Math 5365

Data Mining 1

Homework 11

Mary Barker

Do p. 105(1, 9 ,15) from *Advanced Calculus* by Folland.

1 Find the extreme values of $f(x, y) = 2x^2 + y^2 + 2x$ on the set $\{(x, y) : x^2 + y^2 \leq 1\}$.

The gradient of the function is

$$\nabla f(x, y) = \begin{bmatrix} 4x + 2 \\ 2y \end{bmatrix}$$

- On the boundary $G(x, y) = x^2 + y^2 - 1 = 0$, the critical values are found as follows: $L(\mathbf{x}, \lambda) = f - \lambda G, \mathbf{x} = (x, y)$

$$\frac{\partial}{\partial \mathbf{x}} L = \begin{matrix} 4x + 2 - \lambda 2x \\ 2y - \lambda 2y \end{matrix}$$

From the second row in the equation, $2y - \lambda 2y = 0$, there are two possible choices: $\lambda = 1$, and $y = 0$.

  - If $\lambda = 1$, then

$$4x + 2 - 2x = 0 \implies x = -1$$

And from the boundary constraint, $x = -1 \implies y = 0$. Since this does not create a contradiction with the second equation, there is a value at $(-1, 0)$ At this point, the function has a value of $f(-1, 0) = 2 + 0 - 2 = 0$.

  - The second option is for $y = 0$. From the boundary constraint, there are two possible x-values: $x = \pm 1$.

The point $(-1, 0)$ has already been evaluated, so check $(1, 0)$. From the first equation, $4+2-\lambda 2=0$ which is true whenever $\lambda=3$. Thus the second critical point is $(1,0)$, where the function is evaluated as $f(1,0)=2+0+2=4$.

- Inside the domain, setting the gradient ∇f to 0 gives $x=-\frac{1}{2}, y=0$ as critical point. The value of f at this point is $f(-\frac{1}{2},0)=\frac{1}{2}-1=-\frac{1}{2}$.

Thus the extreme values of the function over the set are

Maximum: 4 at $(1,0)$, Minimum: $-\frac{1}{2}$ at $(-\frac{1}{2},0)$.

9 Find the extreme values of $f(x,y,z)=x^2+2y^2+3z^2$ on the unit sphere $\{(x,y,z): x^2+y^2+z^2=1\}$.

Let $g=x^2+y^2+z^2-1$.

$$\nabla f=\begin{bmatrix}2x\\4y\\6z\end{bmatrix}, \lambda\nabla g=\begin{bmatrix}\lambda 2x\\\lambda 2y\\\lambda 2z\end{bmatrix}$$

And therefore

$$\nabla L=\nabla(f-\lambda g)=\nabla f-\lambda\nabla g=\begin{bmatrix}2x-\lambda 2x\\4y-\lambda 2y\\6z-\lambda 2z\end{bmatrix}$$

Starting with the first row, $2x-\lambda 2x$, there are two possibilities.

- x = 0

  The first row is satisified, and last two rows of $\nabla L=0$ yield:

$$\begin{aligned}4y-\lambda 2y &= 0 \implies \lambda=2\\ 6z-\lambda 2z &= 0 \implies \lambda=3\end{aligned}$$

  This is a contradiction, therefore a critical point is not at $x=0$ on the boundary.

- $\lambda = 1$

  The second two rows of $\nabla L = 0$ with $\lambda = 1$ yield:

  $$\begin{aligned} 4y - 2y &= 0 \quad \Longrightarrow \; y = 0 \\ 6z - 2z &= 0 \quad \Longrightarrow \; z = 0 \end{aligned}$$

  Using the constraint $g = 0$ gives

  $$x^2 + 0^2 + 0^2 - 1 = 0 \implies x^2 = 1$$

  Therefore the critical points are $(\pm 1, 0, 0)$. with values:

  $f(1, 0, 0) = 1$

  $f(-1, 0, 0) = 1$

  This means that the function is constant valued on g.

15 The two planes $x + z = 4$ and $3x - y = 6$ intersect in a line L. Use Lagrange's method to find the point on L that is closest to the origin. (*Hint:* Minimize the square of the distance.)

The line L is given by the parametric equation $L(t) = (4 - t, -6 + 3t, t)$ The Euclidean distance from the origin of any point on the line is therefore given by $D(t) = \sqrt{(4 - t^2) + (-6 + 3t)^2 + t^2} = \sqrt{11t^2 - 44t + 52}$

This quadratic is not factorable since $44^2 - 4(11)(52) < 0$. Therefore no point on the line L ever crosses the origin. However, the point that is the closest to the origin can be found by computing the smallest value of the distance function $D(t)$. This will be at a solution to $\frac{d}{dt} D(t) = 0$.

For simplicity, the square of the distance $D^2(t)$ is used, as it has the same critical point as D (As a positive-valued function, D has the property that the minimum value of D is also the minimum value of D^2).

The solution to $\frac{d}{dt} D^2(t) = 22t - 44$ is at $t = 2$. Therefore the point closest to the origin on L is $L(2) = (2, 0, 2)$

1. Let $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^m, y \in \mathbb{R}^n$. Show that

$$\frac{\partial}{\partial x} x\prime Ay = Ay$$

By definition,

$$x\prime Ay = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i A_{i,j} y_j$$

And $Ay \in \mathbb{R}^{n\times 1}$ where $(Ay)_i = \sum\limits_{j=1}^{n} A_{i,j}y_j$. And therefore

$$\begin{aligned} \tfrac{\partial}{\partial x} x' Ay &= \tfrac{\partial}{\partial x} \sum_{i=1}^{m}\sum_{j=1}^{n} x_i A_{i,j} y_j \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} \tfrac{\partial}{\partial x} x_i A_{i,j} y_j \end{aligned}$$

But since $\frac{\partial}{\partial x} x_j = \hat{e}_j$ where $\hat{e}_j$ is the j-th column in $I_{m\times m}$, the sum reduces to

$$\begin{bmatrix} \sum\limits_{j=1}^{n} A_{1,j}y_j \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \sum\limits_{j=1}^{n} A_{2,j}y_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} + ... + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \sum\limits_{j=1}^{n} A_{m,j}y_j \end{bmatrix} = \begin{bmatrix} \sum\limits_{j=1}^{n} A_{1,j}y_j \\ \sum\limits_{j=1}^{n} A_{2,j}y_j \\ \vdots \\ \sum\limits_{j=1}^{n} A_{m-1,j}y_j \\ \sum\limits_{j=1}^{n} A_{m,j}y_j \end{bmatrix} = Ay$$

2. if $\Sigma \in \mathbb{R}^{n\times n}$ is symmetric, show that

$$\frac{\partial}{\partial x} x\prime \Sigma x = 2\Sigma x$$

The multiplication can be written

$$x'\Sigma x = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \Sigma_{i,j} x_j = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \Sigma_{i,j}$$

And using the definition from problem 1 $\frac{\partial}{\partial x} x_i = \hat{e}_i$,

$$\begin{aligned}
\tfrac{\partial}{\partial x}(x'\Sigma x) &= \tfrac{\partial}{\partial x}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \Sigma_{i,j}\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \tfrac{\partial}{\partial x}(x_i x_j \Sigma_{i,j}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} (\hat{e}_i x_j \Sigma_{i,j} + \hat{e}_j x_i \Sigma_{i,j})
\end{aligned}$$

Using vector notation, this can be written

$$\begin{bmatrix} x_1\Sigma_{1,1} + x_2\Sigma_{1,2} + ... + x_n\Sigma_{1,n} \\ x_1\Sigma_{2,1} + x_2\Sigma_{2,2} + ... + x_n\Sigma_{2,n} \\ \vdots \\ x_1\Sigma_{n,1} + x_2\Sigma_{n,2} + ... + x_n\Sigma_{n,n} \end{bmatrix} + \begin{bmatrix} x_1\Sigma_{1,1} + x_2\Sigma_{2,1} + ... + x_n\Sigma_{n,1} \\ x_1\Sigma_{1,2} + x_2\Sigma_{2,2} + ... + x_n\Sigma_{n,2} \\ \vdots \\ x_1\Sigma_{1,n} + x_2\Sigma_{2,n} + ... + x_n\Sigma_{n,n} \end{bmatrix}$$

And since Σ is symmetric, the result is

$$2\begin{bmatrix} x_1\Sigma_{1,1} + x_2\Sigma_{1,2} + ... + x_n\Sigma_{1,n} \\ x_1\Sigma_{2,1} + x_2\Sigma_{2,2} + ... + x_n\Sigma_{2,n} \\ \vdots \\ x_1\Sigma_{n,1} + x_2\Sigma_{n,2} + ... + x_n\Sigma_{n,n} \end{bmatrix} = 2\Sigma x$$

3. Let $X \in \mathbb{R}^{n\times p}$ be a full rank matrix, and let $Y \in \mathbb{R}^n$. Find the vector $\gamma \in \mathbb{R}^p$ that minimizes $\|Y - X\gamma\|^2$. (Note that this is not a Lagrange multipliers problem, because there is no constraint on γ.)

   There are three options for this problem.

   (a) if n = p

   Then, since X has full rank, it is invertible, and the desired value for γ is the solution to $Y - X\gamma = 0$ or $X\gamma = Y$, in which case,

$$\gamma = X^{-1}Y$$

(b) if n > p

Then, since X has full rank (rank p), $X'X \in \mathbb{R}^{p\times p}$ is invertible. And therefore the desired value for γ is the solution to $Y - X\gamma = 0$ or $X'Y - X'X\gamma = 0$, which gives

$$\gamma = (X'X)^{-1}X'Y$$

(c) if n < p

There is no way of solving this.

4. Bonus: Let $\Sigma_{11} \in \mathbb{R}^{p_1\times p_1}$ and $\Sigma_{22} \in \mathbb{R}^{p_2\times p_2}$ be positive definite matrices. Also, let $\Sigma_{12} \in \mathbb{R}^{p_1\times p_2}$ and define $\Sigma_{21} = \Sigma_{12}\prime$. If

$$c_1 = \max\left\{x'\Sigma_{12}y \mid x'\Sigma_{11}x = 1, y'\Sigma_{22}y = 1\right\}$$

is attained at $(\mathbf{x}, \mathbf{y})$, show that

$$\begin{pmatrix} -c_1\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -c_1\Sigma_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = 0$$

This is the main result from the theory of canonical correlations. This maximum is

$$c_1 = \max\left\{cov(x'U, y'V) \mid Var(x'U) = 1, Var(y'V) = 1\right\}.$$

Where U and V are random vectors with joint covariance matrix.

$$\text{cov}\left[\begin{pmatrix} U \\ V \end{pmatrix}\right] = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix}$$

First note that since Σ_{11}, Σ_{22} are positive definite, for any $x, y \neq 0$ vectors of the appropriate length, $x'\Sigma_{11}x > 0$, and $y'\Sigma_{22}y > 0$.

$$\begin{pmatrix} \mathbf{x}' & \mathbf{y}' \end{pmatrix} \begin{pmatrix} -c_1\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -c_1\Sigma_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{x}'(-c_1)\Sigma_{11}\mathbf{x} + \mathbf{x}'\Sigma_{12}\mathbf{y} \\ \mathbf{y}'\Sigma_{21}\mathbf{x} + \mathbf{y}'(-c_1)\Sigma_{22}\mathbf{y} \end{pmatrix}$$

And since c_1 are scalars, the multiplication can be reordered as follows.

$$\begin{pmatrix} -c_1\mathbf{x}'\Sigma_{11}\mathbf{x} + \mathbf{x}'\Sigma_{12}\mathbf{y} \\ \mathbf{y}'\Sigma_{21}\mathbf{x} - c_1\mathbf{y}'\Sigma_{22}\mathbf{y} \end{pmatrix}$$

But since $\mathbf{x}'\Sigma_{11}\mathbf{x} = 1$ and $\mathbf{y}'\Sigma_{22}\mathbf{y} = 1$, this is

$$\begin{pmatrix} -c_1 + \mathbf{x}'\Sigma_{12}\mathbf{y} \\ \mathbf{y}'\Sigma_{21}\mathbf{x} - c_1 \end{pmatrix}$$

And since $\Sigma_{12} = \Sigma_{21}$, and $\mathbf{y}'\Sigma 21\mathbf{x}$ is a scalar, the transpose is the same value.

$$(\mathbf{y}'\Sigma_{21}\mathbf{x})(\mathbf{y}'\Sigma_{21}\mathbf{x})' = (\mathbf{x}'\Sigma_{12}\mathbf{y}) = c_1$$

And so

$$\begin{pmatrix} \mathbf{x}' & \mathbf{y}' \end{pmatrix} \begin{pmatrix} -c_1\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -c_1\Sigma_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} -c_1 + c_1 \\ c_1 - c_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$