1、A^+：**若A满列秩**，则$A^+=L=(A^HA)^{-1}A^H$，**左伪逆矩阵。若A满行秩**，则$A^+=R=A^H(AA^H)^{-1}$，**右伪逆矩阵。**

因为$Range(A)=Range(AA^+A)\subset Range(AA^+)\subset Range(A)$，所以$Range(AA^+)=Range(A)$，同时由于$AA^+AA^+=AA^+$是幂等矩阵，所以$AA^+$是**$A$列空间的投影**。同理，$Range(A^H)=Range(A^H(A^+)^HA^H)\subset Range(A^H(A^+)^H)\subset Range(A^H)$，又因为$Range(A^H(A^+)^H)=Range(A^+A)$，所以$Range(A^+A)=Range(A^H)$，同时由于$A^+AA^+A=A^+A$是幂等矩阵，所以$A^+A$是$A^H$列空间的投影，即**$A$行空间的投影。**

2、右$Kronecker$积：$A\otimes B=[a_1B,\ldots,a_nB]$。**左$Kronecker$积**：$B\otimes A=[A\otimes B]_{left}=[Ab_1,\ldots,Ab_q]$。**广义$Kronecker$积**：$\{A_i\}_N\otimes B=[A_1b^1,..,A_Nb^N]^T$。

3、奇异值分解：$A=U\Sigma V^T$，**左奇异向量**：$(AA^T)u_i=\lambda_iu_i$、**右奇异向量**：$(A^TA)v_i=\lambda_iv_i$、**奇异值矩阵**：$\sigma_i=\frac{Av_i}{u_i}$。其中$U^TU=I,V^TV=I$。

几何意义：V表示原始域的标准正交基，U表示经过A变换后的辅助域的标准正交基。矩阵A的作用是将一个向量X先在V这组正交基向量的空间作旋转，然后对每个方向进行一定的缩放（缩放因子就是各个特征值），最后在U这组正交基向量的空间再次作旋转。

4、偏导数：$D_X=\frac{\partial}{\partial X^T}$，$D_Xf(X)=\frac{\partial f(X)}{\partial X^T}$。

$Jacobian$**矩阵**：$D_Xf(X)=\frac{\partial f(X)}{\partial X^T}$，$D_{vecX}f(X)=\frac{\partial f(X)}{\partial vecX^T}$，$D_XF(X)=\frac{\partial vec(F(X))}{\partial (vecX)^T}$。

梯度：$\nabla_X=\frac{\partial}{\partial X}$。

微分：$df(X)=tr(AdX)\Leftrightarrow D_Xf(X)=A$。

$[\frac{\partial X^TAX}{\partial X^T}]_i=\frac{\partial}{\partial X_i}\sum_{k=1}^n\sum_{l=1}^n a_{kl}x_kx_l=\sum_{k=1}^n a_{kl}x_k+\sum_{l=1}^n a_{il}x_l$，$Df(X)=X^T(A+A^T)$。
$[\frac{\partial f(X)}{\partial X^T}]_{ij}=\frac{\partial f(X)}{\partial x_{ji}}=\sum_{k=1}^m\sum_{l=1}^m\sum_{p=1}^n\frac{\partial(a_kx_{kp}x_{lp}b_l)}{\partial x_{ji}}=\sum_{k=1}^m\sum_{l=1}^m\sum_{p=1}^n[a_kx_{lp}b_l\frac{\partial x_{kp}}{\partial x_{ji}}+a_kx_{kp}b_l\frac{\partial x_{lp}}{\partial x_{ji}}]$

$=\sum_{l=1}^m a_jx_{li}b_l+\sum_{k=1}^m a_kx_{ki}b_j=[X^Tb]_ia_j+[X^Ta]_ib_j$，$D_Xf(X)=X^T(ba^T+ab^T)$。

$[\frac{\partial tr(XB)}{\partial X^T}]_{ij}=\frac{\partial}{\partial x_{ji}}\sum_{p=1}^m\sum_{l=1}^n x_{pl}b_{lp}=\sum_{p=1}^m\sum_{l=1}^n b_{lp}\frac{\partial x_{pl}}{\partial x_{ji}}=b_{ij}$，$D_Xtr(XB)=B$。

$\frac{\partial f_{kl}}{\partial x_{ij}}=\frac{\partial(AXB)_{kl}}{\partial x_ij}=\frac{\partial(\sum_{u=1}^m\sum_{v=1}^m a_{ku}x_{uv}b_{vl})}{\partial x_{ij}}=a_{kl}b_{jl}$，$D_X(AXB)=B^T\otimes A$。$D_X(AX^TB)=(B^T\otimes A)K_{mn}$。

5、奇异值分解应用令$M_{n\times p}$和$N_{n\times k}$两个矩阵，考虑约束极小化问题$min_{A\in R^{p\times k}}||M-NA^T||_F^2\quad s.tA^TA=I_{k\times k}$。假设$M^TN$的截尾SVD为$UDV^T$，其中$D=diag(d_1,d_2,\ldots,d_r),r=rank(M^TN)$，则优化问题的最优解为$\hat{A}=UV^T$。**证明**：由于$A^A=I$，所以$||M-NA^T||_F^2=tr((M-NA^T)^T(M-NA^T))=tr(M^TM)-2tr(M^TNA^T)+tr(N^TN)$。于是，极小化$||M-NA^T||_F^2$只需极大化$tr(M^TNA^T)$即可。利用$M^TN$的SVD有$tr(M^TNA^T)=tr(UDV^TA^T)=tr(UDA^{*T})=tr(A^{*T}UD)$，其中$A^*=AV$。由于$V$是$k\times r$列正交的且$A^TA=I$，所以$A^{*T}A^*=I_{p\times p}$，即$A^*$的列正交。既然$D$为正对角元的对角矩阵，所以$tr(A^{*T}UD)$为$A^{*T}U$的对角元的非负线性组合，从而当$A^{*T}U$的对角元均为极大时，$tr(A^{*T}UD)$才达到极大。而$A^{*T}U$的对角元分别为$a_1^*u_1,\ldots,a_r^*u_r$.由$Cauchy-Schwartz$不等式得$a_j^*u_j\le||a_j^*||\times||u_j||=1\times1=1$，当且仅当$a_j^*=u_j$时等号成立。即当$A^*=U$时$A^{*T}U$的对角元均达到极大值1。于是，$A^*=AV$得优化问题的极小解即$tr(M^TNA^T)$的极大解为$\hat{A}=A^*V^T=UV^T$。

6、L_1正则化最小二乘问题：$min_xJ(\lambda,x)=\frac{1}{2}||y-Ax||_2^2+\lambda||x||_1$目标函数梯度为
$\nabla_XJ(\lambda,x)=\frac{\partial J(\lambda,x)}{\partial x_i}=-A^T(y-AX)+\lambda\nabla x||x||_1=-c+\lambda\nabla_x||x||_1,\nabla_x||x||_1=\frac{\partial||x||_1}{\partial x_i}=+1,[-1,+1],-1,$
$c_i=+\lambda,[-\lambda,+\lambda],-\lambda$

7、非负矩阵分解：①平方$Eculidean$距离最小化乘法：$min_{A,S}D_E(X||AS)=\frac{1}{2}||X-AS||_F^2$，其梯度下降算法为：
$a_{ik}\leftarrow a_{ik}-\mu_{ik}\frac{\partial D_E(X||AS)}{\partial a_{ik}}$，$S_{kj}\leftarrow S_{kj}-\eta_{kj}\frac{\partial D_E(X||AS)}{\partial S_{kj}}$，其中$\frac{\partial D_E(X||AS)}{\partial a_{ik}}$为$\frac{\partial D_E(X||AS)}{\partial A}=-(X-AS)S^T$第$(i,k)$元素，
$\frac{\partial D_E(X||AS)}{\partial s_{kj}}$为$\frac{\partial D_E(X||AS)}{\partial S}=-A^T(X-AS)$第$(k,j)$元素。若$\mu_{ik}=\frac{a_{ik}}{[ASS^T]_{ik}}$，$\eta_{kj}=\frac{S_{kj}}{[A^TAS]_{kj}}$，则有以下乘法运算：
$a_{ik}\leftarrow a_{ik}\frac{[XS^T]_{ik}}{[ASS^T]_{ik}}$，$S_{kj}\leftarrow S_{kj}\frac{[A^TX]_{kj}}{[A^TAS]_{kj}}$。

②KL散度：$D_{KL}(X||AS)=\sum_{i=1}^N\sum_{j=1}^M(x_{ij}ln\frac{x_{ij}}{[AS]_{ij}}-x_{ij}+[AS]_{ij})$的最小化问题。梯度为
$\frac{\partial D_{KL}(X||AS)}{\partial a_{ik}}=\sum_{j=1}^M(-\frac{S_{kj}x_{ij}}{[AS]_{ij}}+S_{kj})$，$\frac{\partial D_{KL}(X||AS)}{\partial S_{kj}}=\sum_{i=1}^N(-\frac{a_{ik}x_{ij}}{[AS]_{ij}}+a_{ik})$，从而其梯度下降为
$a_{ik}\leftarrow a_{ik}-\mu_{ik}\sum_{j=1}^M(-\frac{S_{kj}x_{ij}}{[AS]_{ij}}+S_{kj})$，$S_{kj}\leftarrow S_{kj}-\eta_{kj}\sum_{i=1}^N(-\frac{a_{ik}x_{ij}}{[AS]_{ij}}+a_{ik})$。令$\mu_{ik}=\frac{a_{ik}}{\sum_{j=1}^M S_{kj}}$，$\eta_{kj}=\frac{S_{kj}}{\sum_{i=1}^N a_{ik}}$，则改写为乘法运算$a_{ik}=a_{ik}\frac{\sum_{j=1}^M S_{kj}x_{ij}/[AS]_{ij}}{\sum_{j=1}^M S_{kj}}$，$S_{kj}\leftarrow S_{kj}\frac{\sum_{i=1}^N a_{ik}x_{ij}/[AS]_{ij}}{\sum_{i=1}^N a_{ik}}$。

8、实对称矩阵的特征值是实数，证明：设实对称矩阵A的复数特征值为λ，对应n维非零向量X，使得$AX=\lambda X$成立，设$\bar{\lambda}$为λ的共轭复数，因为A是实对称矩阵，所以$A=\bar{A}=A^T$，则$A\bar{X}=\bar{A}\bar{X}=\bar{AX}=\bar{\lambda X}=\bar{\lambda}\bar{X}$，同时左乘$\bar{X}^TAX=\bar{X}^T\lambda X=\lambda\bar{X}^TX$，又因为$\bar{X}^TAX=\bar{X}^TA^TX=(A\bar{X})^TX=(\bar{\lambda}\bar{X})^TX=\bar{\lambda}\bar{X}^TX$，即$(\lambda-\bar{\lambda})\bar{X}^TX=(\lambda-\bar{\lambda})[\bar{X},X]=0$，又因为非零向量$X$内积$>0$，即$\lambda=\bar{\lambda}$

9、正定矩阵特征值大于0，证明：设正定矩阵A有一个非正的特征值k，对应有非零特征向量x，则$Ax=kx$，但$x^HAx=kx^Hx<=0$，这与A正定矛盾。

10、L矩阵是半正定矩阵，证明：对任意向量f，有
$f^TLf=f^T(D-W)f=f^TDf-f^TWf=\sum_{i=1}^n d_if_i^2-\sum_{i,j=1}^n w_{ij}f_if_j=\frac{1}{2}(\sum_{i=1}^n d_if_i^2-2\sum_{i,j=1}^n w_{ij}f_if_j+\sum_{j=1}^n d_jf_j^2)$

$=\frac{1}{2}[\sum_{i=1}^n(\sum_{j=1}^n w_{ij})f_i^2-2\sum_{i,j=1}^n w_{ij}f_if_j+\sum_{j=1}^n(\sum_{i=1}^n w_{ji})f_j^2]=\frac{1}{2}[\sum_{i=1}^n(\sum_{j=1}^n w_{ij})f_i^2-2\sum_{i,j=1}^n w_{ij}f_if_j+\sum_{j=1}^n(\sum_{i=1}^n w_{ij})f_j^2]$

$=\frac{1}{2}(\sum_{i,j=1}^{n} w_{ij}f_i^2 - 2\sum_{i,j=1}^{n} w_{ij}f_if_j + \sum_{i,j=1}^{n} w_{ij}f_j^2) = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(f_i^2 - 2f_if_j + f_j^2) = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2$，因为矩阵$W$非负，所以$f^TLf \geqslant 0$