha_j \mathbf{d}_j)}{\mathbf{d}_k^T Q \mathbf{d}_k} \\ &= \frac{\mathbf{d}_k^T Q \epsilon_0 + (0)}{\mathbf{d}_k^T Q \delta_k \mathbf{d}_k} \\ &= \delta_k\end{aligned}$$

7.20 (iii)

Keeping in mind that $Q\mathbf{x}^* = \mathbf{b}$,

$$\begin{aligned}
\alpha_k &= \frac{\mathbf{r}_k^T \mathbf{d}_k}{\mathbf{d}_k^T Q \delta_k \mathbf{d}_k} \\
&= \frac{(\mathbf{b} - Q\mathbf{x}_k)^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k} \\
&= \frac{(\mathbf{b} - Q(\epsilon_k + \mathbf{x}^*))^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k} \\
&= \frac{(\mathbf{b} - Q\epsilon_k - Q\mathbf{x}^*)^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k} \\
&= \frac{-(Q\epsilon_k)^T \mathbf{d}_k}{\mathbf{d}_k^T Q \mathbf{d}_k} \\
&= -\delta_k
\end{aligned}$$

7.20 (iv)

By (ii), we know that

$$\begin{aligned}
\epsilon_n &= \epsilon_0 + \sum_{j=0}^{n-1} \alpha_j \mathbf{d}_j \\
&= \sum_{j=0}^{n-1} \delta_j \mathbf{d}_j + \sum_{j=0}^{n-1} \alpha_j \mathbf{d}_j \\
&= \sum_{j=0}^{n-1} -\alpha_j \mathbf{d}_j + \sum_{j=0}^{n-1} \alpha_j \mathbf{d}_j \\
&= 0
\end{aligned}$$

which we know by (iii)

7.21(i)

Let

$$A = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \text{ and } x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Then

$$\begin{aligned}
 f(\mathbf{x}) = A\mathbf{x} &= \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
 &= \begin{bmatrix} a_{11}x_1 + \dots + a_{1m}x_n \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{bmatrix} \\
 \implies Df(\mathbf{x}) &= \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = A
 \end{aligned}$$

7.21 (ii)

We have that

$$\begin{aligned}
 f(\mathbf{x}) = \mathbf{x}^T A &= \begin{bmatrix} x_1 & \dots & x_m \end{bmatrix} \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \\
 &= \begin{bmatrix} a_{11}x_1 + \dots + a_{m1}x_m \\ \vdots \\ a_{1n}x_1 + \dots + a_{mn}x_m \end{bmatrix} \\
 \implies Df(\mathbf{x}) &= \begin{bmatrix} a_{11} & \dots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1m} & \dots & a_{mn} \end{bmatrix} = A^T
 \end{aligned}$$

7.21 (iii)

We have

$$u = \begin{bmatrix} u_1(\mathbf{x}) \\ \vdots \\ u_m(\mathbf{x}) \end{bmatrix} \quad v = \begin{bmatrix} v_1(\mathbf{x}) \\ \vdots \\ v_m(\mathbf{x}) \end{bmatrix}$$

Therefore

$$f = u_1(\mathbf{x})v_1(\mathbf{x}) + \dots + u_m(\mathbf{x})v_m(\mathbf{x})$$

$$\begin{aligned}
\Rightarrow Df(\mathbf{x}) &= u'_1(\mathbf{x})v_1(\mathbf{x}) + \dots + u'_m(\mathbf{x})v_m(\mathbf{x}) + v'_1(\mathbf{x})u_1(\mathbf{x}) + \dots + v'_m(\mathbf{x})u_m(\mathbf{x}) \\
&= \langle v(\mathbf{x}), Du(\mathbf{x}) \rangle + \langle u(\mathbf{x}), Dv(\mathbf{x}) \rangle \\
&= u^T(\mathbf{x})Dv(\mathbf{x}) + v^T(\mathbf{x})D(u(\mathbf{x}))
\end{aligned}$$

21 (iv)

Let $f(x)$ be given by the following

$$[x_1 \dots x_n] \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n-1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1(x_1a_{11} + \dots + x_na_{n1}) + \dots + x_n(x_1a_{1n} + \dots + x_na_{nn})$$

Then

$$\begin{aligned}
Df(x) &= \begin{bmatrix} 2x_1a_{11} + \dots + x_na_{n1} + \dots + x_na_{1n} \\ \vdots \end{bmatrix} = [x_1 \dots x_n] \left(\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} + \begin{bmatrix} a_{11} & \dots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{nn} \end{bmatrix} \right) \\
&= \mathbf{x}^T(A + A^T)
\end{aligned}$$

7.21 (v)

Let

$$B(\mathbf{x}) = \begin{bmatrix} \mathbf{b}_1(\mathbf{x}) \\ \mathbf{b}_2(\mathbf{x}) \\ \vdots \\ \mathbf{b}_m(\mathbf{x}) \end{bmatrix}$$

Then we have that

$$\begin{aligned}
f(x) &= \mathbf{u}^T B \\
&= [u_1(\mathbf{x}) \quad u_2(\mathbf{x}) \quad \dots \quad u_m(\mathbf{x})] \begin{bmatrix} \mathbf{b}_1(\mathbf{x}) \\ \mathbf{b}_2(\mathbf{x}) \\ \vdots \\ \mathbf{b}_m(\mathbf{x}) \end{bmatrix} \\
&= [u_1(\mathbf{x})\mathbf{b}_1(\mathbf{x}) + u_2(\mathbf{x})\mathbf{b}_2(\mathbf{x}) + \dots + u_m(\mathbf{x})\mathbf{b}_m(\mathbf{x})]
\end{aligned}$$

So

$$\begin{aligned}
Df &= [\mathbf{b}_1(\mathbf{x})u'_1(\mathbf{x}) + u_1(\mathbf{x})\mathbf{b}'_1(\mathbf{x}) \quad \mathbf{b}_2(\mathbf{x})u'_2(\mathbf{x}) + u_2(\mathbf{x})\mathbf{b}'_2(\mathbf{x}) \quad \dots \quad \mathbf{b}_m(\mathbf{x})u'_m(\mathbf{x}) + u_m(\mathbf{x})\mathbf{b}'_m(\mathbf{x})] \\
&= B^T D\mathbf{u}(\mathbf{x}) + \sum_{i=1}^m u_i(\mathbf{x})^T D\mathbf{b}_i(\mathbf{x})
\end{aligned}$$

7.22 (i)

We have that

$$f = \mathbf{r}(\mathbf{x})^T \mathbf{r}(\mathbf{x})$$

By 7.21 (iii), we have that,

$$\begin{aligned} Df(\mathbf{x}) &= \mathbf{r}(\mathbf{x})^T D\mathbf{r}(\mathbf{x}) + \mathbf{r}(\mathbf{x})^T D\mathbf{r}(\mathbf{x}) \\ &= 2\mathbf{r}(\mathbf{x})^T D\mathbf{r}(\mathbf{x}) = 2\mathbf{r}^T D\mathbf{r} \end{aligned}$$

7.22 (ii)

By 7.21 part(v),

$$\begin{aligned} D^2 f(\mathbf{x}) &= 2D\mathbf{r}(\mathbf{x})^T D\mathbf{r}(\mathbf{x}) + \sum_{i=1}^m 2r_i(\mathbf{x}) D^2 r_i(\mathbf{x}) \\ &= 2(D\mathbf{r}(\mathbf{x})^T D\mathbf{r}(\mathbf{x}) + \sum_{i=1}^m r_i(\mathbf{x}) D^2 r_i(\mathbf{x})) = 2(D\mathbf{r}^T D\mathbf{r} + \sum_{i=1}^m r_i D^2 r_i) \end{aligned}$$

7.23

$$D^2 f(x) > 0 \implies -D^2 f(x_k)^{-1} < 0 \implies Df(x_k)(-D^2 f(x_k))Df(x_k) < 0$$

With $d_k = -D^2 f(x_k)^{-1} Df(x_k)$ and $Df(x_k) = \lim_{\alpha \rightarrow 0} \left| \frac{f(x_k + \alpha d_k) - f(x_k)}{\alpha d_k} \right|$, it follows that

$$\begin{aligned} &Df(x_k)(-D^2 f(x_k))Df(x_k) < 0 \\ \implies &\lim_{\alpha \rightarrow 0} \left| \frac{f(x_k + \alpha d_k) - f(x_k)}{\alpha d_k} \right| d_k < 0 \\ \implies &\lim_{\alpha \rightarrow 0} \left| \frac{f(x_k + \alpha d_k)}{\alpha d_k} \right| d_k < \lim_{\alpha \rightarrow 0} \left| \frac{f(x_k) d_k}{\alpha d_k} \right| \\ \implies &f(x + \alpha d_k) < f(x) \end{aligned}$$

7.26 (i-iv)

Newton's method takes iterative derivatives to find zeros or minima of a function. Minimization or maximization can be done by this method. It is strong because it is reliable, but weak in that you need to have a good guess for tricky functions. Maximum Likelihood Estimation finds critical points of a certain function. This method is mathematically sound but is a little more difficult to compute. It is unbiased, though,

which is positive. Gauss-Newton finds critical points by way of Jacobian and Hessian matrices. The problem with this method is that if your guess is too far away from the real zero then it doesn't converge.