Computer Vision (600.461/600.661)
Homework 1: Mathematical Background

Instructor: René Vidal

Due 09/11/2014, 11.59PM Eastern

1. **Properties of Symmetric Matrices [20pts].** Let $S \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. For $i = 1, \dots, n$, let $(\lambda_i, \boldsymbol{v}_i)$ be an eigenvalue-eigenvector pair. Show that:

   (a) All the eigenvalues of S are real, i.e., $\lambda_i \in \mathbb{R}$ [4pts].

   ANSWER: Let $(\lambda, \boldsymbol{v})$ be an eigenvalue-eigenvector pair of S. Then $S\boldsymbol{v} = \lambda\boldsymbol{v}$ and $\bar{\boldsymbol{v}}^\top S\boldsymbol{v} = \lambda\|\boldsymbol{v}\|^2$, where $\bar{\boldsymbol{v}}^\top$ is the conjugate transpose of $\boldsymbol{v}$ and $\bar{\lambda}$ is the conjugate of λ. Since $S^\top = S$, we have $\bar{\boldsymbol{v}}^\top S = \bar{\lambda}\bar{\boldsymbol{v}}^\top$ and so $\bar{\boldsymbol{v}}^\top S\boldsymbol{v} = \bar{\lambda}\|\boldsymbol{v}\|^2$. Therefore, $(\lambda - \bar{\lambda})\|\boldsymbol{v}\|^2 = 0$ with $\boldsymbol{v} \neq 0$, and so $\bar{\lambda} = \lambda$, which implies $\lambda \in \mathbb{R}$.

   (b) Eigenvectors corresponding to distinct eigenvalues are orthogonal, i.e., if $\lambda_i \neq \lambda_j$, then $\boldsymbol{v}_i \perp \boldsymbol{v}_j$ [4pts].

   ANSWER: Let $(\lambda_i, \boldsymbol{v}_i)$ and $(\lambda_j, \boldsymbol{v}_j)$ be two eigenvalue-eigenvector pairs of S. Then, $S\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i$, $\boldsymbol{v}_i^\top S = \lambda_i\boldsymbol{v}_i^\top$ and $S\boldsymbol{v}_j = \lambda_j\boldsymbol{v}_j$. Therefore, $\boldsymbol{v}_i^\top S\boldsymbol{v}_j = \boldsymbol{v}_i^\top(S\boldsymbol{v}_j) = \boldsymbol{v}_i^\top(\lambda_j\boldsymbol{v}_j) = \lambda_j(\boldsymbol{v}_i^\top\boldsymbol{v}_j) = \lambda_i(\boldsymbol{v}_i^\top\boldsymbol{v}_j)$. We then have $(\lambda_i - \lambda_j)(\boldsymbol{v}_i^\top\boldsymbol{v}_j) = 0$. If $\lambda_i \neq \lambda_j$, this equality can hold only if $\boldsymbol{v}_i^\top\boldsymbol{v}_j = 0$, i.e., $\boldsymbol{v}_i \perp \boldsymbol{v}_j$.

   (c) There always exist n orthonormal eigenvectors of S, which form a basis of $\mathbb{R}^n$ [4pts].

   ANSWER: Let $(\lambda_1, \boldsymbol{v}_1)$ be any eigenvector-eigenvalue pair of S and let $\boldsymbol{v}_1^\perp$ be its orthogonal complement. Since S is symmetric, $\boldsymbol{v}_1^\perp$ is an S-invariant subspace of $\mathbb{R}^n$, i.e., for all $\boldsymbol{w} \in \boldsymbol{v}_1^\perp$, we have $S\boldsymbol{w} \in \boldsymbol{v}_1^\perp$ because $\boldsymbol{v}_1^\top S\boldsymbol{w} = \lambda_1\boldsymbol{v}_1^\top\boldsymbol{w} = 0$. Thus, there exists an eigenvector of S in $\boldsymbol{v}_1^\top$. Let $\boldsymbol{v}_2$ be such an eigenvector and let λ_2 be its corresponding eigenvalue (which need not be different from λ_1). By the same argument, $\boldsymbol{v}_1^\perp \cap \boldsymbol{v}_2^\perp$ is an S-invariant subspace of $\mathbb{R}^n$, hence there exists an eigenvector of S in $\boldsymbol{v}_1^\perp \cap \boldsymbol{v}_2^\perp$. Finite induction finishes the proof.

   (d) S is positive definite (positive semidefinite) if and only if all of its eigenvalues are positive (non-negative), i.e., $S \succ 0$ ($S \succeq 0$), iff $\forall i = 1, 2, \dots, n$, $\lambda_i > 0$ ($\lambda_i \geq 0$) [4pts].

   ANSWER: By the previous part, we can choose n eigenvalue-eigenvector pairs $\{(\lambda_i, \boldsymbol{v}_i)\}_{i=1}^n$, such that $\boldsymbol{v}_i^\top\boldsymbol{v}_j = 0$ if $i \neq j$ and $\|\boldsymbol{v}_i\| = 1$. Since these eigenvectors form a basis for $\mathbb{R}^n$, any vector $\boldsymbol{x} \in \mathbb{R}^n$ can be written as $\boldsymbol{x} = \sum_{i=1}^n(\alpha_i\boldsymbol{v}_i), \alpha_i \in \mathbb{R}^n, i = 1, \dots, n$. Then $\boldsymbol{x}^\top S\boldsymbol{x} = \sum_{i=1}^n \lambda_i\alpha_i^2$. Note that if all $\lambda_i > 0$, then $\boldsymbol{x}^\top S\boldsymbol{x} > 0, \forall\boldsymbol{x} \in \mathbb{R}^n - \{\boldsymbol{0}\}$. Conversely, if $\boldsymbol{x}^\top S\boldsymbol{x} > 0$, then it must be the case that $\lambda_i > 0$. Otherwise, if e.g., $\lambda_1 \leq 0$, then $\boldsymbol{v}_1^\top S\boldsymbol{v}_1 = \lambda_1 \leq 0$, which gives a contradiction. Now, similarly if all $\lambda_i \geq 0$, then $\boldsymbol{x}^\top S\boldsymbol{x} \geq 0, \forall\boldsymbol{x} \in \mathbb{R}^n$. Conversely, if $\boldsymbol{x}^\top S\boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$, then it must be the case that $\lambda_i \geq 0$, otherwise if $\lambda_1 < 0$, then $\boldsymbol{v}_1^\top S\boldsymbol{v}_1 = \lambda_1 < 0$, which would give us a contradiction.

   (e) If $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the sorted eigenvalues of S , then $\max\limits_{\|\boldsymbol{x}\|_2=1} \boldsymbol{x}^\top S\boldsymbol{x} = \lambda_1$ and $\min\limits_{\|\boldsymbol{x}\|_2=1} \boldsymbol{x}^\top S\boldsymbol{x} = \lambda_n$ [4pts].

   ANSWER: By part (c), we can choose n eigenvectors forming a basis for $\mathbb{R}^n$. Then, any vector $\boldsymbol{x} \in \mathbb{R}^n$ can be written as $\boldsymbol{x} = \sum_{i=1}^n \alpha_i\boldsymbol{v}_i$, where $\alpha_i \in \mathbb{R}, i = 1, \dots, n$. We thus have $\boldsymbol{x}^\top\boldsymbol{x} = \sum_{i=1}^n \alpha_i^2$ and $\boldsymbol{x}^\top S\boldsymbol{x} = \sum_{i=1}^n \lambda_i\alpha_i^2$. Since $\lambda_i \leq \lambda_1$ for $i = 1, \dots, n$, we have $\boldsymbol{x}^\top S\boldsymbol{x} \leq \lambda_1\sum_{i=1}^n \alpha_i^2 = \lambda_1\boldsymbol{x}^\top\boldsymbol{x}$, with equality occurring when $\alpha_1 \neq 0$ and $\alpha_2 = \dots = \alpha_n = 0$. Therefore, $\max\limits_{\|\boldsymbol{x}\|_2=1}(\boldsymbol{x}^\top S\boldsymbol{x}) = \lambda_1$. We can similarly prove that $\min\limits_{\|\boldsymbol{x}\|_2=1}(\boldsymbol{x}^\top S\boldsymbol{x}) = \lambda_n$, by considering that $\lambda_n \leq \lambda_i$ for $i = 1, \dots, n$.

   ANSWER: Using the method of Lagrange multipliers, we build the Lagrangian function $\mathcal{L}(\boldsymbol{x}, \lambda) = \boldsymbol{x}^T S\boldsymbol{x} + \lambda(1 - \boldsymbol{x}^T\boldsymbol{x})$. Setting the derivative of $\mathcal{L}$ to zero we obtain $2S\boldsymbol{x} - 2\lambda\boldsymbol{x} = 0$, which implies that $S\boldsymbol{x} = \lambda\boldsymbol{x}$. Therefore, the solution $(\lambda, \boldsymbol{x})$ should be an eigenvalue-eigenvector pair of S. Moreover, the optimal value is $\boldsymbol{x}^\top S\boldsymbol{x} = \lambda\boldsymbol{x}^\top\boldsymbol{x} = \lambda$. Therefore, if our goal is to maximize $\boldsymbol{x}^\top S\boldsymbol{x}$, then λ should be the largest eigenvalue. Conversely, if we wish to minimize $\boldsymbol{x}^\top S\boldsymbol{x}$, λ must be the smallest eigenvalue. Finally, to verify that these are indeed maximum and minimum, we notice that the Hessian of $\mathcal{L}$ is given by $H = 2(S - \lambda I)$, so that $\boldsymbol{x}^\top H\boldsymbol{x} = 2(\boldsymbol{x}^\top S\boldsymbol{x} - \lambda\boldsymbol{x}^\top\boldsymbol{x})$. When $\lambda = \lambda_1$, we have that $S - \lambda I \preceq 0$, because $\boldsymbol{x}^\top H\boldsymbol{x} = 2(\boldsymbol{x}^\top S\boldsymbol{x} - \lambda_1\boldsymbol{x}^\top\boldsymbol{x}) = 2(\sum_{i=1}^n \lambda_i\alpha_i^2 - \lambda_1\sum_{i=1}^n \alpha_i^2) = \sum_{i=2}^n(\lambda_i - \lambda_1)\alpha_i^2 \leq 0$. When $\lambda = \lambda_n$, we have that $S - \lambda I \succeq 0$, because $\boldsymbol{x}^\top H\boldsymbol{x} = \sum_{i=1}^{n-1}(\lambda_i - \lambda_n)\alpha_i^2 \geq 0$.

2. **Properties of the SVD [20pts].** Let $A = U\Sigma V^\top$ be the SVD of a matrix $A \in \mathbb{R}^{m\times n}$ of rank r. Show that:

   (a) $A\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j$ for $j = 1, \dots, r$ and $A^\top \boldsymbol{u}_j = \sigma \boldsymbol{v}_j$ for $j = 1, \dots, r$. [4pts].
   ANSWER: Multiplying both sides of $A = U\Sigma V^\top$ by V results in $AV = U\Sigma V^\top V$. Using the orthonormality property of V, i.e. $V^\top V = I$, this reduces to $AV = U\Sigma$. Now in terms of the $j - th$ column of V and U, we have $A\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j$. Likewise, $A^\top = \left(U\Sigma V^\top\right)^\top = V\Sigma^\top U^\top$. After multiplying both sides by U and using the orthonormality property of U, we obtain $A^\top U = V\Sigma^\top$, which results in $A^\top \boldsymbol{u}_j = \sigma_j \boldsymbol{v}_j$.

   (b) The range or image of A is spanned by the left singular vectors of A associated with its nonzero singular values, i.e., $\text{range}(A) = \text{span}\{\boldsymbol{u}_i\}_{i=1}^r$. [4pts]
   ANSWER: The range of A is the set of vectors of the form $\boldsymbol{y} = A\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^n$. We can express this set in terms of the SVD of A as $\text{range}(A) = \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{y} = U\Sigma V^\top \boldsymbol{x}, \boldsymbol{x} \in \mathbb{R}^n\}$. Let $\boldsymbol{z} = \Sigma V^\top \boldsymbol{x}$. We notice that all entries of $\boldsymbol{z}$ beyond the r-th are zero because $\sigma_i = 0$ for $i > r$. This means that

$$\boldsymbol{y} = \begin{bmatrix}\boldsymbol{u}_1 & \dots & \boldsymbol{u}_r\end{bmatrix} \begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_r\end{bmatrix} = \begin{bmatrix}\boldsymbol{u}_1 & \dots & \boldsymbol{u}_r\end{bmatrix} \boldsymbol{w}, \tag{1}$$

   and so the range of A is the span of the first r columns of U, i.e., $\text{range}(A) = \text{span}\{\boldsymbol{u}_i\}_{i=1}^r$.

   (c) The kernel or null space of A is spanned by the right singular vectors of A associated with its zero singular values, i.e., $\ker(A) = \text{span}\{\boldsymbol{v}_i\}_{i=r+1}^n$. [4pts]
   ANSWER: The kernel of A is defined as $\ker(A) = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = 0\}$. Now, since $A\boldsymbol{x} = U\Sigma V^\top \boldsymbol{x}$, we notice that $A\boldsymbol{x} = \mathbf{0}$ if and only if $\Sigma V^\top \boldsymbol{x} = 0$, and do $\ker(A) = \{\boldsymbol{x} : \sigma_i \boldsymbol{v}_i^\top \boldsymbol{x} = 0, i = 1, \dots, n\}$. Since $\sigma_i > 0$ for $i = 1, \dots, r$ and $\sigma_i = 0$ otherwise, we further have $\ker(A) = \{\boldsymbol{x} : \boldsymbol{v}_i^\top \boldsymbol{x} = 0, i = 1, \dots, r\}$. Therefore, $\ker(A)$ is the orthogonal complement to the span of $\{\boldsymbol{v}_i\}_{i=1}^r$. Since the vectors $\{\boldsymbol{v}_i\}_{i=1}^n$ form an orthonormal basis for $\mathbb{R}^n$, we have that $\ker(A)$ is the span of $\{\boldsymbol{v}_i\}_{i=r+1}^n$, as claimed.

   (d) The squared Frobenius norm of A is equal to the sum of the squared singular values of A, i.e., $\|A\|_F^2 = \sum_{ij} a_{ij}^2 = \sum_{k=1}^r \sigma_k^2$. [4pts]
   ANSWER: The sum of all the squared elements of a matrix is the trace of $AA^\top$. Moreover, $\text{trace}(AA^\top)$ is equivalent to the sum of all the eigenvalues of $AA^\top$, which are the squared singular values of A. Thus: $\sum_{ij} a_{ij}^2 = \text{trace}(AA^\top) = \sum_k \lambda_k\{AA^\top\} = \sum_{k=1}^r \sigma_k^2$.

   (e) The right singular vector of A associated to its smallest singular value, $\boldsymbol{v}_n$, is a solution to the optimization problem $\min_{\boldsymbol{x}} \|A\boldsymbol{x}\|_2^2$ such that $\|\boldsymbol{x}\|_2 = 1$. [4pts]
   ANSWER: Since $\|A\boldsymbol{x}\|_2^2 = \boldsymbol{x}^\top A^\top A\boldsymbol{x}$, we can rewrite the optimization problem as $\min_{\boldsymbol{x}} \boldsymbol{x}^\top M\boldsymbol{x}$ s.t. $\|\boldsymbol{x}\|_2 = 1$, where $M = A^\top A$. This is a constrained optimization problem whose Lagrangian is given by $\mathcal{L}(\boldsymbol{x}, \lambda) = \boldsymbol{x}^\top M\boldsymbol{x} + \lambda(1 - \boldsymbol{x}^\top \boldsymbol{x})$. The first order condition for optimality is $\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}} = 2M\boldsymbol{x} - 2\lambda\boldsymbol{x} = 0$. Thus, $M\boldsymbol{x} = \lambda\boldsymbol{x}$, and so λ is an eigenvalue of M. Multiplying both sides by $\boldsymbol{x}^\top$ gives $\boldsymbol{x}^\top M\boldsymbol{x} = \lambda\boldsymbol{x}^\top\boldsymbol{x} = \lambda$. Since we are trying to minimize $\boldsymbol{x}^\top M\boldsymbol{x}$, λ should be the smallest eigenvalue of $A^\top A$ and $\boldsymbol{x}$ should be the corresponding eigenvector. In other words, $\boldsymbol{x}$ should be the right singular vector of A associated to its smallest singular value. Finally, the second order condition for a minimum is $\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{x}^2} = 2(M - \lambda I) \succeq 0$. This is the case when $\lambda = \sigma_n^2$ because the eigenvalues of $M - \lambda I$ are $\sigma_i^2 - \sigma_n^2 \geq 0$ for $i = 1 \dots, n$.

3. **Pseudo-Inverse of a Matrix. [10pts]**

   (a) Let $A = U_r\Sigma_r V_r^\top$ be the compact SVD of a matrix A of rank r. Show that the pseudo-inverse of A is given by $A^\dagger = V_r\Sigma_r^{-1}U_r^\top$. [2pts]
   ANSWER: By the definition of the compact SVD, we have $V_r^\top V_r = U_r^\top U_r = I$, and Σ_r and Σ_r^{-1} are diagonal matrices. We then only need to verify all four criteria of the pseudo-inverse matrix as follows:

      i. $AA^\dagger A = U_r\Sigma_r V_r^\top V_r\Sigma_r^{-1}U_r^\top U_r\Sigma_r V_r^\top = U_r\Sigma_r V_r^\top = A$,
      ii. $A^\dagger AA^\dagger = V_r\Sigma_r^{-1}U_r^\top U_r\Sigma_r V_r^\top V_r\Sigma_r^{-1}U_r^\top = V_r\Sigma_r^{-1}U_r^\top = A^\dagger$,
      iii. $(AA^\dagger)^\top = (U_r\Sigma_r V_r^\top V_r\Sigma_r^{-1}U_r^\top)^\top = (U_rU_r^\top)^\top = U_rU_r^\top = AA^\dagger$,
      iv. $(A^\dagger A)^\top = (V_r\Sigma_r^{-1}U_r^\top U_r\Sigma_r V_r^\top)^\top = (V_rV_r^\top)^\top = V_rV_r^\top = A^\dagger A$.

(b) Consider the linear system of equations $A\boldsymbol{x} = \boldsymbol{b}$, where the matrix $A \in \mathbb{R}^{m \times n}$ is of rank $r = \text{rank}(A) = \min\{m, n\}$. Show that $\boldsymbol{x}^* = A^\dagger \boldsymbol{b}$ minimizes $\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2$, where $A^\dagger$ is the pseudo-inverse of A. When is $\boldsymbol{x}^*$ the unique solution? [4pts]

ANSWER: We want to find the $\boldsymbol{x}^*$ that minimizes $\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 = (A\boldsymbol{x} - \boldsymbol{b})^\top (A\boldsymbol{x} - \boldsymbol{b})$.The first order condition for optimality is given by

$$
\begin{aligned}
\frac{\partial[(A\boldsymbol{x} - \boldsymbol{b})^\top (A\boldsymbol{x} - \boldsymbol{b})]}{\partial \boldsymbol{x}} = 0 &\implies A^\top A\boldsymbol{x} = A^\top \boldsymbol{b} \implies (V_r \Sigma_r U_r^\top)(U_r \Sigma_r V_r^\top)\boldsymbol{x} = (V_r \Sigma_r U_r^\top)\boldsymbol{b} \\
&\implies (V_r \Sigma_r \Sigma_r V_r^\top)\boldsymbol{x} = (V_r \Sigma_r U_r^\top)\boldsymbol{b} \quad (\text{since } U_r^\top U_r = I_{r \times r}) \\
&\implies \Sigma_r V_r^\top \boldsymbol{x} = U_r^\top \boldsymbol{b} \quad (\text{pre-multiplying by } \Sigma_r^{-1} V_r^\top \text{ and using } V_r^\top V_r = I_{r \times r}) \\
&\implies V_r^\top \boldsymbol{x} = \Sigma_r^{-1} U_r^\top \boldsymbol{b} \quad (\text{pre-multiplying by } \Sigma_r^{-1})
\end{aligned}
$$

Note that $\boldsymbol{x}^* = (V_r \Sigma_r^{-1} U_r^\top)\boldsymbol{b} = A^\dagger \boldsymbol{b}$ is a solution of the above. When A is not full column rank, i.e., $m < n$, $\boldsymbol{x}^*$ is not the only solution of the minimization problem. In fact any vector $\boldsymbol{x}^* + \boldsymbol{y}$, where $\boldsymbol{y} \in \text{null}(A)$ would be a solution to the problem. When A is column full rank, i.e., $m \geq n$, $\boldsymbol{x}^*$ is unique.

(c) If $\boldsymbol{b} \in \text{range}(A)$, $\boldsymbol{x}^* = A^\dagger \boldsymbol{b}$ is the solution to the optimization problem $\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_2^2$ such that $A\boldsymbol{x} = \boldsymbol{b}$. [4pts]

ANSWER: The Lagrangian is given by $\mathcal{L}(\boldsymbol{x}, \lambda) = \boldsymbol{x}^\top \boldsymbol{x} + \lambda(\boldsymbol{b} - A\boldsymbol{x})$, where $\lambda \in \mathbb{R}^m$ is a vector of Lagrange multipliers. The first order condition is given by $2\boldsymbol{x} - A^\top \lambda = 0$, which yields $\boldsymbol{x} = \frac{1}{2} A^\top \lambda$. Substituting this back into $A\boldsymbol{x} = \boldsymbol{b}$ we obtain $\frac{1}{2} A A^\top \lambda = \boldsymbol{b} = \frac{1}{2} U_r \Sigma_r V_r^\top V_r \Sigma_r U_r^\top \lambda \implies \frac{1}{2} U_r \Sigma_r^2 U_r^\top \lambda = \boldsymbol{b} \implies U_r^\top \lambda = 2\Sigma_r^{-2} U_r^\top \boldsymbol{b}$. Therefore, $\boldsymbol{x} = \frac{1}{2} A^\top \lambda = \frac{1}{2} V_r \Sigma_r U_r^\top \lambda = V_r \Sigma_r \Sigma_r^{-2} U_r^\top \boldsymbol{b} = A^\dagger \boldsymbol{b}$. Finally, the Hessian of $\mathcal{L}$ is $H = 2I \succ 0$, hence $\boldsymbol{x} = A^\dagger \boldsymbol{b}$ is a minimum.