Preface

写在前面的话

此笔记是基于 某个笔记 增删该而来，如有侵权，请联系我哦! 当然有错误也可以联系我哦！

联系方式：发送邮件

最新下载地址:下载地址

Background

Vector

- 线性相关
  存在一组 $a_1, a_2, ...a_k$ 不全为零的数，使得 $a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_k\vec{x}_k = \vec{0}$ ，那么可以称这组向量 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 是线性相关的。
- 线性无关
  当且仅当 $a_1, a_2, ...a_k$ 全都为0时， $a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_k\vec{x}_k = \vec{0}$ 才成立，，那么可以称这组向量 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 是线性无关的。
- 极大线性无关组
  如果线性无关的 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 是向量组 $\vec{x}$ 部分组，且 $\vec{x}$ 中任一向量都可以用 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 表示，那么 $\vec{x}_1, \vec{x}_2, ..., \vec{x}_k$ 就是一个极大线性无关组或最大线性无关组。
- 向量运算
  内积 $\vec{a} \cdot \vec{b} = \sum a_i b_i$
  叉积 $\vec{a} \times \vec{b} = \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{bmatrix}$
- 范数
  - $\|x\|_1 = \sum |x_i|$
  - $\|x\|_2 = \sqrt{\sum_i x_i^2}$
  - $\|x\|_\infty = \max\{|x_i|\}$
  - $\|x\|_p = (\sum |x_i|^p)^{\frac{1}{p}}$
  - 向量范数的性质：
    1. $\|x\| \geq 0$
    2. $\|kx\| = k\|x\|$
    3. $\|x + y\| \leq \|x\| + \|y\|$

Matrix

- 矩阵转置

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}$$

$$(AB)^T = B^T A^T$$

$$(A+B)^T = A^T + B^T$$

$$(kA)^T = kA^T$$

- 共轭转置

$$A^H = (\bar{A})^T$$

eg.

$$\begin{pmatrix} 1 & 2+i \\ 1-i & 2 \end{pmatrix}^H = \begin{pmatrix} 1 & 1+i \\ 2-i & 2 \end{pmatrix}$$

酉矩阵：复数域上的正交矩阵$u_i^H u_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$

`Hermitian` 矩阵：$A^H = A$，例如$U_1 U_1^H = (U_1 U_1^H)^H$

- 伴随矩阵

$$A^* = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}$$

$A_{ij} = (-1)^{i+j} M_{ij}$

- 矩阵的迹

$$tr(A) = \sum_i^n a_{ii} = \sum_i^n \lambda_i$$

$$a = tr(a)$$

$$tr(AB) = tr(BA)$$

$$tr(A+B) = tr(A) + tr(B)$$

$$tr(A) = tr(A^T)$$

$$tr(A^T B) = \sum_{i,j} A_{ij} B_{ij}$$

$$tr(A^T (B \odot C)) = tr((A \odot B)^T C)$$

- 范数

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2} = tr(AA^T)$$

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^TA)} = \delta_{\max}(A)$$

$$\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$$

$$\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$$

$$\|A\|_* = \sum \delta(A)$$

$$\|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p$$

  矩阵范数满足：
  1. $\|A\| \geq 0$
  2. $\|kA\| = k\|A\|$
  3. $\|A+B\| \leq \|A\| + \|B\|$
  4. $\|AB\| \leq \|A\| \cdot \|B\|$
- 谱半径
  $\rho(A) = \max|\lambda(A)|$
  1. 谱半径不是范数
  2. 若A是 `Hermitian` 矩阵,则$\rho(A) = \|A\|_2$
  3. $\rho(A) = \inf_{\|\cdot\|}\|A\|$
  4. $\sum_k A^k$收敛$\Rightarrow A^k \to 0, \rho(A) < 1$
  5. $\left\|A^k\right\|^{\frac{1}{k}} \to \rho(A)$
- 标准正交矩阵$U = [\alpha_1, \alpha_2, \cdots, \alpha_n]$

$$\alpha_i^T\alpha_j = \begin{cases} 1, & i \neq j \\ 0, & i = j \end{cases}$$

  满足如下性质：
  1. $U^{-1} = U^T$
  2. $rank(U) = n$

3. $U^TU = UU^T = E$
4. $\|U \cdot A\| = \|A\|$

- 部分列正交矩阵$U = [U_1, U_2]$
  $U_1 = [u_1, u_2, \cdots, u_r] \in R^{n\times r}$
  $u_i^T u_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$
  $U_1^T U_1 = E_{r\times r}$
  $UU^T = [U_1, U_2]\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = U_1U_1^T + U_2U_2^T = E$
  U_2是正交补
- 正交化
  有一组向量$a_1, a_2, \cdots, a_n$ 寻找$q_1, q_2, \cdots, q_n$使得

$$span\{a_1, a_2, \cdots, a_n\} = span\{q_1, q_2, \cdots, q_n\}$$

  且$Q = [q_1, q_2, \cdots, q_n]$是标准正交矩阵
  `Gram-Schmidt`

1. $span\{q_1\} = span\{a_1\}$，则$q_1 = \frac{a_1}{\|a_1\|}$
2. 假设$span\{a_1, a_2, \cdots, a_k\} = span\{q_1, q_2, \cdots, q_k\}$且满足$q_i \perp q_j, i \neq j, \|q_i\| = 1$
3. 如何构造q_{k+1}使其满足：$\begin{cases} span\{q_1, q_2, \cdots, q_k\} \oplus span\{q_{k+1}\} = span\{a_1, a_2, \cdots, a_{k+1}\} \\ q_{k+1} \perp q_i, i = 1, 2, \cdots, k \\ \|q_{k+1}\| = 1 \end{cases}$

   $a_{k+1} = \sum_{i=1}^{k+1} r_{i,k+1} q_i \Rightarrow q_i^T a_{k+1} = r_{i,k+1} q_i^T q_i$

   $q_{k+1} = \frac{a_{k+1} - \sum_{i=1}^{k} r_{i,k+1} q_i}{\|a_{k+1} - \sum_{i=1}^{k} r_{i,k+1} q_i\|}$, 其中$r_{i,k+1} = q_i^T a_{k+1}, i = 1, 2, 3, \cdots, k$
   从 `QR Decomposition` 看 `Gram-Schmidt` ：

$$[a_1, a_2, \cdots, a_n] = [q_1, q_2, \cdots, q_n]\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}$$

$$\Rightarrow \begin{cases} a_1 = r_{11}q_1 \\ a_2 = r_{12}q_1 + r_{22}q_2 \\ \vdots \end{cases} \Rightarrow \begin{cases} q_1 = \dfrac{a_1}{\|a_1\|} \\ q_2 = \dfrac{a_2 - r_{12}q_1}{\|a_2 - r_{12}q_1\|} \\ \vdots \end{cases}$$

   `Arnoldi` 分解：在$\mathrm{K}_k(A, r_0) = span\{r_0, Ar_0, \cdots, A^{k-1}r_0\}$上运用 `Gram-Schmidt`

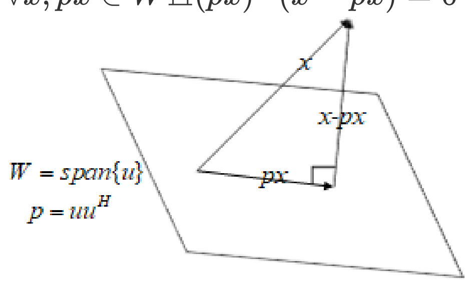
   1. 令$v_1 = \frac{r_0}{\|r_0\|}$
   2. 假设已构造$\{v_1, v_2, \cdots, v_k\}, v_k \in \mathrm{K}_k(A, r_0), q_i \perp q_j, i \neq j, \|v_i\| = 1$

3. 如何构造v_{k+1}使其满足：$\begin{cases} v_{k+1} \in \mathrm{K}_{k+1}(A, r_0) \\ v_{k+1} \perp v_i, i = 1, 2, \cdots, k \\ \|v_{k+1}\| = 1 \end{cases}$
$v_{k+1} \in \mathrm{K}_{k+1}(A, r_0) = \mathrm{K}_k(A, r_0) \oplus span\{Av_k\} = span\{v_1, v_2, \cdots, v_{k+1}\}$
$Av_k = \sum_{i=1}^{k+1} h_{i,k} v_i \Rightarrow v_i^T A v_k = h_{i,k} v_i^T v_i$
$v_{k+1} = \frac{Av_k - \sum_{i=1}^{k} h_{i,k} v_i}{\|Av_k - \sum_{i=1}^{k} h_{i,k} v_i\|}$, 其中$h_{i,k} = v_i^T A v_k, i = 1, 2, 3, \cdots, k$

- A-正交
  $\|X\|_A = \sqrt{X^T A X}$
  有一组向量$a_1, a_2, \cdots, a_n$ 寻找$q_1, q_2, \cdots, q_n$使得

$$span\{a_1, a_2, \cdots, a_n\} = span\{q_1, q_2, \cdots, q_n\}$$

  且$q_1, q_2, \cdots, q_n$与A正交，即$q_i^T A q_j = \begin{cases} \neq 0, & i = j \\ 0, & i \neq j \end{cases}$
  做法同上，这里就不赘述。
- 广义逆
  对$A = U_1 \Sigma_1 V_1^H \in C^{m \times n}$, 记$A^+ = V_1 \Sigma_1^{-1} U_1^H$,满足：
  1. $A^+ A A^+ = A^+$
  2. $A A^+ A = A$
  3. $AA^+, A^+ A$都是 `Hermitian`
- 正交投影
  假设p是子空间W：$span\{U_1\}$的投影，$p = U_1 U_1^H$，U_1是酉阵，则有如下性质：
  1. $p^H = p$
  2. $p^2 = p$
  3. $\forall x, px \in W$且$(px)^H (x - px) = 0$



- 子空间
  - 子空间距离
    - 点到子空间距离和span{x}到子空间的距离
      具体参考正交投影
    - 两个平面以及同维子空间距离
      Z, Y分别是其标准正交基
      $dist(\mathfrak{X}, \mathfrak{Y}) = \sqrt{1 - \sigma_{\min}^2 (Z^H Y)}$
  - Krylov子空间
    $\mathrm{K}_k(A, r_0) = span\{r_0, Ar_0, \cdots, A^{k-1} r_0\}$

Matrix Decomposition

$A \in C^{n\times n}, B \in C^{m\times n}$， U,V为酉阵

QR Decomposition

$$A = [a_1, a_2, \cdots, a_n] = Q \cdot R = [q_1, q_2, \cdots, q_n] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}$$

LU Decomposition

$$A = L \cdot U$$

L为上三角矩阵，U为下三角矩阵

$Ax = b \Rightarrow (LU)x = b \Rightarrow \begin{cases} Ly = b \\ Ux = y \end{cases}$

Shur Decomposition

$$A = URU^H$$

其中U是正交阵，R是上三角阵。

若A是 `Hermitian` 矩阵，则R是对角阵，即$A = U\Lambda U^T$

Singular Value Decompostion

$$B = U\Sigma V^H$$

若$rank(B) = r$，则有

$$A = (u_1, u_2, \cdots, u_m) \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}_{m\times n} (v_1, v_2, \cdots, v_n)^H = U_1\Sigma_1 V_1^H$$

其中，$U_1 = (u_1, u_2, \cdots, u_r), V_1 = (v_1, v_2, \cdots, v_r)$

因此根据$B = U\Sigma V^H$，可以得到

$$Av_i = U\Sigma \begin{pmatrix} v_1^H \\ \vdots \\ v_n^H \end{pmatrix} v_i = \sigma_i u_i$$

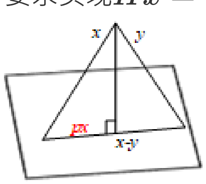
$$u_i^H A = u_i^H \left(u_1, u_2, \cdots, u_m\right) \Sigma V^H = \sigma_i v_i^H$$

思考：什么情况下奇异值与特征值相同？

$A^H A = V\Sigma^H U^H U\Sigma V^H = V\Sigma^H \Sigma V^H$
$AA^H = U\Sigma V^H V\Sigma^H U^H = U\Sigma^H \Sigma U^H$
因此我们可以根据以上两个矩阵的特征值已经特征向量得到矩阵A的奇异值和奇异向量。
例题：$W = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$
$W^H W = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$
特征值：3,1
特征向量$v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$
$WW^H = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$
特征值：3,1,0
特征向量$u_1 = \begin{bmatrix} \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, u_2 = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, u_3 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$
因此$W = (u_1, u_2, u_3) \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} (v_1, v_2)$

Householder Transformation

要求实现$Hx = y, \|x\| = \|y\|$, 要求：$1) H^H = H, 2) H$是酉阵



$y = x - 2\frac{x-y}{2}$
由于$\|x\| = \|y\|$， 因此$x, y, x-y$构成等腰三角形
因此根据正交投影可知$\frac{x-y}{2} = px$， $p = uu^H$是$span\{x-y\}$的正交投影， $u = \frac{x-y}{\|x-y\|}$
$\therefore H = I - 2uu^H$，显然满足上述两点。
利用 `Householder Transformation` 可以将矩阵稀疏化。

$$\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{bmatrix} \xrightarrow{H} \begin{bmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix}$$

注意可以跨行和列去选择变换

$$\begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \xrightarrow{H} \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

同样也能用 `Householder Transformation` 求QR分解

Givens Rotation Transformation

$\begin{bmatrix} C & S \\ -S & C \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sqrt{a^2+b^2} \\ 0 \end{bmatrix}$，其中$C = \frac{a}{\sqrt{a^2+b^2}}, S = \frac{b}{\sqrt{a^2+b^2}}$

同样也能用 `Givens Rotation Transformation` 求QR分解

Application

- 标准正交基
  - $R(A) = \{Ax | x \in R^n\}$
    $\because Ax = U_1 \Sigma_1 V_1^H x = U_1 Z$
    $\therefore R(A) = span\{U_1\}, R(A^H) = span\{V_1\}$
  - $N(A) = \{x | Ax = 0\}$
    $\because Ax = U_1 \Sigma_1 V_1^H x = 0 \Rightarrow V_1^H x = 0$
    $\therefore N(A) = span\{V_2\}, N(N^H) = span\{U_2\}$
- 低秩逼近
  设$rank(A) = r, d < r$, 求$\min_{rank(x)=d} = \|A - x\|_2$。
- 最小二乘法
  $\min_x \|Ax - b\|_2 \Rightarrow \min_{y \in R(A)} \|b - y\|_2$
  $Ax = AA^+ b$的通解为齐次解加上特解，其特解为$A^+ b$,齐次通解为$\sigma = (I - A^+ A)z \in N(A)$
  因此$x = A^+ b + (I - A^+ A)z$

Matrix Differential

导数与微分

- 标量对向量的求导
  对于标量$f, \vec{x}_{(n\times 1)}$ 有：

$$\frac{\partial f}{\partial \vec{x}} = \left[\frac{\partial f}{\partial x_i}\right]$$

- 标量对矩阵的求导
  对于标量$f, X_{(m\times n)}$ 有：

$$\frac{\partial f}{\partial X} = \left[\frac{\partial f}{\partial x_{ij}}\right]$$

我们知道标量对标量的梯度gradient和微分differentiation有这样的关系：

$$df = f'(x)dx$$

$$df = \sum_i \frac{\partial f}{\partial x_i}dx_i = \frac{\partial f}{\partial \vec{x}}^T d\vec{x}$$

那么标量对矩阵也存在：

$$df = \sum_{ij} \frac{\partial f}{\partial x_{ij}}dx_{ij} = tr(\frac{\partial f}{\partial X}^T dX)$$

例子1：已知$f = |X|$，求df？
我们知道

$$|X| = \sum_i x_{ij}A_{ij}(A_{ij}\text{代数余子式})$$

将上式代入得：

$$\frac{\partial f}{\partial X} = \left[\frac{\partial \sum_k x_{kj}A_{kj}}{\partial x_{ij}}\right] = [A_{ij}] = (X^*)^T$$

因此有

$$df = tr(\frac{\partial f}{\partial X}^T dX) = tr(X^* dX) = |X|tr(X^{-1}dX)$$

例子2：求dX^{-1}？
我们知道

$$XX^{-1} = E$$

对等式两边微分有

$$dXX^{-1} = dE$$

$$XdX^{-1}=-X^{-1}dX$$

因此有

$$dX^{-1}=-X^{-1}dXX^{-1}$$

例子3：$f=\vec{a}^TX\vec{b}$，求 $\frac{\partial f}{\partial X}$?

$$df=\vec{a}^TdX\vec{b}=tr(\vec{a}^TdX\vec{b})=tr(\vec{b}\vec{a}^TdX)=tr(\frac{\partial f}{\partial X}^TdX)$$

因此 $\frac{\partial f}{\partial X}=\vec{a}\vec{b}^T$

- 复合法则
  已知$f=g(Y)$，且 $Y=h(X)$，怎么求 $\frac{\partial f}{\partial X}$?(其中g和h都是逐元素的函数)

$$df=tr(\frac{\partial f}{\partial Y}^TdY)=tr(\frac{\partial f}{\partial Y}^T(h'(X)\odot dX))=tr((\frac{\partial f}{\partial Y}\odot h'(X))^TdX)$$

例子4：$loss=-\vec{y}^T\log\ softmax(W\vec{x})$，求 $\frac{\partial\, loss}{\partial W}$。$\vec{y}$ 是只有一个元素为1其余元素为0的向量。

$$softmax(\vec{x})=\frac{e^{\vec{x}}}{\vec{1}^Te^{\vec{x}}}$$

$$loss=-\vec{y}^TW\vec{x}+(\vec{y}^T\vec{1})\log(\vec{1}^Te^{W\vec{x}})\qquad(*)$$

$$d\,loss=-\vec{y}^TdW\vec{x}+\frac{\vec{1}^T(e^{W\vec{x}}\odot dW\vec{x})}{\vec{1}^Te^{W\vec{x}}}\qquad(**)$$

$$d\,loss=-\vec{y}^TdW\vec{x}+\frac{(e^{W\vec{x}})^TdW\vec{x}}{\vec{1}^Te^{W\vec{x}}}$$

$$d\,loss=tr(-\vec{y}^TdW\vec{x}+\frac{(e^{W\vec{x}})^TdW\vec{x}}{\vec{1}^Te^{W\vec{x}}})$$

$$d\,loss=tr(\vec{x}(softmax(W\vec{x})-\vec{y})^TdW)$$

$$\frac{\partial\, loss}{\partial W} = (softmax(W\vec{x}) - \vec{y})\vec{x}^T$$

注意:
(*)式 $\log(\frac{\vec{b}}{c}) = \log\vec{b} - \vec{1}\log c$，且 $\vec{y}^T\vec{1} = 1$
(**)式 $\log(\vec{1}^T e^{W\vec{x}})$ 是标量，$e^{W\vec{x}}$ 是逐元素函数，因此$d\log(\vec{1}^T e^{W\vec{x}}) = \frac{1}{\vec{1}^T e^{W\vec{x}}} \cdot \vec{1}^T(e^{W\vec{x}} \odot dW\vec{x})$

- 向量对向量求导
  对于 $\vec{f}_{(m\times1)}, \vec{x}_{(n\times1)}$ 有：

$$\frac{\partial\vec{f}}{\partial\vec{x}} = \left[\frac{\partial f_i}{\partial x_j}\right]_{(n\times m)}$$

$$d\vec{f} = \left[\frac{\partial f_i}{\partial\vec{x}}^T\right]d\vec{x} = \frac{\partial\vec{f}}{\partial\vec{x}}^T d\vec{x}$$

- 矩阵对矩阵求导
  对于矩阵$F_{(m\times n)}, X_{(p\times q)}$ 有：

$$\frac{\partial F}{\partial X} = \left[\frac{\partial F_{ij}}{\partial x_{kl}}\right]_{(pq\times mn)}$$

矩阵向量化：
对于矩阵$X_{(p\times q)}$，其矩阵向量化$vec(X)_{(pq\times1)} = \left[X_1^T, X_2^T, ..., X_q^T\right]^T, X_i$是$X$的列向量。

$$vec(A+B) = vec(A) + vec(B)$$

$$vec(\vec{a}\vec{b}^T) = \vec{b}\otimes\vec{a}$$

$$X = \sum_i X_i {e_i}^T$$

$$vec((AB)\otimes(CD)) = vec((A\otimes C)(B\otimes D))$$

$$vec(AXB) = vec(\sum_i AX_i{e_i}^TB) = \sum_i vec((AX_i)(B^Te_i)^T) = \sum_i(B^Te_i)\otimes(AX_i) = (B^T\otimes A)vec(X)$$

因此有：

$$\frac{\partial F}{\partial X} = \frac{\partial vec(F)}{\partial vec(X)}_{(pq\times mn)}$$

$$vec(dF) = \frac{\partial F}{\partial X}^T vec(dX)$$

求导时矩阵被向量化，弊端是这在一定程度破坏了矩阵的结构，会导致结果变得形式复杂；好处是多元微积分中关于 `Gradient` 、`Hessian` 矩阵的结论可以沿用过来，只需将矩阵向量化。

应用

1. 紧紧抓住两个转换公式：$df = tr(\frac{\partial f}{\partial X}^T dX), d(trace(f(X))) = trace(df(X))$以及定义，那么几乎所有的导数我们都能求。
2. 泰勒公式
3. 求最优化问题
   - 最小二乘法：
     - $\min_x \|Ax-b\|_2^2$
     - $\min_x \|Ax-b\|_2^2 + \lambda \|x\|_2^2$
   - 有约束问题：
     - $\min_x \|Ax\|_2^2, s.t.\, e^T x = 1$或者$s.t.\ \|x\|_2 = 1$
     - $\min_U tr(U^T AU), s.t. U^H U = I, A$半正定
     - $\min_X [tr(X^T X) - 2tr(X)], s.t. XA = \mathbf{0}$
   - `Locally Linear Embedding`
     给定一组数据$x_i \in R^n$及其邻域数据$x_{i1}, x_{i2}, \cdots, x_{ij}$, 要求将$x_i$降维为$y_i \in R^d$。
     LLE 认为数据局部是线性的$x_i = \sum_{i=1}^j w_{ij} x_{ij}$，且在降维过程中线性不变,且组合系数不变。
     优化目标：$\arg\min_Y \sum_{i=1}^N \left\| y_i - \sum_{j=1}^k w_{ij} y_{ij} \right\|_2^2, s.t.\, YY^T = NI$
4. 总体最小二乘

$$(A+\Delta A)x = (b+\Delta b) \Rightarrow ([A,b] + [\Delta A + \Delta b]) \begin{bmatrix} x \\ -1 \end{bmatrix} = 0$$

令 $B = [A,b], D = [\Delta A + \Delta b], Z = \begin{bmatrix} x \\ -1 \end{bmatrix}$
因此要求x的近似解即求：$\min_{D,x} \|D\|_F^2, s.t.\, (B+D)Z = 0$

- 若B不是列满秩，则存在$Z \neq 0, BZ = 0$，要使$\min_{D,x} \|D\|_F^2$最小，只需$D = 0$
- 若B是列满秩，要使$(B+D)Z = 0$有解，则$r(B+D) \leq n$
  - 若B的奇异值满足$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq \sigma_{n+1}$
    要使要使$\min_{D,x} \|D\|_F^2$最小，只需$D = -\sigma_{n+1} u_{n+1} v_{n+1}^T$
    故$N(B+D) = span\{v_{n+1}\}$

记$v_{n+1}=(v_{n+1,1},v_{n+1,2},\cdots,v_{n+1,n+1})^T$
因此$\begin{pmatrix} x \\ -1 \end{pmatrix}=\begin{pmatrix} -\frac{v_{n+1,1}}{v_{n+1,n+1}} \\ -\frac{v_{n+1,2}}{v_{n+1,n+1}} \\ \vdots \\ -1 \end{pmatrix}$

- 若B的奇异值满足$\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_p\geq\sigma_{p+1}=\cdots=\sigma_{n+1}$
  $N(B+D)=span\{v_{p+1},\cdots,v_{n+1}\}$
  记$V_1=(v_{p+1},\cdots,v_{n+1})^T$
  运用 `Householder`变换 ：

$$HV_1=\begin{pmatrix} \hat{v}_{n+1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Matrix Equation

分裂迭代

方程$Ax=b$
令$A=M-N\Rightarrow Mx=Nx+b$
迭代形式$Mx_k=Nx_{k-1}+b$
故$x_k=M^{-1}(Nx_{k-1}+b)$
收敛性要求：$\rho(M^{-1}N)<1$
若A正定，当且仅当$M+N^H$时，$\rho(M^{-1}N)<1$

- Jacobi迭代
  $A=D-L-U$
  $M=D,N=L+U=D-A$
  故$x_k=D^{-1}((L+U)x_{k-1}+b)=(I-D^{-1}A)x_{k-1}+D^{-1}b$
  收敛性要求：$\rho(I-D^{-1}A)<1$
  适用：A对角占优
- Gauss-Seideld迭代
  $A=D-L-U$
  $M=D-L,N=U$
  故$x_k=(D-L)^{-1}(Ux_{k-1}+b)$
  收敛性要求：$\rho((D-L)^{-1}U)<1$
  适用：元素集中在下三角处
- SOR
  $M=\frac{1}{\omega}D-L,N=((\frac{1}{\omega}-1)D+U)$
  故$(D-\omega L)x_k=((1-\omega)D+\omega U)x_{k-1}+\omega b$
  收敛性要求：$\rho((D-\omega L)^{-1}((1-\omega)D+\omega U))<1,0<\omega<2$
  适用：下三角占优

- SSOR
  交替迭代
  1. $M=\frac{1}{\omega}D-L, N=((\frac{1}{\omega}-1)D+U)$
  2. $M=\frac{1}{\omega}D-U, N=((\frac{1}{\omega}-1)D+L)$

最速下降

$$Ax=b \iff \min_x \psi(x)=\frac{1}{2}(x,x)_A-(b,x)$$

A对称正定
迭代式：$x_{k+1}=x_k+\alpha_k r_k, r_k=b-Ax_k=-\nabla\psi(x_k), \alpha_k=\arg\min_\alpha \|x_k+\alpha r_k-x^*\|$
求α只需满足下式：

$$\frac{d\psi(x_k+\alpha r_k)}{d\alpha}=0$$

$$\psi(x_k+\alpha r_k)=\psi(x_k)+\alpha(-r_k,r_k)+\frac{\alpha^2}{2}(Ar_k,r_k)$$

故$(-r_k,r_k)+\alpha(Ar_k,r_k)=0 \Rightarrow \alpha=\frac{\|r_k\|_2^2}{\|r_k\|_A^2}$
收敛性分析：$\|x_{k+1}-x^*\|_A=\min_\alpha \|x_k+\alpha r_k-x^*\|_A=\min_\alpha \|x_k+\alpha A(x^*-x_k)-x^*\|_A$
$=\min_\alpha \|(I-\alpha A)(x_k-x^*)\|_A$
$\leq \min_\alpha \rho(I-\alpha A)\|x_k-x^*\|_A$
$\leq \frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\|x_k-x^*\|_A$
$\leq (\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1})^k\|x_0-x^*\|_A$

子空间迭代

$A\in C^{n\times n}$
$Ax=b \iff x=A^{-1}b \iff A(x^*-x_0)=b-Ax_0=r_0 \iff x^*=x_0+A^{-1}r_0$
又$A^{-1}=\sum_{i=0}^m C_iA^ir_0$
故$x^*=x_0+\mathrm{K}_{m+1}(A,r_0)$

共轭CG

$x^*=x_0+\mathrm{K}_{m+1}(A,r_0)$
设$g_0,g_1,\cdots,g_m$是$\mathrm{K}_{m+1}(A,r_0)$一组标准A正交基，则

$$x^*=x_0+\sum_{j=0}^m \alpha_j g_j$$

设$x_k = x_0 + \sum_{j=0}^{k-1} \alpha_j g_j$，则$x_{k+1} = x_k + \alpha_k g_k$
如何确定系数α_k和正交基g_k？

1. 求α_k
   显然$x_{m+1} = x^*$，那么$r_{m+1} = b - Ax_{m+1} = b - Ax^* = 0$
   $$\begin{aligned} r_{m+1} &= b - Ax_{m+1} \\ &= b - A(x_m + \alpha_m g_m) \\ &= r_m - \alpha_m A g_m \\ &= b - Ax_m - \alpha_m A g_m \\ &= b - A(x_{m-1} + \alpha_{m-1} g_{m-1}) - \alpha_m A g_m \\ &= r_{m-1} - \alpha_{m-1} A g_{m-1} - \alpha_m A g_m \\ &= \cdots \\ &= r_k - \alpha_k A g_k - \alpha_{k+1} A g_{k+1} - \cdots - \alpha_m A g_m \\ &= 0 \end{aligned}$$
   则$g_k^T r_k = \alpha_k g_k^T A g_k - \alpha_{k+1} g_k^T A g_{k+1} - \cdots - \alpha_m g_k^T A g_m = \alpha_k g_k^T A g_k$
   $\therefore \alpha_k = \frac{g_k^T r_k}{g_k^T A g_k}$
2. 求g_k
   `Gram-shcmit`
   1. 令$g_0 = r_0$
   2. 假设已求得$g_0, g_1, \cdots, g_{k-1}$，有$span\{g_0, g_1, \cdots, g_{k-1}\} = span\{r_0, Ar_0, \cdots, A^{k-1} r_0\}$
   3. 对于r_k有：
   $$\begin{aligned} r_k &= b - Ax_k \\ &= b - A(x_{k-1} + \alpha_{k-1} g_{k-1}) \\ &= r_{k-1} - \alpha_{k-1} A g_{k-1} \\ &= \cdots \\ &= r_0 - \sum_{j=0}^{k-1} \alpha_i A g_i \in \mathrm{K}_{k+1}(A, r_0) \end{aligned}$$
   而$g_k \in \mathrm{K}_{k+1}(A, r_0)$，故$r_k = \beta_0 g_0 + \beta_1 g_1 + \cdots + \beta_k g_k$
   $\therefore g_i^T A r_k = \beta_i g_i^T A g_i,\ \ i = 0, 1, \cdots, k$
   $\therefore g_k$可求

收敛速度大于最速下降法

A非对称

- 直观法
  将$Ax = b$转化为$A^T A x = b$
  收敛速度远小于A正定时收敛速度
- 广义最小残量法
  前面在 `Arnoldi` 分解我们得到$\mathrm{K}_k(A, r_0)$的标准正交基$v_1, v_2, \cdots, v_k$

令$V_k=(v_1,v_2,\cdots,v_k)$，则$AV_k=V_kH_k+h_{k+1,k}v_{k+1}e_k^T=V_{k+1}\hat{H}_k$
$b-Ax_k=b-A(x_0+V_ky)=r_0-V_{k+1}\hat{H}_ky=V_{k+1}(\|r_0\|e_1-\hat{H}_ky)$
$\therefore \|b-Ax_k\|_2=\left\|\|r_0\|e_1-\hat{H}_ky\right\|_2$
$y=\arg\min_y\left\|\|r_0\|e_1-\hat{H}_ky\right\|_2$
求解上述式子：

1. 最小二乘法
2. Givens旋转变换，对$\hat{H}_k$进行QR分解

Convex Optimization

次梯度

g^T 次梯度

1. 向量：$f(y)\geq f(x)+g^T(y-x)$
2. 矩阵：$f(Y)\geq f(X)+trace(g^T(Y-X))$

例题：求和范数的次梯度$\partial\|X\|_*$
$f(X)\geq f(0)+trace(g^TX)\Rightarrow\sum\sigma_i\geq trace(g^TX)$
$tr(\Sigma_1)\geq tr(g^TU_1\Sigma_1V_1^T)\Rightarrow tr(V_1\Sigma_1V_1^T)\geq tr(g^TU_1\Sigma_1V_1^T)$
$\therefore \partial\|X\|_*=\{U_1V_1^T+Y|U_1^TY=0,YV_1=0,\|X\|_2\leq 1\}$

Solving Characteristic Value