Assignment 7, due April 7, 2pm

About homework assignments

- Upload one PDF file with homework solutions to the Brightspace page for the appropriate assignment.
- You may write assignments on paper then photograph or scan them. If you do that, please collect all the images into a single pdf file to upload. You may use handwriting or typing on a tablet and upload it in pdf format. You may use LaTeX, but this is not encouraged because LaTeX is less expressive than handwriting and (for me) takes longer to prepare.
- Solutions must be uploaded before class starts on the day assignments are due.
- Solutions must be uploaded before class starts on the day assignments are due.
- Please check the Brightspace forum corresponding to the assignment before you start working on the assignment and from time to time while you are working on it. There may be questions, comments, or (alas!) corrections that will help you.
- Please post any comments or questions or possible corrections on the Brightspace forum for the assignment.
- Please email the instructor director directly with personal matters including requests for a homework deadline extension.
- Be follow the **academic integrity policies** that apply to this class as explained on the **class web page**. In particular, do not submit solutions prepared by AI tools.

1. (*Second derivative approximations for multi-variate functions*) [Much of the material in this exercise will be familiar from other classes, possibly in different notation.] Taylor series for multi-variate functions can be more complicated than they are for functions of one variable. A function of one variable has one derivative of every order: one first derivative, one second derivative, etc. A function of n variables has n first derivatives, which are the components of the gradient. It has $\frac{1}{2}n(n+1)$ second derivatives, and so on. The shape of the graph of a function of one variable depends on the signs of its derivatives. The first derivative says where it's increasing or decreasing. The second derivative says where it's convex or concave, etc. A *stationary point* is a place where the first derivative is zero. It is a local max or min if the second derivative is negative (local max) or positive (local min). A stationary point is local max or local min unless the second derivative is also zero, which is uncommon. A function of n variables has a gradient vector, which (see below) determines the directions in which it increases or decreases. It has a second derivative matrix, called the *hessian* matrix, which has n eigenvalues that are all real (see below). A stationary point (gradient equal to zero) is a local max or local min if all the eigenvalues of the hessian are negative (max) or positive (min). It's also possible that some eigenvalues are positive and others negative. This makes the stationary point a *saddle point*.

   Here is some notation and background for this exercise. $\phi(x) = \phi(x_1, \cdots, x_n)$ is a function of n

variables. For Exercise 2, we use $f(x)$ to denote the gradient vector:

$$f(x) = f(x_1, \cdots, x_n) = \begin{pmatrix} f_1(x) \\ \\ f_2(x) \\ \\ \vdots \\ \\ f_n(x) \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \phi}{\partial x_1} \\ \dfrac{\partial \phi}{\partial x_2} \\ \vdots \\ \dfrac{\partial \phi}{\partial x_n} \end{pmatrix}$$

This vector is often denoted by $\mathrm{grad}(\phi)$ or $\nabla \phi$. The hessian matrix is $H(x) = H(x_1, \cdots, x_n)$. The (j,k) entry of H is $h_{jk}(x)$:

$$h_{jk}(x) = \frac{\partial^2 \phi}{\partial x_j \, \partial x_k} \; .$$

If $v \in \mathbb{R}^n$ is an n component vector, we denote its *euclidean* norm by

$$\|v\| = \left(v^T v \right)^{\frac{1}{2}} = \left(\sum_{j=1}^{n} v_j^2 \right)^{\frac{1}{2}} \; .$$

We call v a *direction vector* or a *direction* if $\|v\| = 1$. The *inner product* (or, particularly in 2D or 3D, *dot product*) of column vectors u and v is written in various ways:

$$\langle u, v \rangle = u^T v = \sum_{j=1}^{n} u_j v_j \; . \tag{1}$$

The *Cauchy Schwarz inequality* is the inequality[1]

$$|\langle u, v \rangle| \le \|u\| \, \|v\| \; .$$

The angle between vectors (not a concept we use much in formulas) is written $\theta(u,v)$. It is defined by

$$\cos(\theta(u,v)) = \frac{\langle u, v \rangle}{\|u\| \, \|v\|} \; . \tag{2}$$

According to the Cauchy Schwarz inequality, right side is a number between -1 and 1. In 2D, this is the geometric angle between vectors u and v. In more than 2D is the the geometric angle in the plane determined by the origin, and u and v. From the geometry (and the proof, which is not given here) it is clear that $\langle u, v \rangle = \|u\| \, \|v\|$ if $\theta = 0$, which means that u and v point in the same direction. In other words,

$$\langle u, v \rangle = \|u\| \, \|v\| \quad \text{if and only if there is a } \textit{positive} \text{ number } s \text{ with } u = sv \; .$$

The situation corresponding to angle $\theta = \pi$ is $\langle u, v \rangle = -\|u\| \, \|v\|$. For this, it is *necessary and sufficient* that u and v point in exactly opposite directions. In other words

$$\langle u, v \rangle = -\|u\| \, \|v\| \quad \text{if and only if there is a } \textit{negative} \text{ number } s \text{ with } u = sv \; .$$

A symmetric $n \times n$ matrix has n real eigenvalues, not necessarily distinct, and n real orthogonal eigenvectors, not necessarily unique. That is, there are real numbers $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ and vectors $v_1, \cdots, v_n$ with

$$A v_j = \lambda_j v_j \qquad \langle v_j, v_k \rangle = v_j^T v_k = 0 \text{ if } j \ne k \; .$$

[1] The algebraic proof/derivation is not given here, but you can find one in Wikipedia or in a good linear algebra book.

The eigenvectors may be normalized so that $\|v_j\| = 1$. If $u \in \mathbb{R}^n$ is any vector, then u may be represented as a linear combination of eigenvectors as

$$u = \sum_{j=1}^{n} y_j v_j \quad , \qquad y_j = \langle v_j , u \rangle \; .$$

The *quadratic form*, $u^T A u$, may be represented in terms of the coefficients y_j and the eigenvalues λ_j as[2]

$$u^T A u = \sum_{j=1}^{n} \lambda_j y_j^2 \; . \tag{3}$$

A symmetric matrix is *positive definite* if $u^T A u > 0$ whenever $u \neq 0$. The representation formula (3) implies that A is positive definite if and only if all its eigenvalues are positive. It is *negative definite* if $u^T A u < 0$ whenever $u \neq 0$. This is equivalent to all the eigenvalues being negative. It is *indefinite* if there are eigenvalues of both signs.

(a) A first derivative approximation assumes Δx is small:

$$\Delta\phi \approx \langle f, \Delta x \rangle = \sum_{k=1}^{n} \frac{\partial \phi}{\partial x_k} \, \Delta x_k \; . \tag{4}$$

This formula may be written in other ways, including

$$\Delta\phi \approx \nabla\phi^T \Delta x \quad , \qquad \text{or} \quad \phi(x + \Delta x) \approx \phi(x) + \sum_{j=1}^{n} \frac{\partial \phi}{\partial x_j}(x) \, \Delta x_j \; .$$

What Δx with $\|\Delta x\| = r$ gives the greatest decrease in ϕ, according to the first derivative approximation? Assume that $f(x) \neq 0$. In formulas, solve the minimization problem:

$$\min_{\|\Delta x\| = r} \langle f, \Delta x \rangle \; .$$

The direction of this Δx is the *steepest descent* direction.

(b) A point x_* is a *local minimum* if $\Delta\phi > 0$ if Δx is small enough. A *critical point* (also called *stationary point*) is an x_* with $f(x_*) = 0$. Use the first derivative approximation (4) to show that a local minimum or local maximum is a critical point.

(c) The second derivative approximation is

$$\Delta\phi \approx f^T \Delta x + \frac{1}{2} \Delta x^T H \Delta x \; . \tag{5}$$

Show that a critical point is a local maximum or local minimum (using the second derivative approximation) if the hessian at the critical point is negative or positive definite respectively. *Comment* This is the multi-variate version of the "second derivative test" from one variable calculus. *Comment.* The second derivative approximation is accurate only when Δx is small. Therefore, the eigenvalues of the hessian only determine a local max or min.

(d) Show that if $H(x_*)$ is indefinite (both positive and negative eigenvalues) then x_* (a critical point) is neither a local maximum nor a local minimum. Such a critical point is called a *saddle point*.

[2] This formula is easy to derive, Feel free to do it as an exercise (not to hand in) or look it up in a good linear algebra book.

2. (*Gradient flow*) The *gradient flow* for a "potential function" ϕ is a differential equation in which $\dot{x}$ is a steepest descent direction.

$$\dot{x} = -f(x) = -\nabla\phi(x) \ . \tag{6}$$

   (a) Show that if $x(t)$ satisfies (6) then

$$\frac{d}{dt}\phi(x(t)) \leq 0 \ , \quad \text{and} \quad \frac{d}{dt}\phi(x(t)) < 0 \ , \quad \text{unless } \nabla\phi(x(t)) = 0 \ .$$

   (b) Use part (a) to show that a gradient flow cannot have a limit cycle. *Hint.* ϕ cannot keep going down return to an earlier value, despite the famous illustration where that happens.

   (c) Show that the linearization of (6) near a critical point never has a non-real eigenvalue. The phase plane near a critical point can be a source, a sink, or a saddle, but never a spiral.

   (d) Show that a critical point is stable (by linearized analysis) if and only if it is a local minimum.

3. (*A phase plane analysis*) Consider the ODE system

$$\begin{aligned} \dot{x} &= x - xy \\ \dot{y} &= y + 2x \end{aligned}$$

   Do an analysis that results in a sketch of the phase plane that includes

   - The critical points
   - The local behavior near critical points, as found by linearization and eigenvalue/eigenvector analysis for each linearization
   - Enough other vectors to get a clear picture of the global behavior.

4. (*The phase plane for a gradient flow*) Consider the "potential" function

$$\phi(x,y) = \Big[\,(x+y)^2 - 1\,\Big]^2 + (x-y)^2 \ .$$

   (a) Calculate $f(x,y) = \nabla\phi(x,y)$. *Comment* The formulas are a little messy.

   (b) Find the three critical points and classify them as local maxima, minima or saddle points using eigenvalues of the hessian at those points. *Hint.* It might be easier if you express ϕ in terms of the variables $\xi = x + y$ and $\eta = x - y$.

   (c) Do a phase plane analysis of the gradient flow equations

$$\begin{aligned} \dot{x} &= -\frac{\partial\phi}{\partial x} \\ \dot{y} &= -\frac{\partial\phi}{\partial y} \ . \end{aligned}$$

   Repeat the steps of Exercise 3 and determine the basin of attraction of each local minimum.

5. (*Taylor series, radius of convergence*) It can be hard to decide whether an infinite sum converges or not. Fortunately, Taylor series (also called power series representations) are simpler Let

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \ . \tag{7}$$

A series like this has a *radius of convergence*, r. If $|z| > r$ (we say that z is outside the radius of convergence), then the individual terms $|a_n z^n|$ tend to infinity,[3] as $n \to \infty$. Inside the radius of convergence, where $|z| < r$, the series converges "geometrically". This means that there is some constant C and some positive ratio $s < 1$ so that $|a_n z^n| \leq Cs^n$. Since the sum

$$\sum_{n=0}^{\infty} Cs^n$$

converges absolutely, the sequence of smaller numbers (7) also converges absolutely. To summarize: the power series converges like a geometric series inside the radius of converges and diverges outside. The borderline case, where $|z| = r$ is unpredictable. The power series might converge or diverge. Whether it converges or diverges might depend on the value of θ in $z = re^{i\theta}$. This exercise will be concerned only with the simpler cases where z is strictly inside ($|z| < r$) or strictly outside ($|z| > r$).

It might happen that the radius of convergence is $r = 0$. For example, the power series with $a_n = n!$ diverges for any $z \neq 0$. It might be that the radius of convergence is infinite. For example, the series with $a_n = \frac{1}{n!}$ converges for any z. Calculus books often refer to the *ratio test* to find the radius of convergence:

$$\text{Not always!} \qquad\qquad r = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \qquad\qquad \text{Not always!}$$

The problem is that this limit of ratios may not exist, as in example (b) below. Since the odd numbered a_n are zero, the ratios either are zero or undefined. However, every power series has a radius of convergence, even if the ratio limit does not exist.

There is a beautiful theorem in *complex analysis* that says that the radius of convergence is the "distance to the nearest singularity" in the complex plane. Here is a rough description of "nearest singularity". It should be enough for examples (a) and (b). In complex analysis, you learn that that $f(z)$ is differentiable inside the radius of convergence, and that if $f(z)$ is differentiable whenever $|z| < r'$, then $r > r'$. Therefore, you can find the radius of convergence by finding the z with smallest $|z|$ where $f(z)$ is undefined or not differentiable. That is the "distance to the nearest singularity.

In each example below, find the power series representation of f, either by the Taylor series formula

$$a_n = \frac{1}{n!} \frac{d^n f}{dx^n}(0)$$

or by writing f in the form $C\frac{1}{1-u} = C\sum_{n=0}^{\infty} u^n$. If you use the Taylor series formula, start computing the terms for $n = 0$, $n = 1$, $n = 2$, $n = 3$, etc., until you see the pattern. Identify the radius of convergence, either using the ratio test if it applies or by finding what z values have $|u| < 1$. Using the explicit formula for f, show that the radius of convergence found by analyzing the coefficients a_n is equal to the distance to the nearest singularity.

(a) $f(z) = \log(1 + z)$.

(b) $f(z) = \frac{1}{4+z^2}$.

[3]This is an over-simplification. In order to for an infinite sum converge, the terms in the sum have to tend to zero as n goes to infinity. This does not happen if $|z| > r$. It might not be strictly true that $|a_n z^n| \to \infty$ as $n \to \infty$. But it is true in a *subsequence*. There are terms n_k with $n_k \to \infty$ as $k \to \infty$ so that the n_k terms in the sum do go to infinity: $|a_{n_k} z^{n_k}| \to \infty$ as $k \to \infty$. For example, the power series representation of part (b) has $a_n = 0$ if n is odd. The even numbered terms do go to infinity as $n \to \infty$ if $|a| > r$. The subsequence is the even numbered terms, which we express through the subsequence $n_k = 2k$.

6. Download and run the demo Python file `power_demo.py`. Play with the parameters (trial and error) to get a feel for how many terms are needed to achieve a given accuracy. What range of x gets three digit accuracy using only terms up to x^2? How many terms are needed to get five digit accuracy when $x = 5$ or $x = -5$?