AI501 Homework 1

20200130 Yujun Kim

March 2023

1 Exercise #1.2

We want to prove
$$\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$$
If $x \in \mathcal{N}(A^T)$, then $A^T x = 0$. Thus, $a_i^T x = 0$ for all columns a_i of A. Hence $x \in \mathcal{R}(A)^\perp$. i.e. $\mathcal{N}(A^T)$ is a subspace of $\mathcal{R}(A)^\perp$. By rank nullity theorem, $dim(\mathcal{N}(A^T)) = m - r = m - dim(\mathcal{R}(A)) = dim(\mathcal{R}(A)^\perp)$, where $A \in \mathcal{M}_{m\times n}$ and r is the rank of A. Hence, $\mathcal{R}(A)^\perp, \mathcal{N}(A^T)$ has same dimension which proves $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$.

Now, apply above equality to A^T. Then,
$$\mathcal{R}(A^T)^\perp = \mathcal{N}((A^T)^T) = \mathcal{N}(A)$$
Take orthogonal space on both side, which the second desired equality follows.
$$(\mathcal{R}(A^T)^\perp)^\perp = \mathcal{R}(A^T) = \mathcal{N}(A)^\perp$$
Notice that the fact that $(W^\perp)^\perp = W$ for some subspace W of V, where V is finite dimensional vector space also follows by the dimension argument: $dim(W^\perp) = dim(V) - dim(W)$.

2 Exercise #1.5

Before the proof, observe following facts:

- Kronecker product is well defined between matrix of arbitrary size.
- If $A \in \mathcal{M}_{m\times n}, B \in \mathcal{M}_{p\times q}$, then $A \otimes B \in \mathcal{M}_{np\times mq}$.
- Suppose $A \in \mathcal{M}_{m\times n}, B \in \mathcal{M}_{p\times q}$. By definition of the Kronecker product, for any $i_1 \in [n], j_1 \in [m], i_2 \in [p], j_2 \in [q]$, we have
$$(A \otimes B)_{p(i_1-1)+i_2, q(j_1-1)+j_2} = a_{i_1,j_1} b_{i_2,j_2}$$

- $i_1 \in [n], j_1 \in [m], i_2 \in [p], j_2 \in [q]$ fully characterize each element of $A \otimes B \in \mathcal{M}_{np\times mq}$ by $(A\otimes B)_{p(i_1-1)+i_2,q(j_1-1)+j_2}$. In other words, for any $i_3 \in [np], j_3 \in [mq]$, there exists $i_1 \in [n], j_1 \in [m], i_2 \in [p], j_2 \in [q]$ with $(i_3, j_3) = (p(i_1-1)+i_2, q(j_1-1)+j_2)$

Now for the following equalities, let the dimension of every operation on the LHS is well given(i.e. dimension of matrices that should be added is same). Using the second observation, it follows that every operation on RHS is well defined(i.e. dimension of matrices that should be added is same). Also, let indices i, j play around appropriate range(e.g. $i \in [n], j \in [m]$ to describe element $a_{i,j}$ of $A \in \mathcal{M}_{m\times n}$).

2.1 $A \otimes (B + C) = A \otimes B + A \otimes C$

Trivial as $a_{i,j}(B + C) = a_{i,j}B + a_{i,j}C$.

2.2 $(B + C) \otimes A = B \otimes A = C \otimes A$

Trivial as $(b_{i,j} + c_{i,j})A = b_{i,j}A + c_{i,j}A$.

2.3 $A \otimes B \neq B \otimes A$

Take $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, B = I_2$. Then,

$$A \otimes B = \begin{bmatrix} 0 & I_2 \\ I_2 & 0 \end{bmatrix} \neq \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} = B \otimes A$$

2.4 $((A \otimes B) \otimes C = A \otimes (B \otimes C)$

Let $A \in \mathcal{M}_{m\times n}, B \in \mathcal{M}_{p\times q}, C \in \mathcal{M}_{r,s}$. Then,

$$\begin{aligned}
&(A \otimes B) \otimes C)_{r(p(i_1-1)+i_2-1)+i_3,s(q(j_1-1)+j_2-1)+j_3} \\
&= (A \otimes B)_{p(i_1-1)+i_2,q(j_1-1)+j_2} c_{i_3,j_3} \\
&= (a_{i_1,j_1} b_{i_2,j_2}) c_{i_3,j_3} \\
&= a_{i_1,j_1} (b_{i_2,j_2} c_{i_3,j_3}) \\
&= a_{i_1,j_2} (B \otimes C) r(i_2 - 1) + i_3, s(j_2 - 1) + j_3 \\
&= (A \otimes (B \otimes C))_{pr(i_1-1)+r(i_2-1)+i_3,qs(j_1-1)+s(j_2-1)+j_3}
\end{aligned}$$

Note that $(r(p(i_1 - 1) + i_2 - 1) + i_3, s(q(j_1 - 1) + j_2 - 1) + j_3) = (pr(i_1 - 1) + r(i_2 - 1) + i_3, qs(j_1 - 1) + s(j_2 - 1) + j_3)$. i.e. indices in the first and last lines of above equalities are same. Thus, $(A \otimes B) \otimes C = A \otimes (B \otimes C)$

2.5 $(A \otimes B)^T = A^T \otimes B^T$

$$(A \otimes B)^T = \begin{bmatrix} a_{11}B^T & \cdots & a_{m1}B^T \\ \vdots & \ddots & \vdots \\ a_{n1}B^T & \cdots & a_{mm}B^T \end{bmatrix} = A^T \otimes B^T$$

2.6 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

Let $A \in \mathcal{M}_{n,n}, B \in \mathcal{M}_{p,p}$ which is necessary for the existance of inverse of A, B. Let $A^{-1} = C$. Then,

$$\begin{aligned}(A \otimes B)(A^{-1} \otimes B^{-1}) &= \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{bmatrix} \begin{bmatrix} c_{11}B & \cdots & c_{1n}B \\ \vdots & & \vdots \\ c_{n1}B & \cdots & c_{nn}B \end{bmatrix} \\ &= \begin{bmatrix} I_p & 0 & \cdots & 0 \\ 0 & I_p & cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & I_p \end{bmatrix} = I_{np}\end{aligned}$$

This is because

$$[a_{i1}B \cdots a_{in}B] \begin{bmatrix} c_{1j}B^{-1} \\ \vdots \\ c_{nj}B^{-1} \end{bmatrix} = (a_{i1}c_{1j} + \cdots a_{in}c_{nj})BB^{-1} = \delta_{ij}I_p$$

Similarly, $(A^{-1} \otimes B^{-1})(A \otimes B) = I_{np}$ and so $A \otimes B$ is invertible and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

3 Exercise #1.6

We want to prove $(A \otimes B)(C \otimes D) = AC \otimes BD$. For AC, BD to be well defined, let $A \in \mathcal{M}_{m\times n}, C \in \mathcal{M}_{n\times l} B \in \mathcal{M}_{p\times q}, D \in \mathcal{M}_{q\times r}$. Then, $A \otimes B \in \mathcal{M}_{mp\times nq}, C \otimes D \in \backslash\mathrm{II} \times \updownarrow\nabla$. Thus, matrix multiplication of LHS is well defined and the resulting matrices of both sides have same dimension($mp \times lr$)

For $(i_1, j_1) \in [m] \times [l], (i_2, j_2) \in [p] \times [r]$,

$$\begin{aligned}(AC \otimes BD)_{p(i_1-1)+i_2, r(j_1-1)+j_2} &= (AC)_{i_1,j_1}(BD)_{i_2,j_2} \\ &= (\sum_{k=1}^{n} a_{i_1k}c_{kj_1})(\sum_{h=1}^{q} b_{i_2h}d_{hj_2})\end{aligned}$$

$$\begin{aligned}((A\otimes B)(C\otimes D))_{p(i_1-1)+i_2,r(j_1-1)+j_2} &= \sum_{k=1}^{nq}(A\otimes B)_{p(i_1-1)+i_2,k}(C\otimes D)_{k,r(j_1-1)+j_2} \\ &= \sum_{k=1}^{n}\sum_{h=1}^{q}(A\otimes B)_{p(i_1-1)+i_2,q(k-1)+h}(C\otimes D)_{q(k-1)+h,r(j_1-1)+j_2} \\ &= \sum_{k=1}^{n}\sum_{h=1}^{q}(a_{i_1k}b_{i_2h})(c_{kj_1}d_{hj_2}) \\ &= \sum_{k=1}^{n}\sum_{h=1}^{q}(a_{i_1k}c_{kj_1})(b_{i_2h}d_{hj_2})\end{aligned}$$

Thus, the statement holds.

4 Exercise #1.7

Let $A \in \mathcal{M}_{m\times n}, B \in \mathcal{M}_{n\times l}, C \in \mathcal{M}_{p\times m}$. Note that $CAB \in \mathcal{M}_{p\times n}$ to that $Vec(CAB) \in \mathbb{R}^{pl}$. $B^T \otimes C \in \mathcal{M}_{lp\times nm}, Vec(A) \in \mathbb{R}^{nm}$ so that $(B^T\otimes C)Vec(A) \in \mathbb{R}^{pl}$.

Observe that for any matrix $X \in \mathcal{M}_{m\times n}$,

$$Vec(X)_{m(j-1)+i} = x_{ij}$$

By the observation, for $(i,j) \in [p]\times[l]$

$$Vec(CAB)_{p(j-1)+i} = (CAB)_ij = \sum_{a=1}^{m}\sum_{b=1}^{n} c_{ia}a_{ab}b_{bj}$$

$$\begin{aligned}(B^T\otimes C)Vec(A)_{p(j-1)+i} &= \sum_{a=1}^{m}\sum_{b=1}^{n}(B^T\otimes C)_{p(j-1)+i,m(b-1)+a}Vec(A)_{m(b-1)+a} \\ &= \sum_{a=1}^{m}\sum_{b=1}^{n}(B^T)_{jb}C_{ia}A_{ab} \\ &= \sum_{a=1}^{m}\sum_{b=1}^{n}c_{ia}a_{ab}b_{bj}\end{aligned}$$

Thus, the statement holds.

5 Exercise #1.11

(1) Reltive entropy $f(x,y) = y\ln y - y\ln x$. Then,

$$A := \nabla^2 f = \begin{bmatrix} \frac{y}{x^2} & -\frac{1}{x} \\ -\frac{1}{x} & \frac{1}{y}\end{bmatrix}$$

Then,$det(A - \lambda I = \lambda^2 - (\frac{1}{y} + \frac{y}{x^2})\lambda + \frac{2}{x^2}$. This shows that all eigenvalues of $\nabla^2 f$ is positive so that $\nabla f \geq 0$. By exercise 1.12(see below), f is convex.

(2) Indicator Let $\mathcal{C}$ be a convex set. For $x, y \in \mathcal{H}$. Let $k \in [0,1], z = (1-k)x + ky$. Clearly if $k = 0$ or $k = 1$, then $f(z) \leq (1-k)f(x) + kf(y)$. Now suppose $k \in (0,1)$

- If $x \notin \mathcal{C}$ or $y \notin \mathcal{C}$, then, $\iota_C(x) = \infty$ or $\iota_C(y) = \infty$. Thus, the $(1-k)\iota_C(x) + k\iota_C(y) = \infty$ so that the inequality for convex holds.
- If $x, y \in \mathcal{C}$, then $z \in \mathcal{C}$. Thus, $\iota_C(z) = 0 \leq (1-k)\iota_C(x) + k\iota_C(y)$.

(3) Support function For each $y \in C$, $x \mapsto < y, x >$ is a linear map so it is convex. Supremum of convex functions is convex. Thus, support functions are convex.

(4) p-norm $p - norm$ is a norm. Thus,

$$||(1-k)x+ky||_p \leq ||(1-k)x||_p + ||ky||_p = |1-k|||x||_p + |k|||y||_p = (1-k)||x||_p + k||y||_p$$

i.e. $|| \cdot ||_p$ is convex

(5) Max function Each $x \mapsto x_i$ is a linear map so it is convex. Thus, maximum of those function, which is the supremum of convex functions is convex. Hence, max function is convex.

6 Exercise #1.12

We show $(a) \Leftrightarrow (b), (b) \Leftrightarrow (c), (b) \Rightarrow (d), (d) \Rightarrow (c)$. This suffices.

[(a) $\Rightarrow$ (b)] For $k \in [0,1]$,

$$f(x+k(y-x)) - f(x) = f((1-k)x+ky) - f(x) \leq (1-k)f(x) + kf(y) - f(x) = k(f(y) - f(x))$$

Thus,

$$\begin{aligned} < y - x, \nabla f(x) > = f'(x, y - x) &= \lim_{k \to 0} \frac{f(x + k(y-x) - f(x)}{k} \\ &= \lim_{k \to 0+} \frac{f(x + k(y-x) - f(x)}{k} \leq f(y) - f(x) \end{aligned}$$

Hence (b) holds.

[(b) $\Rightarrow$ (a)] Let $x, y \in \mathcal{H}$, $t \in [0,1]$ be given. Let $z(t) = (1-t)x + ty$. By (b),

$$f(x) \geq f(z(t)) + t < y - x, \nabla f(z(t)) \quad (1)$$
$$f(y) \geq f(z(t)) + (1-t) < y - x, \nabla f(z(t)) \quad (2)$$

$(1-t) \times (1) + t \times (2)$ gives

$$(1-t)f(x) + tf(y) \geq f(z(t)) = f((1-t)x + ty)$$

Hence (a) holds.

[(b) $\Rightarrow$ (c)]

$$f(y) \geq f(x) + < y - x, \nabla f(x) > \tag{3}$$
$$f(x) \geq f(y) + < x - y, \nabla f(y) > \tag{4}$$

(1) + (2) gives

$$0 \geq < y - x, \nabla f(x) > + < x - y, \nabla f(y) > = < y - x, \nabla f(x) - \nabla f(y) >$$
$$\Leftrightarrow 0 \leq < y - x, \nabla f(y) - \nabla f(x) >$$

[(c) $\Rightarrow$ (b)] Let $x, y \in \mathcal{H}, k \in [0,1]$. Take $z(t) = (1-t)x + ty$ as above, $g(t) = f(z(t))$.

$$f(y) = g(1) = g(0) = \int_0^1 g'(t)dt = f(x) + \int_0^1 < y - x, \nabla f(z(t)) > dt$$

By (c), $< y - x, \nabla f(z(t)) - \nabla f(x). = \frac{1}{t} < z(t) - x, \nabla f(z(t)) - \nabla f(x) > \geq 0$. Thus, $< y - x, \nabla f(z(t)) > \geq < y - x, \nabla f(x) >$. Put last inequality in the above equality and obtain

$$f(y) \geq f(x) + \int_0^1 < y - x, \nabla f(x) > dt = f(x) + < y - x, \nabla f(x) >$$

[(b) $\Rightarrow$ (d)] By Taylor,

$$f(x + \epsilon) = f(x) +, \epsilon, \nabla f(x) > + < \epsilon, \nabla^2 f(x) \epsilon > + o(\epsilon^2)$$

By (b), $f(x + \epsilon) \geq f(x) + < \epsilon, \nabla f(x) >$. Thus, $< \epsilon, \nabla^2 f(x) \epsilon > \geq 0$ for all sufficiently small ϵ. i.e. $\nabla^2 f(x) \geq 0$.

[(d) $\Rightarrow$ (c)] Let $g(t) = \nabla f(z(t))$. Then,

$$g(1) - g(0) = \int_0^1 \frac{d}{dt} g(t) dt = \int_0^1 \nabla^2 f(z(t))(y - x) dt$$

Then,

$$< y - x, \nabla f(y) - \nabla f(x) > = < y - x, g(1) - g(0) > = \int_0^1 < y - x, \nabla^2 f(z(t))(y - x) > dt \geq 0$$

7 Exercise #1.13

$$\partial f(x) = \begin{cases} \{-1\} \text{ if } x < 0 \\ \{1\} \text{ if } x > 0 \\ [-1, 1] \text{ if } x = 0 \end{cases}$$

(1) For $x > 0$, we have $|y| = f(y) \geq y = x + (y - x) = f(x) + < y - x, 1 >$. If $u \neq 1$, either $|x + 1| \ngeq |x| + < 1, u >$ if $u > 1$ and $|x/2| \ngeq |x| + < x/2 - x, u >$ if $u < 1$. Hence, $\partial f(x) = \{1\}$.

(2) By symmetry, $\partial f(x) = \{-1\}$ for $x < 0$.
(3) Let $x = 0$. If $u \in [-1, 1]$,

$$|y| \geq |0| + < y - 0, u >= uy$$

Thus, $u \in \partial f$. If $|u| > 1$, then for $y = sign(u) = |u|/u$,

$$|y| = 1 \ngeq |0| + < y - 0, u >= |u|$$

Thus, we have $\partial f(0) = [-1, 1]$.

8 Exercise #1.16

$$\begin{aligned}
(\alpha f)^*(u) &= \sup_{x \in \mathcal{H}} \{< u, x > -\alpha f(x)\} \\
&= \sup_{x \in \mathcal{H}} \{\alpha < \frac{u}{\alpha}, x > -\alpha f(x)\} \\
&= \alpha \sup_{x \in \mathcal{H}} \{< \frac{u}{\alpha}, x > -f(x)\} \\
&= \alpha f^*(\frac{u}{\alpha})
\end{aligned}$$

Thus, $(\alpha f)^*(\cdot) = \alpha f^*(\frac{\cdot}{\alpha})$.

9 Exercise #1.17

$$f * (u) = \sup_x \{< u, x > -f(x)\} = \sup_x \{< u, x > -(\frac{1}{2} x^T x - \begin{bmatrix} 1 \\ 1 \end{bmatrix})^T x\}$$

Let $g_u(x) =< u, x > -(\frac{1}{2} x^T x - \begin{bmatrix} 1 \\ 1 \end{bmatrix})^T x$. Solve $\nabla_x g_u(x) = u - x + \begin{bmatrix} 1 \\ 1 \end{bmatrix}$=0. The solution is $x^* = \begin{bmatrix} u_1 + 1 \\ u_2 + 1 \end{bmatrix}$. Then,

$$f^*(u) = g_u(x^*) = \frac{1}{2}((u_1 + 1)^2 + (u_2 + 1)^2)$$

This is the convex conjugate of f.

10 Exercise #1.20

First, find f^*. Let $p_u(x) =< u, x > -||y - x||_2^2$. Then, $\nabla_x p(x) = u - 2(x - y) = 0$ gives $x = y + u/2$. Thus, $f^*(u) = \sup_x \{< u, x > -||y - x||_2^2\} =< u, y + u/2 > -\frac{1}{4} u^T u =< u, y > + \frac{1}{4} u^T u$.

Next, find g^*. Let $q_u(x) =< u, x > -||x||_1$. $g^*(u) = \sup_x q_u(x) = \begin{cases} 0 & \text{if } ||u||_\infty \le 1 \\ -\infty & \text{if } ||u||_\infty > 1 \end{cases}$
By example in p.24 of the textbook, the dual of the given primal problem is

$$\begin{aligned}(D) :&= -\min_u f^*(A^T u) + g^*(-u) \\ &= -\min_u \frac{1}{4} u^T A A^T u + < A^T u, y > + g^*(u) \\ &= -\min_u \frac{1}{4} u^T A A^T u + y^T A u \text{ with respect to } ||u||_\infty \le 1\end{aligned}$$