1. matrix decomposition
   singular value decomposition:
   $A = U\Sigma V^T$
   $Ax = U\Sigma V^T x$

2. eigenvalue decomposition
   For $A = A^T$:
   Let S^n be the set of real symmetric metrices of size n.
   $Av = \lambda v \implies \lambda$ is a eigenvalue of A, v is eigenvector of A
   $A = Q\Sigma Q^T$, Q is a matrix with columns of eigenvectors

3. real symmetric matrices have real eigenvalues:
   proof:
$$\begin{aligned} Av &= \lambda v \\ v^* Av &= v^* \lambda v = \lambda \|v\|_2^2 \\ (v^* Av)^* &= v^* A^* v = \lambda^* \|v\|_2^2 \\ &\implies \lambda = \lambda^* \end{aligned}$$

4. positive (semi)-definite: $A \geq 0$
   A real symmetric, $(\forall v) v^T Av \geq 0 \implies A$ PSD
   A real symmetric, $(\forall v) v^T Av > 0 \implies A$ PD

5. $A \in S_+^n$ iff all eigenvalues are non-negative
   proof:
$$\begin{aligned} v^T AV &= v^T Q\Sigma Q^T v \\ v^T AV &= \tilde{v} Q\Sigma \tilde{v} = \sum \lambda_i \|v_i\|^2 \\ (\forall i)\lambda_i \geq 0) &\implies (\forall v) v^T Av \geq 0 \end{aligned}$$

6. for PSD matrices, we can take square root
   $A^{\frac{1}{2}} = Q\Sigma^{\frac{1}{2}} Q^T$, $A^{\frac{1}{2}} A^{\frac{1}{2}} = A = Q\Sigma Q^T$

7. functions of matrices $f : S^n_{++} \rightarrow \mathbb{R}$
   how to compute gradient
   $f(x) = log(detX)$

$$\begin{aligned}
f(X+\delta X) &= logdet(X+\delta X)\\
&= logdet(X^{\frac{1}{2}}(I + X^{-\frac{1}{2}})\delta X X^{-\frac{1}{2}})X^{\frac{1}{2}})\\
&= logdet(X) + logdet(I + X^{\frac{1}{2}}\delta X X^{-\frac{1}{2}})\\
&\text{let } M = X^{\frac{1}{2}}\delta X X^{-\frac{1}{2}}\\
&= logdet(X) + logdet(I+M)
\end{aligned}$$

claim eigenvalues of $I + M$: $1 + \lambda_i$

$$\begin{aligned}
Mv_i &= \lambda_i v_i\\
(I+M)v_i &= (1+\lambda_i)v_i\\
det(M) &= \prod_i (1+\lambda_i)\\
f(X+\delta X) &= logdet(X) + log\prod_i(1+\lambda_i)\\
&= logdet(X) + \sum_i log(1+\lambda_i)\\
&\approx logdet(X) + \sum_i \lambda_i\\
&\approx logdet(X) + trace(X^{-\frac{1}{2}}\delta X X^{\frac{1}{2}})\\
&\approx logdet(X) + trace(X^{-1}\delta X)\\
trace(X^{-1}\delta X) &= < X^{-T}, \delta X >\\
f(X+\delta X) &= f(X) + < \nabla f(X), \delta X > \implies \nabla f(X) = X^{-1}\\
logdet(X) &= log(X) \implies f'(X) = \frac{1}{X}
\end{aligned}$$

8. 2nd order approximation

$$f(X+\delta X) = f(X) + < \nabla f(X), \delta X > +1/2 < \delta X, \nabla^2 f(x) \delta X >$$

first look at first order approximation: $g(X) = X^{-1}$

$$\begin{aligned} g(X+\delta X) &= (X+\delta X)^{-1} = (X^{\frac{1}{2}}(I + X^{-\frac{1}{2}}\delta X X^{-\frac{1}{2}})X^{\frac{1}{2}})^{-1} \\ &= X^{-\frac{1}{2}}(I + X^{-\frac{1}{2}}\delta X X^{-\frac{1}{2}})^{-1} X^{-\frac{1}{2}} \\ &\text{for small A(small eigenvalues): } (I+A)^{-1} \approx I - A \\ &= X^{-\frac{1}{2}}(I - X^{-\frac{1}{2}}\delta X X^{-\frac{1}{2}}) X^{-\frac{1}{2}} \\ &= X^{-1} - X^{-1}\delta X X^{-1} \end{aligned}$$

$$logdet(X+\delta X) = logdet(X) + tr(X^{-1}\delta X) - \frac{1}{2} tr(\delta X X^{-1} \delta X X^{-1})$$

analogy:

$$\begin{aligned} g(x) &= \frac{1}{X} \\ g'(x) &= \frac{-1}{X^2} = -X^{-2} \end{aligned}$$

9. convex sets
affine sets: a set $C \subseteq \mathbb{R}^n$ is affine if $(\forall x_1, x_2 \leq C)(\forall \theta \in \mathbb{R}) \implies \theta x_1 + (1-\theta)x_2 \in C$
can use this definition to check sets for convexity property
eg: set of solution to a set of linear equations is affine
affine combination: $\sum_i \theta_i x_i, \forall \theta_i \in \mathbb{R}, \sum_i \theta_i = 1$

10. convex sets
a set $C \subseteq \mathbb{R}^n$ is convex if $(\forall x_1, x_2 \in C)(\forall \theta \in \mathbb{R}) 0 \leq \theta \leq 1 \implies \theta x_1 + (1-\theta)x_2 \in C$
convex combination: $\sum_i \theta_i x_i, \forall \theta_i \in \mathbb{R}, \sum_i \theta_i = 1, \theta_i \geq 0$
convex hull: the set of all convex combinations of points in C, the hull is convex
any affine set is convex, since convex property is a constrained version of the affine property

11. hyperplane

$$C = \{x : a^T x = b\}, a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R}$$

12. halfspaces

$$\begin{aligned} C &= \{x : a^T x \leq b\}, a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R} \\ &\text{let } a^T x_c = b \\ &= \{x : a^T (x - x_c) \leq 0\}, a \in \mathbb{R}^n, a \neq 0 \end{aligned}$$