1. project info
   proposal: mar.6
   presentation: apr.3
   report: 5 pages, IEEE style, apr.13

2. dual norm
   $\|z\|_*:= \sup_x\{z^T x : \|x\|_p \leq 1\}$

   dual norm of L1-norm:
   $\|z\|_*:= \sup_x\{z^T x : \|x\|_1 \leq 1\}$
   max $\sum_i z_i x_i$,
   subject to : $\sum_i \|x_i\| \leq 1$
   select x_i corresponding to z_i with maximum absolute value
   equivalent to $\|z\|_* = \|z\|_\infty$

   dual norm of L-∞-norm:
   $\|z\|_*:= \sup_x\{z^T x : \|x\|_\infty \leq 1\}$
   max $\sum_i z_i x_i$,
   subject to : $\|x_i\| \leq 1, \forall i$
   choose $x_i = 1$ if $z_i \geq 0$ and $x_i = 0$ if $z_i < 0$
   equivalent to $\|z\|_* = \|z\|_1$

   in general the dual norm of L_p-norm is L_q-norm where $1/p + 1/q = 1$

3. calculus
   consider $f : \mathbb{R}^n \to \mathbb{R}$
   gradient of f: $\nabla f(x) = \begin{bmatrix} \partial f/\partial x_i \\ .. \end{bmatrix}$
   $f(x) = a^T x \implies \nabla f(x) = a$
   $f(x) = x^T P x, P = P^T \implies \nabla f(x) = 2Px$
   $f(x) = x^T P x \implies \nabla f(x) = 2(\frac{P^T+P}{2})x = (P^T + P)x$

   approximation, Taylor expansion:
   $f(x) \approx f(x_0) + \nabla^T f(x_0)(x - x_0) + o((x - x_0)^2)$
   $f(x + \delta x) \approx f(x_0) + \nabla^T f(x)\delta x + o((\delta x)^2)$

   chain rule:
   $f : \mathbb{R} \to \mathbb{R}, g : \mathbb{R} \to \mathbb{R}, h(x) = f(g(x))$
   $\nabla h(x) = g'(f(x))\nabla f(x)$

$g : \mathbb{R}^m \to \mathbb{R}, g(x) = f(Ax + b)$
$\nabla g(x) = A^T \nabla f(Ax + b)$

2nd derivative:
$\nabla^2 f(x) = \begin{bmatrix} \partial^2 f/\partial x_1 \partial x_1 & ... \\ .. & \partial^2 f/\partial x_n \partial x_n \end{bmatrix}$
$\nabla f(x) = Px + g$
$\nabla^2 f(x) = P$
Hessian gives the 2nd order approximation:
$f(x) \approx f(x_0) + \nabla^T f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$

4. matrices
   $A \in \mathbb{R}^{m \times n}$: set of all real matrices
   inner product: $\sum_i \sum_j x_{ij} y_{ij} = trace(XY^T) = trace(Y^T X) = \sum_i (XY)_{ii}$
   note trace has cyclic property
   frobenius norm: $\|X\|_F = (\sum_i \sum_j X_{ij}^2)^{\frac{1}{2}}$
   range: $R(A) = \{Ax : x \in \mathbb{R}^n\} = \sum_i a_i x_i$, where a_i is ith column (column space of A)
   null space: $N(A) = \{x : Ax = 0\}$

5. matrix decomposition
   SVD:
   $A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T$
   U and V are left and right eigenvector matrixes
   U and V are orthogonal matrixes($BB^T = B^T B = I$)
   Σ is rectangular diagonal matrix of eigenvalues

   rank:
   number of nonzero eigenvalues

   $A_{m \times n} x_n$
   linear transformation: $U \Sigma V^T x$
   rotation - scaling - rotation