Maths Problem Set 4

Tim Munday
OSM Boot Camp 2018

July 11, 2018

Ex 6.6

For a point to be a critical point, it must have first derivative of zero in all arguments.

Differentiating with respect to x, and separately with respect to y, and setting these FOCs equal to zero, gives us four critical points.

For a (strict) local maximum, the Hessian evaluated at the point in question must be negative definite, for a (strict) local minimum, positive definite, and a saddle point if the Hessian is neither. We determine whether the Hessian is positive or negative definite by considering the eigenvalues. If both eigenvalues are positive then the matrix is positive definite, vice versa for negative definite. One positive and one negative eigenvalue indicates a saddle point.

$$A : x = \frac{-1}{3}, y = 0 \quad | \quad \lambda_1 = \frac{1}{3}, \lambda_2 = -3 \Rightarrow \textit{Saddle Point}$$
$$B : x = \frac{-1}{9}, y = \frac{-1}{12} \quad | \quad \lambda_1 = -0.309, \lambda_2 = -1.08 \Rightarrow \textit{Maximum}$$
$$C : x = 0, y = \frac{-1}{4} \quad | \quad \lambda_1 = 1, \lambda_2 = -1 \Rightarrow \textit{Saddle Point}$$
$$D : x = 0, y = 0 \quad | \quad \lambda_1 = -2, \lambda_2 = 0.5 \Rightarrow \textit{Saddle Point}$$

Ex. 6.7

(i) First we show that Q is symmetric.

$$Q^T = (A^T + A) = A + A^T = Q^T$$

Therefore Q is symmetric.
Furthermore:

$$\begin{aligned} x^T Q x &= x^T (A^T + A) x \\ &= x^T A^T x + x^T A x \qquad = 2x^T A x \end{aligned}$$

Where the last line follows from the fact that since x is an n by 1 vector and A n by n, that multiply to produce a scalar, so we can rearrange the order somewhat.

(ii)

If we take the first differential of the equation we have:

$$\begin{aligned} f'(x) &= x^T Q - b^t = 0 \\ &= x^T Q = b^T \\ &= Q^T x = b \end{aligned}$$

We can see that since the function is quadratic with a positive term on the squared term, that the point at which the first derivative is zero will be a minimum.

(iii) First we show that a solution only exists if Q is positive definite. For the equation $Q^T x = b$ to have a solution, Q^T must be invertible. Since eigenvalues are preserved under transposition, this means that Q must be invertible. Therefore, Q must be either positive or negative definite. Since we are looking for a minimum, it must be positive definite, for the second derivatives to be positive. This is a sufficient condition for a minimum.

Ex. 6.11

The Newton method's first iteration gives:

$$x_1 = x_0 - \frac{f'(x0)}{f''(x0)}$$

Where f is a quadratic approximation (in this case the actual function).

Evaluating this:

$$x_1 = \frac{-b}{2a} x_0$$

For x_1 to be a minimiser (we know it is unique as the function itself is quadratic), we need the first derivative to be zero, and the second to be positive.

$$\begin{aligned} f'(x_1) &= 2ax_0 + b \\ &= -b + b \\ &= 0 \end{aligned}$$

$$f''(x_1) = 2a > 0$$

Therefore we have found the unique minimiser in one iteration. We could have expected this, because the Newton method approximates using a quadratic function, so the approx of a quadratic function is the function itself.

6.15

See notebook.

Ex 7.1

$H - Conv(S)$ is convex if:

$$\lambda H_1 + (1-\lambda)H_2 \in H$$

Expanding out, noting that H_i is a linear combination of the members of V, i.e. the x's by definition. We have:

$$\begin{aligned}\lambda H_1 + (1-\lambda)H_2 &= \lambda(\theta_1 x_1 + \theta_2 x_2 ... + \theta_k x_k) + (1-\lambda)(\rho_1 x_1 + ... + \rho_k x_k)\\ &= (\lambda\theta_i + (1-\lambda)\rho_i)x_i \in H\end{aligned}$$

Where the final line follows from the fact that the linear weighting of two series that sum to one will also sum to one, and as such we are essentially just applying a different weighting to the x's, so they must be in H by definition of the convex hull.

Ex 7.2

(i) A Hyperplane is a set of the form:

$$P = \{x \in V | \langle a, x\rangle = b\}$$

Taking two points in the hyper plane: X and Y and taking a linear combination of them, satisfies the first condition, since they are both in V, and the second is also satisfied:

$$\begin{aligned}\lambda\langle a, X\rangle + (1-\lambda)\langle a, Y\rangle &= b\\ \lambda b + (1-\lambda)b &= b\end{aligned}$$

(ii) We can follow the same procedure for a half-space:

$$H = \{x \in V | \langle a, x\rangle \leq b\}$$

Taking two points in the hyper plane: X and Y and taking a linear combination of them, satisfies the first condition, since they are both in V, and the second is also satisfied:

$$\begin{aligned}\lambda\langle a, X\rangle + (1-\lambda)\langle a, Y\rangle &\leq b\\ \lambda p + (1-\lambda)q &\leq= b\\ As\ p &\leq b\\ and\ q &\leq b\end{aligned}$$

7.4

(i)

$$\begin{aligned}
||x-y||^2 &= ||x-p-(y-p)||^2 \\
&= \langle (x-p)-(y-p), (x-p)-(y-p) \rangle \\
&= \langle x-p, x-p \rangle + \langle p-y, p-y \rangle + 2\langle x-p, p-y \rangle \\
&= ||x-p||^2 + ||p-y||^2 + 2\langle x-p, p-y \rangle
\end{aligned}$$

(ii)

From the definition of a projection we know that $2\langle x-p, p-y \rangle \geq 0$. We also know that the norm of order two is always positive so $||p-y||^2 \geq 0$. So, using the expression above we have:

$$||x-y||^2 = ||x-p||^2 + K$$

Where K is strictly positive since $y \neq p$. Thus we have (square rooting both sides):

$$||x-y|| > ||x-p||$$

(iii)

Using the equation we proved in (i), we can write:

$$\begin{aligned}
||x-z||^2 &= ||x-p||^2 + ||p-z||^2 + 2\langle x-p, p-z \rangle \\
&= ||x-p||^2 + ||p-\lambda y + (1-\lambda)p||^2 + 2\langle x-p, p-(\lambda y + (1-\lambda)p) \rangle \\
&= ||x-p||^2 + 2\lambda\langle x-p, p-y \rangle + \lambda^2||y-p||^2
\end{aligned}$$

(iv)

From (ii) we know that:

$$||x-z||^2 > ||x-p||^2$$

Then we can rearrange (7.15) to get:

$$0 \leq 2\langle x-p, p-y \rangle + \lambda||y-p||^2$$

Since we know that this must hold even when λ is 0, we have the result:

$$\langle x-p, p-y \rangle \geq 0$$

Ex 7.8

We are told f is convex. i.e.:

$$f(\lambda X_1 + (1-\lambda)X_2) \leq \lambda f(X_1) + (1-\lambda)f(X_2)$$

Then, since this holds for all X, we can denote X_1 as $Ax_1 + b$, and X_2 as $Ax_2 + b$.

$$\begin{aligned} f(\lambda Ax_1 + (1-\lambda)Ax_2 + b) &\leq \lambda f(Ax_1 + b) + (1-\lambda)f(Ax_2 + b) \\ g(\lambda x_1 + (1-\lambda)x_2) &\leq \lambda g(x_1) + (1-\lambda)f(x_2) \end{aligned}$$

Where the last line shows that g is convex.

Ex 7.12

(i) The set of positive definite matrices is convex if any linear combination of those matrices also lies within the set. Lets assume A_1 and A_2 are positive definite, i.e. $xAx^T > 0$ for all x.

$$\begin{aligned} A_3 &= \lambda A_1 + (1-\lambda)A_2 \\ xA_3x^T &= x(\lambda A_1 + (1-\lambda)A_2)x^T \\ &= \lambda xA_1x^T + (1-\lambda)xA_2x^T > 0 \end{aligned}$$

Since λ is positive, we know the right hand side of the last line is strictly positive, so A_3 is positive semi definite, and thus lies within the set of positive semi definite matrices.
(ii)
(a)

$$f(\lambda A + (1-\lambda)B) = g(\lambda) \leq \lambda g(1) + (1-\lambda)g(0) = \lambda f(A) + (1-\lambda)f(B)$$

Therefore f is convex.
(b)
A is normal, so we can write:

$$XAX^-1 = D$$

Where X is a unitary matrix, and D a diagonal matrix. Since X is unitary we know that $X^-1 = X^H$.

$$\begin{aligned} XAX^-1 &= D \\ AX^-1 &= X^-1D \\ AX^H &= X^HD \\ A &= X^HDX \\ &= (X^HD^0.5)(D^0.5X) \\ &= S^HS \end{aligned}$$

Now we show that the expression given for g is equal to $f(tA+(1-t)B)$, by evaluating the argument:

$$\begin{aligned} S^H(tI+(1-t)(S^H)^-1BS^-1)S& \\ &= S^H tIS + S^H(1-t)(S^H)-1BS^1S \\ &= S^H tS + S^H(1-t)(S^H)^-1B \\ &= tA+(1-t)B \end{aligned}$$

Now we must show the second part of the question. First of all we note that the determinant of the product of two matrices, is the product of the two determinants.

$$\begin{aligned} -log(det(S^H(tI+(1-t)(S^H)^-1BS^-1)S))& \\ &= -log(det(S^HS)det(tI+(1-t)(S^H)^-1BS^-1) \\ &= -log(det(A)det(tI+(1-t)(S^H)^-1BS^-1) \\ &= -(log(det(A))+log(det(tI+(1-t)(S^H)^-1BS^-1))) \\ &= -log(det(A))-log(det(tI+(1-t)(S^H)^-1BS^-1)) \end{aligned}$$

(c)

First we note that the matrix $(S^H)^-1BS^-1$ is full rank. Then if we take some eigenvector of $(S^H)^-1BS^-1$, which we denote x_i, we can write:

$$((S^H)^-1BS^-1)x_i = \lambda_i x_i$$

Therefore, we can also write:

$$\begin{aligned} (tI+(1-t)((S^H)^-1BS^-1))x_i &= tIx_i+(1-t)\lambda_i x_i \\ &= (t+(1-t)\lambda_i)x_i \end{aligned}$$

The last line tells us that $(t+(1-t)\lambda_i)$ represent the i eigenvalues of the matrix $tI+(1-t)((S^H)^-1BS^-1)$.

Since the determinant of a matrix is the product of the eigenvalues, and we can split up log products into additions, so we can now write:

$$\begin{aligned} -log(det(A))-log(det(tI+(1-t)(S^H)^-1BS^-1))& \\ &= -log(det(A))-log(\prod_i(t+(1-t)\lambda_i)) \\ &= -log(det(A))-\sum_i log(t+(1-t)\lambda_i) \end{aligned}$$

(d) Taking the second derivative we get:

$$g'' = \sum_i (1-\lambda_i)^2/(t+(1-t)\lambda_i)^2$$

Since both top and bottom are positive the whole thing is positive.

Ex 7.13

Let's assume f is bounded by M. That is $f(x) \leq M \ \forall x$. Take two points x_1 and x_2 in the domain of f. If $f(x1)$ and $f(x2)$ are not equal, then the line through them must cross $y = M$ at some point. Since f is convex, we know that $f(x)$ must lie on or above the line between $f(x1)$ and $f(x2)$ for all values of x, x_1, x_2. This is clearly not possible since the line will always intersect M unless the function is constant. Therefore the function is constant.

Ex 7.20

If f is convex then we know that:

$$\lambda f(x_1) + (1-\lambda)f(x_2) \geq f(\lambda x_1 + (1-\lambda)x_2)$$

If $-f$ is convex then we know that:

$$\lambda f(x_1) + (1-\lambda)f(x_2) \leq f(\lambda x_1 + (1-\lambda)x_2)$$

Therefore:

$$\lambda f(x_1) + (1-\lambda)f(x_2) = f(\lambda x_1 + (1-\lambda)x_2)$$

So the function is affine.

Ex. 7.21

Let $x^* \in \mathbb{R}^n$ be a local minimizer of f. Then $f(x^*) \leq f(x)$ for all $x \in \mathcal{N}_r(x^*)$, where $\mathcal{N}_r(x^*)$ is an open ball around x^* of radius $r > 0$. Since ϕ is monotonically increasing, $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in \mathcal{N}_r(x^*)$. Thus, x^* is a local minimizer of $\phi \circ f$. Now let x^* be a local minimizer of $\phi \circ f$. Then $\phi(f(x^*)) \leq \phi(f(x))$ for all $x \in \mathcal{N}_r(x^*)$, and since ϕ is monotonically increasing, this implies that $f(x^*) \leq f(x)$ for all $x \in \mathcal{N}_r(x^*)$. Thus, x^* is a local minimizer of f.