Homework set 0

$x^T A$

$\cancel{A}x + b^T$

1.

a) $[\nabla_x f(x)]_k \overset{\Sigma}{=} \frac{\partial}{\partial x_k}\left[\frac{1}{2}x_i A_{ij} x_j + b_i x_i\right] \overset{\Sigma}{=} \left[\frac{1}{2}\frac{\partial}{\partial x_k}\left(\sum_{i\neq k} x_i A_{ij} x_j + \sum_{j\neq k} x_i A_{ij} x_j + \sum_{i,j\neq k} x_i A_{ij} x_j\right)\right.$

$\left. + A_{kk} x_k + b_k\right] \overset{\Sigma}{=} \left[\frac{1}{2}\sum_{i\neq k} x_i A_{ik} + \frac{1}{2}\sum_{i\neq k} A_{kj} x_j + 0 + A_{kk} x_k + b_k\right]$

$\overset{\Sigma}{=} \frac{1}{2}\sum_j A_{kj} x_j + b_k \Rightarrow \underline{\nabla_x f(x)} = \underline{\underline{Ax + b}}$

b) $\underline{\nabla_x f(x)} = \underline{\underline{\frac{\partial g}{\partial h} \cdot \nabla_x h(x)}}$. To be more exact: $\underline{\partial_{x_k} f} = \underline{\underline{\frac{\partial g}{\partial h} \cdot \frac{\partial h}{\partial x_k}}}$.

c) $\nabla_x^2 f(x) = \nabla_x (\nabla_x f(x)) \Rightarrow [\nabla_x^2 f]_{kl} = \frac{\partial}{\partial x_l}(A_{kj} x_j + b_k) = \underline{\underline{A_{kl}}}$

$\Rightarrow \underline{\nabla_x^2 f(x)} = \underline{\underline{A}}$

d) La $f(x) \equiv g(h(x))$, $h(x) = a^T x$.

- $\underline{\nabla_x f(x)} = \frac{\partial g}{\partial h} \cdot \nabla_x h(x) = \frac{\partial g}{\partial h} \cdot a \Leftrightarrow \underline{[\nabla_x f]_k} = \underline{\underline{\frac{\partial g}{\partial h} a_k}}$
- $\underline{\nabla_x^2 f(x)} = \nabla_x (\nabla_x f) = \underline{\underline{\frac{\partial^2 g}{\partial h^2} \cdot a a^T}}$

$\Leftrightarrow \underline{[\nabla_x^2 f(x)]_{kl}} = \frac{\partial^2 g}{\partial h^2} a_k \cdot \frac{\partial a_i x_i}{\partial x_l} = \underline{\underline{\frac{\partial^2 g}{\partial h^2} a_k a_l}}$

2.

a) If $A = zz^T$, then $x^TAx = x^Tz\, z^Tx = |\langle x|z\rangle|^2 \geq 0$.

b) rank $(A) = 1$, $\underline{N(A)} = \{x : Ax = 0\} = \{x : z(z^Tx) = \underline{0}\}$

$= \{x : z^Tx = 0\} = \underline{\underline{\{x : x \perp z\}}}$. $\underline{|N(A)| = n-1}$

c) $x^TBAB^Tx = b^TAb$, $b = B^Tx$. Since b is just another vector, we must have that also BAB^T is psd.

3.

a) $A = T\Lambda T^{-1} \iff AT = T\Lambda \iff AT = [t^{(i)}\lambda_i] \iff [At^{(i)}] = [t^{(i)}\lambda_i] \Rightarrow \underline{At^{(i)} = t^{(i)}\lambda_i}$

b) If U is orthogonal, it must also be invertible (in fact, $U^{-1} = U^T$). Thus this problem is simply a repeat of a.

c) This is obviously true if A is symmetric: $A = U\Lambda U^T \iff U^TAU = \Lambda$, which means (per 2c) Λ is psd when A is. And for Λ to be psd, we must have that every eigenvalue is ≥ 0, because $u^T\Lambda u = \sum_i \lambda_i u_i^2$ is ≥ 0 for all u_i only if $\lambda_i \geq 0 \; \forall i$.

4

a) Y is the aggregate of the elements in the random vector X.
$E(Y) = E(X_1) + E(X_2) + \ldots = \sum_{i=1}^n \mu_i$ and $\operatorname{Var} Y = \operatorname{Var} X_1 + \operatorname{Var} X_2 + \ldots + \operatorname{Var} X_n$
$+ 2\operatorname{Cov}(X_1, X_2 + X_3 + \ldots + X_n) + 2\operatorname{Cov}(X_2, X_3 + X_4 + \ldots X_n) + \ldots + 2\operatorname{Cov}(X_{n-1}, X_n)$.

This distribution is not necessarily known nor simple. But, if Σ is diagonal, it will be a Gaussian. And if the covariances (off-diagonal elements of Σ) aren't too large relative to the variances ($\operatorname{diag}\Sigma$) and n is very large, by CLT Y will also approach a Gaussian.

b) $\underline{E[x^T\Sigma^{-1}x]} = E[x_i(\Sigma^{-1})_{ij}x_j] = (\Sigma^{-1})_{ij}\overbrace{E(x_ix_j)}^{=\Sigma_{ij} + \mu_i\mu_j} = (\Sigma^{-1})_{ij}(\overbrace{\Sigma_{ij}}^{=\Sigma_{ji}} + \mu_i\mu_j)$, $\mu_k = E(x_k)$

$= tr(\overbrace{\Sigma^{-1}\Sigma}^{=I}) + tr(E(x)^T\Sigma^{-1}E(x)) = \underline{\underline{n + E(x)^T\Sigma^{-1}E(x)}}$.

$= E[tr(x^Tx\Sigma^{-1})]$
$= tr[\underbrace{E(x^Tx)}_{\Sigma + \mu^T\mu}\Sigma^{-1}]$
$= tr[\Sigma\Sigma^{-1} + \mu^T\mu\Sigma^{-1}] = n + tr(\mu^T\mu\Sigma^{-1}) = n + tr[\mu^T\Sigma^{-1}\mu] = \underline{n + \mu^T\Sigma^{-1}\mu}$

Alternatively
$E(x^T\Sigma^{-1}x) = E$