Monsoon Semester (Aug-Nov)
Topics in Applied Optimization (TAO)

Tutorial I

Instructor: Dr. Pawan Kumar

INSTRUCTIONS:

The codes can be written in either C, C++, MATLAB, Octave, FORTRAN, or Python. Write useful comments, and do proper indentation. Solution to the assignment problems will be discussed in Tutorials (to be decided after add/drop) every week.

1. **(Inner Product Spaces)** Given a vector space V (over $\mathbb{R}$) equipped with an inner product denoted by $\langle u, v\rangle$, prove the following:
   1. **(triangle inequality)** $\|u+v\| \leq \|u\| + \|v\|, \quad \forall u, v \in V.$
   2. **(parallelogram law)** $\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2)$, $\forall u, v \in V$.
   3. Show that $\langle u+v, u-v\rangle = \|u\|^2 - \|v\|^2, \quad \forall u, v \in V.$
   4. Suppose $u, v \in V$. Prove that $\langle u, v\rangle = 0$ if and only if $\|u\| \leq \|u + av\|$.
   5. Suppose $u, v \in V$. Prove that $\|au+bv\| = \|bu+av\|$ for all $a, b \in \mathbb{R}$ if and only if $\|u\| = \|v\|$.
   6. Prove that
   $$16 \leq (a+b+c+d)(1/a + 1/b + 1/c + 1/d)$$
   for all positive numbers a, b, c, d.
   7. Suppose $u, v \in V$ are such that
   $$\|u\| = 3, \quad \|u+v\| = 4, \quad \|u-v\| = 6.$$
   What number does $\|v\|$ equal?
   8. Consider the vector space of set of all $n \times n$ matrices:
   $$V = \{A \in \mathbb{R}^{n\times n}\}.$$
   Show that
   $$\text{trace } (A) = \sum_{i=1}^{n} A_{ii}$$
   defines an inner product
   $$\langle A, B\rangle = \text{trace}(B^T A)$$
   on V. Here A_{ii} denote the entries on the diagonal of A.
   9. Prove or disprove (or see proofs in some textbook) the following:
      (a) $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$.
      (b) row rank and column rank of a matrix is always same.

(c) eigenvalues of a symmetric matrix is always real.
(d) if determinant of a matrix is zero, then A is rank deficient.
(e) sum of two SPD (symmetric positive definite) matrices is SPD.
(f) eigenvalues of SPD matrices are ≥ 0.

2. **(Multivariable Calculus)**

   1. Give an example of (non-zero and non-constant) function $f : X \to Y$, where
      (a) $X = \mathbb{R}^n, \quad Y = \mathbb{R}$.
      (b) $X = \mathbb{R}, \quad Y = \mathbb{R}^n$.
      (c) $X = \mathbb{R}^m, \quad Y = \mathbb{R}^n$.
      (d) $X = \mathbb{R}^{m\times n}, \quad Y = \mathbb{R}^n$.
      (e) $X = \mathbb{R}^{m\times n}, \quad Y = \mathbb{R}$.
      (f) $X = \mathbb{R}^{m\times n}, \quad Y = \mathbb{R}^{m\times n}$.
   2. For the following functions $f : X \to Y$, find suitable X and Y. In the following, $w \in \mathbb{R}^n, A \in \mathbb{R}^{n\times n}$ are constant vector and matrices.
      (a) $f(x) = Ax$.
      (b) $f(x) = x^T Ax$.
      (c) $f(x) = Ax - w$
   3. Find the derivatives of the following:
      (a) $f(x) = \sin(x) + \cos(x)$.
      (b) $f(x) = \exp(x)\sin(x)$.
      (c) $f(x) = \dfrac{\exp x}{x}$.
      (d) $f(x) = \log(\sin(x))$.
      (e) $f(x) = \exp(\sin x^2)/x^2 + x/\sin(x)$.
   4. Check whether the following functions are linear and if they are linear then find the matrix A such that $f(x) = Ax$.
      (a)
      $$f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \sum_{i=1}^{n} x_i.$$
      (b)
      $$f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \\ \vdots \\ x_n - x_1. \end{bmatrix}.$$
      (c) If the linear function f is bijective, what can you conclude about rank of matrix A?
   5. **(Numerical Differentiation)** The numerical derivative is obtained from the definition of derivative in terms of the limit. We know that:
      $$\frac{df}{dx} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

When h is "small enough", we can approximate the derivative numerically. To obtain the derivative of the function at the point $x = \alpha$, we choose h small enough (not smaller than machine precision!) such that we have:

$$\frac{df}{dx}|_{x=\alpha} \approx \frac{f(\alpha + h) - f(\alpha)}{h}. \quad (1)$$

This method is called automatic differentiation (in short, *autodiff* or *autograd*, however, autodiff is a bit different in that one may apply chain rule to compute derivatives) in the sense that it can be automated by writing a computer program. For the following functions, compute the numerical derivative and derivative by hand, and then plot both the derivatives. For example, for $f(x) = \sin(x), x \in [0, 2\pi]$ do the following steps:

(a) Compute derivative by hand, it comes out to be $\cos(x)$.

(b) Now compute numerical derivatives. Since the interval $[0, 2\pi]$ has infinitely many points, we need to choose finitely many out of these. Let us choose the discrete points $\{0, 0.01, 0.02, ..., 2\pi\}$, i.e., points with spacing of 0.01; hence, $h = 0.01$ (spacing between consecutive points), which is needed for numerical derivatives (1) above. For the following functions, plot the derivatives computed by hand, and the numerical derivative in the same figure. Do they match?

   i. $f(x) = 1/x, \ x \in [0.1, 0.2]$. Take 500 equally spaced points.

   ii. $f(x) = x \sin x, \ x \in [-\pi, \pi]$. Take 1000 equally spaced points.

6. **(Gradient)** The gradient of a single valued multi variable finction

$$f : \mathbb{R}^n \to \mathbb{R}$$

is defined as follows:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

Here the partial derivative $\frac{\partial f}{\partial x_i}$ simply means taking derivative of f with respect to variable x_i, and treating the rest constant. Compute gradients and draw level curves for the following functions:

(a) $f \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + y^2$.

(b) $f \begin{bmatrix} x \\ y \end{bmatrix} = \sin x + \cos y$.

(c) $f \begin{bmatrix} x \\ y \end{bmatrix} = \exp(x * y)$

Can you already see how to compute numerical derivative of these multivariable functions?