Appendix A

Matrix Formulas

A.1 Basics

Let A, B, and C be arbitrary matrices, and suppose that the number of rows and columns are chosen appropriately. Then we can use the following relations:

$$\begin{aligned}
A(B+C) &= AB+AC, && \text{(A.1)}\\
(A+B)^\top &= A^\top + B^\top, && \text{(A.2)}\\
(AB)^\top &= B^\top A^\top, && \text{(A.3)}\\
(A^{-1})^\top &= (A^\top)^{-1}. && \text{(A.4)}
\end{aligned}$$

Let $\lambda_1, ..\lambda_n$ be the eigenvalues of matrix A that are sorted by descending order. We can use the following relations as for the determinant $|A|$:

$$\begin{aligned}
|AB| &= |A||B|, && \text{(A.5)}\\
|A^{-1}| &= \frac{1}{|A|}, && \text{(A.6)}\\
|BAB^{-1}| &= |B||A|\frac{1}{|B|} = |A|, && \text{(A.7)}\\
|A| &= \prod_i \lambda_i, && \text{(A.8)}\\
|A^\top| &= |A|, && \text{(A.9)}\\
|aA| &= a^n|A|, && \text{(A.10)}\\
|-A| &= (-1)^n|A|. && \text{(A.11)}
\end{aligned}$$

As for the trace $\mathrm{tr}(A)$, we can use

$$\begin{aligned}
\mathrm{tr}(A+B) &= \mathrm{tr}(A)+\mathrm{tr}(B), && \text{(A.12)}\\
\mathrm{tr}(A) &= \sum_i \lambda_i, && \text{(A.13)}\\
\mathrm{tr}(ABC) &= \mathrm{tr}(BCA)=\mathrm{tr}(CAB). && \text{(A.14)}
\end{aligned}$$

As for the rank and condition number of matrix A, there exist the following relations:

$$\begin{aligned}
\mathrm{rank}(A) &= \mathrm{rank}(A^\top A)=\mathrm{rank}(AA^\top), && \text{(A.15)}\\
\mathrm{condition\,number}(A) &= \sqrt{\frac{\lambda_1}{\lambda_n}}. && \text{(A.16)}
\end{aligned}$$

A.2 Differential Formulas of Matrices

Suppose $x = [x_1, ..., x_n]^\top \in \mathbf{R}^n$ and $X = [x_{ij}]$ is $n \times m$ matrix. We define the following notations to denote partial differential of each vector (matrix) element.

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

$$\frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1m}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial x_{n1}} & \cdots & \frac{\partial f}{\partial x_{nm}} \end{bmatrix}$$

A.2.1 Formulas

$$\begin{aligned}
\frac{\partial}{\partial X}\mathrm{tr}(X^\top A) &= A && \text{(A.17)}\\
\frac{\partial}{\partial X}\mathrm{tr}(X^\top AX) &= (A+A^\top)X && \text{(A.18)}\\
\frac{\partial}{\partial X}\mathrm{tr}(X^\top AXB) &= AXB + A^\top XB^\top && \text{(A.19)}\\
\frac{\partial}{\partial X}\log|X| &= (X^\top)^{-1} && \text{(A.20)}\\
\frac{\partial}{\partial X}|X| &= (X^\top)^{-1}|X| && \text{(A.21)}
\end{aligned}$$

A.2.2 Examples

Differential Formulas of Vectors

From Equation (A.17) and Equation (A.18), we obtain the following vector formulas:

$$\begin{aligned}\frac{\partial}{\partial x}x^\top a &= a\\ \frac{\partial}{\partial x}x^\top Ax &= (A+A^\top)x = 2Ax\end{aligned}$$

Useful Formulas

$$\begin{aligned}\frac{\partial}{\partial X}||AX+B||^2 &= \frac{\partial}{\partial X}\mathrm{tr}((AX+B)^\top(AX+B)) \quad &(A.22)\\ &= \frac{\partial}{\partial X}\mathrm{tr}(X^\top A^\top AX + X^\top A^\top B + B^\top AX + B^\top B) \quad &(A.23)\\ &= 2A^\top AX + 2A^\top B \quad &(A.24)\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial X}\mathrm{tr}(A^\top XX^\top A) &= \frac{\partial}{\partial X}\mathrm{tr}(X^\top AA^\top X) \quad &(A.25)\\ &= 2AA^\top X \quad &(A.26)\end{aligned}$$

Appendix B

Estimation of the Transition Matrices and Bias Vectors of Linear Dynamical Systems

In this appendix, we show how to estimate the transition matrix $F^{*(i)}$ and bias vector $g^{*(i)}$ from the internal state sequence $x_b^{(i)}, .., x_e^{(i)}$ in temporal range $[b, e]$, which is represented by linear dynamical system D_i. The results described here corresponds to Equation (3.5) and Equation (3.6).

As we introduced in Subsection 3.3.3, we use the following notations:

$$X_0^{(i)} \triangleq [x_b^{(i)}, ..., x_{e-1}^{(i)}], \quad X_1^{(i)} \triangleq [x_{b+1}^{(i)}, ..., x_e^{(i)}] \tag{B.1}$$

$$m_0^{(i)} \triangleq \frac{1}{l-1} \sum_{t=b}^{e-1} x_t^{(i)} = \frac{1}{l-1} X_0^{(i)} [\underbrace{1, ..., 1}_{l-1}]^\top \tag{B.2}$$

$$m_1^{(i)} \triangleq \frac{1}{l-1} \sum_{t=b+1}^{e} x_t^{(i)} = \frac{1}{l-1} X_1^{(i)} [\underbrace{1, ..., 1}_{l-1}]^\top, \tag{B.3}$$

where $l = e - b + 1$. Using the notations above, we can rewrite Equation (3.3) as the following:

$$\sum_{t=b+1}^{e} ||\epsilon_t||^2 = ||X_1^{(i)} - (F^{(i)}X_0^{(i)} + g^{(i)}\underbrace{[1,...,1]}_{l-1})||^2 \tag{B.4}$$

$$= \text{tr}\left(X_1^{(i)} - (F^{(i)}X_0^{(i)} + g^{(i)}[1,...,1])\right)^\top \left(X_1^{(i)} - (F^{(i)}X_0^{(i)} + g^{(i)}[1,...,1])\right)$$

$$= \text{tr}(X_1^{(i)\top}X_1^{(i)}) + \text{tr}\left((F^{(i)}X_0^{(i)} + g^{(i)}[1,...,1])^\top(F^{(i)}X_0^{(i)} + g^{(i)}[1,...,1])\right)$$

$$-\text{tr}\left(X_1^{(i)\top}(F^{(i)}X_0^{(i)} + g^{(i)}[1,...,1])\right) - \text{tr}\left((F^{(i)}X_0^{(i)} + g^{(i)}[1,...,1])^\top X_1^{(i)}\right)$$

Using Equation (B.2) and Equation (B.3) with $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$ and $(AB)^\top = B^\top A^\top$ (see Appendix A), we obtain

$$\begin{aligned}\sum_{t=b+1}^{e} ||\epsilon_t||^2 = \; & \text{tr}(X_1^{(i)\top}X_1^{(i)}) + \text{tr}(F^{(i)}X_0^{(i)}X_0^{(i)\top}F^{(i)\top}) + (l-1)\text{tr}(g^{(i)\top}g^{(i)}) \\ & +(l-1)\text{tr}(F^{(i)\top}g^{(i)}m_0^{(i)\top}) + (l-1)\text{tr}(m_0^{(i)}g^{(i)\top}F^{(i)}) \\ & -\text{tr}(F^{(i)}X_0^{(i)}X_1^{(i)\top}) - (l-1)\text{tr}(g^{(i)}m_1^{(i)\top}) \\ & -\text{tr}(X_1^{(i)}X_0^{(i)\top}F^{(i)\top}) - (l-1)\text{tr}(g^{(i)\top}m_1^{(i)}).\end{aligned}$$

If $l-1 >= n$ (i.e., the number of samples is equal to or greater than the dimensionality of state vectors), we can estimate transition matrix $F^{*(i)}$ and bias vector $g^{*(i)}$ by solving the least squares problem of Equation (3.4). To solve this minimization problem, we first differentiate $\sum_{t=b+1}^{e} ||\epsilon_t||^2$ with respect to each of $F^{(i)}$ and $g^{(i)}$:

$$\begin{aligned}\frac{\partial}{\partial F^{(i)}}\sum_{t=b+1}^{e} ||\epsilon_t||^2 &= 2\left\{F^{(i)}X_0^{(i)}X_0^{(i)\top} + (l-1)g^{(i)}m_0^{(i)\top} - X_1^{(i)}X_0^{(i)\top}\right\}, \\ \frac{\partial}{\partial g^{(i)}}\sum_{t=b+1}^{e} ||\epsilon_t||^2 &= 2(l-1)\left\{g^{(i)} + F^{(i)}m_0^{(i)} - m_1^{(i)}\right\}.\end{aligned}$$

Then, we set zero for all the differentiated elements and obtain the following equations:

$$F^{*(i)}X_0^{(i)}X_0^{(i)\top} + (l-1)g^{*(i)}m_0^{(i)\top} - X_1^{(i)}X_0^{(i)\top} = O, \tag{B.5}$$

$$g^{*(i)} + F^{*(i)}m_0^{(i)} - m_1^{(i)} = 0. \tag{B.6}$$

From Equation (B.6), we obtain

$$g^{*(i)} = m_1^{(i)} - F^{*(i)} m_0^{(i)}. \tag{B.7}$$

Substituting this equation into Equation (B.5), we obtain

$$F^{*(i)} X_0^{(i)} X_0^{(i)\top} + (l-1)(m_1^{(i)} - F^{*(i)} m_0^{(i)}) m_0^{(i)\top} - X_1^{(i)} X_0^{(i)\top} = O. \tag{B.8}$$

Using Equation (B.2) and Equation (B.3) again, Equation (B.8) can be replaced by

$$F^{*(i)} X_0^{(i)} X_0^{(i)\top} + (X_1^{(i)}[1,...,1]^\top - F^{*(i)} X_0^{(i)}[1,...,1]^\top) m_0^{(i)\top} - X_1^{(i)} X_0^{(i)\top} = O.$$

Let $\hat{X}_0^{(i)}$ be centerized $X_0^{(i)}$:

$$\hat{X}_0^{(i)} \triangleq X_0^{(i)} - m_0^{(i)}[1,...,1] = [x_b^{(i)} - m_0^{(i)}, ..., x_{e-1}^{(i)} - m_0^{(i)}], \tag{B.9}$$

then we obtain

$$F^{*(i)} X_0^{(i)} \hat{X}_0^{(i)\top} = X_1^{(i)} \hat{X}_0^{(i)\top}. \tag{B.10}$$

Here, we can replace $X_0^{(i)} \hat{X}_0^{(i)\top}$ as $\hat{X}_0^{(i)} \hat{X}_0^{(i)\top}$ using Equation (B.2):

$$\begin{aligned}
\hat{X}_0^{(i)} \hat{X}_0^{(i)\top} &= (X_0^{(i)} - m_0^{(i)}[1,...,1]) \hat{X}_0^{(i)\top} \\
&= X_0^{(i)} \hat{X}_0^{(i)\top} - m_0^{(i)}[1,...,1](X_0^{(i)} - m_0^{(i)}[1,...,1])^\top \\
&= X_0^{(i)} \hat{X}_0^{(i)\top} - (l-1) m_0^{(i)} m_0^{(i)\top} + (l-1) m_0^{(i)} m_0^{(i)\top} \\
&= X_0^{(i)} \hat{X}_0^{(i)\top}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\hat{X}_1^{(i)} \hat{X}_0^{(i)\top} &= (X_1^{(i)} - m_1^{(i)}[1,...,1]) \hat{X}_0^{(i)} \\
&= X_1^{(i)} \hat{X}_0^{(i)} - m_1^{(i)}[1,...,1](X_0^{(i)} - m_0^{(i)}[1,...,1])^\top \\
&= X_1^{(i)} \hat{X}_0^{(i)\top} - (l-1) m_1^{(i)} m_0^{(i)\top} + (l-1) m_1^{(i)} m_0^{(i)\top} \\
&= X_1^{(i)} \hat{X}_0^{(i)\top}
\end{aligned}$$

Thus, we finally get

$$\begin{aligned}
F^{*(i)} \hat{X}_0^{(i)} \hat{X}_0^{(i)\top} &= \hat{X}_1^{(i)} \hat{X}_0^{(i)\top}, \\
\therefore \quad F^{*(i)} &= \hat{X}_1^{(i)} \hat{X}_0^{(i)\top} (\hat{X}_0^{(i)} \hat{X}_0^{(i)\top})^{-1}
\end{aligned} \tag{B.11}$$

On the other hand, if $l-1 < n$ (i.e., the number of samples is smaller than the dimensionality of state vectors), the solution of the minimization problem does not fix. From Equation (B.4), the special solution of the minimization problem becomes

$$\begin{aligned} F^{*(i)} &= X_1^{(i)}(X_0^{(i)\top}X_0^{(i)})^{-1}X_0^{(i)\top} \\ &= X_1^{(i)}(I-A)((I-A)X_0^{(i)\top}X_0^{(i)}(I-A))^{-1}(I-A)X_0^{(i)\top} \\ &= \hat{X}_1^{(i)}(\hat{X}_0^{(i)\top}\hat{X}_0^{(i)})^{-1}\hat{X}_0^{(i)\top}, \quad \text{(B.12)} \\ g^{*(i)} &= m_1^{(i)} - F^{*(i)}m_0^{(i)}, \quad \text{(B.13)} \end{aligned}$$

where $A = [1, ..., 1]^\top[1, ..., 1]/(l-1)$. The total prediction error becomes zero, if we use the above solution.

Note that both Equation (B.11) and (B.12) satisfy Equation (3.5), which is described by Moore-Penrose generalized inverse.

Appendix C

Gershgorin's Theorem

The Gershgorin's theorem is a well known method to describe a region in the complex plane $\{z \,|\, z \in \mathbf{C}\}$ that contains all the eigenvalues of a complex square matrix [Iri03].

Theorem: Let $A = [a_{ij}]$ be an arbitrary $n \times n$ complex square matrix, and define r_i as :

$$r_i \triangleq \sum_{j=1, j\neq i}^{n} |a_{ij}| \quad (i = 1, 2, ..., n).$$

Then, all the eigenvalues of matrix A exist in the union of circles $\cup_{i=1}^{n} C_i$, where

$$C_i = \{z \,|\, z \in \mathbf{C}, |z - a_{ii}| \leq r_i\}.$$

Corollary: Let $A = [a_{ij}]$ be an arbitrary $n \times n$ complex square matrix, and let $\lambda_i \, (i = 1, ..., n)$ be the eigenvalues of matrix A. The maximum absolute value (spectral radius) of these eigenvalues satisfies the following relation:

$$\max_i |\lambda_i| \leq \max_i \sum_{j=1}^{n} |a_{ij}|. \tag{C.1}$$

Appendix D

Active Appearance Model

The AAM-based feature point tracking consists of two stages. We first build an AAM model using a training set of face images and its feature points given manually. Then, we use the model to extract facial feature points in novel images.

To AAM build the model, we require a training set of images marked with feature points. Figure 4.4 (a) shows an example of a face image labeled with 58 feature points. Let s be a shape vector that represents the coordinate value of feature points. Let g be a grey-level vector that represents the intensity information from the shape-normalized image over the region covered with the mean shape. In the first step, the method applies principal component analysis (PCA) to the data. Any example image can then be approximated using:

$$s = \bar{s} + U_s c_s \ , \ g = \bar{g} + U_g c_g, \tag{D.1}$$

where $\bar{s}$ and $\bar{g}$ are the corresponding sample mean vectors, U_s and U_g are matrices of column eigenvectors of the shape and grey-level, and c_s and c_g are vectors of shape and grey-level parameters, respectively. In the second step, because there may be correlations between the shape and grey-level variation, the method concatenates the vectors c_s and c_g, applies PCA, and obtains a model of the form

$$\begin{bmatrix} W_s c_s \\ c_g \end{bmatrix} = c = \begin{bmatrix} V_s \\ V_g \end{bmatrix} d = Vd, \tag{D.2}$$

where W_s is a diagonal matrix of weights for each shape parameter, allowing for the difference in units between the shape and grey-level models, V is a matrix of column eigenvectors, and d is a vector of appearance parameters controlling both the shape and grey-levels of the model.

Note that the linear nature of the model allows us to express the shape vector s and grey-level vector g directly as functions of d:

$$s = \bar{s} + U_s W_s^{-1} V_s d \ , \ g = \bar{g} + U_g V_g d. \tag{D.3}$$

An example image can be synthesized for a given d by generating the shape-free grey-level image from the vector g and warping it using the feature points described by s. During a training phase we learn the relationship between model parameter displacements and the residual errors induced between a training image and a synthesized image.

The matching process for tracking the feature points is provided as an optimization problem in which we minimize the difference between a target image and an image synthesized by the model.