

Variational Bayesian Mixture of Factor Analysers

Angermueller Christof

August 31, 2014

Contents

1	Factor analysis	2
1.1	Joint distribution	2
1.2	Conditional distributions	2
1.3	Likelihood	2
2	Mixture of factor analysers	3
3	Variational Bayesian mixture of factor analysers	4
3.1	Prior distribution	4
3.2	Joint distribution	4
4	Variational approximation	6
4.1	$Q(\boldsymbol{\pi})$	6
4.2	$Q(\nu_q^s)$	6
4.3	$Q(\tilde{\Lambda}_p^s)$	7
4.4	$Q(\boldsymbol{x}^n s^n)$	8
4.5	$Q(s^n)$	8
	Appendices	9
A	Expectations	9

1 Factor analysis

The factor analysis model describes a P dimensional data vector \mathbf{y} as a linear function of a Q dimensional vector of latent factors \mathbf{x} plus some noise $\boldsymbol{\eta}$:

$$\mathbf{y} = \Lambda \mathbf{x} + \boldsymbol{\mu} + \boldsymbol{\eta} \quad (1)$$

The latent factors are assumed to be standard-normal distributed:

$$P(\mathbf{x}) = \mathcal{N}(\mathbf{x}|0, I_Q) \quad (2)$$

$\boldsymbol{\eta}$ is a normally distributed noise vector with mean zero and diagonal covariance matrix Ψ :

$$P(\boldsymbol{\eta}|\Psi) = \mathcal{N}(\boldsymbol{\eta}|0, \Psi) \quad \Psi = \begin{bmatrix} \Psi_{11} & \cdots & 0 \\ \vdots & \Psi_{ii} & \vdots \\ 0 & \cdots & \Psi_{PP} \end{bmatrix} \quad (3)$$

$\theta = \{\Lambda, \boldsymbol{\mu}, \Psi\}$ are the model parameters. Λ is the $P \times Q$ factor loading matrix, whose columns correspond to the eigenvector of the covariance matrix of \mathbf{y} . Given the latent factors \mathbf{x} , \mathbf{y} is hence normally distributed itself:

$$P(\mathbf{y}|\mathbf{x}, \Lambda, \boldsymbol{\mu}, \Psi) = \mathcal{N}(\mathbf{y}|\Lambda \mathbf{x} + \boldsymbol{\mu}, \Psi) \quad (4)$$

Integrating out \mathbf{x} gives:

$$P(\mathbf{y}|\Lambda, \boldsymbol{\mu}, \Psi) = \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}, \Lambda \Lambda^T + \Psi) \quad (5)$$

$\Lambda \Lambda^T + \Psi$ is therefore a low-rank approximation of the full covariance matrix of \mathbf{y} . The factor analysis model is identical to probabilistic Principle Component Analysis (PCA) if $\Psi = \sigma^2 I_P$. Factor analysis is therefore a generalisation for probabilistic PCA, which allows for non-isotropic noise. However, there is no longer a closed-form solution for the maximum likelihood estimator of Λ .

1.1 Joint distribution

$$P(\mathbf{x}, \mathbf{y}|\theta) = \mathcal{N}\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \mid \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} I_Q & \Lambda^T \\ \Lambda & \Lambda \Lambda^T + \Psi \end{bmatrix}\right) \quad (6)$$

1.2 Conditional distributions

By conditioning on the joint distribution 6, one obtains:

$$P(\mathbf{x}|\mathbf{y}, \theta) = \mathcal{N}(\mathbf{x}|\Lambda^T(\Lambda \Lambda^T + \Psi)(\mathbf{y} - \boldsymbol{\mu})) \quad (7)$$

$$P(\mathbf{y}|\mathbf{x}, \theta) = \mathcal{N}(\mathbf{y}|\Lambda \mathbf{x} + \boldsymbol{\mu}, \Psi) \quad (8)$$

1.3 Likelihood

The log-likelihood of the model parameters θ is:

$$l(\theta) = \log \prod_n P(\mathbf{y}^n|\theta) = \sum_n \log \mathcal{N}(\mathbf{y}^n|\boldsymbol{\mu}^s, \Lambda^s \Lambda^{sT} + \Psi) \quad (9)$$

In contrast to probabilistic PCA, there is no maximum likelihood solution for θ .

2 Mixture of factor analysers

In the mixture of factor analysers model, \mathbf{y} belongs to one of S independent factor analysers with different factor loading matrices $\Lambda = \{\Lambda^s\}_{s=1}^S$ and offset vectors $\boldsymbol{\mu} = \{\boldsymbol{\mu}^s\}_{s=1}^S$:

$$\mathbf{y} = \Lambda^s \mathbf{x} + \boldsymbol{\mu}^s + \boldsymbol{\eta} \quad (10)$$

The indicator s is unknown and therefore integrated out:

$$P(\mathbf{y}|\theta) = \sum_s P(s|\boldsymbol{\pi})P(\mathbf{y}|s, \theta) = \sum_s P(s|\boldsymbol{\pi})\mathcal{N}(\mathbf{y}|\boldsymbol{\mu}^s \Lambda^s \Lambda^{sT} + \Psi) \quad (11)$$

The prior $P(\mathbf{s}|\boldsymbol{\pi})$ is a single draw from a multinomial distribution with mixture-coefficients $\boldsymbol{\pi}$:

$$P(\mathbf{s}|\boldsymbol{\pi}) = \mathcal{M}(\mathbf{s}|1, \boldsymbol{\pi}) \quad (12)$$

The model parameters $\theta = \{\boldsymbol{\pi}, \Lambda, \boldsymbol{\mu}, \Psi\}$ can be estimated by maximum likelihood or Bayesian inference. Maximum likelihood estimation via Expectation Maximisation (EM) is fast, but subject to overfitting since it does not take into account model complexity. Bayesian inference prevents overfitting by imposing a prior on the model parameters, but is often intractable. To make inference tractable, approximate Bayesian inference methods can be used. The following sections will explain one of these methods, Variational Inference, which allows for tractable inference for a Bayesian mixture of factor analysers.

3 Variational Bayesian mixture of factor analysers

3.1 Prior distribution

Bayesian inference requires a prior distribution $P(\theta)$ over the model parameters $\theta = \{\boldsymbol{\pi}, \Lambda, \boldsymbol{\mu}\}$, which can be chosen as follows.

The prior over the mixture proportions $\boldsymbol{\pi}$ is a symmetric Dirichlet distribution with counts $\alpha^* \mathbf{m}^*$:

$$P(\boldsymbol{\pi} | \alpha^* \mathbf{m}^*) = \mathcal{D}(\boldsymbol{\pi} | \alpha^* \mathbf{m}^*), \quad (13)$$

The prior strengths α^* corresponds to the total number of prior counts, and \mathbf{m}^* to the prior distribution, which is supposed to be uniform: $\mathbf{m}^* = [1/S, \dots, 1/S]$. The expectation and variance is therefore:

$$\langle \pi_s \rangle = m_s \quad \text{var}[\pi_s] = \frac{m_s(1 - m_s)}{\alpha^*(\alpha^* + 1)} \quad (14)$$

Each factor loading matrix Λ^s factorizes by columns q , which are normally distributed with mean zero and diagonal covariance matrix $\nu_q^{s-1} I_P$. This is equivalent to saying Λ^s factorizes by rows p , which are normally distributed with mean zero and diagonal covariance matrix $\text{diag}[\boldsymbol{\nu}^s]^{-1}$.

$$P(\Lambda | \boldsymbol{\nu}) = \prod_{s,q} \mathcal{N}(\Lambda_{:,q}^s | 0, \nu_q^{s-1} I_P) \quad (15)$$

$$= \prod_{s,p} \mathcal{N}(\Lambda_p^s | 0, \text{diag}[\boldsymbol{\nu}^s]^{-1}) \quad (16)$$

The factor loading rates $\boldsymbol{\nu}^s$ are independently Gamma distributives with hyperparameter a^* and b^* :

$$P(\boldsymbol{\nu} | a^*, b^*) = \prod_{s,q} \text{Ga}(\nu_q^s | a^*, b^*) \quad (17)$$

The mean vector $\boldsymbol{\mu}^s$ is normally distributed with mean $\boldsymbol{\mu}^*$ and diagonal covariance $\text{diag}[\boldsymbol{\nu}^*]^{-1}$:

$$P(\boldsymbol{\mu} | \boldsymbol{\mu}^*, \boldsymbol{\nu}^*) = \prod_s \mathcal{N}(\boldsymbol{\mu}^s | \boldsymbol{\mu}^*, \text{diag}[\boldsymbol{\nu}^*]^{-1}) \quad (18)$$

In summary, the model parameters θ are

$$\theta = \{\boldsymbol{\pi}, \boldsymbol{\nu}, \Lambda, \boldsymbol{\mu}\}, \quad (19)$$

and the hyperparameters Θ are

$$\Theta = \{\Psi, \alpha^* \mathbf{m}^*, a^*, b^*, \boldsymbol{\mu}^*, \boldsymbol{\nu}^*\}. \quad (20)$$

3.2 Joint distribution

The data D includes the observed variables \mathbf{y}^n , as well as the latent variables \mathbf{x}^n and s^n for $n = 1, \dots, N$ independent samples:

$$D = \{\mathbf{y}^n, \mathbf{x}^n, s^n\}_{n=1}^N \quad (21)$$

The joint distribution of model parameters θ and data D factorises as follows:

$$P(\theta, D | \Theta) = P(\theta | \Theta) P(D | \theta, \Theta) \quad (22)$$

$$= P(\boldsymbol{\pi} | \alpha^* \mathbf{m}^*) P(\boldsymbol{\nu} | a^*, b^*) P(\Lambda | \boldsymbol{\nu}) P(\boldsymbol{\mu} | \boldsymbol{\mu}^*, \boldsymbol{\nu}^*) \quad (23)$$

$$\prod_n P(s^n | \boldsymbol{\pi}) P(\mathbf{x}^n) P(\mathbf{y}^n | \mathbf{x}^n, \Lambda^s, \boldsymbol{\mu}^s, \Psi) \quad (24)$$

The log joint distribution is:

$$\log P(\theta, D|\Theta) = \log P(\theta|\Theta) + \log P(D|\theta, \Theta) \quad (25)$$

$$= \log P(\boldsymbol{\pi}|\boldsymbol{\alpha}^* \mathbf{m}^*) + \log P(\boldsymbol{\nu}|a^*, b^*) + \log P(\Lambda|\boldsymbol{\nu}) + \log P(\boldsymbol{\mu}|\boldsymbol{\mu}^*, \boldsymbol{\nu}^*) \quad (26)$$

$$\sum_n \log P(s^n|\boldsymbol{\pi}) + \log P(\mathbf{x}^n) + \log P(\mathbf{y}^n|\mathbf{x}^n, \Lambda^s, \boldsymbol{\mu}^s, \Psi) \quad (27)$$

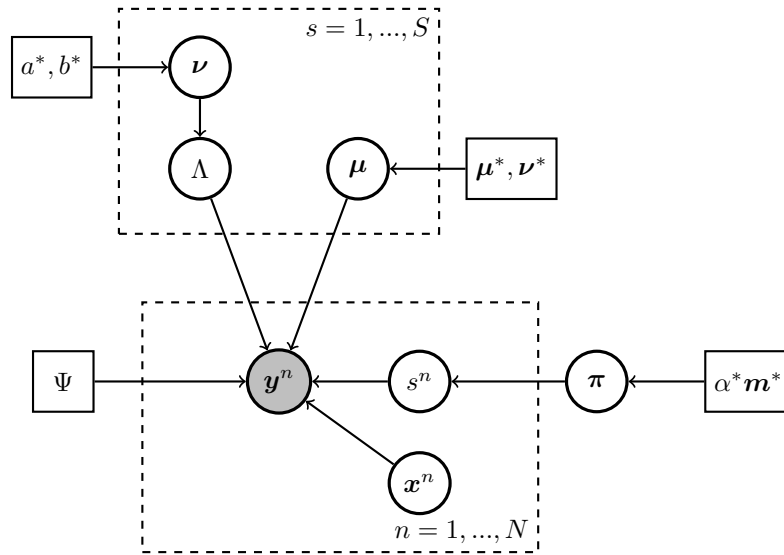


Figure 1: Graphical model

4 Variational approximation

The goal is to approximate the following posterior distribution:

$$P(\boldsymbol{\pi}, \boldsymbol{\nu}, \Lambda, \boldsymbol{\mu}, \mathbf{x}, \mathbf{s} | \mathbf{y}) \approx Q(\boldsymbol{\pi}, \boldsymbol{\nu}, \Lambda, \boldsymbol{\mu}, \mathbf{x}, \mathbf{s}) \quad (28)$$

To render tractable inference and closed for solutions for the approximating factors $Q(\cdot)$, the following factorisation is chosen:

$$Q(\boldsymbol{\pi}, \boldsymbol{\nu}, \Lambda, \boldsymbol{\mu}, \mathbf{x}, \mathbf{s}) = Q(\boldsymbol{\pi}, \boldsymbol{\nu})Q(\Lambda, \boldsymbol{\mu})Q(\mathbf{x}, \mathbf{s}) \quad (29)$$

The choice of this factorisation implies two approximations:

1. Parameters θ and latent variables $\{\mathbf{x}, \mathbf{s}\}$ are independent.
2. $\{\Lambda, \boldsymbol{\mu}\}$ are independent of the remaining parameters.

The optimal $Q(\cdot)$ can be found by computing the expectation of the log joint distribution (eq. 27):

$$\log Q_i \propto \log P(\theta, D | \Theta) >_{Q_{-i}} \quad (30)$$

This leads to the following factorization:

$$Q(\boldsymbol{\pi}, \boldsymbol{\nu})Q(\Lambda, \boldsymbol{\mu})Q(\mathbf{x}, \mathbf{s}) = Q(\boldsymbol{\pi}) \prod_s \prod_q Q(\nu_q^s) \prod_s \prod_p Q(\tilde{\Lambda}_p^s) \prod_n Q(\mathbf{x}^n | s^n) Q(s^n) \quad (31)$$

Here, $\tilde{\Lambda}^s$ is the concatenation of Λ^s and $\boldsymbol{\mu}^s$:

$$\tilde{\Lambda}^s := [\Lambda^s \boldsymbol{\mu}^s]. \quad (32)$$

The factors are distributed as follows:

$$Q(\boldsymbol{\pi}) = \mathcal{D}(\boldsymbol{\pi} | \alpha \mathbf{m}) \quad (33)$$

$$Q(\nu_q^s) = \mathcal{G}(\nu_q^s | a, b_q^s) \quad (34)$$

$$Q(\tilde{\Lambda}_p^s) = \mathcal{N}(\tilde{\Lambda}_p^s | \tilde{\Lambda}_p^s, \tilde{\Gamma}^{p,s}) \quad (35)$$

$$Q(\mathbf{x}^n | s^n) = \mathcal{N}(\mathbf{x}^n | \bar{\mathbf{x}}^{n,s}, \Sigma^s) \quad (36)$$

$$Q(s^n) \text{ is } N \times S \quad (37)$$

4.1 $Q(\boldsymbol{\pi})$

$$Q(\boldsymbol{\pi}) = \mathcal{D}(\boldsymbol{\pi} | \alpha \mathbf{m}) \quad (38)$$

$$\alpha m_s = \alpha^* m_s^* + \sum_n Q(s^n) \quad (39)$$

4.2 $Q(\nu_q^s)$

$$Q(\nu_q^s) = \mathcal{G}(\nu_q^s | a, b_q^s) \quad (40)$$

$$a = a^* + \frac{P}{2} \quad (41)$$

$$b_q^s = b^* + \frac{1}{2} \sum_p \langle \Lambda_{pq}^s \rangle^2 \quad (42)$$

$$= b^* + \frac{1}{2} \sum_p \left(\tilde{\Lambda}_{pq}^s \right)^2 + \tilde{\Gamma}_{qq}^{p,s} \quad (43)$$

Here, we solved $\langle \tilde{\Lambda}_{pq}^s \rangle^2$ by eq. 83.

4.3 $Q(\tilde{\Lambda}_p^s)$

$$Q(\tilde{\Lambda}_p^s) = \mathcal{N}(\tilde{\Lambda}_p^s | \tilde{\Lambda}_p^s, \tilde{\Gamma}^{p,s}) \quad (44)$$

The extended loading matrix $\tilde{\Lambda}^s$ is $P \times (Q + 1)$ and factorizes over rows:

$$\tilde{\Lambda}^s = \left[\begin{array}{ccc|c} \Lambda_{11}^s & \cdots & \Lambda_{1Q}^s & \mu_1^s \\ \vdots & \vdots & \vdots & \vdots \\ \Lambda_{p1}^s & \cdots & \Lambda_{pQ}^s & \mu_p^s \\ \vdots & \vdots & \vdots & \vdots \\ \Lambda_{P1}^s & \cdots & \Lambda_{PQ}^s & \mu_P^s \end{array} \right] \quad (45)$$

The row mean is $(Q + 1) \times 1$:

$$\tilde{\Lambda}_p^s = \begin{bmatrix} \tilde{\Lambda}_p^s \\ \tilde{\mu}_p^s \end{bmatrix} \quad (46)$$

$$\tilde{\Lambda}_p^s = \tilde{\Gamma}_{\Lambda\Lambda}^{p,s} \left(\Psi_{pp}^{-1} \sum_n Q(s^n) y_p^n < \mathbf{x}^n > \right) \quad (47)$$

$$= \tilde{\Gamma}_{\Lambda\Lambda}^{p,s} \left(\Psi_{pp}^{-1} \sum_n Q(s^n) y_p^n \bar{\mathbf{x}}^{n,s} \right) \quad (48)$$

$$\tilde{\mu}_p^s = \tilde{\Gamma}_{\mu\mu}^{p,s} \left(\Psi_{pp}^{-1} \sum_n Q(s^n) y_p^n + \nu_p^* \mu_p^* \right) \quad (49)$$

The row covariance matrix is $P \times (Q + 1)$ and decomposes into 4 blocks:

$$\tilde{\Gamma}^{p,s-1} = \begin{bmatrix} \Sigma_{\Lambda\Lambda}^{p,s-1} & \Sigma_{\Lambda\mu}^{p,s-1} \\ \Sigma_{\mu\Lambda}^{p,s-1} & \Sigma_{\mu\mu}^{p,s-1} \end{bmatrix} \quad (50)$$

$\Sigma_{\Lambda\Lambda}^{p,s}$ is $Q \times Q$:

$$\Sigma_{\Lambda\Lambda}^{p,s-1} = \text{diag}[< \boldsymbol{\nu}^s >] + \Psi_{pp}^{-1} \sum_n Q(s^n) < \mathbf{x}^n \mathbf{x}^{nT} > \quad (51)$$

$$= \text{diag}\left[\frac{a}{b_1^s}, \dots, \frac{a}{b_Q^s}\right] + \Psi_{pp}^{-1} \sum_n Q(s^n) \left(\bar{\mathbf{x}}^{n,s} \bar{\mathbf{x}}^{n,sT} + \Sigma^s \right) \quad (52)$$

Eq. 52 follows from eq. 89.

$\Sigma_{\Lambda\mu}^{p,s-1}$ is $Q \times 1$:

$$\Sigma_{\Lambda\mu}^{p,s-1} = \Psi_{pp}^{-1} \sum_n Q(s^n) < \mathbf{x}^n > \quad (53)$$

$$= \Psi_{pp}^{-1} \sum_n Q(s^n) \bar{\mathbf{x}}^{n,s} \quad (54)$$

$\Sigma_{\mu\mu}^{p,s}$ is 1×1 :

$$\Sigma_{\mu\mu}^{p,s-1} = \nu_p^* + \Psi_{pp}^{-1} \sum_n Q(s^n) \quad (55)$$

4.4 $Q(\mathbf{x}^n | s^n)$

$$Q(\mathbf{x}^n | s^n) = \mathcal{N}(\mathbf{x}^n | \bar{\mathbf{x}}^{n,s}, \Sigma^s) \quad (56)$$

$$\bar{\mathbf{x}}^{n,s} = \Sigma^s \langle \Lambda^{sT} \Psi^{-1} (\mathbf{y}^n - \boldsymbol{\mu}^s) \rangle \quad (57)$$

$$= \Sigma^s \left[\bar{\Lambda}^{sT} \Psi^{-1} (\mathbf{y}^n - \bar{\boldsymbol{\mu}}^s) - \mathbf{v} \right] \quad (58)$$

Eq. 58 follows from eq. 88 and \mathbf{v} is defined as

$$v_q = \text{tr}[\Psi^{-1} \text{cov}[\Lambda_{:q}^s, \boldsymbol{\mu}^s]] \quad (59)$$

$$= \sum_p \Psi_{pp}^{-1} \Sigma_{\Lambda \mu_q}^{p,s}. \quad (60)$$

Eq. 60 follows from $\text{cov}[\Lambda_{pq}^s, \mu_{p'}^s] = 0$ if $p \neq p'$ and $\text{cov}[\Lambda_{pq}^s, \mu_p^s] = \Sigma_{\Lambda \mu_q}^{p,s}$.

$$\Sigma^{s-1} = \langle \Lambda^{sT} \Psi^{-1} \Lambda^s \rangle + I \quad (61)$$

$\langle \Lambda^{sT} \Psi^{-1} \Lambda^s \rangle$ is solved by eq. 87 and $\text{cov}[\Lambda_{pi}^s, \Lambda_{p'j}^s] = 0$ for $p \neq p'$:

$$\langle \Lambda^{sT} \Psi^{-1} \Lambda^s \rangle_{ij} = \langle \Lambda_{:i}^{sT} \Psi^{-1} \Lambda_{:j}^s \rangle \quad (62)$$

$$= \langle \Lambda_{:i}^s \rangle^T \Psi^{-1} \langle \Lambda_{:j}^s \rangle + \sum_p \sum_{p'} \Psi_{pp'}^{-1} \text{cov}[\Lambda_{pi}^s, \Lambda_{p'j}^s] \quad (63)$$

$$= \langle \Lambda_{:i}^s \rangle^T \Psi^{-1} \langle \Lambda_{:j}^s \rangle + \sum_p \Psi_{pp}^{-1} \text{cov}[\Lambda_{pi}^s, \Lambda_{pj}^s] \quad (64)$$

$$= \bar{\Lambda}_{:i}^{sT} \Psi^{-1} \bar{\Lambda}_{:j}^s + \sum_p \Psi_{pp}^{-1} \Sigma_{\Lambda \Lambda}^{p,s} \quad (65)$$

4.5 $Q(s^n)$

$$\ln Q(s^n) = \psi(\alpha m_s) + \frac{1}{2} \ln |\Sigma^s| + \langle \ln P(\mathbf{y}^n | \mathbf{x}^n, s^n, \Lambda^s, \Psi) \rangle \quad (66)$$

$$\begin{aligned} &= \psi(\alpha m_s) + \frac{1}{2} \ln |\Sigma^s| \\ &\quad - \frac{1}{2} (\mathbf{y}^n - \bar{\Lambda}^s \bar{\mathbf{x}}^{n,s})^T \Psi^{-1} (\mathbf{y}^n - \bar{\Lambda}^s \bar{\mathbf{x}}^{n,s}) \\ &\quad - \frac{1}{2} \text{diag}[\Psi^{-1}]^T \boldsymbol{\sigma}^{s,n} \end{aligned} \quad (67)$$

Here, the following definitions were used:

$$\tilde{\mathbf{x}}^n = [\mathbf{x}^n \quad 1]^T \quad (68)$$

$$\boldsymbol{\sigma}^{s,n} = (\bar{\Lambda}^{s2} + C_{\bar{\Lambda}^s}) \text{diag}[\tilde{\Sigma}^s] + C_{\bar{\Lambda}^s} \bar{\mathbf{x}}^{n,s2} \quad (69)$$

$$C_{\bar{\Lambda}^s} = \begin{bmatrix} \text{diag}[\tilde{\Gamma}^{1,s}]^T \\ \dots \\ \text{diag}[\tilde{\Gamma}^{P,s}]^T \end{bmatrix} \quad (70)$$

The expectation in eq. 66 is:

$$\langle \ln P(\mathbf{y}^n | \mathbf{x}^n, s^n, \Lambda^s, \Psi) \rangle \quad (71)$$

$$= -\frac{1}{2} \langle (\mathbf{y}^n - \tilde{\Lambda}^s \tilde{\mathbf{x}}^n)^T \Psi^{-1} (\mathbf{y}^n - \tilde{\Lambda}^s \tilde{\mathbf{x}}^n) \rangle + \text{const} \quad (72)$$

$$= -\frac{1}{2} (\mathbf{y}^n - \bar{\Lambda}^s \bar{\mathbf{x}}^{n,s})^T \Psi^{-1} (\mathbf{y}^n - \bar{\Lambda}^s \bar{\mathbf{x}}^{n,s}) - \frac{1}{2} \text{tr}[\Psi^{-1} \text{var}[\tilde{\Lambda}^s \tilde{\mathbf{x}}^n]] + \text{const} \quad (73)$$

$$\text{var}[\tilde{\Lambda}^s \tilde{\mathbf{x}}^n] = \begin{bmatrix} \text{var}[\tilde{\Lambda}_1^{sT} \tilde{\mathbf{x}}^n] & \cdots & 0 \\ \cdots & \text{var}[\tilde{\Lambda}_p^{sT} \tilde{\mathbf{x}}^n] & \cdots \\ 0 & \cdots & \text{var}[\tilde{\Lambda}_p^{sT} \tilde{\mathbf{x}}^n] \end{bmatrix} \quad (74)$$

$$\text{var}[\tilde{\Lambda}_p^{sT} \tilde{\mathbf{x}}^n] = \sum_q \text{var}[\tilde{\Lambda}_{pq}^s \tilde{x}_q^n] \quad (75)$$

$$= \sum_q \langle \tilde{\Lambda}_{pq}^s \rangle^2 \langle \tilde{x}_q^n \rangle^2 - \langle \tilde{\Lambda}_{pq}^s \rangle^2 \langle \tilde{x}_q^n \rangle^2 \quad (76)$$

$$= \sum_q (\langle \tilde{\Lambda}_{pq}^s \rangle^2 + \text{var}[\tilde{\Lambda}_{pq}^s]) (\langle \tilde{x}_q^n \rangle^2 + \text{var}[\tilde{x}_q^n]) - \langle \tilde{\Lambda}_{pq}^s \rangle^2 \langle \tilde{x}_q^n \rangle^2 \quad (77)$$

$$= \sum_q \langle \tilde{\Lambda}_{pq}^s \rangle^2 \text{var}[\tilde{x}_q^n] + \langle \tilde{x}_q^n \rangle^2 \text{var}[\tilde{\Lambda}_{pq}^s] + \text{var}[\tilde{\Lambda}_{pq}^s] \text{var}[\tilde{x}_q^n] \quad (78)$$

$$= \sum_q \tilde{\Lambda}_{pq}^s \tilde{\Sigma}_{qq}^s + \tilde{x}_q^n \tilde{\Gamma}_{qq}^{p,s} + \tilde{\Gamma}_{qq}^{p,s} \tilde{\Sigma}_{qq}^s \quad (79)$$

$$= \sum_q (\tilde{\Lambda}_{pq}^s \tilde{\Sigma}_{qq}^s + \tilde{\Gamma}_{qq}^{p,s} \tilde{\Sigma}_{qq}^s + \tilde{x}_q^n \tilde{\Gamma}_{qq}^{p,s}) \quad (80)$$

$$\text{var}[\tilde{\Lambda}^s \tilde{\mathbf{x}}^n] = \text{diag}[\boldsymbol{\sigma}^{s,n}] \quad (81)$$

$$\text{tr}[\Psi^{-1} \text{var}[\tilde{\Lambda}^s \tilde{\mathbf{x}}^n]] = \text{diag}[\Psi^{-1}]^T \boldsymbol{\sigma}^{s,n} \quad (82)$$

Appendices

A Expectations

$$\langle \mathbf{x}^2 \rangle = \langle \mathbf{x} \rangle^2 + \text{var}[\mathbf{x}] \quad (83)$$

$$\langle \mathbf{x}^T A \mathbf{x} \rangle = \langle \mathbf{x} \rangle^T A \langle \mathbf{x} \rangle + \text{tr}[A \text{var}[\mathbf{x}]] \quad (84)$$

$$\langle \mathbf{x}^T A \mathbf{y} \rangle = \langle \mathbf{x} \rangle^T A \langle \mathbf{y} \rangle + \text{tr}[A \text{cov}[\mathbf{y}, \mathbf{x}]] \quad (85)$$

$$= \langle \mathbf{x} \rangle^T A \langle \mathbf{y} \rangle + \text{vec}[A]^T \text{vec}[\text{cov}[\mathbf{x}, \mathbf{y}]] \quad (86)$$

$$\langle X A Y \rangle = \langle X \rangle A \langle Y \rangle + V \quad (87)$$

$$V_{ij} = \text{tr}[A \text{cov}[Y_j, X_i]] = \text{vec}[A]^T \text{vec}[\text{cov}[X_i, Y_j]]$$

$$\langle X A \mathbf{y} \rangle = \langle X \rangle A \langle \mathbf{y} \rangle + \mathbf{v} \quad (88)$$

$$\mathbf{v}_i = \text{tr}[A \text{cov}[\mathbf{y}, X_i]] = \text{vec}[A]^T \text{vec}[\text{cov}[X_i, \mathbf{y}]]$$

$$\langle \mathbf{x} \mathbf{x}^T \rangle = \langle \mathbf{x} \rangle \langle \mathbf{x} \rangle^T + \text{cov}[\mathbf{x}] \quad (89)$$

If M and \mathbf{x} are independent:

$$\langle M \mathbf{x} \rangle = \langle M \rangle \langle \mathbf{x} \rangle \quad (90)$$

If \mathbf{x} and \mathbf{y} are independent:

$$\begin{aligned} \text{var}[\mathbf{x}^T \mathbf{y}] &= \sum_i \text{var}[x_i y_i] \\ &= \sum_i \langle x_i^2 \rangle \langle y_i^2 \rangle - \langle x_i \rangle^2 \langle y_i \rangle^2 \end{aligned} \quad (91)$$