1. Exercises from Sections 2.7-2.9

Problem 1. *(Folland 2.7.3) Show that* $|\sin(x) - x + x^3/6| < 0.08$ *for all* $|x| < \pi/2$

(1) Notice that $x^3 - x^3/6$ is the fourth order Taylor approximation of $\sin(x)$ about zero
(2) Since $\sin(x)$ is smooth (therefore C^k for all $k \in \mathbb{N}$), we may use thm. 2.55:

$$R_{a,k}(h) = \frac{h^{k+1}}{k!} \int_0^1 (1-t)^k f^{(k+1)}(a+th)\, dt$$

(3) $d^5(\sin(x))/dx^5 = \cos(x)$, and $|\cos(th)| < 1$ for all $|th| < \pi/2$
(4) Estimating gives:

$$|R_{0,4}(h)| \leq \frac{h^5}{24} \int_0^1 (1-t)^5 |\cos(th)|\, dt < \frac{h^5}{24} \int_0^1 (1-t)^5\, dt = \frac{h^5}{24}\frac{1}{6} < \frac{\pi^5}{2^5 \cdot 144} \sim 0.066 < 0.08$$

Problem 2. *(Folland 2.8.3/4)*

Last week we looked at an example of a non-degenerate critical point (the height function on the sphere). We saw that we can use the eigenvalues of the Hessian to determine whether a function has a maximum, minimum, or a saddle at a critical point. Now let's look at some examples to see how behaviour of a function near a degernate critical point can be more complicated.

1) Degenerate critical points can still be local extrema, $f_1(x,y) = x^2 + y^4$: Compute $\nabla f_1 = 2x\partial_x + 4y^3\partial_y = 0$ if and only if $x = 0$ and $y = 0$, so we have a critical point at the origin. Finding the Hessian:

$$\frac{\partial f_1}{\partial x \partial y} = 0,\ \frac{\partial^2 f_1}{\partial x^2} = 2,\ \frac{\partial^2 f_1}{\partial y^2} = 12y^2$$

So at the origin,

$$\frac{\partial^2 f_1}{\partial x_i \partial x_j} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

Therefore the critical point is degenerate because the Hessian has a zero eigenvalue. We now claim that the origin is a local minimum for $f_1(x,y)$. Indeed, $f_1(0,0) = 0$ while $x^2 > 0$ and $y^2 > 0$ for all $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$.

2) Degenerate critical points can be "saddle-like", $f_2(x,y) = x^2 - y^4$: Doing the same computation as before,

$$\nabla f_2 = 2x\partial_x - 4y^3\partial_y = 0 \Leftrightarrow (x,y) = (0,0)$$
$$\frac{\partial^2 f_2}{\partial x_i \partial x_j} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

So again, the critical point is degenerate. For any fixed x_0, as $y \to \pm\infty$ we have $\lim_{y\to\pm\infty} f_2(x_0, y) = -\infty$, while for any fixed y_0, $\lim_{x\to\pm\infty} f_2(x, y_0) = \infty$. Near the origin, then, $f_2(x,y)$ decreases in either of the y-directions, while it increases in either of the x-directions

3) Degenerate critical points can be isolated, but we can say nothing about the local behaviour, $f_3(x,y) = x^2 - y^3$:

$$\nabla f_3 = 2x\partial_x - 3y^2\partial_y = 0 \Leftrightarrow (x,y) = (0,0)$$

So again, f_3 has an isolated critical point at the origin, and the Hessian at this point is:

$$\frac{\partial^2 f_2}{\partial x_i \partial x_j} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

So the critical point is degenerate. Now notice that if we fix x_0, then $f(x_0, y) \to \infty$ as $y \to -\infty$, while $f(x_0, y) \to -\infty$ as $y \to \infty$. This means that at the origin, $f_3(x,y)$ behaves "non-degenerately" in the x-direction, while in the y-direction, the function can increase or decrease depending on whether or not we choose to increase or decrease y!

4) Degenerate critical points need not be isolated, $f_4(x,y) = x^2$: This is a critical difference between the non-degenerate case and the degenerate case.

$$\nabla f_4 = 2x\partial_x \;\Leftrightarrow\; (x,y) = (0,y)\;\forall\, y \in \mathbb{R}$$

5) Degenerate critical points can look like a minimum in any direction, but still not be a local minimum for the function, $f_5(x,y) = (y - x^2)(y - 2x^2)$: Let (at, bt) be any line through the origin, then

$$g(t) = f_5(at, bt) = (bt - a^2t^2)(bt - 2a^2t^2)$$

$$g'(t) = (b - 2a^2t)(bt - 2a^2t^2) + (bt - a^2t^2)(b - 4a^2t) \Rightarrow g'(0) = 0$$

$$g''(t) = -2a^2(bt - 2a^2t^2) + (b - 2a^2t)(b - 4a^2t) + (b - 2a^2t)(b - 4a^2t) - 4a^2(bt - a^2t^2) \Rightarrow g''(0) = 2b^2 > 0$$

So by the second derivative test, $g(t)$ has a minimum at zero for any line passing through the origin.

Now we can see that $(0,0)$ is not any local minimum for $f_5(x,y)$, even though it is a minimum for any line passing through the origin. First, calculate that $f(0,0) = 0$. Now, see that $f_5(x,y) > 0$ if and only if $y < x^2$ or $y > 2x^2$, and $f_5(x,y) < 0$ if and only if $y > x^2$ and $y < 2x^2$. Let $S = \big\{(x,y) \in \mathbb{R}^2 \mid f_5 > 0 \,\text{or}\, f_5 < 0\big\}$

Draw picture

Recall that a point (x_0, y_0) is a local minimum for f_5 if and only if there exists a neighbourhood U around (x_0, y_0) such that $f(x,y) > f(x_0,y_0)$ for all $(x,y) \in U$. Notice that this could never be the case, for $f(0,0) = 0$ and $(0,0) \in \partial S$ implies that any ball $B_\epsilon(0,0)$ has points in it such that $f(x,y) < 0$.

Exercise 1.1. *Suppose that $S \subseteq \mathbb{R}^n$ is compact, and $f : S \to \mathbb{R}$ is a smooth function with non-degenerate critical points. Explain why f can only have finitely many critical points.*

Problem 3. *(Folland 2.9.19) Let A be a symmetric $n \times n$ matrix, and let $f(x) = x^T Ax$ for $x \in \mathbb{R}^n$. Show that the maximum and minimum of f on the unit sphere $|x|^2 = 1$ are the largest and smallest eigenvalues of A.*

(1) Proceed by method of Lagrange multipliers
(2) Want to find extrema of $f(x)$ subject to the constraint $|x|^2 = 1$
(3) Write $f(x) = \sum_{ij} A_{ij}x_ix_j$ and let $G(x) = 1 - \sum_i x_i^2$
(4) The condition to optimize f subject to G is given by:

$$\nabla f = \lambda \nabla G$$

(5)

$$\nabla f = \partial_k f = \sum_{ij}(A_{ij}\delta_{ik}x_j + A_{ij}\delta_{jk}x_i) = \sum_j A_{kj}x_j + \sum_i A_{ik}x_i = 2\sum_j A_{kj}x_j = 2Ax$$

$$\nabla G = \partial_k G = \sum_i 2x_i\delta_{ik} = 2x_k$$

(6) Then the condition for extremizing f is that $Ax = \lambda x$
(7) Since A is symmetric, it is diagonalizable, so pick an orthonormal basis $\{y_j\}$ of $\mathbb{R}^n$ consisting of eigenvectors of A; so $Ay_j = \lambda_j y_j$.
(8) In this basis, we may write $x = \sum_j c_j y_j$, so that $f(x) = x^T Ax = \sum_i c_i^2 \lambda_i$
(9) Now it is clear that the extrema have values $f(y_j) = \lambda_j$, and thus the global max and min of f are the largest and smallest eigenvalues of A, respectively.