HW 7: Learning Theory II (580.692)

Instructor: René Vidal, Phone: 410-516-7306, E-mail: rvidal@cis.jhu.edu
Grader: Alvina Goh, email: alvina@cis.jhu.edu

Due 11/02/06 beginning of the class

1. Read Chapter 7 until end of 7.1, and Chapter 8 until end of 8.33 of GPCA book. Go to *http://www.vision.jhu.edu/gpcabook/* and submit all the typos you find as well as suggestions you may have to improve the quality and/or readability of the material. **You will receive credit for each interesting typo or suggestion you submit**.

2. **Properties of the Veronese map.**

   Consider the Veronese map $\nu_n : [x_1, \ldots, x_D]^T \mapsto [\ldots, \boldsymbol{x}^n, \ldots]^T$ where $\boldsymbol{x}^n = x_1^{n_1} x_2^{n_2} \ldots x_D^{n_D}$ ranges over all monomials of degree $n = \sum_{i=1}^{D} n_i$ in the variables $x_1, x_2, \ldots, x_D$, sorted in the degree-lexicographic order, and let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^D$.

   (a) **Inner product invariance:** Show that the polynomial kernel $k(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y}^\top \boldsymbol{x})^n$ can be written in terms of the Veronese map as $k(\boldsymbol{x}, \boldsymbol{y}) = \nu_n(\boldsymbol{y})^\top M \nu_n(\boldsymbol{x})$, where $M \in \mathbb{R}^{M_n(D) \times M_n(D)}$ is a diagonal matrix, and its $(n_1, n_2, \ldots, n_D)$th entry is $\frac{n!}{n_1! n_2! \ldots n_D!}$ with $\sum_{i=1}^{D} n_i = n$.
   Hint: Use the Multinomial Theorem.

   (b) **Linear invariance:**

      i. Show that $\nu_n(\alpha \boldsymbol{x} + \boldsymbol{y}) = \sum_{i=0}^{n} \alpha^i f_i(\boldsymbol{x}, \boldsymbol{y})$ where $f_i(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{M_n(D)}$ is a bi-homogenous polynomial of degree i in $\boldsymbol{x}$ and $(n - i)$ in $\boldsymbol{y}$ for $i = 0, \ldots, n$.

      ii. Let S_n be the space of homogeneous polynomials of degree n in D variables. Define the transformation $T : S_n \to S_n$, such that $T(p_n(\boldsymbol{x})) = p_n(A\boldsymbol{x})$, where $A \in \mathbb{R}^{D \times D}$. Show that the transformation T is linear.

      iii. Show that for all $A \in \mathbb{R}^{D \times D}$ there exists an $\tilde{A} \in \mathbb{R}^{M_n(D) \times M_n(D)}$ such that for all $\boldsymbol{x}, \nu_n(A\boldsymbol{x}) = \tilde{A} \nu_n(\boldsymbol{x})$.

   (c) **Rotation invariance:** Show that for $D = 3$ and all $R \in SO(3)$, there exists $\tilde{R} \in SO(M_n(D))$ such that for all $\boldsymbol{x}, \nu_n(R\boldsymbol{x}) = \tilde{R} M^{1/2} \nu_n(\boldsymbol{x})$, where M is the matrix defined in 2(a).
   Hint: Consider $(\boldsymbol{y}^\top R\boldsymbol{x})^n$ in 2(a), and also apply part (b)iii to $\nu_n(R^\top R\boldsymbol{x})$

3. **3-D Reconstruction from Multiple Calibrated Orthographic Views.**

   Let $\{\boldsymbol{X}_p \in \mathbb{R}^3\}_{p=1}^P$ be the coordinates of an *unknown* set of points lying on a rigidly moving object with respect to some fixed coordinate system. Let $(R_f, T_f) \in SE(3)$ be the *unknown* pose of the object at time $f = 1, 2, \ldots, F$ relative to a moving camera observing the object. Let $\boldsymbol{x}_{fp} \in \mathbb{R}^2$ be a *known* measurement of the *orthographic projection* of $\boldsymbol{X}_p \in \mathbb{R}^3$ in frame f. That is, $\boldsymbol{x}_{fp} = M_f \boldsymbol{X}_p + V_f$, where

   $$\begin{bmatrix} M_f & V_f \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_f & T_f \end{bmatrix} \tag{1}$$

   is the projection matrix associated with frame f.

   (a) Show that the optimal solution for the 2-D translation $V_f \in \mathbb{R}^2$ in the sense of minimizing the reprojection error is

   $$V_f = \overline{\boldsymbol{x}_f} = \frac{1}{P} \sum_{p=1}^{P} \boldsymbol{x}_{fp} \qquad f = 1, \ldots, F. \tag{2}$$

Hint: show that one can assume that $\overline{\boldsymbol{X}} = \frac{1}{P}\sum_{p=1}^{P} \boldsymbol{X}_p = 0$ without loss of generality.

(b) Let $\boldsymbol{w}_{fp} = \boldsymbol{x}_{fp} - \overline{\boldsymbol{x}_f}$ be the mean subtracted point correspondences and define a data matrix

$$W = \begin{bmatrix} \boldsymbol{w}_{11} & \cdots & \boldsymbol{w}_{1P} \\ \vdots & & \vdots \\ \boldsymbol{w}_{F1} & \cdots & \boldsymbol{w}_{FP} \end{bmatrix} \in \mathbb{R}^{2F \times P}. \quad (3)$$

Show that the measurement matrix W factors as $W = MS$, where $M = \begin{bmatrix} M_1 \\ \vdots \\ M_F \end{bmatrix} \in \mathbb{R}^{2F\times 3}$ and $S = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_P \end{bmatrix} \in \mathbb{R}^{3\times P}$ are the so-called *motion* and *structure* matrices, respectively. Show that $\text{rank}(W) \leq 3$ and $\text{rank}(M) \geq 2$ and derive conditions on the camera motion and the 3D structure such that $\text{rank}(W) = 3$. Under such conditions, propose an algorithm for computing the motion and structure matrices $M = \tilde{M}K$ and $S = K^{-1}\tilde{S}$ up to an unknown invertible matrix $K \in \mathbb{R}^{3\times 3}$.

(c) Let $Q = KK^T \in \mathbb{R}^{3\times 3}$. Show that the sub-matrix of $\tilde{M}$ consisting of rows $2f-1$ and $2f$, $\tilde{M}_{2f-1:2f} \in \mathbb{R}^{2\times 3}$, is such that

$$\tilde{M}_{2f-1:2f} Q \tilde{M}_{2f-1:2f}^T = I \qquad f = 1, \ldots, F. \quad (4)$$

Propose a linear algorithm to compute Q. What is the minimum number of frames needed? Given Q, show how to compute K up to a rotation. Given such a K show how to compute M, S, R_f and T_f. Is there any ambiguity in the reconstruction?

4. **3-D Motion Segmentation from Multiple Affine Views.**

Let $\{\boldsymbol{X}_p \in \mathbb{R}^3\}_{p=1}^P$ be the coordinates of an *unknown* set of points lying on a collection of n rigidly moving object with respect to some fixed coordinate system. Let $(R_f^i, T_f^i) \in SE(3)$, for $i = 1, \ldots, n$, be the *unknown* poses of the objects at time $f = 1, 2, \ldots, F$ relative to a moving camera observing the objects. Let $\boldsymbol{x}_{fp} \in \mathbb{R}^2$ be a *known* measurement of the *orthographic projection* of $\boldsymbol{X}_p \in \mathbb{R}^3$ in frame f.

(a) Show that the P vectors $(\boldsymbol{x}_{1p}^\top, \boldsymbol{x}_{2p}^\top, \cdots, \boldsymbol{x}_{Fp})^\top \in \mathbb{R}^{2F}$ live in n subspaces of dimension 2, 3, or 4. Explain why when $n > 1$ it is not possible to reduce the dimension of each subspace to 3, as you did in Problem 3.

(b) Use the function `gpca` from Homework 6 to segment the point correspondences of the following five video sequences in the course webpage: i) Kanatani1, ii) Kanatani2, iii) Kanatani3, iv) three-cars, v) can-book. In each case, assume the number of groups is known, plot the grouping of the ordered data given by GPCA, and report the percentage of misclassified points. Use `subplot(5,1,i)` to plot all five graphs on a single figure. Recall that you will need to project the data in $\mathbb{R}^{2F}$ onto a subspace of dimension d. What is the value for d?

(c) Repeat part (b) using the function `ksubspaces` that you implemented in Homework 4. Use the result of GPCA from part (b) to initialize K-subspaces. Use both the data without projection, i.e. the data in $\mathbb{R}^{2F}$, and the projected data in $\mathbb{R}^d$ as the input to K-subspaces. Which one is better, projecting or not, and why?

5. **Face Clustering with Varying Illumination II.**

In HW 4, you assumed that faces are Lambertian for the sake of simplicity and as a consequence, the images of n individuals taken under several illumination conditions live in n 3-dimensional subspaces of $\mathbb{R}^P$, where P is the number of pixels. It follows that clustering a set of images of multiple faces according to which individuals the image belongs to is a subspace clustering problem. On the set of images given in the course web-page, reduce the dimension using PCA to the first 3 principal components. Now, assume the number of groups is known, i.e. $n = 3$, and segment the faces using

(a) GPCA

(b) K-subspaces

(c) K-subspaces initialized by GPCA

Plot the grouping of the ordered data and report the percentage of incorrectly classified images in each case.