1 Lecture Review

1.1 Gradients and Matrix Calculus

1. Suppose $f(x)$ is a function in n variables with m components, that is we can write
$$f(x) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))$$
where $f_1, \ldots, f_m$ are scalar-valued functions. Then
$$df = J \cdot dx$$
where
$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$
is the Jacobian of f and $dx = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

2. Suppose h is a scalar valued function in one variable, W is $n \times p$, $x \in \mathbb{R}^p$ and $b \in \mathbb{R}^n$. If $f(x) = h.(Wx - b)$ then
$$df = (h'.(Wx - b)).^*(W\,dx)$$
where $.^*$ is component-wise product, $h.$ is component-wise evaluation by h, h' is the derivative of h.

3. (Product Rule) If F and G are matrix-valued functions such that matrix product $F \cdot G$ is compatible, then
$$d(FG) = (dF)G + F(dG).$$

4. $d(F^T) = (dF)^T$.

1.2 Determinants

1. Properties:
   (a) Normalization: $\det I = 1$.
   (b) Sign reversal (antisymmetry): when we exchange any two rows, the determinant flips sign.
   (c) The determinant is a linear function of any row, if all other rows are fixed.

2. Computing the determinant: Suppose A is a square matrix.

Step 1: Check if the $(1, 1)$ entry is nonzero. If not, flip two rows so that this is the case. If this is impossible, then the first column is 0 and $\det(A) = 0$.

Step 2: Using row operations, add multiples of the first row to the other rows until all the entries below the $(1, 1)$ entry in the first column are 0.

Step 3: Repeat from the beginning ignoring the first column and row. If this iterates until the matrix is upper triangular, then
$$\det(A) = (-1)^{\text{number of row flips}} \times \text{product of diagonal entries of the upper triangular matrix.}$$

2 Problems

1. Suppose $f(x) = Ax + b$ where $x \in \mathbb{R}^n$, A is $m \times n$ and $b \in \mathbb{R}^m$. Show that $df = A\,dx$.

   Solution. If $h(t) = t$, then $f(x) = h.(Ax + b)$ so that

   $$df = (h'.(Ax + b)).^*(Adx).$$

   Since $h'(t) = 1$, we have $h'.(Ax + b)$ is the vector in $\mathbb{R}^m$ of all ones. Then

   $$df = Adx$$

   since component-wise multiplication of a vector v (in this case Adx) with a vector of all ones doesn't change v. □

2. Suppose W is an $n \times n$ matrix. If $f : \mathbb{R}^n \to \mathbb{R}$ is defined by $f(x) = x^T W x$, show that $df = x^T(W + W^T)dx$.

   Solution. We have

   $$df = d(x^T)Wx + x^T d(Wx) = (dx)^T Wx + x^T W dx.$$

   Note that $(dx)^T Wx$ is a scalar, so in particular it is equal to its transpose:

   $$(dx)^T Wx = ((dx)^T Wx)^T = x^T W^T dx.$$

   This gives

   $$df = x^T W^T dx + x^T W dx = x^T (W^T dx + W dx) = x^T (W^T + W)dx.$$

   □

3. Use the product rule and $AA^{-1} = I$ to show that $d(A^{-1}) = -A^{-1} \cdot (dA) \cdot A^{-1}$. Consider what this means in the case A is 1×1.

   Solution. We have $dI = 0$ (the zero matrix) since I is constant in A. Then

   $$0 = dI = d(AA^{-1}) = (dA)A^{-1} + Ad(A^{-1})$$

   which implies

   $$d(A^{-1}) = -A^{-1}(dA)A^{-1}.$$

   □

4. Suppose C is an $n \times n$ matrix such that $C = C^T$. Show that

   $$d\left(\frac{1}{w^T C w}\right) = -\frac{2w^T C\,dw}{(w^T C w)^2}.$$

Solution. Since $1 = (w^T Cw)\frac{1}{w^T Cw}$, we have

$$0 = d1 = \frac{d(w^T Cw)}{w^T Cw} + w^T Cw\, d\left(\frac{1}{w^T Cw}\right)$$

which implies

$$d\left(\frac{1}{w^T Cw}\right) = -\frac{d(w^T Cw)}{(w^T Cw)^2}.$$

Note

$$d(w^T Cw) = (dw)^T (Cw) + w^T C dw$$

and $(dw)^T (Cw)$ is a scalar so that

$$(dw)^T (Cw) = ((dw)^T (Cw))^T = w^T C^T dw = w^T C dw.$$

Then

$$d(w^T Cw) = 2w^T C dw$$

so that

$$d\left(\frac{1}{w^T Cw}\right) = -\frac{2w^T C\, dw}{(w^T Cw)^2}.$$

□

5. Using row operations, show that $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$

Solution. Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ If $a = 0$ and $c = 0$, then $ad - bc = 0 = \det A$.
If $a = 0$, but $c \neq 0$, then

$$\det \begin{pmatrix} 0 & b \\ c & d \end{pmatrix} = -\det \begin{pmatrix} c & d \\ 0 & b \end{pmatrix} = bc = ad - bc.$$

If $a \neq 0$, then

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det \begin{pmatrix} a & b \\ 0 & d - \frac{bc}{a} \end{pmatrix} = ad - bc$$

where we added $-\frac{c}{a}$ times row 1 to row 2.
This covers all cases. □

6. *(Part of problem 5.1.15 from Strang.)* Use row operations to simplify and compute:

(a) $\det \begin{pmatrix} 1 & 2 & 3 & 0 \\ 2 & 6 & 6 & 1 \\ -1 & 0 & 0 & 3 \\ 0 & 2 & 0 & 7 \end{pmatrix}$

(b) $\det \begin{pmatrix} 1 & t & t^2 \\ t & 1 & t \\ t^2 & t & 1 \end{pmatrix}$

Solution. (a) We have

$$\det\begin{pmatrix}1&2&3&0\\2&6&6&1\\-1&0&0&3\\0&2&0&7\end{pmatrix}=\det\begin{pmatrix}1&2&3&0\\2&6&6&1\\0&2&3&3\\0&2&0&7\end{pmatrix}=\det\begin{pmatrix}1&2&3&0\\0&2&0&1\\0&2&3&3\\0&2&0&7\end{pmatrix}=\det\begin{pmatrix}1&2&3&0\\0&2&0&1\\0&0&3&2\\0&2&0&7\end{pmatrix}$$
$$=\det\begin{pmatrix}1&2&3&0\\0&2&0&1\\0&0&3&2\\0&0&0&6\end{pmatrix}=1\cdot 2\cdot 3\cdot 6=36.$$

where the red highlighted row indicates which rows have been affected by row reduction; in order (1) we add row 1 to row 3, (2) add -2 times row 1 to row 2, (3) add -1 row 2 to row 3, (4) add -1 row 2 to row 4.

(b) We have

$$\det\begin{pmatrix}1&t&t^2\\t&1&t\\t^2&t&1\end{pmatrix}=\det\begin{pmatrix}1&t&t^2\\0&1-t^2&t-t^3\\t^2&t&1\end{pmatrix}=\det\begin{pmatrix}1&t&t^2\\0&1-t^2&t-t^3\\0&t-t^3&1-t^4\end{pmatrix}$$
$$=\det\begin{pmatrix}1&t&t^2\\0&1-t^2&t-t^3\\0&0&1-t^2\end{pmatrix}=(1-t^2)^2$$

where the red highlighted row indicates which rows have been affected by row reduction; in order (1) we add $-t$ row 1 to row 2, (2) add $-t^2$ row 1 to row 3, (3) add $-t$ row 2 to row 3.

□

7. Prove that $\det\left(A^{-1}\right)=(\det A)^{-1}$.

Solution. If A is invertible, then $\det A\neq 0$. Since $\det(AB)=\det(A)\det(B)$ for square matrices A,B of the same size, we have
$$1=\det I=\det(AA^{-1})=\det(A)\det(A^{-1})$$
which implies $\det(A^{-1})=\det(A)^{-1}$. □

8. If U is orthogonal, show that $\det U=\pm 1$.

Solution. Recall that $\det U^T=\det U$. Since
$$1=\det I=\det U^TU=\det(U^T)\det U=\det(U)\det(U)=\det(U)^2$$
we have $\det U=\pm 1$. □

9. *Portfolio optimization.* (Challenging) In this problem, we will consider three financial instruments: Dow Jones Industrial Average index (first coordinate), Activision Blizzard equities (second coordinate) and Fidelity US Bond Index fund (third coordinate) – and will try to figure out what would be the optimal investment in those. Note that this cannot be used to devise your investment strategy, because the expectation of future returns is very hard to estimate. For the purpose of this problem, we will assume that expected returns are proportional to standard deviation of these instruments.

(a) Knowing that standard deviation of gains of Dow Jones is $\sigma_1 = 1$, of Blizzard stock – $\sigma_2 = 2.6$, of US bonds – $\sigma_3 = 0.2$, make a guess which plot corresponds to which instrument.

(b) We write how much we invest in each instrument as a vector $w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$, so here we invest k\$$w_2$ in Activision Blizzard. Note that there is a way to "invest" a negative amount of money in a certain instrument. Interpret this vector w as an investment into a linear combination of the three instruments. We call this vector w the *investment portfolio.*

(c) Covariance is a measure of the joint variability of two random variables. The (approximate) covariance matrix C for the given instruments, and its inverse are the following:

$$C = \begin{pmatrix} 1 & 1.2 & 0 \\ 1.2 & 6.8 & -0.1 \\ 0 & -0.1 & 0.04 \end{pmatrix}, \; C^{-1} = \begin{pmatrix} 1.3 & -0.23 & -0.6 \\ -0.23 & 0.2 & 0.5 \\ -0.6 & 0.5 & 26.2 \end{pmatrix}.$$

The importance of the covariance matrix is that we can compute variance of a linear combination of stocks (portfolio) w easily via the formula $w^T C w$. Note that variance is the square of standard deviation.

Compute the variance of the investment of k\$1 in Blizzard and k\$1 in US bonds. Explain why the diagonal entries of C are exactly the squares of standard deviations: $C_{ii} = \sigma_i^2$.

(d) Modern theory of portfolio optimization says that we want to maximize "Sharpe ratio" – the ratio of expected return to the standard deviation of the portfolio. Assume that the expected return is given by vector $\mu = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{pmatrix}$. Verify that the Sharpe ratio of the portfolio w is given by the following formula:

$$\mathrm{SR} = \frac{\mu^T w}{\sqrt{w^T C w}}.$$

(e) If $f(w)$ is a scalar function of w, i.e. its values are 1×1 matrices – real numbers, compute $\mathrm{d}\left(f(w)^{-1}\right)$ using the product formula for $0 = \mathrm{d}1 = \mathrm{d}\left(f(w)f(w)^{-1}\right)$. Differentials are taken with respect to w.

(f) Compute the differential $\mathrm{d}(\mathrm{SR})$ with respect to w.

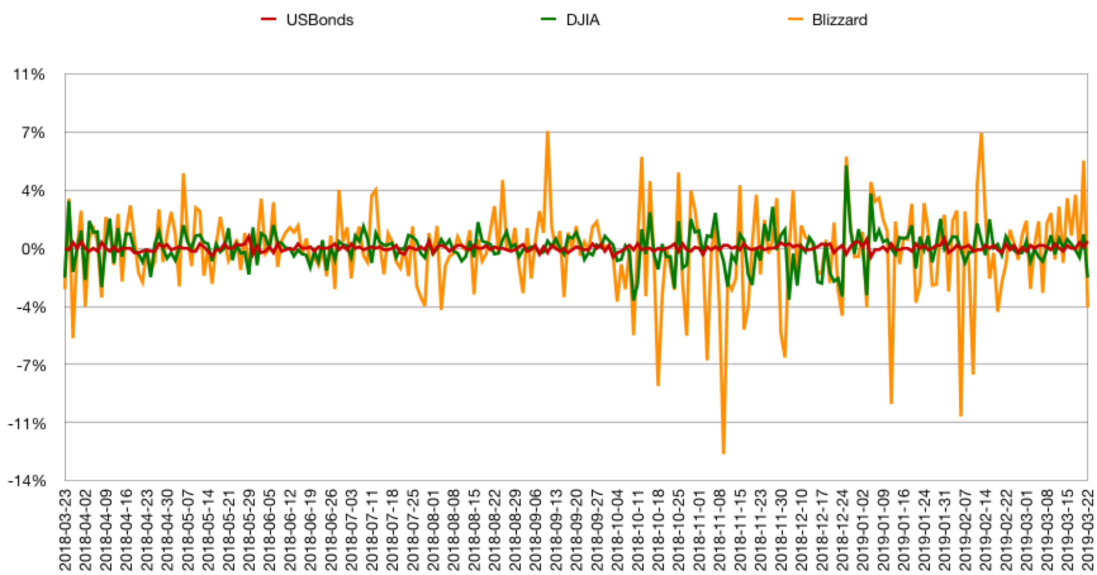
(g) Write the condition for finding a local extremum of SR.

(h) Conclude that w should be parallel to $C^{-1}\mu$, and argue that scaling w does not affect Sharpe ratio, therefore we can take $w = C^{-1}\mu$.

(i) Use the above formula and the given values of μ and C^{-1} to obtain the weights of the optimal portfolio.

Solution.

(a)



(b)

(c) For the first question, our portfolio is $w = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, so variance is

$$w^T C w = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1.2 & 0 \\ 1.2 & 6.8 & -0.1 \\ 0 & -0.1 & 0.04 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.2 & 6.7 & -0.06 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 6.64.$$

(d)

(e)

(f) The computation runs as follows:

$$\begin{aligned} \mathrm{d}(\mathrm{SR}) = \mathrm{d}\frac{\mu^T w}{\sqrt{w^T C w}} &= \frac{\mathrm{d}\left(\mu^T w\right)\sqrt{w^T C w} - \mu^T w \cdot \mathrm{d}\left(\sqrt{w^T C w}\right)}{w^T C w} \\ &= \frac{\mu^T \mathrm{d}w \cdot w^T C w - \mu^T w \cdot \frac{1}{2}\left(\mathrm{d}w^T C w + w^T C \mathrm{d}w\right)}{(w^T C w)^{\frac{3}{2}}} \\ &= \frac{\mu^T \mathrm{d}w \cdot w^T C w - \mu^T w \cdot \frac{1}{2}\left(w^T C \mathrm{d}w + w^T C \mathrm{d}w\right)}{(w^T C w)^{\frac{3}{2}}} \\ &= \frac{\mu^T \mathrm{d}w \cdot w^T C w - \mu^T w \cdot w^T C \mathrm{d}w}{(w^T C w)^{\frac{3}{2}}}. \end{aligned}$$

(g) The condition is $\mu^T w \cdot w^T C = w^T C w \cdot \mu^T$. We can transpose both sides (noting that $w^T C w$ and $\mu^T w$ are scalars) and get $\mu^T w \cdot C w = w^T C w \cdot \mu$. After multiplying both sides by C^{-1} we get $\mu^T w \cdot w = w^T C w \cdot C^{-1} \mu$.

(h) Note again that $w^T C w$ and $\mu^T w$ are scalars, so from the previous part we get that w and $C^{-1}\mu$ are parallel.

(i) $w = C^{-1}\mu = \begin{pmatrix} 1.3 & -0.23 & -0.6 \\ -0.23 & 0.2 & 0.5 \\ -0.6 & 0.5 & 26.2 \end{pmatrix} \begin{pmatrix} 1 \\ 2.6 \\ 0.2 \end{pmatrix} = \begin{pmatrix} 0.582 \\ 0.39 \\ 5.94 \end{pmatrix}$

$\square$