ORF525 - Problem Set 1

Bachir EL KHADIR

February 15, 2016

Q.1

1. The conversion is necessary because otherwise we would have an unwanted order relation. For three similar houses A, B, C in zipcodes 98001, 98002, 98003, a linear model would be forced to affect a price for the house A that lies between the price for house A and C, which is a bug and not a future of the data itself.

2.

Q.2

- $\|Y-\theta\|_2^2 + 4\tau^2\|\theta\|_0 = \sum(y_i-\theta_i)^2 + 4\tau^2 1_{\theta_i\neq 0} = \sum_i f(\theta_i)$ Where $f : \theta \to (y-\theta)^2 + 4\tau^2 1_{\theta\neq 0}$, eg

$$f(\theta) = \begin{cases} y^2 & \text{if } \theta = 0 \\ (y-\theta)^2 + 4\tau^2 & \text{if } \theta \neq 0 \end{cases}$$

  The problem is linearly separable, we can minimize on each variable θ_i independetly:

  - If $|y| > 2\tau$, then $y^2 \geq 4\tau^2$ and $(y-\theta)^2 + 4\tau^2 \geq 4\tau^2 = f(y)$.
  - If $|y| \leq 2\tau$, then $f(0) = y^2 \leq 4\tau^2 \leq y^2 + (y-\theta) = f(\theta) \forall \theta \neq 0$.

  So $\arg\min \|Y-\theta\|_2^2 + 4\tau^2 = \hat{\theta}^{\text{hard}}$.

- $$\|Y-\theta\|_2^2 + 4\tau\|\theta\|_1 = \sum(y_i-\theta_i)^2 + 4\tau|\theta_i| = \sum_i g(\theta_i)$$

  The problem is linearly separable, we can minimize on each variable θ_i independetly. $g : \theta \to (y-\theta)^2 + 4\tau|\theta|$, eg

$$g(\theta) = \begin{cases} g_1(\theta) = (y-\theta)^2 + 4\tau\theta & = (\theta - (-2\tau + y))^2 + 2\tau(\tau + y) & \text{if } \theta \geq 0 \\ g_2(\theta) = (y-\theta)^2 - 4\tau\theta & = (\theta - (2\tau + y))^2 + 2\tau(\tau + y) & \text{if } \theta \leq 0 \end{cases}$$

  - * If $|y| \leq 2\tau$, then g_1 is increasing on $[0, \infty)$, so it has a minimum at 0, and the minimum is $y^2 = g(0)$.
    * g_2 is decreasing on $(-\infty, 0]$ so it has a minimum at 0.
    * c/c $\arg\min g$ in this case is 0.
  - * If $y \geq 2\tau$ then g_1 has minimum at $y - 2\tau > 0$ and the minimum is $y^2 - (2\tau - y)^2$.
    * g_2 is decreasing and has a minimum at 0, $g_2(0) = y^2 \geq g(y-2\tau)$ with equality only if $y - 2\tau = 0$

* c/c $\arg\min g$ in this case is $y - 2\tau$.

- if $y \leq -2\tau$, by a similar argument, $\arg\min g$ in this case is $y - 2\tau$.

So $\arg\min \|Y - \theta\|_2^2 + 4\tau\theta = \hat{\theta}^{\text{hard}}$.

Q.3

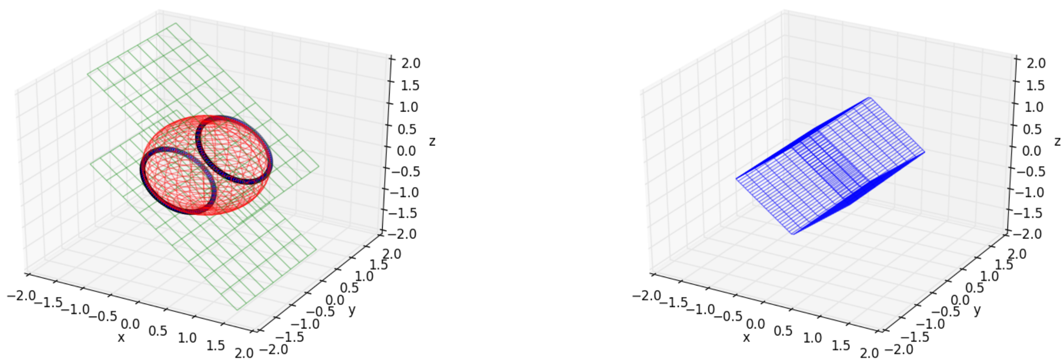

Figure 1: Set

1. (a) $M(x, y, z)$ is of rank 1, so (xy) and (yz) are colinear, so $\exists\lambda \in \mathbb{R}$ $(xy) = \lambda(yz)$ or $(yz) = \lambda(zy)$

    i. Let's assume that $(xy) = \lambda(yz)$, and we can deduce the other case by symmetry. In this case $x = \lambda^2 z, y = \lambda z$. Which means that $y^2 = \lambda^2 zz = xz$.

    ii. When $(yz) = \lambda(xy)$, by a similar argument, $y^2 = xz$

    The op norm is the biggest eigen value in absolute value, in this case since one of the eigen values is 0 (because the determinant is 0) the op norm is: $|tr(M(x, y, z))| = |x + z|$, this is equal to 1 only if $1 = |x + z|^2 = x^2 + z^2 + 2xz = x^2 + z^2 + 2y^2$ Letting $t = \sqrt{y}$, we have that: $\{(x, \sqrt{2}y, z) |\, \text{Rank}(M) = 1, \|M\|_{op} = 1\} \subseteq \{(x, t, z) : x^2 + t^2 + z^2 = 1, |x + z| = 1\}$

    For a matrix $M(x, y, z)$ such that $x^2 + z^2 + 2y^2 = 1$ and $|x + z| = 1$, it is easy to see that:

    - $\|M\|_{op} = |\operatorname{tr}(M)| = |x + z| = 1$
    - $y^2 = \frac{1-x^2+z^2}{2} = \frac{(x+z)^2-x^2-y^2}{2} = xz$, so $\det(M) = 0$, therefore M cannot have rank 2. It cannot have rank 0 either because $x^2 + 2y^2 + z^2 = 1 \implies$ one of the coefficient is not 0.

    Therefore we have: $\{(x, \sqrt{2}y, z) |\, \text{Rank}(M) = 1, \|M\|_{op} = 1\} = \{(x, t, z) : x^2 + t^2 + z^2 = 1, |x + z| = 1\}$

    $x^2 + t^2 + z^2$ describes the unit sphere, $|x + z| = 1$ describe the union of the two hyperplanes $x + z = \pm 1$. Therefore this set is the union of intersection of sphere with two hyperplanes, e.g a union of two circles.

    Or in parametric form:

$$\begin{cases} x &= \frac{1}{2} + \cos\theta \\ t &= \frac{\sqrt{2}}{2}\sin\theta \\ z &= -\frac{1}{2} - \cos\theta \end{cases}$$

$$\begin{cases} x &= -\frac{1}{2} + \cos\theta \\ t &= \frac{\sqrt{2}}{2}\sin\theta \\ z &= -\frac{3}{2} - \cos\theta \end{cases}$$

(b) $M := M(x, y, z)$ is symmetric, let's call σ_1, σ_2 its eigen values, we know that:

$$\begin{aligned} \|M\|^* &= |\sigma_1| + |\sigma_2| \\ tr(M) &= x + z = \sigma_1 + \sigma_2 \\ tr(M^T M) &= x^2 + z^2 - 2y^2 = \sigma_1^2 + \sigma_2^2 \\ \det(M) &= xz - y^2 = \sigma_1 \sigma_2 \end{aligned}$$

Therefore $(\sigma_1 - \sigma_2)^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2 = tr(M^T M) - 2\det(M) = x^2 + z^2 - 2y^2 - 2(xz - y^2) = (x - z)^2$ Since $\|M\|_*^2 = \max\{|\sigma_1 + \sigma_2|^2, |\sigma_1 - \sigma_2|^2\}$, $\|M\|_*^2 = \max\{(x + z)^2, (x - z)^2\}$, and therefore

$$\|M\|_* \leq 1 \iff \begin{cases} (x+z)^2 \leq 1 \\ (x-z)^2 \leq 1 \end{cases} \iff \begin{cases} -1 \leq x + z \leq 1 \\ -1 \leq x - z \leq 1 \end{cases}$$

Which describes the square whose edges are $(1, 0), (0, 1), (-1, 0), (0, -1)$ in the plane (X, Z). In 3d it is the polyhedral obtained by extruding that square according to the Y-axis.

2. If $A = uv^T$ has rank ≤ 1, A has at most one non null singular value, then $\|A\|_{op} = |\operatorname{tr}(A)| = |v^T u| = \|A\|_*$, if $\|A\|_{op} \leq 1$, then $\|A\|_* \leq 1$. The nuclear norm is a norm (fact proven in the next exercise), so the unit ball is convex, and therefore: $\operatorname{conv}\{uv^T : \|uv^T\|_{op} \leq 1\} \subseteq \{X : \|X\|_* \leq 1\}$.

   Let $X \in \mathbb{R}^{d_1 \times d_2}$ st $\|X\|_* \leq 1$ and let $U\Lambda V^T$ be its SVD. Then $\Lambda = \sum_{i=1}^d \sigma_i(X) e_i^T e_i$ where $(e_i)_i$ is the canonical basis of R^{d^2}. Therefore $X = \sum_{i=1}^d \sigma_i(X) \underbrace{U e_i^T e_i V^T}_{\text{Rank}=1} + (1 - \underbrace{\sum_{i=1}^{d} \sigma_i(X)}_{\|X\|_{op} \leq 1})0 \in \operatorname{conv}\{uv^T : \|uv^T\|_{op} \leq 1\}$. c/c $\operatorname{conv}\{uv^T : \|uv^T\|_{op} \leq 1\} = \{X : \|X\|_* \leq 1\}$.

3. Two remarks:

   - For A, B sqaure matrices of the same shape, $\|AB\|_F^2 = \operatorname{tr}(ABB^T A^T) = \operatorname{tr}(A^T ABB^T) = \operatorname{tr}(BB^T A^T A) = \|B^T A^T\|_F^2$
   - If O is orthogonal, the $\|AO\|_F^2 = \operatorname{tr}(AOO^T A^T) = \operatorname{tr}(AA^T) = \|A\|_F^2$, $\|OA\|_F = \|A^T O^T\|_F = \|A^T\|_F = \|A\|_F$.

   $\|X - XZ\|_F^2 = \|X(I - Z)\|_F^2 = \|U\Lambda V^T(I - Z)\|_F^2 = \|\Lambda V^T(I - Z)\|_F^2 = \|\Lambda V^T V(V^T V - V^T Z V)V^T\|_F^2 = \|\Lambda(I - V^T Z V)\|_F^2$.

   Since V is invertible, let's do the change of variable $Y = V^T Z V$. Moreover, V is orthogonal, so the singular values of Y and Z are the same as well as their nuclear norm. The problem can thus be reduced to:

$$\min_Y \|Y\|_* + \frac{\tau}{2}\|\Lambda(I - Y)\|_F^2$$

$$
\begin{aligned}
\|\Lambda(I-Y)\|_F^2 &= \sum_{ij}[e_i^T\Lambda(I-Y)e_j]^2 \\
&= \sum_{ij}[\Lambda_i e_i^T(e_j - Ye_j)]^2 \\
&= \sum_{ij}[\Lambda_i(\delta_{ij} - Y_{ij})]^2 \\
&= \sum_{i\neq j}\underbrace{\Lambda_i^2 y_{ij}^2}_{\geq 0} + \sum_i \Lambda_i^2(1-y_{ii})^2 \\
&\geq \sum_i \Lambda_i^2(1-y_{ii})^2
\end{aligned}
$$

Let $Y = \sum_i \sigma_i(Y)u_iv_i^T$ be the SVD of Y, then $\operatorname{tr}(Y) = \sum_i \sigma_i(Y)\operatorname{tr}(u_iv_i^T) = \sum_i \sigma_i(Y) \underbrace{v_i^T u_i}_{\leq \|u_i\|\|v_i\|\leq 1} \leq \|Y\|_*$ As a result $\|Y\|_* + \frac{\tau}{2}\|\Lambda - \Lambda Y\|_F^2 \geq \sum_i y_{ii} + \frac{\tau}{2}\Lambda_i^2(1-y_{ii})^2$ Minimizing the quadratic function $y \to y + \frac{\tau}{2}\Lambda_i^2(1-y)^2$ leads to $y = (1 - \frac{1}{\tau\Lambda_i^2})^+$ Therefore $Y =$

Q.4

1. (a) Let $Y = U\Lambda V^T$ be the SVD decomposition of Y, then $\langle Y, UVT\rangle = \operatorname{tr}(V\Lambda U^TUV^T) = \operatorname{tr}(V\Lambda V^T) = \operatorname{tr}(\Lambda) = \|Y\|_*$.
   $\langle Y, X\rangle = \operatorname{tr}(Y^TX) = \operatorname{tr}(V\Lambda U^TX) = \operatorname{tr}(\Lambda U^TXV) = \sum \Lambda_{ii}(U^TXV)_{ii} = \sum \Lambda_{ii}\underbrace{u_i^TXv_i}_{\leq \|X\|_{op}} \leq \|X\|_{op}\|\Lambda\|_* \leq \|Y\|_*$.

   c/c $\|Y\|_* = \max_{\|X\|_{op}\leq 1}\langle Y, X\rangle$

2. for $\alpha \in (0,1)$, $\|\alpha Y + (1-\alpha)Z\|_* = \max_{\|X\|_{op}\leq 1}\alpha\langle Y, X\rangle + (1-\alpha)\langle Z, X\rangle \leq \alpha\max_{\|X\|_{op}\leq 1}\langle Y, X\rangle + (1-\alpha)\max\langle Z, X\rangle \leq \alpha\|Y\|_* + (1-\alpha)\|Z\|_*$

3. $\Rightarrow$ Let's suppose $Z \in \partial\|A\|_*$ Then $\|B\|_* \geq \|A\|_* + \langle Z, B - A\rangle$. For $B = Z$, $\|Z\|_* \geq \|A\|_* + \|Z\|_F^2 - \langle Z, A\rangle$. For $B = 0$, $0 \geq \|A\|_* - \langle Z, A\rangle \implies \langle Z, A\rangle \geq \|A\|_*$.

   $\Leftarrow$ Let Z, such that $\|Z\|_{op} = 1$ and $\langle Z, A\rangle = \|A\|_*$, then $\forall X$ $\langle Z, X\rangle \leq \|X\|_*$ and: $\|X\|_* - \|A\|_* \geq \langle Z, X\rangle - \langle Z, A\rangle$, which means that $Z \in \partial\|A\|_*$

4. (a)

$$
\begin{aligned}
Z \in \partial\|A\| &\implies \|Z\|_{op} = 1, \|A\|_* = \|\Lambda\|_* = \langle Z\,,\, A\rangle \\
&\implies \|\Lambda\|_* = \operatorname{tr}(Z^TU\Lambda V^T) = \operatorname{tr}((U^TZV)^T\Lambda) = \sum_i (U^TZV)_{ii}\Lambda_i \\
&\implies \sum_i \underbrace{\Lambda_i(1 - u_i^TZv_i)}_{\geq 0} = 0 \qquad (|u_i^TZv_i| \leq \|u_i\|\|v_i\|\|Z\|_{op} \\
&\implies \forall i\; u_i^TZv_i = 1
\end{aligned}
$$

   Let's complete the family $(u_i)_{i\leq r}$ into $(u_i)_{i\leq d_1}$ an orthonormal basis of $\mathbb{R}^{d_1}$. Then, for $i \leq r$, $1 \geq \|Zv_i\|^2 = \sum_{j=1}^{d_1}(u_j^TZv_i)^2 \geq 1 + \sum_{j\neq i}\underbrace{(u_j^TZv_i)^2}_{=0} \geq 1$

So $u_j^T Z v_i = \delta_{ij}$ and $Zv_i = \sum_{j=1}^{d_1} u_j^T Z v_i u_j = u_i^T Z v_i u_i = u_i$. In matrix form: $ZV = U$, and using a similar arguement $U^T Z = V$.

Let $W := Z - UV^T$, then the last equations can be written as $U = (W + UV^T)V = WV + UV^T V = WV + U \implies WV = 0$ and similarly $U^T W = 0$ Let $x \in \mathbb{R}^{d_2}$, and let $x = x_1 + x_2$ be a decomposition according to $\mathbb{R}^{d_2} = im(V) + im(V)^\perp$, and let $y \in \mathbb{R}^r$ such that $x_1 = Vy$. (note that $\|x\|^2 = \|x_1\| + \|x_2\|^2$)

Then $\|Wx\|^2 = \|\underbrace{ZV}_{U} y + Zx_2 - U \underbrace{V^T V}_{I_r} y - U \underbrace{V^T x_2}_{=0}\|^2 = \|Uy + Zx_2 - Uy\|^2 = \|Zx_2\|^2 \leq \|x_2\|^2 \leq \|x\|^2$

c/c $Z = U^T V + W, \|W\|_{op} \leq 1$.

(b) Now take Z of the form $UV^T + W$, and let's prove that $\|Z\|_{op} \leq 1$ and $\langle Z\,,\, A\rangle = \|A\|_*$.

$$\begin{aligned}
\langle UV^T + W\,,\, A\rangle &= \mathrm{tr}(VU^T U\Lambda V^T) + \mathrm{tr}(W^T U \Lambda V^T) \\
&= \mathrm{tr}(\Lambda) \\
&= \|A\|_*
\end{aligned}$$

Let $x \in \mathbb{R}^{d_2}$, then :

$$\begin{aligned}
\|UV^T x + Wx\|^2 &= \|UV^T x\|^2 + \|Wx\| && \text{(because } im(U) \perp im(W)\text{)} \\
&= \|V^T x\|^2 + \|Wx\|^2 && \text{(Because } U \text{ is an isometrie)}
\end{aligned}$$

Let's write $x = x_1 + x_2$ according to the decomposition $\mathbb{R}^{d_2} = im(V) + im(V)^\perp$, and let $y \in \mathbb{R}^r$ such that $x_1 = Vy$. (note that $\|x_1\| = \|y\|$) $V^T x = V^T x_1 + V^T x_2 = V^T V y = y$, $Wx = WVy + Wx_2 = Wx_2$ so $\|(UV^T + W)x\|^2 = \|y\|^2 + \|Wx_2\|^2 \leq \|x_1\|^2 + \|x_2\|^2 = \|x\|^2$, which proves that $\|UV^T + W\|_{op} \leq 1$.

5.
- Let $Z \in \partial\|A\|_*$, then Z can be written as $Z = UV^T + W$ where $U^T W = WV = 0$ and $\|W\|_{op} \leq 1$ $UV^T = U\frac{V}{2}^T + \frac{U}{2}V^T \in T$ Let's now prove that $W \perp T$. Let $X \in \mathbb{R}^{d_1 \times r}, Y \in \mathbb{R}^{d_2 \times r}$: $\langle UX^T + Y^T V\,,\, W\rangle = \mathrm{tr}(UX^T W) + \mathrm{tr}(Y^T V W) = \mathrm{tr}(WUX^T) = 0$. As a conclusion $\Pi_T(Z) = UV^T, \Pi_{T^c}(Z) = W$, $\|\Pi_{T^c}(Z)\|_{op} = \|W\|_{op} \leq 1$
- Let Z be such that $\Pi_T(Z) = UV^T$ and $\|\Pi_{T^c}(Z)\|_{op} \leq 1$ Let's note $W = Z - UV^T$. To prove that $Z \in \partial\|W\|_*$, it is enough to prove that $W = \Pi_{T^c}(Z)$.