16| (a) $\mathbb{R}^{n\times m} \mapsto \mathbb{R}$

$X \mapsto tr(AX^TB)$

$f(X+\varepsilon) = f(X) + Df(X)(\varepsilon) + o(\varepsilon)$

$$tr(A(X+\varepsilon)^TB) = tr(AX^TB + A\varepsilon^TB)$$
$$= tr(AX^TB) + tr(A\varepsilon^TB)$$
$$= tr(BAX^T) + tr(BA\varepsilon^T)$$
$$= \underbrace{\langle BA, X\rangle_F}_{f(x)} + \underbrace{\langle BA, \varepsilon\rangle_F}_{D_f f(x)\varepsilon} + \underbrace{0}_{O(\|\varepsilon\|)}$$

Differential of f at x $D_x f(x)\varepsilon$: $\langle BA, \varepsilon\rangle_F$

Derivative of f at x : BA

(b) $\mathbb{R}^{n\times n} \mapsto \mathbb{R}^{n\times n}$

$X \mapsto tr(X^n)$

$f(X+\varepsilon) = f(X) + Df(X)\varepsilon + o(\varepsilon)$

First we consider that $(X+\varepsilon)^n$

$(X+\varepsilon)^n = X^n + nX^{n-1}\varepsilon + \frac{n(n-1)}{2}X^{n-2}\varepsilon^2 + \cdots$

taking trace of this term

$$tr((X+\varepsilon)^n) = tr(X^n) + n\, tr(X^{n-1}\varepsilon) + \frac{n(n-1)}{2} tr(X^{n-2}\varepsilon^2) + \cdots$$
$$= tr(X^n) + n\, tr(X^{n-1}\varepsilon) + O(\|\varepsilon\|)$$
$$= \underbrace{tr(X^n)}_{f(x)} + \underbrace{\langle n(X^{n-1})^T, \varepsilon\rangle_F}_{Df(x)\varepsilon} + O(\|\varepsilon\|)$$

Differential of f at x $D_x f(x)\varepsilon$: $\langle n(x^{n-1})^T, \varepsilon\rangle_F$

Derivative of f at x : $n(x^{n-1})$

(c) $\mathbb{R}^{n\times n} \longmapsto \mathbb{R}^{n\times n}$

$X \longmapsto tr(e^X)$

$f(x+\varepsilon) = f(x) + \langle D, \varepsilon \rangle_F + o(\|\varepsilon\|_F)$

$tr(e^{X+\varepsilon}) = tr\left(\sum_{k=0}^{\infty} \frac{1}{k!}(X+\varepsilon)^k\right)$

$= \sum_{k=0}^{\infty} \frac{1}{k!} tr((X+\varepsilon)^k)$

$=$

17 (a) Consider two arbitrary points x and y in P,

where $x_i \geq 0$, $y_i \geq 0$, $\Sigma_i x_i = 1$, $\Sigma_i y_i = 1$

Let $z = \lambda x + (1-\lambda) y$, $0 \leq \lambda \leq 1$

We need to show that z is in P

$z_i = \lambda x_i + (1-\lambda) y_i$

since $x_i \geq 0$ and $y_i \geq 0$, then $\lambda x_i + (1-\lambda) y_i \geq 0$

$\Sigma_i z_i = \Sigma_i \lambda x_i + \Sigma_i (1-\lambda) y_i$

$= \lambda \Sigma_i x_i + (1-\lambda) \Sigma_i y_i$

Since $\Sigma_i x_i = 1$ and $\Sigma_i y_i = 1$

$\Sigma_i z_i = \lambda + 1 - \lambda = 1$

Therefore $\Sigma_i z_i = 1$

$\Rightarrow$ any convex combination of points in P is also in P.

So P is a convex subset of $\mathbb{R}^n$.

(b) Assume that A is a vector $(a_1, \cdots, a_n)$

$E(p) = \sum_{i=1}^{n} p_i \cdot a_i = Ap$

This is an affine function: $p \mapsto Ap$

Since p is convex, then $E(p)$ is convex

(c) $Var(p) := \sum_{i=1}^{n} p_i (a_i - E(p))^2$

$= \sum_{i=1}^{n} p_i a_i^2 - E(p)^2$

Let $h(x) = x^2$

then $h''(x) = 2 \geq 0$

Therefore $h(x) = x^2$ is a convex function

18) (a) For $f(x,y) = \frac{x}{y}$, the Hessian matrix is

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial^2 y} \end{bmatrix}$$

Compute 1st and 2rd derivatives:

$$\frac{\partial f}{\partial x} = \frac{1}{y}, \quad \frac{\partial f}{\partial y} = \frac{-x}{y^2}, \quad \frac{\partial^2 f}{\partial x \partial y} = -\frac{1}{y^2}, \quad \frac{\partial^2 f}{\partial^2 x} = 0, \quad \frac{\partial^2 f}{\partial^2 y} = \frac{2x}{y^3}$$

So $H = \begin{bmatrix} 0 & -\frac{1}{y^2} \\ -\frac{1}{y^2} & \frac{2x}{y^3} \end{bmatrix}$

Now, check the definiteness of H.

for any vector $[a, b]$, if H is positive semidefinite, then

$$[a, b] \, H \begin{bmatrix} a \\ b \end{bmatrix} \geq 0$$

$$\Rightarrow [a \;\; b] \begin{bmatrix} 0 & -\frac{1}{y^2} \\ -\frac{1}{y^2} & \frac{2x}{y^3} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \geq 0$$

$$\Rightarrow \quad -\frac{b^2}{y^2} + \frac{2abx}{y^3} \geq 0$$

consider that $a=1$ and $b=1$

we can get $-\frac{1}{y^2} + \frac{2x}{y^3}$

if y is positive and x is negative, this expression can be negative

Thus, the Hessian matrix is not positive, the function is not convex.

(b) The quadratic function $f: \mathbb{R}^n \to \mathbb{R}$ is convex iff. the Hessian matrix f is positive semidefinite.

$$f(x) = x^T A x + bx + C$$

$$H = \nabla^2 f(x) = 2A$$

$$\Rightarrow \ 2A \succeq 0$$

This means that A should be positive semidefinite or all eigenvalues of A are nonnegative.

(c) $f: P \to \mathbb{R}$ $\quad f(p) := \sum_{i=1}^{n} p_i \log p_i$

Now, let's find $\frac{\partial f}{\partial x_j}$

$$\frac{\partial f}{\partial P_j} = \frac{\partial}{\partial P_j}\left(\sum_{i=1}^{n} P_i \log(P_i)\right)$$

$$= \sum_{i=1}^{n}\left(\frac{\partial}{\partial P_j} P_i \log(P_i)\right)$$

$$= \sum_{i=1}^{n}\left(\frac{\partial P_i}{\partial P_j}\log(P_j) + P_i \frac{\partial \log(P_i)}{\partial P_j}\right)$$

for $i = j$,

$$\frac{\partial f}{\partial P_j} =$$