CS229 - Problem Set 0

You

November 30, 2023

Exercise 1

(a)

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix}$$

For $f(x) = \frac{1}{2}x^T Ax + b^T x, A^T = A$, we'll get

$$\frac{\partial}{\partial x_k} f(x) = \frac{\partial}{\partial x_k}\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i A_{ij} x_j + \frac{1}{2}\sum_{i=1}^{n} x_i b_i\right) = \frac{1}{2}\frac{\partial}{\partial x_k}\left(\sum_{i\neq k}^{n} x_k A_{ki} x_i + \sum_{j\neq k}^{n} x_j A_{jk} x_k + \frac{1}{2}x_k^2 A_{kk}\right) +$$
$$\frac{\partial}{\partial x_k}\sum_{i=1}^{n} x_i b_i = \frac{\partial}{\partial x_k}\sum_{i=1, i\neq k}^{n} x_k A_{ik} x_i + A_{kk} + b_k = \sum_{i=1}^{n} A_{ik} x_i + b_k$$

And we got $\nabla f(x) = Ax + b$

(b)

By the multivariable chain rule

$$\nabla f(x) = \nabla g(h(x)) = g'(h(x))\nabla h(x)$$

(c)

From (a) we got $\frac{\partial}{\partial x_k} f(x) = \sum_{i=1}^{n} A_{ik} x_i + b_k$, so

$$\frac{\partial}{\partial x_k x_t} f(x) = \frac{\partial}{\partial x_t}\left(\sum_{i=1}^{n} A_{ik} x_i + b_k\right) = A_{it}$$

And we got $\nabla^2 f(x) = A$

(d)

Let $f(x) = g(a^T x)$, by (b)

$$\nabla f(x) = g'(a^T x)\nabla h(x) = \nabla(a^T x), \frac{\partial}{\partial x_k} a^T x = \frac{\partial}{\partial x_k}\sum_{i=1}^{n} a_i x_i = x_k$$

So $\nabla f(x) = g'(a^T x)a$

If we'll use the multivariable chain rule on $\nabla f(x) = g'(a^T x)a$, we'll get $\nabla^2 f(x) = g''(a^T x)aa^T$. And without the first calculation

$$\nabla^2 f(x) = g'(a^T)\nabla^2(a^T x) + g''(a^T x)\nabla(a^T x)\nabla(a^T x)^T$$

But $\frac{\partial}{\partial x_k x_t} a^T x = \frac{\partial}{\partial x_t} a_k = 0$, thus $\nabla^2 a^T x = 0$, and by (c) we got $\nabla a^T x = a$, so

$$\nabla^2 g(a^T x) = g''(a^T x)aa^T$$

Exercise 2

(a)

Let $z \in \mathbb{R}^n, A = zz^T$, then $A^T = (zz^T)^T = (z^T)^T z^T = zz^T = A$, and for $v \in \mathbb{R}^n$ we got $\langle Av, v \rangle = \langle zz^T v, v \rangle = \langle z^T v, z^T v \rangle = \|z^T v\| \geq 0$, thus by definition $A \succeq 0$

(b)

I'll show that $\mathcal{N}(A) = W = \{v : \langle z, v \rangle = 0\}$, indeed, if $\langle z, v \rangle = 0$ then $z^T v = 0 \implies Av = 0$, so $W \subseteq \mathcal{N}$, also

$$Av = 0 \implies \|z^T v\| = \langle z^T v, z^T v \rangle = \langle zz^T, v \rangle = zz^T v = 0$$

And thus we got $\mathcal{N} = \{v : \langle z, v \rangle = 0\}$.
From the rank-nullity theorem $\dim \mathrm{Im} T = n - \dim \ker T = n - (n-1) \overset{(1)}{=} 1$

(c)

The statement is true, indeed, if $A \succeq 0$, then $(BAB^T)^T = (B^T)^T A^T B^T = B^T A^B$, and for $v \in \mathbb{R}^n$ we got $\langle B^T ABv, v \rangle = \langle ABv, Bv \rangle \geq 0$, while the last claim is true because $A \succeq 0$

Exercise 3

(a)

$A = T\Lambda T^{-1} \implies AT = \Lambda T$, so for each column of T, $t^{(i)}$ we have $At^{(i)} = [T\Lambda]^{(i)}$, but

$$[T\Lambda]_k^{(i)} = \sum_{j=1}^{n} T_{kj}\Lambda_{ji} = T_{ki}\Lambda_{ii} \implies [T\Lambda]^{(i)} = \Lambda_{ii} t^{(i)}$$

And thus $At^{(i)} = [T\Lambda]^{(i)} \overset{(2)}{=} \lambda_i t^{(i)}$

(b)

By the definition of orthogonal matrix, if U is orthogonal then $U^T = U^{-1}$, so $A = U\Lambda U^{-1}$ and the claim is true from (a)

(c)

Suppose $A \succeq 0$ and suppose that $u^{(i)}$ is an eigenvector of A with eigenvalue λ_i, then

$$0 \leq \langle Au^{(i)}, u^{(i)} \rangle = \langle \lambda_i u^{(i)}, u^{(i)} \rangle = \lambda_i \langle u^{(i)}, u^{(i)} \rangle = \lambda \|u^{(i)}\|$$

So because $\|u^{(i)}\| \geq 0$ we got $\lambda_i \geq 0$