15-496/859X: Computer Science Theory for the Information Age — Spring 2012
Carnegie Mellon University — V. Guruswami & R. Kannan

Problem Set 2 Solution

by Yuan Zhou

1. We can write A as
$$A = \sum_{i=1}^{d} \ell_i \cdot \frac{\boldsymbol{w_i}}{\ell_i} \cdot \boldsymbol{e_i}^T,$$
where $\boldsymbol{e_i}$ is the i-th unit vector with all entries 0 except for the i-th entry being 1.

2. Since the row vectors of A are orthonormal, we have that $AA^T = I$. For square matrix A, this implies that $A^T = A^{-1}$. Since $A^{-1}A = I$, we have $A^T A = I$, which implies that the column vectors of A are also orthonormal.

   When A is not a square matrix (when $A \in \mathbb{R}^{m\times n}$ where $m < n$), the statement is not true. The following matrix is a counterexample,
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

3. (a) Since A has rank n, $A^T A$ also has rank n (full rank). Therefore $A^T A$ is a positive definite matrix, and $(A^TA)^{1/2}, (A^TA)^{-1/2}, (A^TA)^{-1}$ exist. Now, note that
$$\begin{aligned}\|Ax-b\|^2 =&(Ax-b)^T(Ax-b)\\ =&x^T(A^TA)x - 2b^TAx + b^Tb\\ =&((A^TA)^{1/2}x)^T((A^TA)^{1/2}x) - 2((A^TA)^{-1/2}A^Tb)^T((A^TA)^{1/2}x) + \|b\|^2\\ =&\|(A^TA)^{1/2}x - (A^TA)^{-1/2}A^Tb\|^2 - \|(A^TA)^{-1/2}A^Tb\|^2 + \|b\|^2.\end{aligned}$$
   The second term above is a constant (independent of x), while the first term is always nonnegative, and it is 0 only when $x = (A^TA)^{-1}A^Tb$. Therefore, $x = (A^TA)^{-1}A^Tb$ is the unique minimizer of $\|Ax-b\|^2$ (as well as $\|Ax-b\|$), and the minimum value is $(\|b\|^2 - \|(A^TA)^{-1/2}A^Tb\|^2)$ ($\sqrt{\|b\|^2 - \|(A^TA)^{-1/2}A^Tb\|^2}$ correspondingly).

   (b) Fix an x, let $x = \sum_{i=1}^r \alpha_i v_i + x^\perp$ where $x^\perp \perp v_i$ for all i. We also let $b = \sum_{i=1}^r \beta_i u_i + b^\perp$ where $b^\perp \perp u_i$ for all i. Now we have
$$\|Ax-b\|^2 = \left\|\sum_{i=1}^r (\sigma_i\alpha_i - \beta_i)u_i + b^\perp\right\|^2 = \sum_{i=1}^r (\sigma_i\alpha_i - \beta_i)^2 + \|b^\perp\|^2 \geq \|b^\perp\|^2.$$
   Where the equality is achieved when $\alpha_i = \frac{\beta_i}{\sigma_i} = \frac{\langle b, u_i\rangle}{\sigma_i}$ for all i. Therefore,
$$x^* = \sum_{i=1}^r \beta_i v_i = \sum_{i=1}^r \frac{\langle b, u_i\rangle}{\sigma_i} v_i$$
   minimizes $\|Ax-b\|^2$ (which also minimizes $\|Ax-b\|$).

4. (a) The n singular values are $\lambda_1, \lambda_2, \cdots, \lambda_n$.

(b) "If" part: since M is real symmetric, we can assume $v_1, v_2, \cdots, v_n$ is a set of orthonormal eigenvectors. The corresponding eigenvalue $\lambda_i = v_i^T M v_i \geq 0$ for all i. Therefore M is p.s.d. by definition.

"Only if" part: if M is p.s.d., then we can write $M = \sum_{i=1}^n \lambda_i v_i v_i^T$ where $v_1, v_2, \cdots, v_n$ is a set of orthonormal eigenvectors and $\lambda_i \geq 0$ for all i. Now, for any $x \in \mathbb{R}^n$, $x^T M x = \sum_{i=1}^n \lambda_i (v_i^T x)^2 \geq 0$.

(c) For all $x \in \mathbb{R}^n$, $x^T V M V^T x = (V^T x) M (V^T x) \geq 0$ (by part(b)). Therefore, VMV^T is p.s.d. (by part(b) again).

(d) Write $A = U \Sigma V^T$ in its singular value decomposition form. Therefore $A = UV^T V \Sigma V^T = WP$ where we define $W = UV^T$ and $P = V \Sigma V^T$. Observe that $W^T W = VU^T UV^T = I$, $WW^T = UV^T VU^T = I$ and P is p.s.d. by part (c).

5. (a) Note that for all $x \in \mathbb{R}^n$,

$$x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2 \geq 0.$$

Therefore L is p.s.d. .

(b) Let $x = (1, 1, 1, \cdots, x)^T$. We see that $Lx = \mathbf{0}$. Therefore the smallest eigenvalue of L is 0 (since all the eigenvalues are nonnegative).

(c) For all unit vector x,

$$x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2 \leq \sum_{(i,j) \in E} 2(x_i^2 + x_j^2) = 2d \sum_i x_i^2 = 2d.$$

Therefore the largest eigenvalue (which is $\|L\|_2$, since L is p.s.d.) is at most $2d$.