IDIAP COMMUNICATION REPORT

NOTES ON PROBABILISTIC LINEAR DISCRIMINANT ANALYSIS

Chris McCool Laurent El Shafey

Idiap-Com-03-2013

JUNE 2013

Centre du Parc, Rue Marconi 19, P.O. Box 592, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

Notes on Probabilistic Linear Discriminant Analysis

Christopher Steven McCool and Laurent El Shafey

(csmccool79@gmail.com, laurent.el-shafey@idiap.ch)

May 15, 2012

Idiap Research Institute,
Centre du Parc,
PO Box 592,
1920 Martigny,
Switzerland

Contents

1 Introduction

The following are a set of working notes on how to derive and implement probabilistic linear discriminant analysis (PLDA) using Gaussian priors as proposed in [4]. This includes:

- how to derive the update formulae using maximum likelihood (ML),
- how to derive the formulae for estimating the latent variables,
- how to derive the formulae to perform scoring, and
- finally simplifications for memory and computation time.

We start by introducing the PLDA model. Then we go on to describe how it can be derived so that we can train the parameters. After this we deal with how to derive an efficient way to implement the model to perform training and to compute the likelihood of some input samples given this model. For clarity we have provided a set of Appendices which contains several useful definitions and derivations used in the text.

1.1 Acknowledgments

We want to thank many people who helped us in writing down these derivations. First and foremost we would like to thank Sébastien Marcel who gave us the freedom to examine this model in more detail. Other people that we would like to thank are Carl Scheffler and the rest of the people at the Biometrics Group of the Idiap Research Institute.

2 The PLDA Model

The PLDA model of Prince and Elder [4] consists of a mean offset $\boldsymbol{\mu}$, a subspace describing the main directions of identity variation $\mathbf{F}$, a subspace describing the main directions of condition, or session variation, $\mathbf{G}$ and a noise term $\boldsymbol{\epsilon}_{ij}$, where i is the i^{th} identity and j is their j^{th} observation. We can then describe PLDA as the process to represent the D_x dimensional feature vector $\mathbf{x}_{ij}$ such that

$$\mathbf{x}_{ij} = \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \boldsymbol{\epsilon}_{ij}, \tag{1}$$

where $\mathbf{h}_i$ is the latent variable representing identity variation (the weight of the directions for identity variation) and $\mathbf{w}_{ij}$ is the latent variable representing condition or session variation (the weight of the directions for identity variation).

2.1 Setting up the Basics

To solve this set of equations we place certain restrictions upon it. First, we assume that $p(\mathbf{h}_i)$ is a zero mean unit standard deviation multivariate Gaussian ($\mathcal{N}_{\mathbf{h}}[\mathbf{0}, \mathbf{I}]$), that $p(\mathbf{w}_{ij})$ is a zero mean unit standard deviation multivariate Gaussian ($\mathcal{N}_{\mathbf{w}}[\mathbf{0}, \mathbf{I}]$) and that $\boldsymbol{\epsilon}_{ij}$ is assumed to be Gaussian with a diagonal covariance $\boldsymbol{\Sigma}$. We can then write Equation 1 as a conditional probability $Pr(\mathbf{x}_{ij}|\mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\Theta})$ where the parameters for our PLDA model are defined as $\boldsymbol{\Theta} = [\boldsymbol{\mu}, \mathbf{F}, \mathbf{G}, \boldsymbol{\Sigma}]$. We can now write the following,

$$p(\mathbf{h}_i) = \mathcal{N}_{\mathbf{h}}[\mathbf{0}, \mathbf{I}], \tag{2}$$

$$p(\mathbf{w}_{ij}) = \mathcal{N}_{\mathbf{w}}[\mathbf{0}, \mathbf{I}], \tag{3}$$

$$p(\mathbf{x}_{ij}|\mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\Theta}) = \mathcal{N}_{\mathbf{x}}[\boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij}, \boldsymbol{\Sigma}]. \tag{4}$$

2.2 Extending the notation for Multiple Samples of the Same Identity

The notation from the previous section can be extended to include the case of multiple observations for a single identity i. In this case $\tilde{\mathbf{x}}_i = [\mathbf{x}_{i1}^T, \mathbf{x}_{i2}^T, ..., \mathbf{x}_{iJ_i}^T]^T$, $\tilde{\boldsymbol{\mu}} = [\boldsymbol{\mu}^T, \boldsymbol{\mu}^T, ..., \boldsymbol{\mu}^T]^T$, $\tilde{\mathbf{y}}_i = [\mathbf{h}_i^T, \mathbf{w}_{i1}^T, \mathbf{w}_{i2}^T, ..., \mathbf{w}_{iJ_i}^T]^T$ and $\tilde{\boldsymbol{\epsilon}}_i = [\boldsymbol{\epsilon}_{i1}^T, \boldsymbol{\epsilon}_{i2}^T, ..., \boldsymbol{\epsilon}_{iJ_i}^T]$ where J_i is the number of observations for identity i; these are vectors of size $(J_i D_x, 1)$, $(J_i D_x, 1)$, $(D_F + J_i D_G, 1)$ and $(J_i D_x, 1)$ respectively. Using this we can write Equation 1 as,

$$\tilde{\mathbf{x}}_i = \tilde{\boldsymbol{\mu}} + \tilde{\mathbf{A}}\tilde{\mathbf{y}}_i + \tilde{\boldsymbol{\epsilon}}_i. \tag{5}$$

This is the same as Equation 11 in [4]. We take the example of $J_i = 3$ (we will keep the same value of J_i for the rest of the examples in these notes) and demonstrate the structure of the variables we have introduced.

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{F} & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \mathbf{G} \end{bmatrix}, \tag{6}$$

$$\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix}. \tag{7}$$

We can then write the probabilities as,

$$p\left(\tilde{\mathbf{x}}_i | \tilde{\mathbf{y}}_i, \boldsymbol{\Theta}\right) = \mathcal{N}_{\tilde{\mathbf{x}}_i}\left[\tilde{\boldsymbol{\mu}} + \tilde{\mathbf{A}}\tilde{\mathbf{y}}_i, \tilde{\boldsymbol{\Sigma}}\right], \tag{8}$$

$$p\left(\tilde{\mathbf{y}}_i\right) = \mathcal{N}_{\tilde{\mathbf{y}}_i}\left[\mathbf{0}, \mathbf{I}\right]. \tag{9}$$

It can be seen that in this more compact formulation the matrix $\tilde{\mathbf{A}}$ consists of the matrices $\mathbf{F}$ and $\mathbf{G}$ which are repeated J_i times and the matrix $\tilde{\boldsymbol{\Sigma}}$ consists of the matrix $\boldsymbol{\Sigma}$ also repeated J_i times. For later use, we now define another latent variable $\mathbf{y}_{ij} = [\mathbf{h}_i^T, \mathbf{w}_{ij}^T]^T$ which is the latent variable for the j^{th} observation of identity i.

2.3 Completing the Formulation

With the above equations we can now write down $p\left(\tilde{\mathbf{x}}_i\right)$ and $p\left(\tilde{\mathbf{y}}_i | \tilde{\mathbf{x}}_i, \boldsymbol{\Theta}\right)$ by using the identities B42-B45 (from page 689) of [1]. We have rewritten these identities below.

$$p\left(\boldsymbol{y}_B\right) = \mathcal{N}_{\boldsymbol{y}_B}\left[\boldsymbol{\mu}_{\boldsymbol{y}_B}, \boldsymbol{\Sigma}_B\right], \tag{10}$$

$$p\left(\boldsymbol{x}_B | \boldsymbol{y}_B\right) = \mathcal{N}_{\boldsymbol{x}_B}\left[\boldsymbol{A}_B\boldsymbol{y}_B + \boldsymbol{b}_B, \boldsymbol{L}_B^{-1}\right], \tag{11}$$

$$\boldsymbol{\Lambda}_B = \boldsymbol{\Sigma}_B^{-1}. \tag{12}$$

With these we can define the following:

$$p\left(\boldsymbol{x}_B\right) = \mathcal{N}_{\boldsymbol{x}_B}\left[\boldsymbol{A}_B\boldsymbol{\mu}_{\boldsymbol{y}_B} + \boldsymbol{b}_B, \boldsymbol{L}_B^{-1} + \boldsymbol{A}_B\boldsymbol{\Sigma}_B\boldsymbol{A}_B^T\right], \tag{13}$$

$$p\left(\boldsymbol{y}_B | \boldsymbol{x}_B\right) = \mathcal{N}_{\boldsymbol{y}_B}\left[\boldsymbol{\Sigma}^*\left(\boldsymbol{A}_B^T\boldsymbol{L}_B(\boldsymbol{x}_B - \boldsymbol{b}_B) + \boldsymbol{\Sigma}_B^{-1}\boldsymbol{\mu}_{\boldsymbol{y}_B}\right), \boldsymbol{\Sigma}^*\right], \tag{14}$$

$$\boldsymbol{\Sigma}^* = \left(\boldsymbol{\Sigma}_B^{-1} + \boldsymbol{A}_B^T\boldsymbol{L}_B\boldsymbol{A}_B\right)^{-1}. \tag{15}$$

Where we have used the subscript B to clearly indicate that these are the formulae from B42-B45 but with $\boldsymbol{x}$ and $\boldsymbol{y}$ reversed (which has no effect); we do this because it makes it easier to see the relationship for our problem.

2.3.1 Deriving $p(\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i, \mathbf{\Theta})$ and $p(\tilde{\mathbf{x}}_i)$

To derive the equations for $p(\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i, \mathbf{\Theta})$ and $p(\tilde{\mathbf{x}}_i)$ we use the above identities and make the following substitutions.

$$\boldsymbol{\mu}_{\boldsymbol{y}_B} = 0, \tag{16}$$

$$\mathbf{\Sigma}_B = \mathbf{I}, \tag{17}$$

$$\boldsymbol{y}_B = \tilde{\mathbf{y}}_i, \tag{18}$$

$$\boldsymbol{x}_B = \tilde{\mathbf{x}}_i, \tag{19}$$

$$\boldsymbol{A}_B = \tilde{\mathbf{A}}, \tag{20}$$

$$\boldsymbol{b}_B = \tilde{\boldsymbol{\mu}}, \tag{21}$$

$$\mathbf{\Sigma}^* = \left(\mathbf{I}^{-1} + \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{A}}\right)^{-1} = \left(\mathbf{I} + \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{A}}\right)^{-1}, \tag{22}$$

$$\boldsymbol{L}_B^{-1} = \tilde{\mathbf{\Sigma}}. \tag{23}$$

Therefore, we get the following for $p(\tilde{\mathbf{x}}_i)$:

$$p(\tilde{\mathbf{x}}_i) = \mathcal{N}_{\tilde{\mathbf{x}}_i}\left[\tilde{\mathbf{A}}\mathbf{0} + \tilde{\boldsymbol{\mu}}, \tilde{\mathbf{\Sigma}} + \tilde{\mathbf{A}}\mathbf{I}\tilde{\mathbf{A}}^T\right] = \mathcal{N}_{\tilde{\mathbf{x}}_i}\left[\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right]. \tag{24}$$

And we get the following for $p(\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i, \mathbf{\Theta})$:

$$p(\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i, \mathbf{\Theta}) = \mathcal{N}_{\tilde{\mathbf{y}}_i}\left[\mathbf{\Sigma}^*\left(\tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1}\left(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}\right) + \mathbf{I}^{-1}\mathbf{0}\right), \mathbf{\Sigma}^*\right] = \mathcal{N}_{\tilde{\mathbf{y}}_i}\left[\mathbf{\Sigma}^* \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1}\left(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}\right), \mathbf{\Sigma}^*\right], \tag{25}$$

$$p(\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i, \mathbf{\Theta}) = \mathcal{N}_{\tilde{\mathbf{y}}_i}\left[\left(\mathbf{I} + \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{A}}\right)^{-1} \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1}\left(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}\right), \left(\mathbf{I} + \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{A}}\right)^{-1}\right]. \tag{26}$$

2.3.2 Some Sanity Checks on Array and Matrix Sizes

We will make some statements so that if you go through these formulae things will hopefully make a bit more sense to you.

- D_G is associated with the subspace $\mathbf{G}$ (condition or session subspace) which is of size (D_x, D_G)
- D_F is associated with the subspace $\mathbf{F}$ (identity subspace) which is of size (D_x, D_F)
- $\tilde{\mathbf{y}}_i$ is the latent variables for the J_i observations of identity i and is a vector of size $(D_F + J_i D_G, 1)$
- $\tilde{\mathbf{A}}$ is a matrix of size $(J_i D_x, D_F + J_i D_G)$ and its transpose $\tilde{\mathbf{A}}^T$ is a matrix of size $(D_F + J_i D_G, J_i D_x)$
- $\tilde{\mathbf{\Sigma}}$ is a matrix of size $(J_i D_x, J_i D_x)$
- $\tilde{\boldsymbol{\mu}}$ is a vector of size $(J_i D_x, 1)$
- $\tilde{\mathbf{x}}_i$ is a vector of size $(J_i D_x, 1)$

2.4 Sufficient Statistics

In this section we describe how to estimate the sufficient statistics to train the model. These statistics are the first order moment and second order moment of the latent variables. We find these quantities by making use of Equation 26. It will be shown that in both cases they depend upon the quantity $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}$.

First order moment of $\tilde{\mathbf{y}}_i$ From Equation 26 the expected value of $\tilde{\mathbf{y}}_i$ is given by,

$$E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]=\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right). \tag{27}$$

This is the same solution as provided in Appendix 1, Equation 22 of [4].

Second order moment of $\tilde{\mathbf{y}}_i$ From Equation 26 we know $\mathrm{Var}\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]$ and so we have the following,

$$E\left[\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]=\mathrm{Var}\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]+E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]^T. \tag{28}$$

This comes from the fact that $\mathrm{Var}\left[y\right]=E\left[y^2\right]-E\left[y\right]^2$, which is the previous equation but reformed in terms of $E\left[y^2\right]$. Continuing the derivation, we take $E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]$ from Equation 27 and $\mathrm{Var}\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]$ is taken directly from Equation 26. This gives us,

$$E\left[\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]=\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}+E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]^T. \tag{29}$$

This is the same equation as provided in Appendix 1, Equation 23 of [4].

2.4.1 Complexity of the Problem

Given the above formulae, the next question is how to go about actually implementing this.

1. Calculating the first order moment, $E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]=\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)$, implies the calculation of the large matrix $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}$.
2. Calculating the second order moment, $E\left[\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]=\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}+E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]^T$, also implies the calculation of the same matrix.

So the main stumbling block is being able to calculate $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}$. This matrix depends on the number of samples, J_i, for a particular client i. We will describe the derivation later, but below we provide the scalable and exact equations for this problem.

2.4.2 Calculating $E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]$

To calculate $E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]$ the final scalable solution becomes, for $J_i=3$,

$$E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\mathbf{\Theta}\right]=\begin{bmatrix}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}\\ -\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}+\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i1}\\ -\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}+\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i2}\\ -\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}+\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i3}\end{bmatrix}. \tag{30}$$

This gives us that,

$$E\left[\mathbf{h}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right] = \left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}, \tag{31}$$

and

$$E\left[\mathbf{w}_{ij}|\mathbf{h}_i,\mathbf{x}_{ij},\boldsymbol{\Theta}\right] = \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\left(\bar{\mathbf{x}}_{ij}-\mathbf{F}E\left[\mathbf{h}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]\right). \tag{32}$$

Where $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1} = \left(\boldsymbol{\Sigma}+\mathbf{G}\mathbf{G}^T\right)^{-1}$ and $\bar{\mathbf{x}}_{ij} = \mathbf{x}_{ij}-\boldsymbol{\mu}$. This solution will be discussed later in Section 4. Note that we have included $\mathbf{h}_i$ on the right hand side of $E\left[\mathbf{w}_{ij}|\mathbf{h}_i,\mathbf{x}_{ij},\boldsymbol{\Theta}\right]$ as it is assumed to have been pre-calculated, but this is not strictly necessary; this can be calculated each time but in this case we would write $E\left[\mathbf{w}_{ij}|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]$ and imply a greater overhead.

2.4.3 Calculating $E\left[\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]$

To calculate $E\left[\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]$ we need to calculate,

$$E\left[\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right] = \left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1} + E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]^T. \tag{33}$$

However, in practice we do not actually want to do this jointly for all of the J_i observations. We do need to have $E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]$ estimated jointly but we then consider it one sample at a time and so we go to the notation $E\left[\mathbf{y}_{ij}|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]$; we use the notation with $\mathbf{y}_{ij}$ to be explicit. We also need an appropriate expression for $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}$. This leads to the following scalable solution,

$$E\left[\mathbf{y}_{ij}\mathbf{y}_{ij}^T|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right] = \begin{bmatrix}\mathbf{T}_{ul} & \mathbf{T}_{ur}\\ \mathbf{T}_{ll} & \mathbf{T}_{lr}\end{bmatrix} + E\left[\mathbf{y}_{ij}|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]E\left[\mathbf{y}_{ij}|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]^T, \tag{34}$$

where,

$$\mathbf{T}_{ul} = \left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}, \tag{35}$$

$$\mathbf{T}_{ll} = \mathbf{T}_{ur}^T = -\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}, \tag{36}$$

$$\mathbf{T}_{lr} = \left(\mathbf{I}-\mathbf{T}_{ll}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}. \tag{37}$$

This solution will be described in more detail in Section 4.

2.5 Authentication Scores for PLDA

A verification score for PLDA can be derived by examining Figure 2B in [4]. In this figure there are two models M_0 and M_1. For M_0 the latent identity variables are considered to be independent while for M_1 the latent identity variables is shared by all of the observations. Given two observations one for enrolment $\mathbf{x}_i$ of client i and another for testing $\mathbf{x}_t$ of an unknown client t, this can then be expressed as follows.

For M_0 each of the observations, $\mathbf{x}_i$ and $\mathbf{x}_t$, are considered to have independent latent identity variables $\mathbf{h}_i$ and $\mathbf{h}_p$. They are marginalised independently as follows,

$$p\left(\mathbf{x}_i,\mathbf{x}_p|M_0\right) = \int\int p\left(\mathbf{x}_i,\mathbf{h}_i,\mathbf{w}_{i1}\right)p\left(\mathbf{x}_t,\mathbf{h}_p,\mathbf{w}_{p1}\right)d\mathbf{h}_i d\mathbf{w}_{i1}d\mathbf{h}_p d\mathbf{w}_{p1}, \tag{38}$$

$$= \int\int p\left(\mathbf{x}_i,\mathbf{h}_i,\mathbf{w}_{i1}\right)d\mathbf{h}_i d\mathbf{w}_{i1}\int\int p\left(\mathbf{x}_t,\mathbf{h}_p,\mathbf{w}_{p1}\right)d\mathbf{h}_p d\mathbf{w}_{p1}. \tag{39}$$

For M_1 the two observations are considered to have the same latent identity variable $\mathbf{h}_i$ and so are jointly marginalised. However, the session variables are still considered independently,

$$p\left(\mathbf{x}_i, \mathbf{x}_p | M_1\right) = \int\int p\left(\mathbf{x}_i, \mathbf{x}_p, \mathbf{h}_i, \mathbf{w}_{i1}, \mathbf{w}_{it}\right) d\mathbf{h}_i d\mathbf{w}_{i1} d\mathbf{w}_{it}, \tag{40}$$

$$= \int \left[\int p\left(\mathbf{x}_i, \mathbf{h}_i, \mathbf{w}_{i1}\right) d\mathbf{w}_{i1} \int p\left(\mathbf{x}_p, \mathbf{h}_i, \mathbf{w}_{ip}\right) d\mathbf{w}_{it} \right] d\mathbf{h}_i. \tag{41}$$

We can then use these two models to form a discriminant function using the likelihood ratio (LR),

$$LR = \frac{p\left(\mathbf{x}_i, \mathbf{x}_p | M_1\right)}{p\left(\mathbf{x}_i, \mathbf{x}_p | M_0\right)}. \tag{42}$$

The logarithm of this can also be taken to get the log-likelihood ratio (LLR),

$$LLR = \ln\left[p\left(\mathbf{x}_i, \mathbf{x}_p | M_1\right)\right] - \ln\left[p\left(\mathbf{x}_i, \mathbf{x}_p | M_0\right)\right]. \tag{43}$$

3 Learning the Parameters

We now come to the problem of how to learn our parameters $\boldsymbol{\Theta} = [\boldsymbol{\mu}, \mathbf{F}, \mathbf{G}, \boldsymbol{\Sigma}]$. For this purpose, an expectation-maximisation (EM) algorithm can be used and such an approach was described in [4] as indicated by Equation 26 in Appendix 1 of their work. We rewrite this equation below,

$$Q\left(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_{t-1}\right) = \sum_{i=1}^{I} \sum_{j=1}^{J_i} \int p\left(\mathbf{y}_{ij} | \mathbf{x}_{i1}, ..., \mathbf{x}_{iJ_i}, \boldsymbol{\Theta}_{t-1}\right) \ln\left[p\left(\mathbf{x}_{ij} | \mathbf{y}_{ij}, \boldsymbol{\Theta}_t\right) p\left(\mathbf{y}_{ij}\right)\right] d\mathbf{y}_{ij}. \tag{44}$$

The first term is obtained when performing the expectation part (E-step) by keeping the parameters fixed at $t-1$ ($\boldsymbol{\Theta}_{t-1}$), this is solved by calculating the sufficient statistics as described in Section 2.4. The second term, $\ln\left[p\left(\mathbf{x}_{ij} | \mathbf{y}_{ij}, \boldsymbol{\Theta}\right) p\left(\mathbf{y}_{ij}\right)\right]$, is the maximisation part (M-step) of the EM derivation. We rewrite this equation below,

$$\ln\left[p\left(\mathbf{x}_{ij} | \mathbf{y}_{ij}, \boldsymbol{\Theta}\right) p\left(\mathbf{y}_{ij}\right)\right] = \ln\left[p\left(\mathbf{x}_{ij} | \mathbf{y}_{ij}, \boldsymbol{\Theta}\right)\right] + \ln\left[p\left(\mathbf{y}_{ij}\right)\right]. \tag{45}$$

To maximise Equation 45 we differentiate with respect to $\boldsymbol{\Theta}$ and set the left hand side to 0. So let us write it fully to understand what we will get for the different parts of $\boldsymbol{\Theta}$ that we need to maximise,

$$\ln\left[p\left(\mathbf{x}_{ij} | \mathbf{y}_{ij}, \boldsymbol{\Theta}\right) p\left(\mathbf{y}_{ij}\right)\right] = \ln\left[p\left(\mathbf{x}_{ij} | \mathbf{y}_{ij}, \boldsymbol{\Theta}\right)\right] + \ln\left[p\left(\mathbf{y}_{ij}\right)\right]. \tag{46}$$

For the first term we have,

$$\ln\left[p\left(\mathbf{x}_{ij} | \mathbf{y}_{ij}, \boldsymbol{\Theta}\right)\right] = -\frac{D_x}{2} \ln\left(2\pi\right) - \frac{1}{2} \ln\left(\det\left(\boldsymbol{\Sigma}\right)\right) - \frac{1}{2} \left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right), \tag{47}$$

$$\mathbf{A} = \left[\begin{array}{cc} \mathbf{F} & \mathbf{G} \end{array} \right], \tag{48}$$

$$\mathbf{y}_{ij} = \left[\begin{array}{cc} \mathbf{h}_i^T, & \mathbf{w}_{ij}^T \end{array} \right]^T. \tag{49}$$

For the second term we have,

$$\ln\left[p\left(\mathbf{y}_{ij}\right)\right] = \ln\left[\mathcal{N}_{\mathbf{y}}\left[\mathbf{0}|\mathbf{I}\right]\right], \tag{50}$$

$$= \ln\left[(2\pi)^{-\frac{D_F + D_G}{2}} \det(\mathbf{I})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{y}_{ij}^T \mathbf{I} \mathbf{y}_{ij}\right)\right], \tag{51}$$

$$= -\frac{D_F + D_G}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\mathbf{I})) - \frac{1}{2} \mathbf{y}_{ij}^T \mathbf{y}_{ij}. \tag{52}$$

From the above equations we can see that $\ln[p(\mathbf{y}_{ij})]$ does not depend on the parameters $\boldsymbol{\Theta}$, and so we can ignore this part in the derivation. Therefore, we concentrate on differentiating Equation 47 with respect to the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\mathbf{A}$ (which includes $\mathbf{F}$ and $\mathbf{G}$). Using the definitions from Appendix A we go onto to differentiate Equation 47. We take up the derivations for each parameter below.

3.1 Differentiating with Respect to $\boldsymbol{\mu}$

We need to differentiate Equation 47 with respect to $\boldsymbol{\mu}$. We begin this differentiation below:

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ln[p(\mathbf{x}_{ij}|\mathbf{y}_{ij}, \boldsymbol{\Theta}))] =$$
$$\frac{\partial}{\partial \boldsymbol{\mu}} \left(-\frac{D_x}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\boldsymbol{\Sigma})) - \frac{1}{2} (\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}) \right), \tag{53}$$

$$= -0 - 0 - \frac{\partial}{\partial \boldsymbol{\mu}} \left[\frac{1}{2} (\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}) \right]. \tag{54}$$

We make the substitution $\mathbf{a}_{ij} = (\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij})$ and $\frac{\partial \mathbf{a}_{ij}}{\partial \boldsymbol{\mu}} = -1$ so that we can rewrite the equation above as,

$$= \frac{\partial}{\partial \mathbf{a}_{ij}} \frac{\partial \mathbf{a}_{ij}}{\partial \boldsymbol{\mu}} \left(-\frac{1}{2} \mathbf{a}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{a}_{ij} \right) = \frac{\partial}{\partial \mathbf{a}_{ij}} (-1) \left(-\frac{1}{2} \mathbf{a}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{a}_{ij} \right). \tag{55}$$

We then apply the identity in Equation 193 to get,

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ln[p(\mathbf{x}_{ij}|\mathbf{y}_{ij}, \boldsymbol{\Theta})] = (-1) \left(-\frac{1}{2} \right) \left(\boldsymbol{\Sigma}^{-1} + \left(\boldsymbol{\Sigma}^{-1} \right)^T \right) \mathbf{a}_{ij}, \tag{56}$$

$$= \left(\frac{1}{2} \right) \left(2\boldsymbol{\Sigma}^{-1} \right) (\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}), \tag{57}$$

$$= \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}). \tag{58}$$

Above we have used the fact that $\boldsymbol{\Sigma}^{-1} = \left(\boldsymbol{\Sigma}^{-1} \right)^T$. We can now maximise this by setting the left hand side (LHS) to zero and reintroducing the sums to get the following,

$$0 = \sum_{i=1}^{I} \sum_{j=1}^{J_i} \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}), \tag{59}$$

$$0 = \sum_{i=1}^{I} \sum_{j=1}^{J_i} \mathbf{x}_{ij} - \sum_{i=1}^{I} \sum_{j=1}^{J_i} \boldsymbol{\mu} - \sum_{i=1}^{I} \sum_{j=1}^{J_i} \mathbf{A}\mathbf{y}_{ij}, \tag{60}$$

$$\sum_{i=1}^{I} \sum_{j=1}^{J_i} \boldsymbol{\mu} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (\mathbf{x}_{ij} - \mathbf{A}\mathbf{y}_{ij}), \tag{61}$$

$$\boldsymbol{\mu} = \frac{1}{IJ_i} \sum_{i=1}^{I} \sum_{j=1}^{J_i} (\mathbf{x}_{ij} - \mathbf{A}\mathbf{y}_{ij}). \tag{62}$$

The result for $\boldsymbol{\mu}$, Equation 62, does not match the one presented in [4] as they use a simplification that $\boldsymbol{\mu}$ is the mean of the data itself; as such it is not actually updated. The most likely explanation for this, which was never given, is that because the latent variables should be

zero mean then $\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{A}\mathbf{y}_{ij}$ should be 0. If we make this assumption then we get the same equation which is,

$$\boldsymbol{\mu} = \frac{1}{IJ_i}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{x}_{ij}. \tag{63}$$

3.2 Differentiating with Respect to A

We need to differentiate Equation 47 with respect to $\mathbf{A}$. We begin this differentiation below:

$$\frac{\partial}{\partial \mathbf{A}}\ln\left[p\left(\mathbf{x}_{ij}|\mathbf{y}_{ij},\boldsymbol{\Theta}\right)\right] =$$
$$\frac{\partial}{\partial \mathbf{A}}\left(-\frac{D_x}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left(\det\left(\boldsymbol{\Sigma}\right)\right) - \frac{1}{2}\left(\mathbf{x}_{ij}-\boldsymbol{\mu}-\mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_{ij}-\boldsymbol{\mu}-\mathbf{A}\mathbf{y}_{ij}\right)\right), \tag{64}$$

$$= -0 - 0 - \frac{\partial}{\partial \mathbf{A}}\left[\frac{1}{2}\left(\mathbf{x}_{ij}-\boldsymbol{\mu}-\mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_{ij}-\boldsymbol{\mu}-\mathbf{A}\mathbf{y}_{ij}\right)\right]. \tag{65}$$

We now substitute $\boldsymbol{\alpha}_{ij} = \left(\mathbf{x}_{ij}-\boldsymbol{\mu}\right)$ into the above equation so that we can rewrite it as,

$$= -\frac{\partial}{\partial \mathbf{A}}\left[\frac{1}{2}\left(\boldsymbol{\alpha}_{ij}-\mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\alpha}_{ij}-\mathbf{A}\mathbf{y}_{ij}\right)\right]. \tag{66}$$

This takes a form similar to that of the identity of Equation 198 and so we use this identity to get the following:

$$\frac{\partial}{\partial \mathbf{A}}\ln\left[p\left(\mathbf{x}_{ij}|\mathbf{y}_{ij},\boldsymbol{\Theta}\right)\right] = (-2)\left(-\frac{1}{2}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\alpha}_{ij}-\mathbf{A}\mathbf{y}_{ij}\right)\mathbf{y}_{ij}^T. \tag{67}$$

To maximise with respect to $\mathbf{A}$, we set the LHS of this equation to zero and reintroduce the sums. This gives us the following result,

$$0 = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\alpha}_{ij}-\mathbf{A}\mathbf{y}_{ij}\right)\mathbf{y}_{ij}^T, \tag{68}$$

$$0 = \boldsymbol{\Sigma}^{-1}\left[\sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(\mathbf{x}_{ij}-\boldsymbol{\mu}\right)\mathbf{y}_{ij}^T - \sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{A}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\right], \tag{69}$$

$$\mathbf{A}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{y}_{ij}^T = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(\mathbf{x}_{ij}-\boldsymbol{\mu}\right)\mathbf{y}_{ij}^T, \tag{70}$$

$$\mathbf{A} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(\mathbf{x}_{ij}-\boldsymbol{\mu}\right)E\left[\mathbf{y}_{ij}\right]^T\left(\sum_{i=1}^{I}\sum_{j=1}^{J_i}E\left[\mathbf{y}_{ij}\mathbf{y}_{ij}^T\right]\right)^{-1}. \tag{71}$$

This is the same as the update rule provided in [4].

3.3 Differentiating with Respect to Σ

We need to differentiate Equation 47 with respect to $\boldsymbol{\Sigma}$. We begin this differentiation below:

$$\frac{\partial}{\partial \boldsymbol{\Sigma}}\ln\left[p\left(\mathbf{x}_{ij}|\mathbf{y}_{ij},\boldsymbol{\Theta}\right)\right] =$$
$$\frac{\partial}{\partial \boldsymbol{\Sigma}}\left[-\frac{D_x}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left(\det\left(\boldsymbol{\Sigma}\right)\right) - \frac{1}{2}\left(\mathbf{x}_{ij}-\boldsymbol{\mu}-\mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_{ij}-\boldsymbol{\mu}-\mathbf{A}\mathbf{y}_{ij}\right)\right], \tag{72}$$

We substitute $\mathbf{a}_{ij} = (\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij})$ so that we can rewrite the equation above as,

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} \ln\left[p\left(\mathbf{x}_{ij}|\mathbf{y}_{ij}, \boldsymbol{\Theta}\right)\right] = \frac{\partial}{\partial \boldsymbol{\Sigma}}\left[-\frac{D_x}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left(\det\left(\boldsymbol{\Sigma}\right)\right) - \frac{1}{2}\mathbf{a}_{ij}^T\boldsymbol{\Sigma}^{-1}\mathbf{a}_{ij}\right]. \tag{73}$$

We can then use the identities in Equations 194 and 195. Using these identities we get the following:

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} \ln\left[p\left(\mathbf{x}_{ij}|\mathbf{y}_{ij}, \boldsymbol{\Theta}\right)\right] = \left(-0 - \frac{1}{2}\boldsymbol{\Sigma}^{-T} + \frac{1}{2}\boldsymbol{\Sigma}^{-T}\mathbf{a}_{ij}\mathbf{a}_{ij}^T\boldsymbol{\Sigma}^{-1}\right), \tag{74}$$

$$= -\frac{1}{2}\boldsymbol{\Sigma}^{-T} + \frac{1}{2}\boldsymbol{\Sigma}^{-T}\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}. \tag{75}$$

We need to take the above result, set the LHS to zero and reintroduce the sum. This gives us the following result (remembering that we want to find an expression for $\boldsymbol{\Sigma}$).

$$0 = -\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\boldsymbol{\Sigma}^{-T} + \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\boldsymbol{\Sigma}^{-T}\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}, \tag{76}$$

$$\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J_i}1 = \boldsymbol{\Sigma}^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}, \tag{77}$$

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J_i}1 = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}, \tag{78}$$

$$IJ_i = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}, \tag{79}$$

$$IJ_i\boldsymbol{\Sigma} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}, \tag{80}$$

As we want $\boldsymbol{\Sigma}$ to be diagonal, we finally get

$$\boldsymbol{\Sigma} = \frac{1}{IJ_i}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{diag}\left[\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)\left(\mathbf{x}_{ij} - \boldsymbol{\mu} - \mathbf{A}\mathbf{y}_{ij}\right)^T\right]. \tag{81}$$

Everything else that follows is trying to get this expression to match up with what is presented in [4]. It should also provide a more stable estimate of $\boldsymbol{\Sigma}$.

Let's substitute $\mathbf{b}_{ij} = \mathbf{x}_{ij} - \boldsymbol{\mu}$,

$$\boldsymbol{\Sigma} = \frac{1}{IJ_i}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{diag}\left[\left(\mathbf{b}_{ij} - \mathbf{A}\mathbf{y}_{ij}\right)\left(\mathbf{b}_{ij} - \mathbf{A}\mathbf{y}_{ij}\right)^T\right], \tag{82}$$

$$\boldsymbol{\Sigma} = \frac{1}{IJ_i}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{diag}\left[\mathbf{b}_{ij}\mathbf{b}_{ij}^T - \mathbf{b}_{ij}\mathbf{y}_{ij}^T\mathbf{A}^T - \mathbf{A}\mathbf{y}_{ij}\mathbf{b}_{ij}^T + \mathbf{A}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\mathbf{A}^T\right], \tag{83}$$

$$\boldsymbol{\Sigma} = \frac{1}{IJ_i}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{diag}\left[\mathbf{b}_{ij}\mathbf{b}_{ij}^T - 2\mathbf{A}\mathbf{y}_{ij}\mathbf{b}_{ij}^T + \mathbf{A}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\mathbf{A}^T\right]. \tag{84}$$

We now make use of our previous definition of $\mathbf{A}$,

$$\mathbf{A} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(\mathbf{x}_{ij} - \boldsymbol{\mu}\right)E\left[\mathbf{y}_{ij}\right]^T\left(\sum_{i=1}^{I}\sum_{j=1}^{J_i}E\left[\mathbf{y}_{ij}\mathbf{y}_{ij}^T\right]\right)^{-1}. \tag{85}$$

Taking this and substituting in,

$$\boldsymbol{\Sigma} = \frac{1}{IJ_i}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\text{diag}\left[\mathbf{b}_{ij}\mathbf{b}_{ij}^T - 2\mathbf{A}\mathbf{y}_{ij}\mathbf{b}_{ij}^T + \mathbf{A}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\mathbf{A}^T\right]. \tag{86}$$

$$= \frac{1}{IJ_i}\text{diag}\left[\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{b}_{ij}\mathbf{b}_{ij}^T - \sum_{i=1}^{I}\sum_{j=1}^{J_i}2\mathbf{A}\mathbf{y}_{ij}\mathbf{b}_{ij}^T + \sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{A}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\mathbf{A}^T\right], \tag{87}$$

As we will be taking the diagonal of $\boldsymbol{\Sigma}$, this can be further written as,

$$= \frac{1}{IJ_i}\text{diag}\left[\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{b}_{ij}\mathbf{b}_{ij}^T - 2\mathbf{A}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{b}_{ij}^T + \mathbf{A}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\mathbf{A}^T\right]. \tag{88}$$

This is the point at which we put in our definition for $\mathbf{A}$, for the last term only,

$$\boldsymbol{\Sigma} = \frac{1}{IJ_i}\text{diag}\left[... + \left(\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{b}_{ij}^T\right)\left(\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\right)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\mathbf{A}^T\right], \tag{89}$$

$$= \frac{1}{IJ_i}\text{diag}\left[... + \left(\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{b}_{ij}^T\right)\left(\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\right)^{-1}\left(\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{y}_{ij}^T\right)\mathbf{A}^T\right], \tag{90}$$

$$= \frac{1}{IJ_i}\text{diag}\left[... + \left(\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{b}_{ij}^T\right)\mathbf{I}\mathbf{A}^T\right], \tag{91}$$

$$= \frac{1}{IJ_i}\text{diag}\left[\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{b}_{ij}\mathbf{b}_{ij}^T - 2\mathbf{A}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{b}_{ij}^T + \left(\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{b}_{ij}^T\right)\mathbf{I}\mathbf{A}^T\right], \tag{92}$$

$$= \frac{1}{IJ_i}\text{diag}\left[\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{b}_{ij}\mathbf{b}_{ij}^T - 2\mathbf{A}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{b}_{ij}^T + \mathbf{A}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\mathbf{b}_{ij}^T\right], \tag{93}$$

$$= \frac{1}{IJ_i}\text{diag}\left[\sum_{i=1}^{I}\sum_{j=1}^{J_i}\left(\mathbf{x}_{ij} - \boldsymbol{\mu}\right)\left(\mathbf{x}_{ij} - \boldsymbol{\mu}\right)^T - \mathbf{A}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{y}_{ij}\left(\mathbf{x}_{ij} - \boldsymbol{\mu}\right)^T\right]. \tag{94}$$

This provides us with the same solution as Prince and Elder. We can now write,

$$\boldsymbol{\Sigma} = \frac{1}{IJ_i}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\text{diag}\left[\left(\mathbf{x}_{ij} - \boldsymbol{\mu}\right)\left(\mathbf{x}_{ij} - \boldsymbol{\mu}\right)^T - \mathbf{A}\mathbf{y}_{ij}\left(\mathbf{x}_{ij} - \boldsymbol{\mu}\right)^T\right]. \tag{95}$$

4 Implementation Details

The following is a quick summary of some of the implementation details for PLDA. This is not a guide on how to implement PLDA, although it could help considerably.

The first point to note is that there are values that we will use throughout that are probably worth storing somewhere. A non-exhaustive list is:

- $\ln(2\pi)$
- $-\frac{D_\mathbf{x}}{2}\ln(2\pi)$
- $\ln\left[\det\left(\boldsymbol{\Sigma}\right)\right]$

- $\boldsymbol{\Sigma}^{-1}$ because we assume this to be diagonal this is actually a vector inversion.
- $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}$ because it is used for estimating the latent variables and scoring.
- $\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}$ and $\mathbf{F}^T\mathbf{Q}$
- $\left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}$ because this is used for estimating the latent variables and for scoring.

We now take up scalable and exact solutions for the problems of evaluating the likelihood and estimating the sufficient statistics.

4.1 Log-likelihood of $\tilde{\mathbf{x}}$

This involves calculating the likelihood of the following,

$$p\left(\tilde{\mathbf{x}}_i\right) = \mathcal{N}_{\tilde{\mathbf{x}}}\left[\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right]. \tag{96}$$

We can simplify this by calculating the log-likelihood which gives us,

$$\ln\left[p\left(\tilde{\mathbf{x}}_i\right)\right] = -\frac{J_i D_x}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left[\det\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)\right] - \frac{1}{2}\left(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}\right)^T\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)^{-1}\left(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}\right).$$

For each of the three parts of this calculation we note the following:

1. $-\frac{J_i D_x}{2}\ln\left(2\pi\right)$ is trivial to calculate as it depends only on the dimensionality of the feature space (D_x) and the number of samples (J_i),
2. $-\frac{1}{2}\ln\left[\det\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)\right]$ will be difficult to calculate as it involves calculating the determinant of potentially very large matrices, therefore, simplifications need to be used, and
3. $-\frac{1}{2}\left(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}\right)^T\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)^{-1}\left(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}\right)$ will be difficult to calculate as it involves multiplying a very large matrix, therefore, simplifications need to be used.

Because Point 1 can be relatively easily dealt with we only address Points 2 and 3 in the following sections.

4.1.1 Calculating $-\frac{1}{2}\left(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}\right)^T\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)^{-1}\left(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}\right)$ Efficiently: Method 1

We will examine what the matrix $\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)^{-1}$ is. First we derive fully what $\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)$ is,

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{F} & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \mathbf{G} \end{bmatrix}, \tag{97}$$

$$\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \begin{bmatrix} \mathbf{F} & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \mathbf{G} \end{bmatrix}\begin{bmatrix} \mathbf{F}^T & \mathbf{F}^T & \mathbf{F}^T \\ \mathbf{G}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}^T \end{bmatrix}, \tag{98}$$

$$\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \begin{bmatrix} \mathbf{F}\mathbf{F}^T + \mathbf{G}\mathbf{G}^T, & \mathbf{F}\mathbf{F}^T, & \mathbf{F}\mathbf{F}^T \\ \mathbf{F}\mathbf{F}^T, & \mathbf{F}\mathbf{F}^T + \mathbf{G}\mathbf{G}^T, & \mathbf{F}\mathbf{F}^T \\ \mathbf{F}\mathbf{F}^T, & \mathbf{F}\mathbf{F}^T, & \mathbf{F}\mathbf{F}^T + \mathbf{G}\mathbf{G}^T \end{bmatrix}, \tag{99}$$

$$\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} + \mathbf{F}\mathbf{F}^T + \mathbf{G}\mathbf{G}^T, & \mathbf{F}\mathbf{F}^T, & \mathbf{F}\mathbf{F}^T \\ \mathbf{F}\mathbf{F}^T, & \tilde{\boldsymbol{\Sigma}} + \mathbf{F}\mathbf{F}^T + \mathbf{G}\mathbf{G}^T, & \mathbf{F}\mathbf{F}^T \\ \mathbf{F}\mathbf{F}^T, & \mathbf{F}\mathbf{F}^T, & \tilde{\boldsymbol{\Sigma}} + \mathbf{F}\mathbf{F}^T + \mathbf{G}\mathbf{G}^T \end{bmatrix}. \tag{100}$$

This matrix is actually very big $(J_i D_x, J_i D_x)$ and it would need to be inverted so calculating this efficiently is necessary if we want to implement scoring this way. If we take the Woodbury matrix identity,

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}, \tag{101}$$

and apply it in a similar way to [2] page 15 (Section 3.2.3) we would get the following,

$$\mathbf{\Lambda} = \left(\tilde{\mathbf{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)^{-1} = \tilde{\mathbf{\Sigma}}^{-1} - \tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\left(I + \tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}, \tag{102}$$

$$\mathbf{\Lambda} = \tilde{\mathbf{\Sigma}}^{-1} - \mathbf{\Gamma}\mathbf{\Gamma}^T. \tag{103}$$

Where, $\mathbf{\Gamma} = \tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\left(\mathbf{I} + \tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-\frac{1}{2}}$.

Using the above we can now rewrite the quadratic term in the Gaussian that we need to evaluate as being,

$$(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})^T\tilde{\mathbf{\Sigma}}^{-1}(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}) - (\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})^T\mathbf{\Gamma}\mathbf{\Gamma}^T(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}). \tag{104}$$

The first part, $(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})^T\tilde{\mathbf{\Sigma}}^{-1}(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})$, is efficient to compute as $\tilde{\mathbf{\Sigma}}^{-1}$ is diagonal. The second part, $(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})^T\mathbf{\Gamma}\mathbf{\Gamma}^T(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})$, is more complicated. But we can note that $(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})^T\mathbf{\Gamma} = \left(\mathbf{\Gamma}^T(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})\right)^T$ and so we only ever need to evaluate one half of the equation. And that in the end $\mathbf{\Gamma}^T(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})$ is a vector of dimensions $(D_F + J_i D_G, 1)$. Therefore, the final result of the second part is just the magnitude of this vector. This comes from [2].

The above solution is still problematic. This is because $\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}$ is a matrix of size $(J_i D_x, D_F + J_i D_G)$ and is quadratic with the number of samples J_i. So let us expand this out and see if we can reduce the load on the memory somehow. We want to calculate $\mathbf{\Gamma}^T(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}})$ which is given by,

$$\mathbf{\Gamma}^T(\tilde{\mathbf{x}}_i - \tilde{\boldsymbol{\mu}}) = \left[\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\left(\mathbf{I} + \tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-\frac{1}{2}}\right]^T \begin{bmatrix} \bar{\mathbf{x}}_{i1} \\ \bar{\mathbf{x}}_{i2} \\ \bar{\mathbf{x}}_{i3} \end{bmatrix}, \tag{105}$$

$$= \left[\left(\mathbf{I} + \tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-\frac{1}{2}}\right]^T \tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1} \begin{bmatrix} \bar{\mathbf{x}}_{i1} \\ \bar{\mathbf{x}}_{i2} \\ \bar{\mathbf{x}}_{i3} \end{bmatrix}, \tag{106}$$

$$= \tilde{\mathbf{Z}}\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1} \begin{bmatrix} \bar{\mathbf{x}}_{i1} \\ \bar{\mathbf{x}}_{i2} \\ \bar{\mathbf{x}}_{i3} \end{bmatrix}, \tag{107}$$

$$= \tilde{\mathbf{Z}} \begin{bmatrix} \mathbf{F}^T\mathbf{\Sigma}^{-1}, & \mathbf{F}^T\mathbf{\Sigma}^{-1}, & \mathbf{F}^T\mathbf{\Sigma}^{-1} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T\mathbf{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}^T\mathbf{\Sigma}^{-1} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_{i1} \\ \bar{\mathbf{x}}_{i2} \\ \bar{\mathbf{x}}_{i3} \end{bmatrix}, \tag{108}$$

$$= \tilde{\mathbf{Z}} \begin{bmatrix} \sum_{j=1}^{J_i} \mathbf{F}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i2} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{bmatrix}. \tag{109}$$

Where $\tilde{\mathbf{Z}} = \left[\left(\mathbf{I} + \tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-\frac{1}{2}}\right]^T$ and $\bar{\mathbf{x}}_{i1} = \mathbf{x}_{i1} - \boldsymbol{\mu}$.

Obviously the above way of writing it is much more efficient in terms of memory as $\tilde{\mathbf{Z}}$ is a matrix of size $(D_F + J_i D_G, D_F + J_i D_G)$ and the second vector is of size $(D_F + J_i D_G, 1)$. The calculation of the matrix $\tilde{\mathbf{Z}}$ will be a limiting factor in any application as it involves calculating the inverse and a square root. However, this only has to be derived once for a given number of observations J_i. Still, this is not the best way to solve this problem. We describe a better, more efficient way of doing this below.

4.1.2 Calculating $-\frac{1}{2}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)^T\left(\tilde{\boldsymbol{\Sigma}}+\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)^{-1}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)$ Efficiently: Method 2

Starting from where we left off before what we actually want to calculate is,

$$\begin{bmatrix} \sum_{j=1}^{J_i}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i2} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{bmatrix}^T \tilde{\mathbf{Z}}^T\tilde{\mathbf{Z}} \begin{bmatrix} \sum_{j=1}^{J_i}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i2} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{bmatrix}, \tag{110}$$

$$= \begin{bmatrix} \sum_{j=1}^{J_i}\bar{\mathbf{x}}_{ij}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}, & \bar{\mathbf{x}}_{i1}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}, & \ldots, & \bar{\mathbf{x}}_{i3}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{bmatrix}\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}\begin{bmatrix} \sum_{j=1}^{J_i}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i2} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{bmatrix}. \tag{111}$$

This comes from the definition of $\tilde{\mathbf{Z}}$. Now we can use the matrix inversion lemma, using an intermediate form of the Woodbury identity (coming from an LDU decomposition) again to simplify the inversion of $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)$. This leads to the above equation becoming,

$$= \mathbf{a}\begin{bmatrix} I_{D_F} & 0 \\ -D^{-1}C, & I_{J_iD_G} \end{bmatrix}\begin{bmatrix} \left(A-BD^{-1}C\right)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}\begin{bmatrix} I_{D_F} & -BD^{-1} \\ 0 & I_{J_iD_F} \end{bmatrix}\mathbf{b}. \tag{112}$$

Where,

$$\mathbf{a} = \begin{bmatrix} \sum_{j=1}^{J_i}\bar{\mathbf{x}}_{ij}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}, & \bar{\mathbf{x}}_{i1}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}, & \ldots, & \bar{\mathbf{x}}_{i3}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{bmatrix}, \tag{113}$$

and

$$\mathbf{b} = \begin{bmatrix} \sum_{j=1}^{J_i}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i2} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{bmatrix}. \tag{114}$$

We will now proceed to decompose this set of matrix multiplications to derive a more efficient solution. But first we define the elements that we will use below.

$$A = \left[\mathbf{I}+J_i\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\right], \tag{115}$$

$$B = \begin{bmatrix} \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}, & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}, & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{bmatrix}, \tag{116}$$

$$C = \begin{bmatrix} \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} \end{bmatrix} = B^T, \tag{117}$$

$$D = \begin{bmatrix} \mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{bmatrix}. \tag{118}$$

First we note that,

$$\mathbf{a}\begin{bmatrix} I_{D_F}, & 0 \\ -D^{-1}C, & I_{J_iD_G} \end{bmatrix} = \left(\begin{bmatrix} I_{D_F}, & -BD^{-1} \\ 0, & I_{J_iD_G} \end{bmatrix}\mathbf{b}\right)^T. \tag{119}$$

This means that solving one provides us with the solution for the other, provided we take the transpose. For convenience we choose to solve the second one. We note that,

$$-BD^{-1} = -\begin{bmatrix} \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}, & \ldots, & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{bmatrix}\begin{bmatrix} \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{bmatrix}, \tag{120}$$

$$-BD^{-1} = -\left[\begin{array}{cccc} \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}, & \ldots, & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{array}\right]. \tag{121}$$

The matrix $-BD^{-1}$ is of size (D_F, J_iD_G) and $\mathbf{b}$ is of size $(D_F + J_iD_G, J_iD_x)$. Thus, the multiplication of these two terms is given by,

$$\left[\begin{array}{cc} I & -BD^{-1} \\ 0 & I \end{array}\right]\mathbf{b} = \left[\begin{array}{cc} I_{D_F}, & -BD^{-1} \\ 0, & I_{J_iD_G} \end{array}\right]\left[\begin{array}{c} \sum_{j=1}^{J_i}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array}\right], \tag{122}$$

$$= \left[\begin{array}{c} -BD^{-1}\left[\begin{array}{c} \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array}\right] + \sum_{j=1}^{J_i}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array}\right], \tag{123}$$

$$= \left[\begin{array}{c} \left(\sum_{j=1}^{J_i}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} - \sum_{j=1}^{J_i}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij}\right) \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array}\right], \tag{124}$$

$$= \left[\begin{array}{c} \mathbf{F}^T\left(\sum_{j=1}^{J_i}\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} - \sum_{j=1}^{J_i}\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij}\right) \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array}\right], \tag{125}$$

$$= \left[\begin{array}{c} \mathbf{F}^T\left(\sum_{j=1}^{J_i}\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\right)\bar{\mathbf{x}}_{ij}\right) \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array}\right], \tag{126}$$

$$= \left[\begin{array}{c} \sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array}\right]. \tag{127}$$

Where $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1} = \left(\boldsymbol{\Sigma}+\mathbf{G}\mathbf{G}^T\right)^{-1}$. We also denote another term for use later, $\mathbf{L} = \boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}$. Therefore,

$$\mathbf{a}\left[\begin{array}{cc} I_{D_F} & 0 \\ -D^{-1}C & I_{J_iD_G} \end{array}\right] = \left[\begin{array}{c} \sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array}\right]^T, \tag{128}$$

$$= \left[\begin{array}{cccc} \sum_{j=1}^{J_i}\bar{\mathbf{x}}_{ij}^T\mathbf{Q}^T\mathbf{F}, & \bar{\mathbf{x}}_{i1}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}, & \ldots, & \bar{\mathbf{x}}_{i3}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{array}\right]. \tag{129}$$

Now let us work on the central term,

$$\left[\begin{array}{cc} \left(A - BD^{-1}C\right)^{-1} & 0 \\ 0 & D^{-1} \end{array} \right]. \tag{130}$$

First we note that we need to solve $BD^{-1}C = BD^{-1}B^T$. We first use the definition of BD^{-1} from before.

$$BD^{-1} = \left[\begin{array}{ccc} \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}, & \ldots, & \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{array} \right]. \tag{131}$$

Then to find $BD^{-1}B^T$ we get that,

$$BD^{-1}B^T = BD^{-1}\left[\begin{array}{c} \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \end{array} \right], \tag{132}$$

$$= \left[J_i \left(\mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \right) \right] = BD^{-1}C. \tag{133}$$

We use this to continue to solve for $A - BD^{-1}C$,

$$A - BD^{-1}C = \mathbf{I} + J_i\mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{F} - J_i\left(\mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \right), \tag{134}$$

$$A - BD^{-1}C = \mathbf{I} + J_i\mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{F} - J_i\left(\mathbf{F}^T\mathbf{L}\mathbf{F} \right), \tag{135}$$

$$A - BD^{-1}C = \mathbf{I} + J_i\mathbf{F}^T\left(\mathbf{\Sigma}^{-1}\mathbf{F} - \mathbf{L}\mathbf{F} \right), \tag{136}$$

$$A - BD^{-1}C = \mathbf{I} + J_i\mathbf{F}^T\left(\mathbf{\Sigma}^{-1} - \mathbf{L} \right)\mathbf{F}, \tag{137}$$

$$A - BD^{-1}C = \mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}. \tag{138}$$

Therefore,

$$\left[\begin{array}{cc} \left(A - BD^{-1}C\right)^{-1} & 0 \\ 0 & D^{-1} \end{array} \right] = \left[\begin{array}{cc} \left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} & 0 \\ 0 & D^{-1} \end{array} \right]. \tag{139}$$

Now, let us keep going with this. Work out the right hand side which is,

$$\left[\begin{array}{cc} \left(A - BD^{-1}C\right)^{-1} & 0 \\ 0 & D^{-1} \end{array} \right] \left[\begin{array}{cc} I_{D_F} & -BD^{-1} \\ 0 & I_{J_iD_F} \end{array} \right] \mathbf{b} = \left[\begin{array}{cc} \left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} & 0 \\ 0 & D^{-1} \end{array} \right] \left[\begin{array}{c} \sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array} \right], \tag{140}$$

$$= \left[\begin{array}{cc} \left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left[\begin{array}{ccc} \left(\mathbf{I} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\mathbf{I} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{array} \right] \end{array} \right] \left[\begin{array}{c} \sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array} \right], \tag{141}$$

$$= \left[\begin{array}{c} \left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ \left(\mathbf{I} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \left(\mathbf{I} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{array} \right]. \tag{142}$$

Then we incorporate the left hand side and we get,

$$\left[\begin{array}{cccc} \sum_{j=1}^{J_i} \bar{\mathbf{x}}_{ij}^T \mathbf{Q}^T \mathbf{F}, & \bar{\mathbf{x}}_{i1}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}, & \ldots, & \bar{\mathbf{x}}_{i3}^T \boldsymbol{\Sigma}^{-1} \mathbf{G} \end{array} \right] \left[\begin{array}{c} \left(\mathbf{I} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1} \sum_{j=1}^{J_i} \mathbf{F}^T \mathbf{Q} \bar{\mathbf{x}}_{ij} \\ \left(\mathbf{I} + \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}}_{i1} \\ \vdots \\ \left(\mathbf{I} + \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}}_{i3} \end{array} \right] \tag{143}$$

$$= \left[\sum_{j=1}^{J_i} \bar{\mathbf{x}}_{ij}^T \mathbf{Q}^T \mathbf{F} \left(\mathbf{I} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1} \sum_{j=1}^{J_i} \mathbf{F}^T \mathbf{Q} \bar{\mathbf{x}}_{ij} + \sum_{j=1}^{J_i} \bar{\mathbf{x}}_{ij}^T \boldsymbol{\Sigma}^{-1} \mathbf{G} \left(\mathbf{I} + \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}}_{ij} \right]. \tag{144}$$

This is the efficient formula to do full integration log-likelihood scoring for PLDA. It is efficient for the following reasons.

1. Computing $\left(\mathbf{I} + \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}\right)^{-1}$ will be efficient because it is a matrix of size (D_G, D_G). In fact it can be pre-computed and does not depend at all upon the number of samples J_i.
2. $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{G} \left(\mathbf{I} + \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \boldsymbol{\Sigma}^{-1}$ is efficient to compute because $\left(\mathbf{I} + \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \mathbf{G}\right)^{-1}$ is efficient to compute as is $\mathbf{G}^T \boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}^{-1} \mathbf{G}$ because $\boldsymbol{\Sigma}^{-1}$ is diagonal. In fact it can be pre-computed and does not depend at all upon the number of samples J_i.
3. Computing $\left(\mathbf{I} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1}$ is efficient because it is a matrix of size (D_F, D_F) and $\mathbf{Q}$ is efficient to compute (and can even be pre-computed). Also, the matrix does not increase in size depending upon the number of samples J_i. However, it does depend upon the number of samples.
4. $\mathbf{F}^T \mathbf{Q} \bar{\mathbf{x}}_{ij}$ is efficient to compute and is only of size $(D_F, 1)$, also $\mathbf{G}^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}}_{ij}$ is efficient to compute and is only of size $(D_G, 1)$. This also indicates that enrolment should consist of obtaining these quantities and not on keeping the full feature vector.

4.1.3 Determinant of $\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T\right)$

How do we easily find the determinant of this quite large matrix? First we note that,

$$\det\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T\right) = \det\left(\tilde{\boldsymbol{\Sigma}}\right) \det\left(\mathbf{I} + \tilde{\mathbf{A}}^T \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{A}}\right). \tag{145}$$

This particular trick can be found in wikipedia (http://en.wikipedia.org/wiki/Determinant, Sylvester's determinant theorem). This works well because $\tilde{\boldsymbol{\Sigma}}$ is diagonal and so $\det\left(\tilde{\boldsymbol{\Sigma}}\right)$ is trivial to compute,

$$\det\left(\tilde{\boldsymbol{\Sigma}}\right) = \det\left(\boldsymbol{\Sigma}\right)^{J_i}$$

Taking up the second term, we note that $\tilde{\mathbf{A}}^T \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{A}}$ is of size $\left[\left(D_F + J_i D_G\right), \left(D_F + J_i D_G\right)\right]$. However, there should be a structure to $\tilde{\mathbf{A}}^T \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{A}}$, so let us try to further simplify this,

$$\tilde{\mathbf{A}}^T \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{A}} = \left[\begin{array}{ccc} \mathbf{F}^T & \mathbf{F}^T & \mathbf{F}^T \\ \mathbf{G}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}^T \end{array} \right] \left[\begin{array}{ccc} \boldsymbol{\Sigma}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}^{-1} \end{array} \right] \left[\begin{array}{cccc} \mathbf{F} & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \mathbf{G} \end{array} \right], \tag{146}$$

$$= \left[\begin{array}{ccc} \mathbf{F}^T \boldsymbol{\Sigma}^{-1}, & \cdots, & \mathbf{F}^T \boldsymbol{\Sigma}^{-1} \\ \mathbf{G}^T \boldsymbol{\Sigma}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}^T \boldsymbol{\Sigma}^{-1} \end{array} \right] \left[\begin{array}{cccc} \mathbf{F} & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \mathbf{G} \end{array} \right], \tag{147}$$

$$= \begin{bmatrix} J_i\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}, & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}, & \cdots & ,\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} & \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} & \mathbf{0} & \mathbf{0} & \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{bmatrix}, \tag{148}$$

$$\mathbf{I} + \tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{I} + J_i\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}, & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}, & \cdots, & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} & \mathbf{I} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} & \mathbf{0} & \mathbf{0} & \mathbf{I} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{bmatrix}. \tag{149}$$

If need be we could further subdivide the problem. By decomposing the determinant of $\mathbf{I} + \tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}$ into a block matrix we can use the identity,

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A)\det\left(D - CA^{-1}B\right) = \det(D)\det\left(A - BD^{-1}C\right). \tag{150}$$

The terms A, B, C and D are the same as before. So an efficient way to compute this is to note that,

$$\det(D) = \begin{bmatrix} \mathbf{I} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{bmatrix}, \tag{151}$$

$$= \det\left(\mathbf{I} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{J_i}. \tag{152}$$

This is an (D_G, D_G) matrix whose determinant would be efficient to compute. Next, we already have an efficient expression for the second term and that is,

$$A - BD^{-1}C = \mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}. \tag{153}$$

Consequently we need to determine,

$$\det\left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right). \tag{154}$$

This is an (D_F, D_F) matrix and could be computed efficiently.

Finally this gives us the following solution.

$$\det\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right) = \det\left(\tilde{\boldsymbol{\Sigma}}\right)\det\left(\mathbf{I} + \tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right), \tag{155}$$

$$= \left[\det(\boldsymbol{\Sigma})\right]^{J_i}\left[\det\left(\mathbf{I} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)\right]^{J_i}\left[\det\left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)\right]. \tag{156}$$

Taking the logarithm we would get,

$$\ln\left(\det\left(\tilde{\boldsymbol{\Sigma}} + \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)\right) = \ln\left(\left[\det(\boldsymbol{\Sigma})\right]^{J_i}\left[\det\left(\mathbf{I} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)\right]^{J_i}\left[\det\left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)\right]\right), \tag{157}$$

$$= \ln\left(\left[\det(\boldsymbol{\Sigma})\right]^{J_i}\right) + \ln\left(\left[\det\left(\mathbf{I} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)\right]^{J_i}\right) + \ln\left(\left[\det\left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)\right]\right), \tag{158}$$

$$= J_i\ln\left(\det(\boldsymbol{\Sigma})\right) + J_i\ln\left(\det\left(\mathbf{I} + \mathbf{G}^{T-1}\mathbf{G}\right)\right) + \ln\left(\det\left(\mathbf{I} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)\right). \tag{159}$$

Providing a more useful final solution this gives us that,

$$-\frac{1}{2}\ln\left[\det\left(\tilde{\mathbf{\Sigma}}+\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)\right]=-\frac{1}{2}\ln\left(\left[\det\left(\mathbf{\Sigma}\right)\right]^{J_i}\left[\det\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)\right]^{J_i}\det\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)\right), \tag{160}$$

$$=-\frac{1}{2}\left[J_i\ln\left(\det\left(\mathbf{\Sigma}\right)\right)+J_i\ln\left(\det\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)\right)+\ln\left(\det\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)\right)\right], \tag{161}$$

$$=-\frac{J_i}{2}\ln\left(\det\left(\mathbf{\Sigma}\right)\right)-\frac{J_i}{2}\ln\left(\det\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)\right)-\frac{1}{2}\ln\left(\det\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)\right). \tag{162}$$

4.1.4 Final Solution

Now using the above solution we can extend it and find an efficient way to calculate the sufficient statistics. We take this up below. But before we do that, what is the final equation to calculate the log-likelihood?

$$\ln\left[p\left(\tilde{\mathbf{x}}_i\right)\right]=-\frac{J_iD_x}{2}\ln\left(2\pi\right)-\frac{1}{2}\ln\left[\det\left(\tilde{\mathbf{\Sigma}}+\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)\right]-\frac{1}{2}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)^T\left(\tilde{\mathbf{\Sigma}}+\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)^{-1}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right), \tag{163}$$

$$\begin{aligned}\ln\left[p\left(\tilde{\mathbf{x}}_i\right)\right]=&-\frac{J_iD_x}{2}\ln\left(2\pi\right)-\frac{1}{2}\ln\left[\det\left(\tilde{\mathbf{\Sigma}}+\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)\right]\\&-\frac{1}{2}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)^T\tilde{\mathbf{\Sigma}}^{-1}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)+\frac{1}{2}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)^T\mathbf{\Gamma}\mathbf{\Gamma}^T\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right),\end{aligned} \tag{164}$$

$$\begin{aligned}\ln\left[p\left(\tilde{\mathbf{x}}_i\right)\right]=&-\frac{J_iD_x}{2}\ln\left(2\pi\right)-\frac{1}{2}\ln\left[\det\left(\tilde{\mathbf{\Sigma}}+\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)\right]-\frac{1}{2}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)^T\tilde{\mathbf{\Sigma}}^{-1}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)\\&+\frac{1}{2}\sum_{j=1}^{J_i}\bar{\mathbf{x}}_{ij}^T\mathbf{Q}^T\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}+\frac{1}{2}\sum_{j=1}^{J_i}\bar{\mathbf{x}}_{ij}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{ij}.\end{aligned} \tag{165}$$

We further note the following,

$$-\frac{1}{2}\ln\left[\det\left(\tilde{\mathbf{\Sigma}}+\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\right)\right]=-\frac{J_i}{2}\ln\left(\det\left(\mathbf{\Sigma}\right)\right)-\frac{J_i}{2}\ln\left(\det\left(\mathbf{I}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)\right)-\frac{1}{2}\ln\left(\det\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)\right), \tag{166}$$

and

$$-\frac{1}{2}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)^T\tilde{\mathbf{\Sigma}}^{-1}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right)=-\frac{1}{2}\left[\begin{array}{ccc}\bar{\mathbf{x}}_{i1}^T, & \cdots, & \bar{\mathbf{x}}_{i3}^T\end{array}\right]\left[\begin{array}{ccc}\mathbf{\Sigma}^{-1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \ddots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}^{-1}\end{array}\right]\left[\begin{array}{c}\bar{\mathbf{x}}_{i1}\\ \vdots\\ \bar{\mathbf{x}}_{i3}\end{array}\right], \tag{167}$$

$$=-\frac{1}{2}\left[\begin{array}{ccc}\bar{\mathbf{x}}_{i1}^T\mathbf{\Sigma}^{-1}, & \cdots, & \bar{\mathbf{x}}_{i3}^T\mathbf{\Sigma}^{-1}\end{array}\right]\left[\begin{array}{c}\bar{\mathbf{x}}_{i1}\\ \vdots\\ \bar{\mathbf{x}}_{i3}\end{array}\right], \tag{168}$$

$$=-\frac{1}{2}\sum_{j=1}^{J_i}\bar{\mathbf{x}}_{ij}^T\mathbf{\Sigma}^{-1}\bar{\mathbf{x}}_{ij}. \tag{169}$$

Thus, this can all be efficiently computed and even pre-computed for a model. The very final formulation that we have is,

$$\begin{aligned}\ln\left[p\left(\tilde{\mathbf{x}}_i\right)\right] = &-\frac{J_i D_x}{2}\ln\left(2\pi\right) - \frac{J_i}{2}\ln\left(\det\left(\boldsymbol{\Sigma}\right)\right) - \frac{J_i}{2}\ln\left(\det\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)\right) \\ &- \frac{1}{2}\ln\left(\det\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)\right) + \frac{1}{2}\sum_{j=1}^{J_i}\bar{\mathbf{x}}_{ij}^T\mathbf{Q}^T\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ &+ \frac{1}{2}\sum_{j=1}^{J_i}\bar{\mathbf{x}}_{ij}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} - \frac{1}{2}\sum_{j=1}^{J_i}\bar{\mathbf{x}}_{ij}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij}. \quad (170)\end{aligned}$$

4.2 Sufficient Statistics

The sufficient statistics for the PLDA model are $E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]$ and $E\left[\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]$. These both involve calculations with the matrix $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}$. Above, we showed that for the log-likelihood we can exploit the structure of this matrix to find a scalable and exact solution. We will now show how to do the same thing for the sufficient statistics.

4.2.1 Solving for $E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]$

First, we reuse the matrix $-BD^{-1}$,

$$-BD^{-1} = -\left[\begin{array}{cccc}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}, & \ldots, & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\end{array}\right], \quad (171)$$

and its transpose,

$$-D^{-1}C = -\left[\begin{array}{c}\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} \\ \vdots \\ \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\end{array}\right]. \quad (172)$$

We then note that,

$$\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\left(\tilde{\mathbf{x}}_i-\tilde{\boldsymbol{\mu}}\right) = \left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}\left[\begin{array}{c}\sum_{j=1}^{J_i}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{ij} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i2} \\ \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3}\end{array}\right], \quad (173)$$

$$= \left[\begin{array}{cc}I_{D_F} & 0 \\ -D^{-1}C, & I_{J_iD_G}\end{array}\right]\left[\begin{array}{cc}\left(A-BD^{-1}C\right)^{-1} & 0 \\ 0 & D^{-1}\end{array}\right]\left[\begin{array}{cc}I_{D_F} & -BD^{-1} \\ 0 & I_{J_iD_F}\end{array}\right]\mathbf{b}. \quad (174)$$

We have already solved most of this so we take up the solution halfway through by noting that we need to find,

$$= \left[\begin{array}{cc}I_{D_F} & 0 \\ -D^{-1}C, & I_{J_iD_G}\end{array}\right]\left[\begin{array}{c}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3}\end{array}\right], \quad (175)$$

$$= \begin{bmatrix} I_{D_F} & & 0 \\ -\begin{bmatrix} \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} \\ \vdots \\ \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F} \end{bmatrix}, & \begin{bmatrix} I_{D_G} & 0 & 0 \\ 0 & I_{D_G} & 0 \\ 0 & 0 & I_{D_G} \end{bmatrix} \end{bmatrix} \begin{bmatrix} \left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{bmatrix}, \tag{176}$$

$$= \begin{bmatrix} \left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ -\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}+\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i1} \\ \vdots \\ -\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}+\left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}_{i3} \end{bmatrix}. \tag{177}$$

This provides us with a very interesting solution because we have now separated the factors and all the matrices we have to invert are small. The final solution is obviously,

$$E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right] = \begin{bmatrix} \left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij} \\ \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\left[\bar{\mathbf{x}}_{i1}-\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}\right] \\ \vdots \\ \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\left[\bar{\mathbf{x}}_{i3}-\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}\right] \end{bmatrix}. \tag{178}$$

This leads to,

$$E\left[\mathbf{h}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right] = \left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}, \tag{179}$$

$$E\left[\mathbf{w}_{ij}|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right] = \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\left[\bar{\mathbf{x}}_{ij}-\mathbf{F}\left(\mathbf{I}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\sum_{j=1}^{J_i}\mathbf{F}^T\mathbf{Q}\bar{\mathbf{x}}_{ij}\right], \tag{180}$$

$$= \left(\mathbf{I}+\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\left[\bar{\mathbf{x}}_{ij}-\mathbf{F}E\left[\mathbf{h}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]\right]. \tag{181}$$

4.2.2 Solving for $E\left[\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]$

To solve this we use the equation,

$$E\left[\tilde{\mathbf{y}}_i\tilde{\mathbf{y}}_i^T|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right] = \left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1} + E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]E\left[\tilde{\mathbf{y}}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\Theta}\right]^T. \tag{182}$$

This implies that we have to store the full matrix $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}$ whose size depends on the number of samples, however, in fact we use this on a per latent variable basis. That is, we find $\mathbf{y}_{ij} = [\mathbf{h}_i, \mathbf{w}_{ij}]$ for each $j = [1, 2, ..., J_i]$ and they are treated separately. This is because it is only used in Equation 71. This suggests that it might be possible to do this efficiently if we can find an expression for $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}$. We take this problem up below.

The matrix $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}$ can be defined, and calculated efficiently, by using the matrix inversion lemma. That is,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} \left(A - BD^{-1}C\right)^{-1} & -\left(A - BD^{-1}C\right)^{-1} BD^{-1} \\ -D^{-1}C\left(A - BD^{-1}C\right)^{-1}, & D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1} BD^{-1} \end{pmatrix}, \tag{183}$$

and

$$\left(\mathbf{I} + \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{A}}\right) = \begin{bmatrix} \mathbf{I} + J_i \mathbf{F}^T \mathbf{\Sigma}^{-1} \mathbf{F}, & \mathbf{F}^T \mathbf{\Sigma}^{-1} \mathbf{G}, & \cdots, & \mathbf{F}^T \mathbf{\Sigma}^{-1} \mathbf{G} \\ \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{F} & \mathbf{I} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{F} & \mathbf{0} & \mathbf{0} & \mathbf{I} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G} \end{bmatrix}. \tag{184}$$

We take some definitions from the appendix and reproduce them here.

$$\left(A - BD^{-1}C\right)^{-1} = \left(\mathbf{I}_{D_F} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1}, \tag{185}$$

$$-D^{-1}C\left(A - BD^{-1}C\right)^{-1} = -\begin{bmatrix} \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{F} \left(\mathbf{I}_{D_F} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1} \\ \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{F} \left(\mathbf{I}_{D_F} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1} \\ \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{F} \left(\mathbf{I}_{D_F} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1} \end{bmatrix}, \tag{186}$$

$$-\left(A - BD^{-1}C\right)^{-1} BD^{-1} = -\left[D^{-1}C\left(A - BD^{-1}C\right)^{-1}\right]^T, \tag{187}$$

and

$$\begin{aligned} &D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1} BD^{-1} \\ &\qquad = \begin{bmatrix} \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1} + \mathbf{RP}, & \mathbf{RP} & \mathbf{RP} \\ \mathbf{RP} & \ddots & \mathbf{RP} \\ \mathbf{RP} & \mathbf{RP} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1} + \mathbf{RP} \end{bmatrix}, \end{aligned} \tag{188}$$

where

$$\mathbf{P} = \mathbf{F}^T \mathbf{\Sigma}^{-1} \mathbf{G} \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1}, \tag{189}$$

and

$$\mathbf{R} = \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{F} \left(\mathbf{I}_{D_F} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1}. \tag{190}$$

From the above definitions we can now define our matrix $\left(\mathbf{I} + \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{A}}\right)^{-1}$ as being,

$$\begin{bmatrix} \left(\mathbf{I}_{D_F} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1} & -\begin{bmatrix} \mathbf{R}^T & \mathbf{R}^T & \mathbf{R}^T \end{bmatrix} \\ -\begin{bmatrix} \mathbf{R} \\ \mathbf{R} \\ \mathbf{R} \end{bmatrix} & \begin{bmatrix} \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1} + \mathbf{RP}, & \mathbf{RP} & \mathbf{RP} \\ \mathbf{RP} & \ddots & \mathbf{RP} \\ \mathbf{RP} & \mathbf{RP} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1} + \mathbf{RP} \end{bmatrix} \end{bmatrix}. \tag{191}$$

Using the above definition it is easy to note that if we want to work out $E\left[\mathbf{y}_{ij}\mathbf{y}_{ij}^T | \tilde{\mathbf{x}}_i, \mathbf{\Theta}\right]$ we would use the following subrepresentation of $\left(\mathbf{I} + \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{A}}\right)^{-1}$,

$$\begin{bmatrix} \mathbf{T}_{ul} & \mathbf{T}_{ur} \\ \mathbf{T}_{ll} & \mathbf{T}_{lr} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{I}_{D_F} + J_i \mathbf{F}^T \mathbf{Q} \mathbf{F}\right)^{-1} & -\mathbf{R}^T \\ -\mathbf{R} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T \mathbf{\Sigma}^{-1} \mathbf{G}\right)^{-1} + \mathbf{RP} \end{bmatrix}. \tag{192}$$

A Rules and Identities for Differentiation

In this section we provide some of the useful identities for differentiation.

$$\frac{\partial \mathbf{x}^T\mathbf{B}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B}+\mathbf{B}^T)\mathbf{x}, \tag{193}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}}{\partial \mathbf{X}} = -\mathbf{X}^{-T}\mathbf{a}\mathbf{b}^T\mathbf{X}^{-T}, \tag{194}$$

$$\frac{\partial \left(\det(\mathbf{X}^k)\right)}{\partial \mathbf{X}} = k\det(\mathbf{X}^k)\mathbf{X}^{-T}, \tag{195}$$

$$\frac{\partial \ln|\det(\mathbf{X})|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^T = (\mathbf{X}^T)^{-1}, \tag{196}$$

$$\frac{\partial (\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s})}{\partial \mathbf{s}} = -2\mathbf{A}^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}), \tag{197}$$

$$\frac{\partial (\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s})}{\partial \mathbf{A}} = -2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s})\mathbf{s}^T, \tag{198}$$

$$\frac{\partial (\mathbf{x}-\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{s})}{\partial \mathbf{s}} = -2\mathbf{W}(\mathbf{x}-\mathbf{s}). \tag{199}$$

Equations 197 and 198 are valid if the matrix W is symmetric. These identities come from the matrix cookbook [3].

B Matrix Identities

B.1 Inverse of a Product of Matrices

$$(AB)^{-1} = B^{-1}A^{-1}, \tag{200}$$

B.2 Block Matrix Inversion

The matrix inversion lemma which is an instance of using the Schur complement can be derived in several ways. First we present the solution that we are interested in below,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A-BD^{-1}C\right)^{-1} & -\left(A-BD^{-1}C\right)^{-1}BD^{-1} \\ -D^{-1}C\left(A-BD^{-1}C\right)^{-1}, & D^{-1}+D^{-1}C\left(A-BD^{-1}C\right)^{-1}BD^{-1} \end{bmatrix}. \tag{201}$$

This identity consists of several parts which occur frequently. These parts are,

$$\left(A-BD^{-1}C\right)^{-1}, \tag{202}$$

$$\left(A-BD^{-1}C\right)^{-1}BD^{-1}. \tag{203}$$

In addition to the above identity there is another useful intermediate form of this inversion identity. This intermediate form is obtained by using the LDU decomposition (B.4) and taking the inverse. We then use the identity given in Equation 200 to get,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \left(\begin{bmatrix} I, & BD^{-1} \\ 0, & I \end{bmatrix}\begin{bmatrix} A-BD^{-1}C, & 0 \\ 0, & D \end{bmatrix}\begin{bmatrix} I, & 0 \\ D^{-1}C, & I \end{bmatrix}\right)^{-1}, \tag{204}$$

$$= \begin{bmatrix} I, & 0 \\ D^{-1}C, & I \end{bmatrix}^{-1}\begin{bmatrix} A-BD^{-1}C, & 0 \\ 0, & D \end{bmatrix}^{-1}\begin{bmatrix} I, & BD^{-1} \\ 0, & I \end{bmatrix}^{-1}. \tag{205}$$

Tackling each term separately we derive their inverse by solving the equation $AA^{-1} = I$, where A^{-1} is the unknown. Using this we get the following,

$$\begin{bmatrix} I, & BD^{-1} \\ 0, & I \end{bmatrix}^{-1} = \begin{bmatrix} I, & -BD^{-1} \\ 0, & I \end{bmatrix}, \tag{206}$$

$$\begin{bmatrix} A - BD^{-1}C, & 0 \\ 0, & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A - BD^{-1}C\right)^{-1}, & 0 \\ 0, & D^{-1} \end{bmatrix}, \tag{207}$$

and

$$\begin{bmatrix} I & 0 \\ D^{-1}C, & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -D^{-1}C, & I \end{bmatrix}. \tag{208}$$

Finally this yields the intermediate form of,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -D^{-1}C, & I \end{bmatrix} \begin{bmatrix} \left(A - BD^{-1}C\right)^{-1}, & 0 \\ 0, & D^{-1} \end{bmatrix} \begin{bmatrix} I, & -BD^{-1} \\ 0, & I \end{bmatrix}. \tag{209}$$

B.3 Block LU Decomposition

For the LU decomposition we begin as follows. For the L part we want,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ x_1 & x_2 \end{bmatrix} = \begin{bmatrix} A + Bx_1 & Bx_2 \\ C + Dx_1 & Dx_2 \end{bmatrix}, \tag{210}$$

with $C + Dx_1 = 0$ and $Dx_2 = I$. To achieve this,

$$Dx_1 = -C, \tag{211}$$

$$x_1 = -D^{-1}C, \tag{212}$$

and

$$Dx_2 = I, \tag{213}$$

$$x_2 = D^{-1}. \tag{214}$$

First, we substitute for x_1, this would give us,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I, & 0 \\ -D^{-1}C, & x_2 \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C, & Bx_2 \\ C - DD^{-1}C, & Dx_2 \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C, & Bx_2 \\ 0, & Dx_2 \end{bmatrix}, \tag{215}$$

Second, we substitute for x_2, this would give us,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I, & 0 \\ -D^{-1}C, & D^{-1} \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C, & BD^{-1} \\ 0, & DD^{-1} \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C, & BD^{-1} \\ 0, & I \end{bmatrix}. \tag{216}$$

This would finally give us the LU decomposition.

B.4 Block LDU Decomposition

For the LDU decomposition, the following identity holds,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I, & BD^{-1} \\ 0, & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C, & 0 \\ 0, & D \end{bmatrix} \begin{bmatrix} I, & 0 \\ D^{-1}C, & I \end{bmatrix}. \tag{217}$$

This can be proven by direct computation, noting that,

$$\begin{bmatrix} I, & BD^{-1} \\ 0, & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C, & 0 \\ 0, & D \end{bmatrix} \begin{bmatrix} I, & 0 \\ D^{-1}C, & I \end{bmatrix} = \begin{bmatrix} I, & BD^{-1} \\ 0, & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C, & 0 \\ C, & D \end{bmatrix}, \tag{218}$$

$$\begin{bmatrix} I, & BD^{-1} \\ 0, & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C, & 0 \\ C, & D \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C + BD^{-1}C, & B \\ C, & D \end{bmatrix} = \begin{bmatrix} A, & B \\ C, & D \end{bmatrix}. \tag{219}$$

B.5 Square Root of a Matrix

In general, a matrix can have several square root. However, a positive-definite matrix has precisely one positive-definite square root. The square root of a matrix can be found by first diagonalising it (when this is possible). The process is described below, but can also be found in more detail here [http://en.wikipedia.org/wiki/Square_root_of_a_matrix (By diagonalization)]. If we consider a matrix M that has the following decomposition, D being diagonal,

$$M = ZDZ^{-1}, \tag{220}$$

and if we find the square root $D^{\frac{1}{2}}$ of the diagonal matrix D, which means that$D = D^{\frac{1}{2}}D^{\frac{1}{2}}$, then we can compute a square root $M^{\frac{1}{2}}$ of M as follows,

$$M^{\frac{1}{2}} = ZD^{\frac{1}{2}}Z^{-1}. \tag{221}$$

B.6 Log Determinant of a Diagonal Matrix

The log determinant of a diagonal matrix is easy to compute. Assuming a diagonal matrix Σ, we have,

$$\ln|\det(\Sigma)| = \ln|\prod_{i=1}^{I}\sigma_{ii}|. \tag{222}$$

We can change this product around to be the sum of a log. This gives us,

$$\ln|\det(\Sigma)| = \sum_{i=1}^{I}\ln|\sigma_{ii}|. \tag{223}$$

B.7 Log Determinant of a Symmetric Real Matrix

A symmetric real matrix is diagonalisable by orthognal matrix, i.e., given a real symmetric matrix A, Q^TAQ is diagonal for some orthogonal matrix Q. Furthermore, the log determinant of a symmetric real matrix can be found using its eigenvalue decomposition. Assuming such a symmetric real matrix A, its eigenvalue decomposition is given by,

$$A = Q\Lambda Q^T, \tag{224}$$

where Λ is the diagonal matrix containing the eigenvalue of A, and Q is orthogonal, which means that $QQ^T = I$. Futhermore,

$$\ln|\det(A)| = \ln|\det(Q\Lambda Q^T)| = \ln|\det(\Lambda)\det(QQ^T)| = \ln|\det(\Lambda)\det(I)| = \ln|\det(\Lambda)| \tag{225}$$

There is a built-in function called slogdet(), shipped with recent versions of SciPy/NumPy, which performs such a computation.

B.8 Inverse of a Block Diagonal Matrix

The inverse of a block diagonal matrix is a block diagonal matrix of the inverse blocks. That does not make a lot of sense so let us write it down,

$$\begin{bmatrix} A_1 & 0 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 & 0 \\ 0 & 0 & A_3 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & A_n \end{bmatrix}^{-1} = \begin{bmatrix} A_1^{-1} & 0 & 0 & 0 & 0 \\ 0 & A_2^{-1} & 0 & 0 & 0 \\ 0 & 0 & A_3^{-1} & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & A_n^{-1} \end{bmatrix}, \tag{226}$$

where $[A_1, A_2, A_3, \ldots, A_n]$ are a set of block matrices.

C Intermediate Solutions for PLDA: Gaussian Priors

In this appendix we provide several intermediate solutions and representations that we use for PLDA when we have Gaussian priors.

C.1 Matrix Inversion: $\left(\mathbf{I} + \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{A}}\right)^{-1}$

The matrix $\left(\mathbf{I} + \tilde{\mathbf{A}}^T \tilde{\mathbf{\Sigma}}^{-1} \tilde{\mathbf{A}}\right)^{-1}$ is central to many calculations for PLDA with a Gaussian prior. We provide a series of simplifications that can be used. We will tackle all of these in blocks or parts. To derive simplifications using this matrix we use its explicit form given by the block matrix inversion identity which we covered in the previous section.

For PLDA with a Gaussian prior the block matrix to invert is defined by the following,

$$A = \left[\mathbf{I}_{D_F} + J_i \mathbf{F}^T \mathbf{\Sigma}^{-1} \mathbf{F}\right], \tag{227}$$

$$B = \left[\begin{array}{ccc} \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}, & \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}, & \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G} \end{array}\right], \tag{228}$$

$$C = \begin{bmatrix} \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \end{bmatrix} = B^T, \tag{229}$$

$$D = \begin{bmatrix} \left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right) \end{bmatrix}, \tag{230}$$

$$D^{-1} = \begin{bmatrix} \left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{bmatrix}. \tag{231}$$

$$BD^{-1} = \left[\begin{array}{ccc} \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}, & \ldots, & \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{array}\right]. \tag{232}$$

$$BD^{-1}B^T = BD^{-1} \begin{bmatrix} \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \end{bmatrix}, \tag{233}$$

$$= \left[J_i \left(\mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\right)\right] = BD^{-1}C. \tag{234}$$

C.1.1 Finding $\left(A - BD^{-1}C\right)^{-1}$

We now define the inverted matrix $\left(A - BD^{-1}C\right)^{-1}$ as this is used in several places.

$$\left(A - BD^{-1}C\right)^{-1} = \left[\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{F} - J_i\left(\mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\right)\right]^{-1}, \tag{235}$$

$$= \left[\mathbf{I}_{D_F} + J_i\left(\mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{F} - \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\right)\right]^{-1}, \tag{236}$$

$$= \left[\mathbf{I}_{D_F} + J_i\mathbf{F}^T\left(\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\right)\mathbf{F}\right]^{-1}, \tag{237}$$

$$= \left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}. \tag{238}$$

Where,

$$\mathbf{Q} = \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}. \tag{239}$$

C.1.2 Finding $-D^{-1}C\left(A - BD^{-1}C\right)^{-1}$

We now need to find the final form for the matrix $-D^{-1}C\left(A - BD^{-1}C\right)^{-1}$. This is given by,

$$-D^{-1}C\left(A - BD^{-1}C\right)^{-1} = -D^{-1}C\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}, \tag{240}$$

$$= -D^{-1}\begin{bmatrix}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\end{bmatrix}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}, \tag{241}$$

$$= -D^{-1}\begin{bmatrix}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\end{bmatrix}, \tag{242}$$

$$= -\begin{bmatrix}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} & \dots & \mathbf{0}\\ \vdots & \ddots & \vdots\\ \mathbf{0} & \dots & \left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\\ \vdots\\ \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\end{bmatrix}, \tag{243}$$

$$= -\begin{bmatrix}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\\ \left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\\ \left(\mathbf{I}_{D_G} + \mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\end{bmatrix}. \tag{244}$$

C.1.3 Finding $D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1}BD^{-1}$

The last term that we are interested in is in the bottom right hand corner which is given by $D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1}BD^{-1}$. We have the following expression for this matrix,

$$D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1}BD^{-1} = D^{-1} - D^{-1}C\left(A - BD^{-1}C\right)^{-1}\left(-BD^{-1}\right), \tag{245}$$

$$= D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1}BD^{-1}, \tag{246}$$

$$= D^{-1} + \begin{bmatrix} \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} \\ \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} \\ \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} \end{bmatrix} BD^{-1}, \tag{247}$$

$$= D^{-1} + \begin{bmatrix} \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} \\ \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} \\ \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} \end{bmatrix}$$
$$\begin{bmatrix} \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}, & \ldots, & \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{bmatrix}, \tag{248}$$

We now need to make some substitutions. Let,

$$\mathbf{P} = \mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}, \tag{249}$$

$$\mathbf{R} = \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}, \tag{250}$$

Substituting these back in gives us,

$$D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1}BD^{-1} = D^{-1} + \begin{bmatrix} \mathbf{R} \\ \mathbf{R} \\ \mathbf{R} \end{bmatrix}\begin{bmatrix} \mathbf{P} & \mathbf{P} & \mathbf{P} \end{bmatrix}, \tag{251}$$

$$= D^{-1} + \begin{bmatrix} \mathbf{RP}, & \mathbf{RP}, & \mathbf{RP} \\ \mathbf{RP}, & \mathbf{RP}, & \mathbf{RP} \\ \mathbf{RP}, & \mathbf{RP}, & \mathbf{RP} \end{bmatrix}, \tag{252}$$

$$= \begin{bmatrix} \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{RP}, & \mathbf{RP}, & \mathbf{RP} \\ \mathbf{RP}, & \mathbf{RP}, & \mathbf{RP} \\ \mathbf{RP}, & \mathbf{RP}, & \mathbf{RP} \end{bmatrix}, \tag{253}$$

$$= \begin{bmatrix} \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} + \mathbf{RP} & \mathbf{RP} & \mathbf{RP} \\ \mathbf{RP} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} + \mathbf{RP} & \mathbf{RP} \\ \mathbf{RP} & \mathbf{RP} & \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} + \mathbf{RP} \end{bmatrix}. \tag{254}$$

We now define each entry by noting that,

$$\mathbf{RP} = \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}, \tag{255}$$

and

$$\left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} + \mathbf{RP} =$$
$$\left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1} + \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}, \tag{256}$$

$$= \left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\left[\mathbf{I}_{D_F} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}\left(\mathbf{I}_{D_F} + J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1}\mathbf{F}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G} + \mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G}\right)^{-1}\right]. \tag{257}$$

C.2 LDU Decomposition for $\left(\mathbf{I}+\tilde{\mathbf{A}}^T\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{A}}\right)^{-1}$

If we consider the LDU decomposition for the matrix inversion we need to define the subparts for three block matrices. These matrices are given in Equation 209,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -D^{-1}C, & I \end{bmatrix} \begin{bmatrix} \left(A-BD^{-1}C\right)^{-1}, & 0 \\ 0, & D^{-1} \end{bmatrix} \begin{bmatrix} I, & -BD^{-1} \\ 0, & I \end{bmatrix}. \tag{258}$$

We now define each of three matrices below,

$$\begin{bmatrix} I & 0 \\ -D^{-1}C, & I \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{D_F}, & \mathbf{0}_{(D_F,J_iD_G)} \\ -\begin{bmatrix} \left(\mathbf{I}_{D_G}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \\ \left(\mathbf{I}_{D_G}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \\ \left(\mathbf{I}_{D_G}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{F} \end{bmatrix} & \mathbf{I}_{J_iD_G} \end{bmatrix}, \tag{259}$$

$$\begin{bmatrix} \left(A-BD^{-1}C\right)^{-1}, & 0 \\ 0, & D^{-1} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{I}_{D_F}+J_i\mathbf{F}^T\mathbf{Q}\mathbf{F}\right)^{-1} & \mathbf{0}_{(D_F,J_iD_G)} \\ \mathbf{0}_{(J_iD_G,D_F)}, & \begin{bmatrix} \left(\mathbf{I}_{D_G}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \left(\mathbf{I}_{D_G}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{bmatrix} \end{bmatrix}, 1 \tag{260}$$

and

$$\begin{bmatrix} I, & -BD^{-1} \\ 0, & I \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{D_F}, & -\left[\begin{array}{ccc} \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}, & \dots, & \mathbf{F}^T\mathbf{\Sigma}^{-1}\mathbf{G}\left(\mathbf{I}_{D_G}+\mathbf{G}^T\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} \end{array}\right] \\ \mathbf{0}_{(J_iD_G,D_F)}, & \mathbf{I}_{J_iD_G} \end{bmatrix}. \tag{261}$$

References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[2] P. Li and S. J. D. Prince. *Advance in Face Image Analysis: Techniques and Technologies*, chapter Probabilistic Methods for Face Registration and Recognition (In Press). Idea Group Publishing, 2010.

[3] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook, Version: November 14, 2008*. Technical University of Denmark, 2008.

[4] S. J. D. Prince and J. H. Elder. Probabilistic linear discriminant analysis for inferences about identity. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007.