ORF523 - Problem Set 1

Bachir EL KHADIR

February 5, 2016

Problem 1

Q.1

1. Let λ be an eigen value of $A^T A$ corresponding to an eigen vector $u \neq 0$, then $0 \leq ||Au||^2 = u^T A^T A u = \lambda ||u||^2$, therefore $\lambda > 0$.

2. Let λ be an eigen value of A corresponding to an eigen vector u, the $A^T A u = A(Au) = \lambda^2 u$, so λ^2 is an eigen value of $A^T A$. Since A has n eigen values (accounting for multplicity), the eigen values of $A^T A$ are exactly the squares of the eigen values of A, and therefore the singular values of A are the absolute values of the eigen values of A.

3. $$u_i^T u_j = u_i^T \frac{A^T A u_j}{\lambda_j} = \frac{u_i^T A^T A}{\lambda_j} u_j$$

   Since $\lambda_i \neq \lambda_j$, $u_i^T u_j = 0$

Q.2

- The L_2 norm for vectors is unitarly invariant: Let O unitary matrix and X a vector, then $||OX||^2 = X^T O^T O X = X^T X = ||X||^2$.

- Since O is invertible, the application $S \to S, X \to OX$, where S is the L_2 sphere, is a bijection. So $\{x, ||x||_2 = 1\} = \{Ox, ||x||_2 = 1\}$

- The L_2 norm for matrices is unitarly invariant. If A a matrix, then $||AO|| = \max_{||x||_2=1} ||AOx|| = \max_{||Ox||_2=1} ||AOx|| = \max_{||y||_2=1} ||Ay|| = ||A||$ and $||OA|| = \max_{||x||_2=1} ||OAx|| = \max_{||x||_2=1} ||Ax|| = ||A||$.

- Let B be a matrix of rank at most k. $||A - B|| = ||U(\Sigma - U^T B V)V^T|| = ||\Sigma - U^T B V||$ Let $U'\Sigma' V'$ be the SVD of B, by a similar argument: $||A - B|| = ||\Sigma - \Sigma'||$.

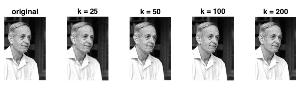
  $rank(B) = rank(\Sigma') \leq k$, so Σ' can be written as $\Sigma'^{(k)} = diag(\sigma_1', \ldots, \sigma_k', 0 \ldots)$. $||A - B|| = \sqrt{\sum_{i=1\ldots k} (\sigma_i - \sigma_i')^2 + \sum_{i=k+1\ldots n} \sigma_i^2} = \sqrt{\sum_{i=1\ldots k} (\sigma_i - \sigma_i')^2 + ||A - A^{(k)}||^2} \geq ||A - A^{(k)}||$, Since $rank(A^{(k)}) = k$ we have proved the result.

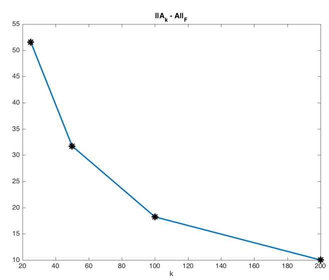
Q.3

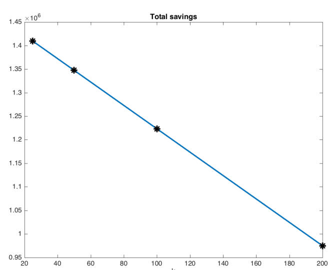
1.

k	$\|\|A - A_{(k)}\|\|_F$
25	51.5669
50	31.7256
100	18.2247
200	10.0059

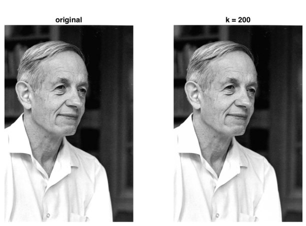
2. To store $A^{(k)}$ we need only to store $\Sigma^{(k)}$ (k parameters), $U^{(k)}, V^{(k)}$ (each of them has only at most k column not set to zero, so they take $nk + mk$), while A takes nm numbers to store. In conclusion, we save $(nm - k(n + m + 1))$.

    - 

    - 

    - 

3. Comparing the original and the compressed image for $k = 200$



Code

```
%cd Documents/Princeton/ORF523/hw1/

%% Load Image
A=imread('nash.jpg');
A=im2double (A) ;
A=rgb2gray(A) ;
[m,n] = size(A);

%% SVD decomposition
[U,S,V] = svd(A);

%% arrays initialization
k_range = [25 50 100 200];
diff_norm = zeros('like', k_range);
total_savings = zeros('like', k_range);

%% Plot the original image
subplot(1,length(k_range)+1, 1);
imshow(A);
title('original')

for i = 1:length(k_range)
    k = k_range(i);
    Uk = U; Sk = S; Vk = V;
    Sk(k:end, k:end) = 0;
    Uk(:, k:end) = 0;
    Vk(:, k:end) = 0;
    Ak = Uk * Sk * Vk';
    diff_norm(i) = norm(A - Ak,'fro');
    total_savings(i) = prod(size(A)) ... % size of A
    - k*m ...% size of Uk
    - k*n ... % size of Vk
    - k; % size of Sigma

    subplot(1,length(k_range)+1, i+1);
    imshow(Ak);
    title(['k = ' int2str(k)])
end
print('compressed','-dpng')

%% Plot ||A_k - A||_F
figure
plot(k_range, diff_norm, ...
     '-*',...
     'LineWidth',2,...
     'MarkerEdgeColor','k',...
     'MarkerFaceColor',[.49 1 .63],...
```

```
     'MarkerSize',10)
title('||A_k - A||_F')
xlabel('k')
print('diffnorm','-dpng')

%% Plot total savings
figure
plot(k_range, total_savings,...
     '-*',...
     'LineWidth',2,...
     'MarkerEdgeColor','k',...
     'MarkerFaceColor',[.49 1 .63],...
     'MarkerSize',10)

title('Total savings')
xlabel('k')
print('totalsavings','-dpng')

%% Plot original and compressed for k = 200
figure
subplot(1,2, 1);
imshow(A);
title('original')
subplot(1,2, 2);
imshow(Ak);
title('k = 200')
print('compare','-dpng')
```

Problem 2

1. Let $u = x_1 + x_2$, it is easy to verify that $f(x_1, x_2) = \underbrace{\frac{u^2}{2} - 2u}_{g(u)} + \underbrace{-2x_2^2 + 3x_2 + \frac{1}{3}x_2^3}_{h(x_2)} = g(u) + h(x_2)$

   Since this transformation is a diffeomorphisme, it preserves neighbourhoods, and therefore it preserves local minimizers/maximizers.

   A point (u, x_2) is a local maximizer (resp. minimizer) of $g(u) + h(x_2)$ if and only if u is a local maximizer (resp. minimizer) of g and x_2 is local maximizer (resp. minimizer) of h.

   - $g(u) = \frac{1}{2}(u-2)^2 - 2$ has one local(and global) minimum $u = 2$
   - $h'(x_2) = x_2^2 - 4x_2 + 3 = (x_2 - 1)(x_2 - 3)$, $h''(x_2) = 2x_2 - 4$

   The candidates are $x_2 = 1$ and $x_2 = 3$. Since $h''(1) = -2 < 0, h''(3) = 2 > 0$, 1 is local maximum and 3 is a local minimum.

   In conclusion, $(u, x_2) = (2, 1)$ is the only local optimizer (it is a local min) of $g(u) + h(x_2)$, and $(x_1 = -3, x_2 = 1)$ is the only local optimzer (minimum) of f.

2. By Taylor expansion, since f is a polynomial of order 2: $f(x) - f(\bar{x}) = \nabla f(\bar{x})(x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla^2 f(x - \bar{x})(x - \bar{x})$

   a)($\Rightarrow$) Let $\bar{x}$ be a local min, then $\nabla f(\bar{x}) = 0$. If $\nabla^2 f$ is not positive semi-definite, let $d \in \mathbb{R}^n d^T \nabla f(\bar{x}) =: -\lambda < 0$, therefore $f(x + \alpha d) - f(x) = -\alpha^2 \lambda < 0$, and for any neighbourhood of $\bar{x}$, there exist α such that $\bar{x} + \alpha d$ is in that neighbourhood, and we have a contradiction.

   ($\Leftarrow$) In this case: $f(x) - f(\bar{x}) = \frac{1}{2}(x - \bar{x})\nabla^2 f(x - \bar{x}) \geq 0$, which proves that $\bar{x}$ is local min.

   b)($\Rightarrow$) In this case: $f(x) - f(\bar{x}) = \frac{1}{2}(x - \bar{x})\nabla^2 f(x - \bar{x}) > 0$, which proves that $\bar{x}$ is strict local min.

   ($\Leftarrow$) In this case: $f(x) - f(\bar{x}) = \frac{1}{2}(x - \bar{x})\nabla^2 f(x - \bar{x}) > 0$, which proves that $\bar{x}$ is strict local min.

   Counterexamples: $f : \mathbb{R} \to \mathbb{R}, x \to x^3$, $f'(0) = 0, f''(0) = 0$, but 0 is not a local min. $f : \mathbb{R} \to \mathbb{R}, x \to x^4$, $f'(0) = 0, f''(0) = 0$, but 0 is a strict local min.

3. using the chain rule: $g'(\alpha) = \frac{d}{||d||}\nabla f(x + \alpha\frac{d}{||d||})$, $g'(0)\frac{d}{||d||}\nabla f(x)$, using Cauchy-Shwarz, this scalar product is minimal/ maximal when d and $\nabla f(x)$ are colinear.

Problem 3

$Q > 0$, so it admist an eigenvalue decomposition and all the eigen values are positive, let $Q = U\Sigma U^T$ be that decomposition, with $\Sigma = diag(\sigma_1, \ldots, \sigma_n)$, and Let $\sqrt{\Sigma} = diag(\sqrt{\sigma_1}, \ldots, \sqrt{\sigma_n})$, then $Q = \underbrace{U\sqrt{\Sigma}U^T}_{\sqrt{Q}} U\sqrt{\Sigma}U^T = \sqrt{Q}^2$

$f(x) = \sqrt{x^T\sqrt{Q}\sqrt{Q}x} = \sqrt{||\sqrt{Q}x||^2} = ||\sqrt{Q}x||$

1. f is a norm because:

   - $$f(\lambda x) = \sqrt{(\lambda x)^T Q(\lambda x)} = \sqrt{\lambda^2} f(x) = |\lambda| f(x)$$

- $$f(x+y) = ||\sqrt{Q}x + \sqrt{Q}y|| \leq ||\sqrt{Q}x|| + ||\sqrt{Q}y|| \leq f(x) + f(y)$$
- $$f(x) = 0 \iff \sqrt{Q}x = 0 \iff x = 0$$

By Riesz Representation theorem, we indentify a vector x with the linear form $y \to x^T y$. Let g be the dual norm of f, then

$$\begin{aligned}
g(x) &= \sup_{y \neq 0} \frac{x^T y}{f(y)} \\
&= \sup_{u \neq 0} \frac{x^T \sqrt{Q}^{-1} u}{||u||} && (y \to u = \sqrt{Q}y \text{ is a bijection}) \\
&= \sup_{u \neq 0} x^T \sqrt{Q}^{-1} \frac{u}{||u||} \\
&= x^T \sqrt{Q}^{-1} \frac{(\sqrt{Q}^{-1} x)}{||\sqrt{Q}^{-1} x||} && \text{(Cauchy shwarz)} \\
&= \frac{x^T Q^{-1} x}{\sqrt{x^T Q^{-1} x}} \\
&= \sqrt{x^T Q x^{-1}}
\end{aligned}$$

3. $A^T A \geq 0$, Let $U \Sigma U^T$ be an eigen value decomposition where $\Sigma = diag(\sigma_1, \ldots, \sigma_n)$, and $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$.

$$\begin{aligned}
||A||_2 &= \sup_{||x||=1} x^T (A^T A) x \\
&= \sup_{||x||=1} ||\sqrt{A^T A} x|| \\
&= ||\sqrt{A^T A}||_2 \\
&= ||\Sigma||_2 \\
&= \sup_{||x||=1} x^T (\Sigma) x \\
&= \sup_{||x||=1} \sum_i x_i^2 \sigma_i \leq \sum_i x_i^2 \sigma_1 \\
&= \sigma_1 \\
&= ||\Sigma e_1||
\end{aligned}$$

So $||A||_2 = \sigma_1 = \sqrt{\lambda_{\max}(A^T A)}$