ME5114 Sp22 **Homework # 1**
Assigned Jan 14, 2022 Due: Jan 22 , 2022

- This homework is mainly about the required math and Python coding background of this course. Depending on your previous training, you may need to read related materials online. Some background materials and tutorials are uploaded to Blackboard and course website
- You can type your solution in Latex (encouraged but not required) or just hand write the solution on paper. Submit your homework (pdf) through Blackboard
- To receive credits, please write down all the necessary steps leading to the final answer.

---

1. **Python**: Please go over the Python tutorial posted on the course website. Make sure you can write basic python codes.

2. **System Installation**: In this class, we will be using Drake (a robot simulator) in class, and also for homework and projects. Drake runs on Mac-OS or Ubuntu. If you do not have a Mac computer, please install Ubuntu either directly on your computer or through Windows Subsystem for Linux (WSL). Some instructions are given in the following video.

   `https://www.bilibili.com/video/BV1Hb4y1U7fo/`

   You don't have to follow all the steps in the video. As long as you have a working Ubuntu system and Python environment, you are good to go. For this homework, please test your installation using the "CartPoleExample.ipynb" file on `https://github.com/GitWeiZhang/teaching/tree/main/ME424_F21/Drake_Examples`. Please attach a few snapshots of the Meshcat visualization of your simulation results.

3. **Lipschitz Continuity**

   (a) Please state the formal definition of continuous functions

   (b) Please state the formal definitions of Lipschitz continuity and locally Lipschitz continuity.

   Solution:
   (a): Suppose X and Y are metric spaces, $E \in X$, $p \in E$, and f maps E into Y. Then f is said to be continuous at p if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

   $$d_Y(f(x), f(p)) < \varepsilon \tag{1}$$

   for all points $x \in E$ for which $d_X(x, p) < \delta$. d_X denotes the metric on X.
   If f is continuous at every point of E, then f is said to be continuous on E.

   (b): Suppose X and Y are metric spaces, $E \in X$, $p \in E$, and f maps E into Y. Then f is said to be Lipschitz continuous if there exists a real constant $L \geq 0$ such that

   $$d_Y(f(x_1), f(x_2)) \leq L d_X(x_1, x_2) \tag{2}$$

   for all x_1 and x_2 in E.
   f is called locally Lipschitz continuous if for every x in E there exists a neighborhood U of x such that f restricted to U is Lipschitz continuous.

   □

4. **Matrix calculus**

(a) Let $f : \mathbb{R}^{n\times m} \to \mathbb{R}$ be a scalar function of matrix variable. Please write a tutorial paragraph explaining (in your own words) the meaning of $\frac{\partial}{\partial X} f(X)$.

(b) Let $A \in \mathbb{R}^{n\times m}, X \in \mathbb{R}^{m\times n}$. Derive an expression for $\frac{\partial}{\partial X} tr(AX)$ (show your derivation steps; your derivation should be directly from the definition of matrix derivatives)

(c) Derive an expression for $\frac{\partial}{\partial x} f(x)$, where $f(x) = x^T Qx + tr(xx^T)$ and $x \in \mathbb{R}^n$

Solution:

(a): Here we use the "broad definition" in [1] for $\frac{\partial}{\partial X} f(X)$.

$$\frac{\partial}{\partial X} f(X) = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1m}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2m}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{n1}} & \frac{\partial f}{\partial x_{n2}} & \cdots & \frac{\partial f}{\partial x_{nm}} \end{bmatrix} \tag{3}$$

or as

$$\frac{\partial}{\partial X} f(X) = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{21}} & \cdots & \frac{\partial f}{\partial x_{n1}} \\ \frac{\partial f}{\partial x_{12}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{n2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{1m}} & \frac{\partial f}{\partial x_{2m}} & \cdots & \frac{\partial f}{\partial x_{nm}} \end{bmatrix} \tag{4}$$

For (3), every column of $\frac{\partial}{\partial X} f(X)$ can be regarded as the gradient of $f(X)$ in the corresponding column of X.

(b):

$$tr(AX) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} x_{ji} \tag{5}$$

combine (5) and (3) (or (4)), it is obvious that

$$\frac{\partial}{\partial X} tr(AX) = A^T \tag{6}$$

(c):

$$x^T Qx = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_3 \end{bmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i q_{ij} x_j \tag{7}$$

$$tr(xx^T) = \sum_{i=1}^{n} x_i^2 \tag{8}$$

So

$$\frac{\partial}{\partial x}f(x) = \begin{bmatrix} \sum_{i=1}^{n} q_{1i}x_i + \sum_{i=1}^{n} q_{i1}x_i + 2x_1 \\ \sum_{i=1}^{n} q_{2i}x_i + \sum_{i=1}^{n} q_{i2}x_i + 2x_2 \\ \vdots \\ \sum_{i=1}^{n} q_{ni}x_i + \sum_{i=1}^{n} q_{in}x_i + 2x_n \end{bmatrix} = Qx + Q^T x + 2x \tag{9}$$

□

5. **Inner product**

   (a) Describe the way to calculate the angle between two vectors $x, y \in \mathbb{R}^n$ using inner product

   (b) Trace can be used to define inner products for matrices. Let $A, B \in \mathbb{R}^{m\times n}$, then $\langle A, B\rangle = tr(A^T B)$. Compute the angle between the following two matrices

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad B = \begin{bmatrix} -1 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

Solution:

(a):

$$\alpha = \arccos\left(\frac{\langle x, y\rangle}{||x||||y||}\right) \tag{10}$$

(b):

$$\alpha = \arccos\left(\frac{\langle A, B\rangle}{\sqrt{\langle A, A\rangle}\sqrt{\langle B, B\rangle}}\right) = \arccos\left(\frac{tr(A^T B)}{tr(A^T A)tr(B^T B)}\right) = \frac{\pi}{2} \tag{11}$$

□

6. **Some linear algebra**

   (a) State the condition on A such that $Ax = b$ has at least one solution

   (b) Let $A = [a_1, a_2, a_3, a_4]$, where $a_i \in \mathbb{R}^n$ are columns of A. Suppose a_1, a_2 are linearly independent, and $a_3 + a_1 = a_2$ and $a_4 - a_3 = a_1$. Compute $rank(A)$ and $Null(A)$

   (c) Given a vector $y \in \mathbb{R}^n$ and a matrix $A \in \mathbb{R}^{n\times m}$, find an expression of the projection of y onto the column space of A

Solution:

(a): $Ax = b$ has at least one solution if and only if b is a linear combination of the columns of A ($b \in C(A)$).

(b): We know that a_3 and a_4 are both the linear combination of a_1 and a_2. So the dimension of its column space is 2, then

$$rank(A) = 2 \tag{12}$$

On the other hand, $a_1 - a_2 + a_3 = 0$, $a_1 + a_3 - a_4 = 0$ can be rewritten as

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} = \mathbf{0} \tag{13}$$

and

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix} = \mathbf{0} \tag{14}$$

So the $Null(A)$ is span$\{c_1, c_2\}$ where $c_1 = \begin{bmatrix} 1 & -1 & 1 & 0 \end{bmatrix}^T$ and $c_2 = \begin{bmatrix} 1 & 0 & 1 & -1 \end{bmatrix}^T$.

(c): A linear mapping $P : X \to X$ is called a projection if it satisfies

$$P^2 = P \tag{15}$$

where X is a linear space. Further, if the projection is self-adjoint which means $\langle x, Py \rangle = \langle Px, y \rangle$, the projection is called orthogonal projection. When we do not specify the metric of a linear space, it is usually considered a Euclidean space. And the "projection" always means "orthogonal projection".

If X is a linear space, V is a subspace of X and $V^\perp$ is the orthogonal complement of V. This can be denoted as

$$X = V \oplus V^\perp$$

where $\oplus$ means direct addition.
The orthogonal projection of X onto V can be defined as

$$P_V : X \longrightarrow X \qquad P_V : v + v_\perp \longmapsto v \tag{16}$$

where $v \in V$ and $v_\perp \in V^\perp$.

Now let us deal with the question (c). Column space of A ($C(A)$) is a subspace of $\mathbb{R}^n$, if we want to find the projection of y onto $C(A)$, according to (16), the projection can be given by

$$P_{C(A)}(y) = y_\parallel \tag{17}$$

where $y = y_\parallel + y_\perp$, $y_\parallel \in C(A)$ and $y_\perp \in C(A)^\perp$. So we need to find the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_k$ such that

$$y_\parallel = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_k u_k \tag{18}$$

where $C(A) = \text{span}\{u_1, u_2, \dots, u_k\}$.

We can get the coefficients in (18) as follows:
From $y = y_\parallel + y_\perp$ we can know that

$$\begin{aligned}
\alpha_1\langle u_1, u_1\rangle + \alpha_2\langle u_2, u_1\rangle + \cdots + \alpha_k\langle u_k, u_1\rangle + \langle y_\perp, u_1\rangle &= \langle y, u_1\rangle \\
\alpha_1\langle u_1, u_2\rangle + \alpha_2\langle u_2, u_2\rangle + \cdots + \alpha_k\langle u_k, u_2\rangle + \langle y_\perp, u_2\rangle &= \langle y, u_2\rangle \\
&\vdots \\
\alpha_1\langle u_1, u_k\rangle + \alpha_2\langle u_2, u_k\rangle + \cdots + \alpha_k\langle u_k, u_k\rangle + \langle y_\perp, u_k\rangle &= \langle y, u_k\rangle
\end{aligned} \tag{19}$$

or in "einstein notation"

$$\alpha_j\langle u_j, u_i\rangle + \langle y_\perp, u_i\rangle = \langle y, u_i\rangle \qquad i = 1, 2, \dots, k \tag{20}$$

If you are very familiar with "einstein notation", the solution for α_j is straight forward. Here we show the rest of solution from (19). Using $x^T y = \langle x, y\rangle$ we can get

$$\begin{bmatrix} u_1^T u_1 & u_1^T u_2 & \cdots & u_1^T u_k \\ u_2^T u_1 & u_2^T u_2 & \cdots & u_2^T u_k \\ \vdots & \vdots & \ddots & \vdots \\ u_k^T u_1 & u_k^T u_2 & \cdots & u_k^T u_k \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix} = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_k^T \end{bmatrix} y \tag{21}$$

We set $U = \begin{bmatrix} u_1 & u_2 & \cdots & u_k \end{bmatrix}$ and we can rewrite (21) in matrix form as

$$U^T U \alpha = U^T y \tag{22}$$

Because $U^T U$ is invertible and combine (18) we know

$$P_{C(A)}(y) = y_\parallel = U\alpha = U(U^T U)^{-1} U^T y \tag{23}$$

So if A is full column rank, the projection is given by

$$P_A = A(A^T A)^{-1} A^T \tag{24}$$

Now let's proof that So the expression of the projection of y onto the column space of A is

$$P_A(y) = A(A^T A)^{-1} A^T y \tag{25}$$

If the A is not full column rank, let $A' = [u_1 \quad \dots \quad u_k]$ where $u_1, u_2, ..., u_k$ is a basis of the column space of A. Then the expression of the projection of y onto the column space of A is given by

$$P_A(y) = A'(A'^T A')^{-1} A'^T y \tag{26}$$

Or you can use $P_A(y) = \arg\ \min\{\|y - p\| : p \in C(A)\}$. $\square$

7. **Ellipsoids:** Ellipsoid in $\mathbb{R}^n$ have two equivalent representations: (i) $E_1(P, x_c) = \{x \in \mathbb{R}^n : (x - x_c)P^{-1}(x - x_c) \leq 1\}$ and (ii) $E_2(A, x_c) = \{Au + x_c : \|u\|^2 \leq 1\}$. Given an eillipsoid $E_1(P, x_c)$ with P positive definite, its volume is $\nu_n\sqrt{det(P)}$ where ν_n is the volume of unit ball in $\mathbb{R}^n$, its semi-axes directions are given by the eigenvectors of P and the lengths of semi-axes are $\sqrt{\lambda_i}$, where λ_i are eigenvalues of P.

   (a) Given an Ellipsoid $E_1(P, x_c)$, find the corresponding (A, b) (in terms of P and x_c) such that $E_2(A, b)$ represents the same ellipsoid as $E_1(P, x_c)$

   (b) Draw the ellipse $E_1(P, x_c)$ with $P = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$ and $x_c = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ by hand.

Solution:

(a): For representation (i), the positive definite P can be decomposed as $P = LL^T$ where L is invertible. So

$$(x - x_c)^T P^{-1}(x - x_c) = (x - x_c)^T {L^{-1}}^T L^{-1}(x - x_c) \leq 1 \tag{27}$$

which leads

$$||L^{-1}(x - x_c)|| \leq 1 \tag{28}$$

If we let $L^{-1}(x - x_c) = u$, it is easily to know

$$Lu + x_c = x \tag{29}$$

where $||u|| \leq 1$ (because of (28)).

Now we can find that (29) has the same formulation with the representation (ii). So give a ellipsoid $E_1(P, x_c)$, the another representation is $E_2(L, x_c)$ where $P = LL^T$.

(b):

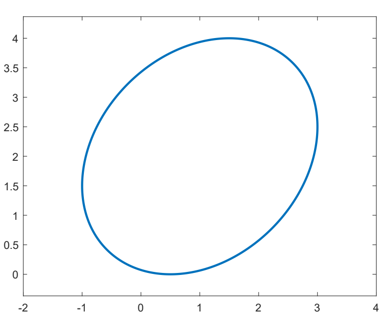

Figure 1: The ellipse $E_1(P, x_c)$

□

8. **Linear System Solution**: Consider the following linear control system:

$$\dot{x}(t) = Ax(t) + Bu(t), \quad \text{with } x(0) = x_0$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}^m$ is the control input. Show

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau$$

is a solution to the above control system.

Solution:

We have

$$\begin{aligned} \frac{d}{dt}e^{At} &= \frac{d}{dt}(I + At + \frac{A^2}{2!}t^2 + \cdots) \\ &= A + A^2t + \frac{A^3}{2!}t^2 + \cdots \\ &= A(I + At + \frac{A^2}{2!}t^2 + \cdots) \\ &= Ae^{At} \end{aligned} \tag{30}$$

$$\frac{\partial}{\partial t}\int_{t_0}^t f(t,\tau)d\tau = \int_{t_0}^t \left(\frac{\partial}{\partial t}f(t,\tau)\right)d\tau + f(t,\tau)|_{\tau=t} \tag{31}$$

Now we show

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau$$

is a solution to the above control system:

$$\begin{aligned} \frac{d}{dt}x(t) &= Ae^{At}x_0 + \int_0^t Ae^{A(t-\tau)}Bu(\tau)d\tau + e^{A(t-\tau)}Bu(\tau)|_{\tau=t} \\ &= A\left(e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau\right) + Bu(\tau) \\ &= Ax(t) + Bu(\tau) \end{aligned} \tag{32}$$

And $x(0) = e^{\mathbf{0}}x_0 = x_0$ which is consistent with the initial condition.

□

References

[1] Jan R. Magnus. On the concept of matrix derivative. *Journal of Multivariate Analysis*, 101(9):2200–2206, 2010.