Solutions to Homework 4

1) a) Let $P(x) = \frac{x^T A x}{x^T x}$. Since A is symmetric, we have that

$A = V \Lambda V^T$, V orthogonal & Λ diagonal. Consequently,

$$P(x) = \frac{x^T V \Lambda V^T x}{x^T x} = \frac{y^T \Lambda y}{y^T y} = \frac{\sum \lambda_i y_i^2}{\sum y_i^2} \quad \text{where } y = V^T x$$

Note that by the chain rule,

$$\nabla_x P(x) = \nabla_x P(y(x)) = V^T \nabla_y P(y) = 0 \iff \nabla_y P(y) = 0 \text{ since}$$

V is orthogonal (and hence invertible).

Thus, it is sufficient to find the stationary pts of $P(y)$.

To that end,

$$\frac{\partial P}{\partial y_k} = \frac{2\lambda_i y_k \sum y_i^2 - 2y_i \sum \lambda_i y_i^2}{\left(\sum y_i^2\right)^2} = \frac{2\lambda_i y_k}{\sum y_i^2} - \frac{2y_i}{\sum y_i^2} \cdot \underbrace{\left(\frac{\sum \lambda_i y_i^2}{\sum y_i^2}\right)}_{P(y).}$$

Now suppose $\exists\, y \neq 0$ s.t. $\nabla P(y) = 0$. Then

$\exists k$ s.t. $y_k \neq 0$ so $\frac{2y_k \lambda_k}{\sum y_i^2} = \frac{2y_k P(y)}{\sum y_i^2} \Rightarrow P(y) = \lambda_k$.

Consequently, $\nabla_y P(y) = 0 \Leftrightarrow P(y) = \lambda_k$ for some eigenvalue λ_k.

Furthermore, $\nabla_y P(y) = \frac{(\Lambda - P(y) I)y}{\|y\|^2}$ so $\nabla_y P(y) = 0 \Leftrightarrow (\Lambda - P(y) I)y = 0$

But $P(y) = \lambda_k$ so y s.t. $(\Lambda - \lambda_k I)y = 0 \Rightarrow y$ is an eigenvector

corresponding to λ_k. Finally, since $\nabla_x P(x) = 0 \Leftrightarrow \nabla_y P(y) = 0$ we

have that $x = Vy$ are all the stationary pts of $P(x)$

where y is an eigenvector of Λ. In that case

$Ax = V\Lambda V^T V y = V\Lambda y = \lambda V y = \lambda x$. so x is an eigenvector

of A.

b) Note that $\nabla_x^2 \rho = V^T \nabla_y \rho V$ so by Sylvester's law of inertia, $\nabla_x^2 \rho$ has exactly the same number of +ve, -ve and 0 eigenvalues as $\nabla_y \rho$. Since for classifying stationary pts., only the sign of the eigenvalues is consequential, we may switch to classifying the stationary pts. of $\rho(y) = \frac{y^T \Lambda y}{y^T y}$

Moreover, we may assume $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ is +ve def. otherwise let $\mu > \max |\lambda_i|$ and observe that

$$g(y) = \frac{y^T(\Lambda + \mu I)y}{y^T y} = \frac{y^T \Lambda y}{y^T y} + \frac{\mu y^T y}{y^T y} = \rho(y) + \mu \quad \text{so}$$

ρ has exactly the same extrema (including types of extrema) as g. Now, to that end, let λ_k be any non-extremal eigenvalue of Λ and let y_k be the unit norm eigenvector. Let v be the unit norm e. vec. for λ_{max}, the maximal eigenvalue.

Then let $f(\alpha) = P(y_k + \alpha v) = \dfrac{(y_k+\alpha v)^T \Lambda (y_k + \alpha v)}{(y_k + \alpha v)^T (y_k + \alpha v)}$

$$= \frac{y_k^T \Lambda y_k + \underbrace{2\alpha y_k^T \Lambda v}_{0 \text{ because } v \perp y_k \text{ (spectral thm)}} + \alpha^2 v^T \Lambda v}{y_k^T y_k + 2\alpha y_k^T v + v^T v}$$

$$= \frac{\lambda_k \|y_k\|^2 + \alpha^2 \lambda_{max} \|v\|^2}{\|y_k\|^2 + \alpha^2 \|v\|^2}$$

$$= \frac{\lambda_k + \alpha^2 \lambda_{max}}{1 + \alpha^2}$$

$$= 1/\lambda_{max} \left(\frac{\underbrace{\lambda_k/\lambda_{max}}_{\varepsilon_k < 1.} + \alpha^2}{1+\alpha^2} \right)$$

$$= (1/\lambda_{max}) \left(\frac{\varepsilon_k + \alpha^2}{1+\alpha^2} \right).$$

Now, $f'(0) = v^T \nabla P(y_k) = 0$ since y_k is a critical pt. So we must check 2nd derivative @ $\alpha = 0$.

$$f'(\alpha) = (1/\lambda_{max}) \left(\frac{2\alpha(1+\alpha^2) - 2\alpha(\varepsilon_k + \alpha^2)}{(1+\alpha^2)^2} \right) = \frac{2\alpha}{\lambda_{max}} \left(\frac{1}{1+\alpha^2} - \frac{\varepsilon_k + \alpha^2}{(1+\alpha^2)^2} \right)$$

and $f''(\alpha) = \left(1/\lambda_{max}\right)\left(\frac{2\alpha}{(1+\alpha^2)^4} f(\alpha) + 2.\left(\frac{1}{1+\alpha^2} - \frac{\varepsilon_k + \alpha^2}{(1+\alpha^2)^2}\right)\right)$

& $f''(0) = 2(1-\varepsilon_k) > 0$ since $\varepsilon_k < 1$.

Thus along v, f'' is concave up so f decreases to

~~α~~ $f(0)$ along v; Moreover if instead we take v to

be the eigenvector u corresponding to λ_{min}, then since

$\lambda_{min} > 0$ and $\lambda_k > \lambda_{min}$ we have that $\varepsilon_k > 1$ so the same

computation reveals that when $f = P(y_k + \alpha u)$, $f''(0) < 0$

so f increases along u to $f(0)$. Thus, y_k must be a saddle whenever it is not an extremal eigenvector.

Finally, note that

$$\lambda_{min} = \frac{\lambda_{min} \sum y_i^2}{\sum y_i^2} = \frac{\sum \lambda_{min} y_i^2}{\sum y_i^2} \leq \frac{\sum \lambda_i y_i^2}{\sum y_i^2} = \rho(y) \leq$$

$$\frac{\sum \lambda_{max} y_i^2}{\sum y_i^2} = \lambda_{max}.$$ Furthermore, ρ achieves these minima

and maxima: When $y = v$, $\rho(v) = \lambda_{max}$ & $\rho(u) = \lambda_{min}$.

Thus, ρ has a unique max v, unique min u and

the rest of the non-extremal eigenvectors are saddles.

2) We proceed by ~~the~~ backwards induction on the rank of M,
To that end, let $A : \mathbb{R}^d \longrightarrow \mathbb{R}^n$ linear and fix the rank of A to be r. If $\operatorname{rank} M = k = r$ then clearly $\underset{rk(M) \leq r}{\operatorname{argmin}} \|A - M\| =$

$A = U_r \Sigma_r V_r^T = U_k \Sigma_k V_k^T$ where $\|\cdot\|$ is some arbitrary Ky-Fan norm.
Now suppose $\underset{\operatorname{rank} M \leq k}{\operatorname{argmin}} \|A - M\| = U_k \Sigma_k V_k^T$; for the inductive step,

assume $\operatorname{rank}(M) = k-1$. Then $M : \operatorname{Ker} M^\perp \rightarrow \operatorname{Img}(M) \subseteq \mathbb{R}^n$ is an isomorphism. Denoting $\mathcal{V}_j = \operatorname{span}\{v_1, \dots, v_j\}$ (and similarly $\mathcal{U}_j = \operatorname{span}\{u_1, \dots, u_j\}$), we note that $\operatorname{Ker} M^\perp \cong \mathcal{V}_{k-1}$ and $\operatorname{Img}(M) \cong \mathcal{U}_{k-1}$ where the isomorphisms are given by VV^T and UU^T respectively;
In particular, we may view $M : \mathcal{V}_{k-1} \longrightarrow \operatorname{Ker} M^\perp \longrightarrow \operatorname{Im}(M) \longrightarrow \mathcal{U}_{k-1}$
(Indeed there is a bijection between rank $k-1$ operators $\mathbb{R}^d \longrightarrow \mathbb{R}^n$ and isomorphisms from $\mathcal{V}_{k-1}$ to $\mathcal{U}_{k-1}$). Since M has rank $k-1$,
$M = \sum_{i=1}^{k-1} \delta_i^M u_i^M (v_i^M)^T$ s.t $u_i^M \in \mathcal{U}_{k-1}$ & $v_i^M \in \mathcal{V}_{k-1}$.

Explicitly, we may write $M = U_{k-1} U_{k-1}^T U_M \Sigma_M V_M^T V V_{k-1}^T$

Now, let $M_\varepsilon = \sum_{i=1}^{k-1} \sigma_i^M u_i^M (v_i^M)^T + \varepsilon u_{k-1} v_{k-1}^T$. Now, M_ε has rank k since $u_{k-1} \perp u_i^M$ & $v_{k-1} \perp v_i^M$; consequently, from the inductive hypothesis we have that $\min_{M_\varepsilon} \|A - M_\varepsilon\| = f((\sigma_k - \varepsilon), \sigma_{k+1}, \dots, \sigma_r)$ where f is an L^p norm. Finally, note that

$$\min_{rk(M)=k-1} \|A - M\| = \min_{rk M = k-1} \lim_{\varepsilon \to 0} \|A - M_\varepsilon\| = \lim_{\varepsilon \to 0} \min_{M = k-1} \|A - M_\varepsilon\| = \lim_{\varepsilon \to 0} f((\sigma_k - \varepsilon), \sigma_{k+1}, \dots$$

$$\sigma_r) = f(\sigma_k, \dots, \sigma_r)$$; furthermore the minimizer is $M = U_k \Sigma_k V_k^T$.

The induction is complete and thus $\operatorname{argmin}_{rk(M) \leq k} \|A - M\| = U_k \Sigma_k V_k^T$.

$\forall k \leq r$.