Chapter 7

Summary and Outlook

7.1 Summary

In this thesis we present a theoretical framework to investigate prototype-based training prescriptions. The analysis is performed using concepts from statistical physics which allow for an exact mathematical description of the system in terms of characteristic quantities, so-called order parameters. The set of order parameters fully describes the training dynamics or equilibrium learning states. The model scenario we utilize to compare LVQ algorithms is described in Chapter 2.

In chapters 3 and 4, we examine on-line training and observe the learning dynamics of (unsupervised) VQ and (supervised) LVQ. The learning behavior is monitored in terms of the evolution of the order parameters described by ordinary differential equations. In the unsupervised data clustering analysis, we demonstrate the advantages of Neural Gas (NG) over Winner-Takes-All algorithms. First, NG can improve convergence speed in comparison to Winner-Takes-All for poor initialization of prototypes. Furthermore, NG achieves robustness with respect to initial conditions. However, the Neural Gas may still converge into local minima and does not always obtain the best possible quantization error.

Among the supervised LVQ schemes, we compare in detail the advantages and drawbacks of various window-based example selection schemes including LVQ 2.1, Learning From Mistakes (LFM), Generalized LVQ (GLVQ) and Robust Soft LVQ (RSLVQ). The sensitivity of the learning curves with respect to parameters is studied for all algorithms. Surprisingly, LVQ 2.1 produces the optimal linear decision boundary yielding optimal generalization ability. We also find learning plateaus in the learning stage in multi-prototype systems.

In chapters 5 and 6 we apply statistical physics of off-learning to analyse LVQ algorithms. Training is interpreted as a stochastic minimization of the cost function on the data set \( D \), where the formal temperature \( T \) controls the degree of randomness. We investigate the equilibrium properties of WTA and rank-based VQ systems using the high temperature limit and the so-called annealed approximation. We find that Neural Gas is more robust than WTA with respect to initialization, which
agrees well with the results obtained from on-line learning analysis. We find phase transitions in the learning phase: a critical number of examples must be presented to the system before it can identify any underlying structure within the data. The nature of transition is continuous in two-prototype systems and discontinuous in three-prototype systems. This is highly relevant from a practical point of view: any optimization strategy will ultimately fail completely if too few example data are available and metastable states for three-prototype systems may cause long delays in learning.

Finally, we analyse the cost function of LFM, LVQ 2.1 and RSLVQ in the high temperature limit. For this learning problem, the performance of LFM is unexpectedly poor compared to the optimum achievable error. Given properly chosen weight decay, LVQ 2.1 exhibits better generalization ability than LFM and RSLVQ for both two- and three-prototype systems. In three-prototype systems, we find continuous phase transitions between prototype configurations. We observe that critical sizes of the training set are required to effectively utilize all available prototypes and find substructures within the data. Treatment of systems with more prototypes may demonstrate additional phase transitions.

### 7.2 Relation between on-line and off-line analyses

While the general approaches of on-line and off-line analyses are technically different, the emphasis of both techniques highly complement each other. On-line analysis allows for investigation of convergence speed depending on learning parameters and initial conditions, which are essentially absent in off-line analysis. Furthermore it allows the study of heuristic practical LVQ prescriptions which are not based on cost functions. Additionally, unbounded cost functions such as LVQ 2.1 can lead to highly divergent behaviors which cannot be treated using the off-line analysis. From on-line analysis, we can approximate their non-trivial asymptotic learning behaviors. Alternatively, off-line analysis rigorously evaluates the characteristics of the cost function landscape in order to find all possible equilibrium states, without explicitly considering convergence times. The stable and metastable states uncover the presence of fixed points that may be encountered in on-line analysis.

Qualitatively, the resemblance between the findings of both methods is already apparent. Our on-line analysis of unsupervised learning in Chapter 3 reveals that learning slows down considerably for WTA with poor initialization and demonstrates the advantage of rank-based NG algorithms. In off-line learning, this corresponds to suboptimal equilibrium states discussed in Section 5.4.1 where trivial minima exist at large prototype lengths. We also demonstrate in off-line NG how
the rank parameter $\lambda$ changes the energy landscape, which will affect typical escape times from the suboptimal configuration. From both analyses we demonstrate that NG is more robust than WTA. In on-line learning, the asymptotic configurations of NG are independent of initial conditions. This is mirrored in off-line learning by the smooth energy landscape where the metastable states disappear, given sufficiently large $\lambda$.

In general, the results of both analyses become identical at long learning times $\tilde{\alpha} \to \infty$ for on-line and large training sets $\hat{\alpha} \to \infty$ for off-line training. We compare supervised problems in Chapter 4 and Chapter 6 and confirm identical results for RSLVQ algorithms. The dependence of asymptotic configurations on the initial conditions is explained by the degeneracy of the cost function minima. The comparison of LFM is more restricted, because the prototypes coincide in the aforementioned limits. We observe that its behavior at large $\tilde{\alpha}$ approximates that of large $\hat{\alpha}$.

In order to fully compare both approaches, several imposed limits have to be resolved. For instance, we assume a small learning rate in the on-line analysis. However this would correspond to low temperature off-line analysis which requires a full treatment of the replica method. In this work, this can only be approached with the annealed approximation. Conversely, off-line analysis allows investigation of learning from restricted data sets and we can observe distinct training and generalization errors at finite temperatures, see the annealed approximation results in Section 5.4.3. In on-line analysis, this corresponds to recycling examples from a finite data set which imposes correlations between training examples and the system. The subject of on-line learning for finite training sets in neural networks are investigated in, for example, (Barber and Sollich 1998, Rae et al. 1999).

The presence of symmetries produces learning behaviors which carry over to both on-line and off-line analysis. This appears during the specialization phase of prototypes in unsupervised learning. Permutation symmetry between two or more prototypes, i.e. equivalent configurations from exchange of prototypes, greatly inhibits the learning process in the thermodynamic limit. In on-line learning, prototypes require long learning times to escape from symmetry-induced fixed points, which result in plateaus in the learning curves. In off-line learning, symmetries create competition of states with varying entropy, where the unspecialized state is favored at small training sets. This in turn produces retarded learning, where specialization only occurs for training sets larger than a critical size. In supervised learning problems, permutation symmetries are broken by different class assignments to prototypes, but they remain among prototypes with the same class. Hence, learning plateaus and retarded learning are observed only in multi-prototype LVQ systems.
7.3 Outlook

While our analysis focuses on an idealized scenario for this training scheme, the results uncover underlying non-trivial effects which would carry on to any learning problems. Nevertheless, other effects may be discovered in more realistic models, which may actually become more relevant in practical situations. In this outlook, we explore several possibilities to analyse more complex models and provide more general learning characteristics.

We have shown in this thesis that it is possible to extend the model to accommodate arbitrarily many isotropic clusters, provided that the number of clusters is small compared to the dimensionality. However, we assume that the features of the data, i.e. the dimensions, are completely uncorrelated, which would be unreasonable in practical situations. An important step would include learning systems with anisotropic clusters. Along these lines, we can also interpret data with highly correlated components as non-isotropic clusters with uncorrelated components, e.g. after processing by principal component analysis, or other methods. This can be studied along with the selection of distance measures, which are critical to successful learning. A straightforward extension is weighing each dimension separately, leading to the so-called Relevance LVQ, see e.g. (Hammer and Villmann 2002). By providing additional order parameters, one can quantify the overlap of the relevance vectors to the discriminative orientation. The evolution of the overlap during learning is subsequently analysed by ordinary differential equations.

The statistical physics framework in this thesis will be useful to study various prototype-based learning methods, among others the popular Self Organising Maps (SOM) (Kohonen 1997). In principle, SOMs can be treated in a similar manner to our analysis of Neural Gas and would require systems with very many prototypes. While numerically these problems are solvable within our framework, the computations become exceedingly expensive, as the number of order parameters grows and higher dimensional numerical integrations are required. Alternative approaches may be called upon to analyse many-prototype systems, for instance by treating the prototypes as densities instead of individuals. These important extensions will provide meaningful insights into prototype-based learning schemes.
Appendix A

Statistics of the projections

For convenience, we combine the projections $h_S = w_S \cdot \xi$ and $b_\sigma = B_\sigma \cdot \xi$ defined in (3.10) into a $D$-dimensional vector, where $D = K + M$, as

$$
x = \begin{pmatrix} h_1^\mu & h_2^\mu & \ldots & h_K^\mu & b_1^\mu & b_2^\mu & \ldots & b_M^\mu \end{pmatrix}^T \quad (A.1)
$$

In our analysis of on-line learning, we assume that $\xi$ is statistically independent from $w_S$, because $\xi^\mu$ is uncorrelated to all previous data and $w_S^{\mu-1}$. Therefore we observe that $h_S$ and $b_\sigma$ become correlated Gaussian random quantities following the Central Limit Theorem and can be fully described by their first and second moments, i.e. its conditional averages $\mu_\sigma = \langle x \rangle_\sigma$ and conditional covariance matrix $C_\sigma = \langle x \cdot x^T \rangle_\sigma$. We compute these averages in the following.

A.1 First order statistics

We compute the averages of the components of $x$ as follows:

$$
\langle h_i \rangle_\sigma = \int_\mathbb{R}^N \xi \cdot w_i p(\xi|\sigma) d\xi = w_i \cdot \int_\mathbb{R}^N \xi p(\xi|\sigma) d\xi = w_i \cdot \ell_\sigma B_\sigma = \ell_\sigma R_i \sigma \quad (A.2)
$$

$$
\langle b_\tau \rangle_\sigma = \int_\mathbb{R}^N \xi \cdot B_\tau p(\xi|\sigma) d\xi = B_\tau \cdot \int_\mathbb{R}^N \xi p(\xi|\sigma) d\xi = B_\tau \cdot \ell_\sigma B_\sigma = \ell_\sigma T_\tau \sigma \quad (A.3)
$$

with $T_\tau \sigma = B_\tau \cdot B_\sigma$. To a large extent, we utilize orthonormal cluster center vectors, i.e. $B_\tau \cdot B_\sigma = \delta_\tau \sigma$ where $\delta$ is the Kronecker delta. The conditional first order moments $\mu_\sigma = \langle x \rangle_\sigma$ can be expressed in terms of order parameters as

$$
\mu = \ell_\sigma \begin{pmatrix} R_1 \sigma & R_2 \sigma & \ldots & R_K \sigma & T_1 \sigma & T_2 \sigma & \ldots & T_M \sigma \end{pmatrix}^T \quad (A.4)
$$
A.2 Second order statistics

To compute the conditional variance $\langle x_i x_m \rangle_\sigma - \langle x_i \rangle_\sigma \langle x_m \rangle_\sigma$ we first look at the average

$$\langle h_i h_j \rangle_\sigma = \left( \sum_{k=1}^{N} (w_i)_k (\xi)_k \right) \left( \sum_{l=1}^{N} (w_j)_l (\xi)_l \right)$$

$$= \left( \sum_{k=1}^{N} (w_i)_k (w_j)_k (\xi)_k (\xi)_l + \sum_{k=1}^{N} \sum_{l=1, l \neq k}^{N} (w_i)_k (w_j)_l (\xi)_k (\xi)_l \right)$$

$$= \sum_{k=1}^{N} (w_i)_k (w_j)_k \langle ((\xi)_k (\xi)_l) \rangle_\sigma + \sum_{k=1}^{N} \sum_{l=1, l \neq k}^{N} (w_i)_k (w_j)_l \ell^2_\sigma (B_\sigma)_k (B_\sigma)_l$$

$$= \sum_{k=1}^{N} (w_i)_k (w_j)_k (v_\sigma + \ell^2_\sigma (B_\sigma)_k (B_\sigma)_k) + \sum_{k=1}^{N} \sum_{l=1, l \neq k}^{N} (w_i)_k (w_j)_l \ell^2_\sigma (B_\sigma)_k (B_\sigma)_l$$

$$= \sum_{k=1}^{N} (w_i)_k (w_j)_k + \sum_{k=1}^{N} \sum_{l=1}^{N} (w_i)_k (w_j)_l (B_\sigma)_k (B_\sigma)_l$$

$$= \sum_{k=1}^{N} (w_i)_k (w_j)_k + \ell^2_\sigma (B_\sigma)_k (B_\sigma)_k$$

Here we exploit the following

$$\langle ((\xi)_k (\xi)_l) \rangle_\sigma = v_\sigma + \ell^2_\sigma (B_\sigma)_k (B_\sigma)_k$$

Hence we obtain the conditional second order moment, from Eqs. (A.5) and (A.2),

$$\langle h_i h_j \rangle_\sigma - \langle h_i \rangle_\sigma \langle h_j \rangle_\sigma = v_\sigma Q_{ij} + \ell^2_\sigma R_{ij} - \ell_\sigma R_{ij} \ell_\sigma R_{ij} = v_\sigma Q_{ij}$$

Analogously, we get the second order statistics of $b$ and the covariance as follows:

$$\langle h_i b_{\rho} \rangle_\sigma - \langle h_i \rangle_\sigma \langle b_{\rho} \rangle_\sigma = v_\sigma T_{\sigma \rho} + \ell^2_\sigma T_{\sigma \rho} - \ell_\sigma T_{\sigma \rho} \ell_\sigma T_{\sigma \rho} = v_\sigma T_{\sigma \rho}$$

$$\langle h_i b_{\tau} \rangle_\sigma - \langle h_i \rangle_\sigma \langle b_{\tau} \rangle_\sigma = v_\sigma R_{\sigma \tau} + \ell^2_\sigma R_{\sigma \tau} - \ell_\sigma R_{\sigma \tau} \ell_\sigma R_{\sigma \tau} = v_\sigma R_{\sigma \tau}$$

The conditional covariance matrix $C_\sigma = \langle x \cdot x^T \rangle_\sigma$ can be written in terms of order parameters as

$$C_\sigma = v_\sigma \begin{pmatrix} Q_{11} & \cdots & Q_{1K} & R_{11} & \cdots & R_{1M} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ Q_{K1} & \cdots & Q_{KK} & R_{K1} & \cdots & R_{KM} \\ R_{11} & \cdots & R_{K1} & T_{11} & \cdots & T_{1M} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ R_{1M} & \cdots & R_{KM} & T_{M1} & \cdots & T_{MM} \end{pmatrix}$$
Appendix B

Averages

B.1 Form of the Differential Equations

In order to perform the ordinary differential equations described in (3.14), we need to plug in the values of \( \langle f_S \rangle \), \( \langle x_n f_S \rangle \) and \( \langle f_S f_T \rangle \) (B.1)

Note that \( \langle f_S f_T \rangle \) is not required in the limit \( \eta \to 0 \), where terms proportional to \( \eta^2 \) can be neglected. We write the forms for the following algorithms: LVQ 2.1, LFM-W, GLVQ and RSLVQ.

LVQ 2.1

The general modulation function for LVQ 2.1 is described in Eq. (4.6) as

\[
f_S = \chi(c_S, y^\mu) \sum_{T: c_T \neq c_S} (\Theta_S^\delta - \Theta_T^\delta) \prod_{U \neq S, T} \Theta_U \Theta_T \]

with \( \chi(c_S, y^\mu) = 1 \) if \( c_S = y^\mu \) and \( \chi(c_S, y^\mu) = -1 \) else. We can rewrite

\[
\Theta_S^\delta = \Theta (d_T - d_S - \delta)
  = \Theta (-2w_T \cdot \xi^\mu + w_S^2 + w_S \cdot \xi^\mu - w_S^2 - \delta)
  = \Theta (-2h_T^S + 2h_S^T + Q_{TT} - Q_{SS} - \delta)
  = \Theta (\alpha_{ST} \cdot x - \beta_{ST}^\delta),
\]

with \( \alpha_{ST} = (0, \ldots, 2, \ldots, -2, \ldots, 0) \) at \( S \) and \( \beta_{ST}^\delta = Q_{SS} - Q_{TT} - \delta \) at \( T \).

For two prototype systems with labels \( w_S \) and \( w_T \), we can simplify the above as

\[
f_S = \chi(c_S, y^\mu) (\Theta_S^\delta - \Theta_T^\delta). \quad \text{(B.3)}
\]
And the required averages over the joint density (B.1) are calculated as

\[ \langle f_S \rangle = \langle \chi(c_S, y_\sigma^w) \left( \Theta_{ST}^\delta - \Theta_{ST}^{+\delta} \right) \rangle = \sum_{\sigma=1}^{M} p_\sigma \chi(c_S, y_\sigma) \langle \Theta_{ST}^\delta - \Theta_{ST}^{+\delta} \rangle_\sigma \]

\[ \langle x_n f_s \rangle = \sum_{\sigma=1}^{M} p_\sigma \chi(c_S, y_\sigma) \langle x_n (\Theta_{ST}^\delta - \Theta_{ST}^{+\delta}) \rangle_\sigma \]

\[ \langle f_s f_s \rangle = \left( \chi(c_S, y_\sigma^w) \right)^2 \langle \Theta_{ST}^\delta - \Theta_{ST}^{+\delta} \rangle_\sigma^2 = \sum_{\sigma=1}^{M} p_\sigma \langle \Theta_{ST}^\delta - \Theta_{ST}^{+\delta} \rangle_\sigma \]

\[ \langle f_s f_T \rangle = \chi(c_S, y_\sigma^w) \chi(c_T, y_\sigma^w) \langle \Theta_{ST}^\delta - \Theta_{ST}^{+\delta} \rangle_\sigma^2 = -\sum_{\sigma=1}^{M} p_\sigma \langle \Theta_{ST}^\delta - \Theta_{ST}^{+\delta} \rangle_\sigma \]

The quantities \( \langle (\Theta_{ST}^\delta - \Theta_{ST}^{+\delta}) \rangle_\sigma \) and \( \langle x_n (\Theta_{ST}^\delta - \Theta_{ST}^{+\delta}) \rangle_\sigma \) are calculated in Appendix B.2.

**LFM-W**

The general modulation function for LFM-W is described in Eq. (4.7) as

\[ f_S = \left\{ \begin{array}{ll}
\sum_{K:s_K \neq y_\sigma^w} (\Theta_{KS} - \Theta_{KS}^\dagger) \psi(S, K) & \text{if } c_S = y_\sigma^w \\
\sum_{J:s_J = y_\sigma^w} (\Theta_{SJ} - \Theta_{SJ}^\dagger) \psi(J, S) & \text{else.}
\end{array} \right. \quad \text{(B.5)} \]

with \( \psi(J, K) = \prod_{T:s_T = y_\sigma^w} \Theta_{JT} \prod_{U:s_U \neq y_\sigma^w} \Theta_{KU} \). With only two prototypes, both \( w_S \) and \( w_T \) are winners of their respective class, thus \( \psi(.) = 1 \) and the averages are

\[ \langle f_S \rangle = \sum_{\sigma, y_\sigma^w = c_S} p_\sigma \langle \Theta_{TS} - \Theta_{TS}^\dagger \rangle_\sigma + \sum_{\sigma, y_\sigma^w \neq c_S} p_\sigma \langle \Theta_{ST} - \Theta_{ST}^\dagger \rangle_\sigma \]

\[ \langle x_n f_s \rangle = \sum_{\sigma, y_\sigma^w = c_S} p_\sigma \langle x_n (\Theta_{TS} - \Theta_{TS}^\dagger) \rangle_\sigma + \sum_{\sigma, y_\sigma^w \neq c_S} p_\sigma \langle x_n (\Theta_{ST} - \Theta_{ST}^\dagger) \rangle_\sigma \quad \text{(B.6)} \]

**GLVQ**

The general modulation function for GLVQ is described in Eq. (4.13) as

\[ f_S = \left\{ \begin{array}{ll}
\sum_{K:s_K \neq y_\sigma^w} \left( \frac{2}{v_G} \phi \left( \frac{d_S - d_K}{v_G} \right) \right) \psi(S, K) & \text{if } c_S = y_\sigma^w \\
-\sum_{J:s_J = y_\sigma^w} \left( \frac{2}{v_G} \phi \left( \frac{d_J - d_S}{v_G} \right) \right) \psi(J, S) & \text{else.}
\end{array} \right. \quad \text{(B.7)} \]
For two prototypes,

\[
\langle f_S \rangle = \sum_{\sigma: y_\sigma = e_S} p_\sigma \frac{2}{v_G} \langle \phi (\alpha_{ST} \cdot x - \beta_{ST}) \rangle_\sigma - \sum_{\sigma: y_\sigma \neq e_S} p_\sigma \frac{2}{v_G} \langle \phi (\alpha_{ST} \cdot x - \beta_{ST}) \rangle_\sigma
\]

\[
\langle x_n f_S \rangle = \sum_{\sigma: y_\sigma = e_S} p_\sigma \frac{2}{v_G} \langle \phi (\alpha_{ST} \cdot x - \beta_{ST}) \rangle_\sigma - \sum_{\sigma: y_\sigma \neq e_S} p_\sigma \frac{2}{v_G} \langle \phi (\alpha_{ST} \cdot x - \beta_{ST}) \rangle_\sigma
\]

(B.8)

with \(\alpha_{ST} = \{\ldots, -\frac{1}{v_G}, \ldots, +\frac{1}{v_G}, \ldots, 0, 0\}\), \(\beta_{ST} = \frac{-Q_{SS} - Q_{TT}}{v_G}\).

The quantities \(\langle \phi (\alpha_{ST} \cdot x - \beta_{ST}) \rangle_\sigma\) are found in Eq. (B.26) in Appendix B.2.

RSLVQ

With one prototype representing each class, (4.15) become

\[
P_\sigma (S|\xi^\mu) = \frac{\exp \left( -\left(\xi^\mu - w_S^\mu \right)^2 / 2v_{soft} \right)}{\exp \left( -\left(\xi^\mu - w_S^\mu \right)^2 / 2v_{soft} \right)} = 1
\]

\[
P(S|\xi^\mu) = \frac{\exp \left( -\left(\xi^\mu - w_S^\mu \right)^2 / 2v_{soft} \right)}{\sum_{T=1}^{K} \exp \left( -\left(\xi^\mu - w_T^\mu \right)^2 / 2v_{soft} \right)}
\]

\[
= \frac{1}{1 + \sum_{T \neq S}^{K} \exp \left( \frac{1}{2v_{soft}} \left( -2\xi^\mu w_S^\mu + \left( w_S^\mu \right)^2 + 2\xi^\mu w_T^\mu - \left( w_T^\mu \right)^2 \right) \right)}
\]

\[
= \frac{1}{1 + \sum_{T \neq S}^{K} \exp \left( \frac{1}{2v_{soft}} \left( -2h_S + Q_{SS} + 2h_T - Q_{TT} \right) \right)}
\]

(B.9)

where we defined

\[
\alpha_{ST} = \{\ldots, \frac{1}{v_{soft}} \ldots, +\frac{1}{v_{soft}} \ldots, 0, 0\}, \quad \beta_{ST} = \frac{-Q_{SS} - Q_{TT}}{2v_{soft}}
\]

Therefore the RSLVQ modulation function becomes

\[
f_S = \frac{1}{v_{soft}} (\delta(c_S, y_\mu^S) - \Omega_S)
\]

(B.10)
where $\delta(x, y)$ is the Kronecker delta and

$$
\Omega_S = \frac{1}{1 + \sum_{T \neq S}^K \exp (\alpha_{ST} \cdot x - \beta_{ST})}
$$

We obtain the averages

$$
\langle f_S \rangle = \frac{1}{v_{soft}} \langle \delta(c_S, y_\sigma) - \Omega_S \rangle = \frac{1}{v_{soft}} \left( \sum_{\sigma, y_\sigma = c_S} p_\sigma - \sum_\sigma p_\sigma \langle \Omega_S \rangle_\sigma \right)
$$

$$
\langle x_n f_S \rangle = \begin{cases} 
\frac{1}{v_{soft}} \left( \sum_{\sigma, y_\sigma = c_S} p_\sigma \langle h_n \rangle_\sigma - \sum_\sigma p_\sigma \langle x_n \Omega_S \rangle_\sigma \right) & \text{if } n \leq K \\
\frac{1}{v_{soft}} \left( \sum_{\sigma, y_\sigma = c_S} p_\sigma \langle b_{n-K} \rangle_\sigma - \sum_\sigma p_\sigma \langle x_n \Omega_S \rangle_\sigma \right) & \text{if } n > K
\end{cases}
$$

The required quantities $\langle \Omega_S \rangle_\sigma$ and $\langle x_n \Omega_S \rangle_\sigma$ are supplied in Appendix B.2.
B.2 Gaussian Averages

B.2.1 Two prototypes

For generic functions $f_{ab} \equiv f(\alpha_{ab} \cdot x - \beta_{ab})$, the quantities $\langle f_{ab} \rangle_{\sigma}$ and $\langle x_n f_{ab} \rangle_{\sigma}$ are computed as follows:

$$
\langle f_{ab} \rangle_{\sigma} = \frac{1}{(2\pi)^{D/2}(\det(C_{\sigma}))^{1/2}} \int_{\mathbb{R}^D} f(\alpha_{ab} \cdot x - \beta_{ab}) \\
\times \exp \left( -\frac{1}{2}(x - \mu_{\sigma})^T C_{\sigma}^{-1}(x - \mu_{\sigma}) \right) \, dx
$$

$$
= \frac{1}{(2\pi)^{D/2}(\det(C_{\sigma}))^{1/2}} \int_{\mathbb{R}^D} f(\alpha_{ab} \cdot x' + \alpha_{ab} \cdot \mu_{\sigma} - \beta_{ab}) \\
\times \exp \left( -\frac{1}{2}(x')^T C_{\sigma}^{-1}(x') \right) \, dx',
$$

(B.13)

with the substitution $x' = x - \mu_{\sigma}$. Because the covariance matrix $C_{\sigma}$ is positive definite, $C_{\sigma}^{1/2}$ exists. Defining $x' = C_{\sigma}^{1/2} y$, we obtain $x'^T C_{\sigma}^{-1} x' = y^2$, $dx' = (\det C_{\sigma})^{1/2} dy$ and

$$
\langle f_{ab} \rangle_{\sigma} = \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} f(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \exp \left( -\frac{1}{2} y^2 \right) \, dy
$$

(B.14)

with $\tilde{\beta}_{ab,\sigma} = \alpha_{ab} \cdot \mu_{\sigma} - \beta_{ab}$. Since $\exp(-\frac{1}{2}y^2)$ has rotational invariance, it is possible to rotate the orthonormal coordinate system so that one axis, say $\tilde{y}$, is aligned with vector $\alpha_{ab} C_{\sigma}^{1/2}$. The remaining $(D - 1)$ dimensions can be integrated over with $\int \exp(-\frac{1}{2}y^2) dy = \sqrt{2\pi}$ and we obtain

$$
\langle f_{ab} \rangle_{\sigma} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\|\alpha_{ab} C_{\sigma}^{1/2}\| \tilde{y} + \tilde{\beta}_{ab,\sigma}) \exp \left( -\frac{1}{2} \tilde{y}^2 \right) \, d\tilde{y}
$$

$$
= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\tilde{\alpha}_{ab,\sigma} \tilde{y} + \tilde{\beta}_{ab,\sigma}) \exp \left( -\frac{1}{2} \tilde{y}^2 \right) \, d\tilde{y}
$$

(B.15)

with $\tilde{\alpha}_{ab,\sigma} = \|\alpha_{ab} C_{\sigma}^{1/2}\|$. Next we calculate the quantity

$$
\langle x_n f_{ab} \rangle_{\sigma} = \frac{1}{(2\pi)^{D/2}(\det(C_{\sigma}))^{1/2}} \int_{\mathbb{R}^D} x_n f(\alpha_{ab} \cdot x - \beta_{ab}) \\
\exp \left( -\frac{1}{2}(x - \mu_{\sigma})^T C_{\sigma}^{-1}(x - \mu_{\sigma}) \right) \, dx
$$

$$
= \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} (C_{\sigma}^{1/2} y + \mu_{\sigma})_n f(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \exp \left( -\frac{1}{2} y^2 \right) \, dy
$$
B. Averages

\[
\frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} \left( C_{\sigma}^{1/2} y \right)_n f \left( \alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma} \right) \exp \left( -\frac{1}{2} y_j^2 \right) dy
+ (\mu_{\sigma})_n \langle f_{ab} \rangle_{\sigma}
\]

\[
= \frac{1}{(2\pi)^{D/2}} I + (\mu_{\sigma})_n \langle f_{ab} \rangle_{\sigma} \tag{B.16}
\]

where \( I \) is an integral to be computed. Consider \( I = (1/(2\pi)^{D/2}) \sum_{j=1}^{D} I_j \), we calculate

\[
I_j = \int_{\mathbb{R}} \left( C_{\sigma}^{1/2} \right)_n (y)_j f \left( \alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma} \right) \exp \left( -\frac{1}{2} (y_j)^2 \right) d(y)_j \tag{B.17}
\]

by applying integration by parts \( \int u dv = uv - \int v du \) with

\[
u = \left( -\left( C_{\sigma}^{1/2} \right)_n \right) \exp \left( -\frac{1}{2} (y_j)^2 \right)
\]

\[
\begin{align*}
\text{du} &= \left( \alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma} \right) \left( \frac{\partial}{\partial (y_j)} \alpha_{ab} C_{\sigma}^{1/2} y \right) d(y)_j \\
&= \sum_{i=1}^{D} (\alpha_{ab})_i \left( C_{\sigma}^{1/2} \right) ij f' \left( \alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma} \right) \exp \left( -\frac{1}{2} (y_j)^2 \right) d(y)_j
\end{align*}
\]

\[
\text{dv} = \left( C_{\sigma}^{1/2} \right)_n (y)_j \exp \left( -\frac{1}{2} (y_j)^2 \right) d(y)_j. \tag{B.18}
\]

Hence we have,

\[
I_j = \left[ (-) f \left( \alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma} \right) \left( C_{\sigma}^{1/2} \right)_n \jmath \exp \left( -\frac{1}{2} (y_j)^2 \right) \right]_{-\infty}^{\infty}
- \left[ (-) \int_{\mathbb{R}} \left( C_{\sigma}^{1/2} \right)_n \sum_{i=1}^{D} (\alpha_{ab})_i \left( C_{\sigma}^{1/2} \right) ij f' \left( \alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma} \right) \exp \left( -\frac{1}{2} (y_j)^2 \right) d(y)_j \right]
= \left( C_{\sigma}^{1/2} \right)_n \sum_{i=1}^{D} (\alpha_{ab})_i \left( C_{\sigma}^{1/2} \right) ij \int_{\mathbb{R}} f' \left( \alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma} \right) \exp \left( -\frac{1}{2} (y_j)^2 \right) d(y)_j. \tag{B.19}
\]
Summing all \( j \), we obtain

\[
I = \frac{1}{(2\pi)^{D/2}} \sum_{j=1}^{D} (C_{\sigma}^{1/2})_{nj} \sum_{i=1}^{D} (\alpha_{ab})_{ij} (C_{\sigma}^{1/2})_{ij} \int f'(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \exp \left( -\frac{1}{2} y^2 \right) dy
\]

\[
= \frac{1}{(2\pi)^{D/2}} (C_{\sigma} \alpha_{ab}) n \int_{\mathbb{R}^D} f'(\tilde{\alpha}_{ab,\sigma} y + \tilde{\beta}_{ab,\sigma}) \exp \left( -\frac{1}{2} y^2 \right) dy
\]

(B.20)

Rotating the coordinate system similar to Eq. (B.15), we obtain the final form

\[
\langle x_n f_{ab} \rangle_{\sigma} = \frac{(C_{\sigma} \alpha_{ab}) n}{\sqrt{2\pi}} \int_{\mathbb{R}^D} f'(\tilde{\alpha}_{ab,\sigma} y + \tilde{\beta}_{ab,\sigma}) \exp \left( -\frac{1}{2} y^2 \right) dy + (\mu_{ab})_{n} \langle f_{ab} \rangle_{\sigma} .
\]

(B.21)

**LVQ 2.1, LFM-W**

The following quantities are required for two prototype LVQ2.1 and LFM-W:

\[
\langle \Theta_{ab}^\delta - \Theta_{ab}^\gamma \rangle_{\sigma} \quad \text{and} \quad \langle x_n (\Theta_{ab}^\delta - \Theta_{ab}^\gamma) \rangle_{\sigma} .
\]

Averages of the form \( \langle \Theta_{ab} \rangle_{\sigma} \) and \( \langle x_n \Theta_{ab} \rangle_{\sigma} \) can be performed analytically; we refer to (Biehl et al. 2004) for the details and calculations and write down the final forms:

\[
\langle \Theta_{ab} \rangle_{\sigma} = \Phi \left( \frac{\tilde{\beta}_{ab,\sigma}}{\alpha_{ab,\sigma}} \right),
\]

(B.22)

\[
\langle x_n \Theta_{ab} \rangle_{\sigma} = \frac{(C_{\sigma} \alpha_{ab}) n}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\tilde{\beta}_{ab,\sigma}}{\alpha_{ab,\sigma}} \right)^2 \right] + (\mu_{ab})_{n} \Phi \left( \frac{\tilde{\beta}_{ab,\sigma}}{\alpha_{ab,\sigma}} \right)
\]

(B.23)

where \( \Phi(\cdot) \) is the standard normal cumulative distribution function. Using the window rule, we obtain the required averages for LVQ 2.1 and LFM:

\[
\langle \Theta_{ab}^\delta - \Theta_{ab}^\gamma \rangle_{\sigma} = \Phi \left( \frac{\tilde{\beta}_{ab,\sigma}^\delta}{\alpha_{ab,\sigma}} \right) - \Phi \left( \frac{\tilde{\beta}_{ab,\sigma}^\gamma}{\alpha_{ab,\sigma}} \right)
\]

\[
\langle x_n (\Theta_{ab}^\delta - \Theta_{ab}^\gamma) \rangle_{\sigma} = \frac{(C_{\sigma} \alpha_{ab}) n}{\sqrt{2\pi}} \left\{ \exp \left[ -\frac{1}{2} \left( \frac{\tilde{\beta}_{ab,\sigma}^\delta}{\alpha_{ab,\sigma}} \right)^2 \right] - \exp \left[ -\frac{1}{2} \left( \frac{\tilde{\beta}_{ab,\sigma}^\gamma}{\alpha_{ab,\sigma}} \right)^2 \right] \right\}
\]

\[
+ (\mu_{ab})_{n} \Phi \left( \frac{\tilde{\beta}_{ab,\sigma}^\delta}{\alpha_{ab,\sigma}} \right) - \Phi \left( \frac{\tilde{\beta}_{ab,\sigma}^\gamma}{\alpha_{ab,\sigma}} \right).
\]

(B.24)
GLVQ

For GLVQ, the quantities $\langle \phi_{ab} \rangle_\sigma$ and $\langle x_n \phi_{ab} \rangle_\sigma$ are required to compute Eq. (B.8). From Eq. (B.15), we calculate the averages as follows:

$$\langle \phi_{ab} \rangle_\sigma = \frac{1}{\sqrt{2\pi}} \int_\mathbb{R} \phi(\tilde{\alpha}_{ab,\sigma} \tilde{y} + \tilde{\beta}_{ab,\sigma}) \exp\left( -\frac{1}{2} \tilde{y}^2 \right) d\tilde{y}$$

$$= \frac{1}{\sqrt{2\pi}} \int_\mathbb{R} \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} \left( \tilde{\alpha}_{ab,\sigma} \tilde{y} + \tilde{\beta}_{ab,\sigma} \right)^2 \right) \exp\left( -\frac{1}{2} \tilde{y}^2 \right) d\tilde{y}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} \tilde{\beta}_{ab,\sigma}^2 \right) \int_\mathbb{R} \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} (\tilde{\alpha}_{ab,\sigma} + 1) \tilde{y}^2 - (\tilde{\alpha}_{ab,\sigma} \tilde{\beta}_{ab,\sigma}) \tilde{y} \right) d\tilde{y}.$$  \hspace{1cm} \text{(B.25)}

We use the substitution $\int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} ax^2 + bx \right) = \frac{1}{\sqrt{a}} \exp\left( \frac{b^2}{2a} \right)$ to obtain

$$\langle \phi_{ab} \rangle_\sigma = \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} \tilde{\beta}_{ab,\sigma}^2 \right) \frac{1}{\sqrt{\tilde{\alpha}_{ab,\sigma}^2 + 1}} \exp\left( \frac{\tilde{\alpha}_{ab,\sigma}^2 \tilde{\beta}_{ab,\sigma}^2}{2(\tilde{\alpha}_{ab,\sigma}^2 + 1)} \right)$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\tilde{\alpha}_{ab,\sigma}^2 + 1}} \exp\left( -\frac{1}{2} \tilde{\beta}_{ab,\sigma}^2 \left( 1 - \frac{\tilde{\alpha}_{ab,\sigma}^2}{\tilde{\alpha}_{ab,\sigma}^2 + 1} \right) \right)$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\tilde{\alpha}_{ab,\sigma}^2 + 1}} \exp\left( -\frac{1}{2} \tilde{\beta}_{ab,\sigma}^2 \right).$$  \hspace{1cm} \text{(B.26)}

The remaining average to be computed is

$$\langle x_n \phi_{ab} \rangle_\sigma = \frac{(C_n \alpha_{ab} n)}{\sqrt{2\pi}} \int_\mathbb{R} \phi'(\tilde{\alpha}_{ab,\sigma} \tilde{y} + \tilde{\beta}_{ab,\sigma}) \exp\left( -\frac{1}{2} \tilde{y}^2 \right) d\tilde{y} + (\mu_n) \langle \phi_{ab} \rangle_\sigma.$$  \hspace{1cm} \text{(B.27)}

Using $\frac{\partial}{\partial \tilde{y}} \exp(-y^2/2) = -y \exp(-y^2/2)$, we obtain

$$\langle x_n \phi_{ab} \rangle_\sigma = \frac{(C_n \alpha_{ab} n)}{\sqrt{2\pi}} \int_\mathbb{R} -\left( \tilde{\alpha}_{ab,\sigma} \tilde{y} + \tilde{\beta}_{ab,\sigma} \right) \phi' \left( \tilde{\alpha}_{ab,\sigma} \tilde{y} + \tilde{\beta}_{ab,\sigma} \right) \exp\left( -\frac{1}{2} \tilde{y}^2 \right) d\tilde{y}$$

$$+ (\mu_n) \langle \phi_{ab} \rangle_\sigma$$

$$= -\frac{(C_n \alpha_{ab} n)}{\sqrt{2\pi} \alpha_{ab,\sigma}} \int_\mathbb{R} \left( z + \tilde{\beta}_{ab,\sigma} \right) \phi \left( z + \tilde{\beta}_{ab,\sigma} \right) \exp\left( -\frac{1}{2} \tilde{y}^2 \right) d\tilde{y}$$

$$+ (\mu_n) \langle \phi_{ab} \rangle_\sigma$$

(with $z = \tilde{\alpha}_{ab,\sigma} \tilde{y}$ and $dz = \tilde{\alpha}_{ab,\sigma} d\tilde{y}$).
\[ B.2. \text{Gaussian Averages} \]

\[
\langle x_n \phi_{ab} \rangle_{\sigma} = -\frac{(C_\sigma \alpha_{ab})_n}{\sqrt{2\pi \tilde{\alpha}_{ab,\sigma}}} \exp \left( -\frac{1}{2} \tilde{\beta}_{ab,\sigma} \right) \left[ \int_{\mathbb{R}} \frac{z + \tilde{\beta}_{ab,\sigma}}{\sqrt{2\pi}} \exp \left( \frac{1}{2} \frac{z^2}{\tilde{\alpha}_{ab,\sigma}} - z \tilde{\beta}_{ab,\sigma} \right) dz \right] + (\mu_\sigma)_n \langle \phi_{ab} \rangle_{\sigma}. \tag{B.28}
\]

Now we use the substitutions

\[
\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} ax^2 + bx \right) = \frac{b}{\sqrt{a}} \exp \left( \frac{b^2}{2a} \right)
\]

\[
\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} ax^2 + bx \right) = \frac{1}{\sqrt{a}} \exp \left( \frac{b^2}{2a} \right).
\]

Hence,

\[
\langle x_n \phi_{ab} \rangle_{\sigma} = -\frac{(C_\sigma \alpha_{ab})_n}{\sqrt{2\pi \tilde{\alpha}_{ab,\sigma}}} \exp \left( -\frac{1}{2} \tilde{\beta}_{ab,\sigma} \right) \exp \left( -\frac{\tilde{\beta}_{ab,\sigma}^2}{2 \left( 1 + 1/\tilde{\alpha}_{ab,\sigma}^2 \right)} \right)
\]

\[
\times \left[ -\frac{\tilde{\beta}_{ab,\sigma}}{\sqrt{1 + 1/\tilde{\alpha}_{ab,\sigma}^2}} + \frac{\tilde{\beta}_{ab,\sigma}}{\sqrt{1 + 1/\tilde{\alpha}_{ab,\sigma}^2}} \right] + (\mu_\sigma)_n \langle \phi_{ab} \rangle_{\sigma}, \tag{B.29}
\]

and we obtain the final form

\[
\langle x_n \phi_{ab} \rangle_{\sigma} = -\frac{(C_\sigma \alpha_{ab})_n}{\sqrt{2\pi \tilde{\alpha}_{ab,\sigma}}} \exp \left( -\frac{\tilde{\beta}_{ab,\sigma}^2}{2 (1 + 1/\tilde{\alpha}_{ab,\sigma}^2)} \right)
\]

\[
\times \left[ -\frac{\tilde{\beta}_{ab,\sigma}}{\sqrt{1 + 1/\tilde{\alpha}_{ab,\sigma}^2}} + \frac{\tilde{\beta}_{ab,\sigma}}{\sqrt{1 + 1/\tilde{\alpha}_{ab,\sigma}^2}} \right] + (\mu_\sigma)_n \langle \phi_{ab} \rangle_{\sigma}. \tag{B.30}
\]

Note that both \( \langle \phi_{ab} \rangle_{\sigma} \) and \( \langle x_n \phi_{ab} \rangle_{\sigma} \) can be calculated analytically.

**RSLVQ**

For RSLVQ, the quantities \( \langle \Omega_{ab} \rangle_{\sigma} \) and \( \langle x_n \Omega_{ab} \rangle_{\sigma} \) are required.

\[
\langle \Omega_{ab} \rangle_{\sigma} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{1 + \exp(\tilde{\alpha}_{ab,\sigma} \tilde{y} + \tilde{\beta}_{ab,\sigma})} \exp \left( -\frac{1}{2} \tilde{y}^2 \right) d\tilde{y}. \tag{B.31}
\]

This one-dim. integration has to be solved numerically. Analogously, we obtain

\[
\langle x_n \Omega_{ab} \rangle_{\sigma} = -\frac{(C_\sigma \alpha_{ab})_n}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{\exp \left( \tilde{\alpha}_{ab,\sigma} \tilde{y} + \tilde{\beta}_{ab,\sigma} \right)}{(1 + \exp(\tilde{\alpha}_{ab,\sigma} \tilde{y} + \tilde{\beta}_{ab,\sigma}))^2} \exp \left( -\frac{1}{2} \tilde{y}^2 \right) d\tilde{y}
\]

\[ + (\mu_\sigma)_n \langle \Omega_{ab} \rangle_{\sigma}, \tag{B.32} \]

which is also solved numerically.
B.2.2 Three prototypes

For generic function \( f_{ab} f_{cd} \equiv f(\alpha_{ab} \cdot x - \beta_{ab}) f(\alpha_{cd} \cdot x - \beta_{cd}) \), the quantities \( \langle f_{ab} f_{cd} \rangle_k \) and \( \langle x_n f_{ab} f_{cd} \rangle_k \) are required. For instance, the averages \( \langle \Theta_{ab}\Theta_{cd}\rangle_\sigma \) are calculated as follows,

\[
\langle \Theta_{ab}\Theta_{cd}\rangle_\sigma = \frac{1}{(2\pi)^{D/2}(\det C_\sigma)^{1/2}} \int_{R^D} \Theta(\alpha_{ab} \cdot x - \beta_{ab}) \Theta(\alpha_{cd} \cdot x - \beta_{cd}) \times \exp\left(-\frac{1}{2}(x - \mu_\sigma)^T C_\sigma^{-1}(x - \mu_\sigma)\right) dx
\]

\[
= \frac{1}{(2\pi)^{D/2}(\det C_\sigma)^{1/2}} \int_{R^D} \Theta(\alpha_{ab} \cdot x' + \alpha_{ab} \cdot \mu_\sigma - \beta_{ab}) \times \Theta(\alpha_{cd} \cdot x' + \alpha_{cd} \cdot \mu_\sigma - \beta_{cd}) \exp\left(-\frac{1}{2}x'^T C_\sigma^{-1}x'\right) dx' \quad (B.33)
\]

(with the substitution \( x' = x - \mu_\sigma \)).

Because the covariance matrix \( C_\sigma \) is positive definite, \( C_\sigma^{1/2} \) exists. Defining \( x' = C_\sigma^{1/2} y \), we obtain \( x'^T C_\sigma^{-1} x' = y^2 \), \( dx' = (\det C_\sigma)^{1/2} dy \) and

\[
\langle \Theta_{ab}\Theta_{cd}\rangle_\sigma = \frac{1}{(2\pi)^{D/2}} \int_{R^D} \Theta(\alpha_{ab} C_\sigma^{1/2} y + \alpha_{ab} \cdot \mu_\sigma - \beta_{ab}) \times \Theta(\alpha_{cd} C_\sigma^{1/2} y + \alpha_{cd} \cdot \mu_\sigma - \beta_{cd}) \exp\left(-\frac{1}{2}y^2\right) dy
\]

\[
= \frac{1}{(2\pi)^{D/2}} \int_{R^D} \Theta(\alpha_{ab} C_\sigma^{1/2} y + \tilde{\beta}_{ab,\sigma}) \times \Theta(\alpha_{cd} C_\sigma^{1/2} y + \tilde{\beta}_{cd,\sigma}) \exp\left(-\frac{1}{2}y^2\right) dy \quad (B.34)
\]

(where \( \tilde{\beta}_{ab,\sigma} = \alpha_{ab} \cdot \mu_\sigma - \beta_{ab} \)).

Since \( \exp(-\frac{1}{2}y^2) \) has rotational invariance, it is possible to rotate the orthonormal coordinate system \( y = (y_1 e_1 + y_2 e_2 + \ldots + y_N e_N) \) into \( y' = (y_1' e_1' + y_2' e_2' + \ldots + y_N' e_N') \) where one axis, \( e_1' \), is aligned with \( \alpha_{ab} C_\sigma^{1/2} \) and another axis, \( e_2' \), lies on the plane spanned by \( \alpha_{ab} C_\sigma^{1/2} \) and \( \alpha_{cd} C_\sigma^{1/2} \). This is done by the Gram-Schmidt orthonormal transformation:

\[
e_1' = \frac{\alpha_{ab} C_\sigma^{1/2}}{\|\alpha_{ab} C_\sigma^{1/2}\|}
\]

\[
e_2' = \frac{\alpha_{cd} C_\sigma^{1/2} - (\alpha_{ab} C_\sigma^{1/2} \cdot e_1') e_1'}{\|\alpha_{cd} C_\sigma^{1/2} - (\alpha_{ab} C_\sigma^{1/2} \cdot e_1') e_1'\|} \cdot (B.35)
\]

The other axes \( \{e_1', e_2', \ldots, e_N'\} \) are orthogonal to both \( \alpha_{ab} C_\sigma^{1/2} \) and \( \alpha_{cd} C_\sigma^{1/2} \) and can be integrated over using the substitution \( \frac{1}{\sqrt{2\pi}} \int_{R} \exp\left(-\frac{1}{2}z^2\right)dz = 1 \). We obtain from
B.2. Gaussian Averages

Eq. (B.34),
\[
\langle \Theta_{ab} \Theta_{cd} \rangle_{\sigma} = \frac{1}{(2\pi)} \int_{\mathbb{R}^2} \Theta\left(\alpha_{ab} C_{\sigma}^{1/2} \cdot y'_1 e'_1 + \tilde{\beta}_{ab,\sigma}\right) \times \Theta\left(\alpha_{cd} C_{\sigma}^{1/2} \cdot (y'_2 e'_2 + \tilde{\beta}_{cd,\sigma}) \exp\left(-\frac{1}{2}(y''_1^2 + y''_2^2)\right) \right) dy'_1 dy'_2.
\]

(B.36)

We examine the Heaviside functions \(\Theta(x) = 1\) if \(x > 0\); \(0\) else. \(\Theta(\alpha_{ab} C_{\sigma}^{1/2} y' + \tilde{\beta}_{ab,\sigma}) = 1\) and \(\Theta(\alpha_{cd} C_{\sigma}^{1/2} y' + \tilde{\beta}_{cd,\sigma}) = 1\) if the following conditions are satisfied

\[
y'_1 > y'_1^* \quad \text{with} \quad y'_1^* = -\frac{\tilde{\beta}_{ab,\sigma}}{\alpha_{ab,\sigma}} - \alpha_{ab,\sigma} C_{\sigma} y'_1
\]

\[
y'_2 > y'_2^* \quad \text{with} \quad y'_2^* = -\frac{\tilde{\beta}_{cd,\sigma} \alpha_{ab,\sigma} + (\alpha_{cd,\sigma} y'_1)}{\sqrt{\alpha_{cd,\sigma}^2 \alpha_{ab,\sigma}^2 - (\alpha_{cd,\sigma} \alpha_{ab})^2}}
\]

where we defined \(\tilde{\alpha}_{ab,\sigma} = ||\alpha_{ab} C_{\sigma}^{1/2}||\). Substituting the conditions into Eq. (B.36), we get

\[
\langle \Theta_{ab} \Theta_{cd} \rangle_{\sigma} = \frac{1}{(2\pi)} \int_{y'_1}^{\infty} \int_{y'_2}^{\infty} \exp\left(-\frac{1}{2}(y''_1^2 + y''_2^2)\right) dy'_2 dy'_1
\]

\[
= \frac{1}{(2\pi)} \int_{y'_1}^{\infty} \exp\left(-\frac{1}{2}y''_1^2\right) \left(\int_{y'_2}^{\infty} \exp\left(-\frac{1}{2}y''_2^2\right) dy'_2\right) dy'_1.
\]

(B.37)

We get the final result in closed form as

\[
\langle \Theta_{ab} \Theta_{cd} \rangle_{\sigma} = \frac{1}{\sqrt{2\pi}} \int_{-\tilde{\beta}_{ab,\sigma}}^{\infty} \exp\left(-\frac{1}{2}y''_1^2\right) \Phi\left(\frac{\tilde{\beta}_{cd,\sigma} \alpha_{ab,\sigma} + (\alpha_{cd,\sigma} \alpha_{ab}) y'_1}{\sqrt{\alpha_{cd,\sigma}^2 \alpha_{ab,\sigma}^2 - (\alpha_{cd,\sigma} \alpha_{ab})^2}}\right) dy'_1
\]

(B.38)

with \(\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}d^2) \, da\). The one-fold integration in Eq. (B.38) has to be performed numerically.

The remaining average to be computed is

\[
\langle (x)_{n} \Theta_{ab} \Theta_{cd} \rangle_{\sigma} = \frac{1}{(2\pi)^{D/2} (\det C_{\sigma})^{1/2}} \int_{\mathbb{R}^D} (x)_{n} \Theta(\alpha_{ab} \cdot x - \beta_{ab}) \Theta(\alpha_{cd} \cdot x - \beta_{cd}) \times \exp\left(-\frac{1}{2}(x - \mu_\sigma)^T C_{\sigma}^{-1} (x - \mu_\sigma)\right) dx.
\]

(B.39)
Similar to Eq. (B.34), we obtain the form

\[
\langle (x)_n \Theta_{ab} \Theta_{cd} \rangle \sigma = \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} (C_{\sigma}^{1/2} y)_n \Theta (\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \times \Theta (\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma}) \exp(-\frac{1}{2} y^2) dy + (\mu_\sigma)_n (\Theta_{ab} \Theta_{cd}) \sigma
\]

\[
= I + (\mu_\sigma)_n (\Theta_{ab} \Theta_{cd}) \sigma \tag{B.40}
\]

(where \( I \) is an integral to be computed).

Consider the integrals contributing to \( I \)

\[
I_j = \int_{\mathbb{R}} (C_{\sigma}^{1/2}) n_j(y)_j \Theta (\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \Theta (\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma}) \times \exp(-\frac{1}{2} (y_j^2)) d(y)_j \tag{B.41}
\]

we perform integration by parts \( \int u dv = uv - \int v du \) with

\[
u = (C_{\sigma}^{1/2}) n_j(y)_j \exp(-\frac{1}{2} (y_j^2)),
\]

\[
u = (C_{\sigma}^{1/2}) n_j(y)_j \exp(-\frac{1}{2} (y_j^2)) \]

\[
\]

and we obtain

\[
I_j = \left[-\Theta (\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \Theta (\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma}) (C_{\sigma}^{1/2}) n_j \exp(-\frac{1}{2} (y_j^2)) \right]_{-\infty}^{\infty}
\]

\[
+ \int_{\mathbb{R}} (C_{\sigma}^{1/2}) n_j \exp(-\frac{1}{2} (y_j^2)) \left[ \frac{\partial}{\partial y_j} \left( \Theta (\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \right) \right] d(y)_j
\]

we perform integration by parts \( \int u dv = uv - \int v du \) with

\[
u = (C_{\sigma}^{1/2}) n_j(y)_j \exp(-\frac{1}{2} (y_j^2)) \]

\[
u = (C_{\sigma}^{1/2}) n_j(y)_j \exp(-\frac{1}{2} (y_j^2)) \]

and we obtain

\[
I_j = \left[-\Theta (\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \Theta (\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma}) (C_{\sigma}^{1/2}) n_j \exp(-\frac{1}{2} (y_j^2)) \right]_{-\infty}^{\infty}
\]

\[
+ \int_{\mathbb{R}} (C_{\sigma}^{1/2}) n_j \exp(-\frac{1}{2} (y_j^2)) \left[ \frac{\partial}{\partial y_j} \left( \Theta (\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \right) \right] d(y)_j
\]
B.2. Gaussian Averages

\[ (C_{\sigma}^{1/2})_{nj} \left( \int_{\mathbb{R}^D} \frac{\partial \Theta(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma})}{\partial y_j} \Theta(\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma}) \right) \]

\[ \times \exp\left(-\frac{1}{2}y_j^2\right) \int_{\mathbb{R}^D} \Theta(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \frac{\partial \Theta(\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma})}{\partial y_j} \]

\[ \times \exp\left(-\frac{1}{2}y_j^2\right) d(y_j). \]  

(B.42)

The sum over \( j \) gives

\[ I = \frac{1}{(2\pi)^{D/2}} \sum_{j=1}^{D} (C_{\sigma}^{1/2})_{nj} \left( \int_{\mathbb{R}^D} \frac{\partial \Theta(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma})}{\partial y_j} \Theta(\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma}) \right) \]

\[ \times \exp\left(-\frac{1}{2}y_j^2\right) d(y_j) \int_{\mathbb{R}^D} \Theta(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \frac{\partial \Theta(\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma})}{\partial y_j} \]

\[ \times \exp\left(-\frac{1}{2}y_j^2\right) d(y_j). \]  

(B.43)

\[ = \frac{1}{(2\pi)^{D/2}} \left( C_{\sigma} \alpha_{ab} \right)_n \int_{\mathbb{R}^D} \delta(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \Theta(\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma}) \]

\[ \times \exp\left(-\frac{1}{2}y^2\right) d(y) + \left( C_{\sigma} \alpha_{cd} \right)_n \int_{\mathbb{R}^D} \delta(\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma}) \]

\[ \times \Theta(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \exp\left(-\frac{1}{2}y^2\right) d(y). \]  

(B.44)

where \( \delta(\cdot) \) is the Dirac-delta function. In the last step we have used

\[ \frac{\partial}{\partial y_j} \left( \Theta(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \right) = \sum_{i=1}^{D} (\alpha_{ab})_i (C_{\sigma}^{1/2})_{ij} \delta(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}). \]

Calculating the first term only,

\[ I_{ab} = \frac{1}{(2\pi)^{D/2}} \left( C_{\sigma} \alpha_{ab} \right)_n \int_{\mathbb{R}^D} \delta(\alpha_{ab} C_{\sigma}^{1/2} y + \tilde{\beta}_{ab,\sigma}) \Theta(\alpha_{cd} C_{\sigma}^{1/2} y + \tilde{\beta}_{cd,\sigma}) \exp\left(-\frac{1}{2}y^2\right) d(y), \]
we rotate the coordinate system as in Eq. (B.36) and obtain the following

\[ I_{ab} = \frac{1}{2\pi} (C\sigma\alpha_{ab})n \int_{-\infty}^{\infty} \delta \left( \alpha_{ab} C_{\sigma}^{1/2} \cdot y_1' e_1' + \tilde{\beta}_{ab,\sigma} \right) \times \Theta (\alpha_{cd} C_{\sigma}^{1/2} \cdot (y_1' e_1' + y_2' e_2') + \tilde{\beta}_{cd,\sigma}) \exp \left( -\frac{1}{2} (y_1'^2 + y_2'^2) \right) dy_1' dy_2' \]

\[ = \frac{(C\sigma\alpha_{ab})n}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta \left( \tilde{\alpha}_{ab,\sigma} y_1' + \tilde{\beta}_{ab,\sigma} \right) \exp \left( -\frac{1}{2} y_1'^2 \right) \times \Phi \left( \frac{\tilde{\beta}_{cd,\sigma} \tilde{\alpha}_{ab,\sigma} + (\alpha_{cd} C_{\sigma} \alpha_{ab}) y_1'}{\tilde{\alpha}_{cd,\sigma} \tilde{\alpha}_{ab,\sigma} - (\alpha_{cd} C_{\sigma} \alpha_{ab})^2} \right) dy_1'. \]

Substituting \( z = \tilde{\alpha}_{ab,\sigma} y_1' \),

\[ I_{ab} = \frac{(C\sigma\alpha_{ab})n}{\sqrt{(2\pi)\tilde{\alpha}_{ab,\sigma}}} \int_{-\infty}^{\infty} \delta \left( z + \tilde{\beta}_{ab,\sigma} \right) \exp \left( -\frac{1}{2} \left( \frac{z}{\tilde{\alpha}_{ab,\sigma}} \right)^2 \right) \times \Phi \left( \frac{\tilde{\beta}_{cd,\sigma} \tilde{\alpha}_{ab,\sigma} + (\alpha_{cd} C_{\sigma} \alpha_{ab}) z}{\tilde{\alpha}_{cd,\sigma} \tilde{\alpha}_{ab,\sigma} - (\alpha_{cd} C_{\sigma} \alpha_{ab})^2} \right) \frac{dz}{\tilde{\alpha}_{ab,\sigma}}. \]

(B.45)

Analogously we compute the second term in Eq. (B.44) and obtain the final form

\[ \langle (x)_n \Theta_{ab} \Theta_{cd} \rangle \sigma = \frac{(C\sigma\alpha_{ab})n}{\sqrt{(2\pi)\tilde{\alpha}_{ab,\sigma}}} \exp \left( -\frac{1}{2} \frac{\tilde{\beta}_{ab,\sigma}^2}{\tilde{\alpha}_{ab,\sigma}} \right) \Phi \left( \frac{\tilde{\beta}_{cd,\sigma} \tilde{\alpha}_{ab,\sigma}^2 - \tilde{\beta}_{ab,\sigma} (\alpha_{cd} C_{\sigma} \alpha_{ab})}{\tilde{\alpha}_{cd,\sigma} \tilde{\alpha}_{ab,\sigma}^2 - (\alpha_{cd} C_{\sigma} \alpha_{ab})^2} \right) \]

\[ + \frac{(C\sigma\alpha_{cd})n}{\sqrt{(2\pi)\tilde{\alpha}_{cd,\sigma}}} \exp \left( -\frac{1}{2} \frac{\tilde{\beta}_{cd,\sigma}^2}{\tilde{\alpha}_{cd,\sigma}} \right) \Phi \left( \frac{\tilde{\beta}_{cd,\sigma} \tilde{\alpha}_{cd,\sigma}^2 - \tilde{\beta}_{cd,\sigma} (\alpha_{ab} C_{\sigma} \alpha_{cd})}{\tilde{\alpha}_{cd,\sigma} \tilde{\alpha}_{cd,\sigma}^2 - (\alpha_{cd} C_{\sigma} \alpha_{ab})^2} \right) \]

\[ + (\mu_{\sigma})_n \langle \Theta_{ab} \Theta_{cd} \rangle \sigma. \]

(B.46)

Window

With the addition of a window, these quantities are required:

\[ \langle (\Theta_{ab}^\phi - \Theta_{ab}^\phi) \Theta_{cd} \rangle \sigma = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left( -\frac{1}{2} \frac{\tilde{\beta}_{ab,\sigma}^2}{\tilde{\alpha}_{ab,\sigma}} \right) \Phi \left( \frac{\tilde{\beta}_{cd,\sigma} \tilde{\alpha}_{ab,\sigma} + (\alpha_{cd} C_{\sigma} \alpha_{ab}) y_1'}{\tilde{\alpha}_{cd,\sigma} \tilde{\alpha}_{ab,\sigma} - (\alpha_{cd} C_{\sigma} \alpha_{ab})^2} \right) dy_1'. \]
\[ \langle x \rangle_n (\Theta_{ab}^\delta - \Theta_{ab}^\gamma) \Theta_{cd} \sigma 
\]
\[ = \frac{(C_\sigma \alpha_{ab})}{\sqrt{(2\pi)\alpha_{ab,\sigma}}} \exp \left(-\frac{1}{2} \frac{\tilde{\beta}_{ab,\sigma}^2 \alpha_{ab,\sigma}}{\tilde{\alpha}_{ab,\sigma}} \right) \Phi \left( \frac{\tilde{\beta}_{cd,\sigma} \sqrt{\tilde{\alpha}_{cd,\sigma}}}{\tilde{\alpha}_{ab,\sigma}} \tilde{\alpha}_{ab,\sigma} - (\alpha_{cd} C_\sigma \alpha_{ab}) \right) 
\]
\[ + \frac{(C_\sigma \alpha_{cd})}{\sqrt{(2\pi)\tilde{\alpha}_{cd,\sigma}}} \exp \left(-\frac{1}{2} \frac{\tilde{\beta}_{cd,\sigma}^2 \alpha_{cd,\sigma}}{\tilde{\alpha}_{cd,\sigma}} \right) \Phi \left( \frac{\tilde{\beta}_{ab,\sigma} \sqrt{\tilde{\alpha}_{cd,\sigma}}}{\tilde{\alpha}_{ab,\sigma}} \tilde{\alpha}_{ab,\sigma} - (\alpha_{cd} C_\sigma \alpha_{ab}) \right) 
\]
\[ - \frac{(C_\sigma \alpha_{ab})}{\sqrt{(2\pi)\tilde{\alpha}_{ab,\sigma}}} \exp \left(-\frac{1}{2} \frac{\tilde{\beta}_{ab,\sigma}^2 \alpha_{ab,\sigma}}{\tilde{\alpha}_{ab,\sigma}} \right) \Phi \left( \frac{\tilde{\beta}_{cd,\sigma} \sqrt{\tilde{\alpha}_{cd,\sigma}}}{\tilde{\alpha}_{ab,\sigma}} \tilde{\alpha}_{ab,\sigma} - (\alpha_{cd} C_\sigma \alpha_{ab}) \right) 
\]
\[ - \frac{(C_\sigma \alpha_{cd})}{\sqrt{(2\pi)\tilde{\alpha}_{cd,\sigma}}} \exp \left(-\frac{1}{2} \frac{\tilde{\beta}_{cd,\sigma}^2 \alpha_{cd,\sigma}}{\tilde{\alpha}_{cd,\sigma}} \right) \Phi \left( \frac{\tilde{\beta}_{ab,\sigma} \sqrt{\tilde{\alpha}_{cd,\sigma}}}{\tilde{\alpha}_{ab,\sigma}} \tilde{\alpha}_{ab,\sigma} - (\alpha_{cd} C_\sigma \alpha_{ab}) \right) 
\]
\[ + \frac{(\mu_\sigma)_n (\Theta_{ab}^\delta - \Theta_{ab}^\gamma) \Theta_{cd} \sigma}{\sqrt{(2\pi)\tilde{\alpha}_{ab,\sigma}}} \Phi \left( \frac{\tilde{\beta}_{cd,\sigma} \sqrt{\tilde{\alpha}_{cd,\sigma}}}{\tilde{\alpha}_{ab,\sigma}} \tilde{\alpha}_{ab,\sigma} - (\alpha_{cd} C_\sigma \alpha_{ab}) \right) 
\]
\[ + (\mu_\sigma)_n (\Theta_{ab}^\delta - \Theta_{ab}^\gamma) \Theta_{cd} \sigma. \quad (B.47) \]
B.3 Generalization error

Two prototypes

We compute the generalization error from Eq. (4.20) as follows. For two prototypes \( \mathbf{w}_p \) and \( \mathbf{w}_- \), we calculate \( \epsilon_g = \sum_{\sigma} p_{\sigma} \epsilon_{g, \sigma} \) with

\[
\epsilon_{g, \sigma} = \langle \Theta_{-\sigma\sigma} \rangle_+ = \Phi \left( \frac{\tilde{\beta}_{-\sigma\sigma}}{\tilde{\alpha}_{-\sigma\sigma}} \right)
\]

with \( \tilde{\alpha}_{ST, \sigma} = \sqrt{\alpha_{ST} \alpha_{ST}} \) and \( \tilde{\beta}_{ST, \sigma} = \alpha_{ST} \mu_{\sigma} = \beta_{ST} \). We refer the calculations to (Biehl et al. 2004). Plugging in the values, we obtain

\[
\epsilon_{g, \sigma} = \Phi \left( \frac{Q_{\sigma\sigma} - Q_{-\sigma\sigma} - 2\ell_{\sigma}(R_{\sigma\sigma} - R_{-\sigma\sigma})}{2\sqrt{\nu_{\sigma}} \sqrt{Q_{\sigma\sigma} - 2Q_{-\sigma\sigma} + Q_{-\sigma\sigma}} - \Delta_\sigma} \right)
\]

(B.49)

By using \( Z_{\sigma} = Q_{\sigma\sigma} - Q_{-\sigma\sigma} - 2\ell_{\sigma}(R_{\sigma\sigma} - R_{-\sigma\sigma}) \) and \( \Delta_q = \sqrt{Q_{++} - 2Q_{+-} + Q_{--}} \), we can calculate the derivative of the generalization error with respect to the order parameters \( \mathbf{O} = \{ R_{++}, R_{+-}, R_{-+}, R_{--}, Q_{++}, Q_{+-}, Q_{--} \}^T \) as follows:

\[
\frac{d\epsilon_{g, \sigma}}{d\mathbf{O}} = \frac{1}{\sqrt{2\pi} 2\sqrt{\nu_{\sigma}}} \exp \left( -\frac{1}{2} \left[ \frac{Z_{\sigma}}{2\sqrt{\nu_{\sigma}} \Delta_q} \right]^2 \right) \frac{d Z_\sigma}{d\mathbf{O}} \Delta_q
\]

(B.50)

where we used \( d\Phi(\tau)/d\tau = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} \tau^2) \). Derivations with respect to the order parameters yield

\[
\begin{align*}
\frac{d}{d\mathbf{O}} \Delta_q^+ \begin{bmatrix}
-2\ell/\Delta_q \\
0 \\
+2\ell/\Delta_q \\
1/\Delta_q - Z_+/(2\Delta_q^3)
\end{bmatrix}
\end{align*}
\]

and

\[
\begin{align*}
\frac{d}{d\mathbf{O}} \Delta_q^- \begin{bmatrix}
0 \\
+2\ell/\Delta_q \\
0 \\
-1/\Delta_q - Z_-/(2\Delta_q^3)
\end{bmatrix}
\end{align*}
\]

(B.51)

In the special case of \( p_+ = p_- = 0.5 \) and \( v_+ = v_- = v \), one obtains

\[
\frac{d\epsilon_{g, \sigma}}{d\mathbf{O}} = \sum_{\sigma} \frac{d\epsilon_{g, \sigma}}{d\mathbf{O}} = \frac{1}{2\sqrt{2\pi} \sqrt{\nu}} \exp \left( -\frac{1}{2} \left[ \frac{Z}{2\sqrt{\nu} \Delta_q} \right]^2 \right) \frac{d Z}{d\mathbf{O}} \Delta_q
\]

(B.52)
B.4 Quantization error

Three prototypes
To compute the generalization error in systems with three prototypes \( w_S, w_T, w_U \), we require the quantity

\[
\epsilon_g,\sigma = \sum_{S : c_S \neq y_\sigma} K \sum_{S=1}^{K} \langle \Theta_{ST} \Theta_{SU} \rangle_\sigma , \tag{B.53}
\]

where the averages are written in Eq. (B.38).

B.4 Quantization error

For two prototypes \( w_1 \) and \( w_2 \), we compute the quantization error

\[
E(W) = \frac{1}{2} \sum_{S=1}^{K} \left( \prod_{T \neq S} \Theta_{ST} \right) Q_{SS} - \sum_{S=1}^{K} \langle h_S \prod_{T \neq S} \Theta_{ST} \rangle \tag{B.54}
\]
as follows. We calculate \( \epsilon_g = \sum_\sigma p_\sigma \epsilon_{g,\sigma} \) with

\[
E_\sigma(W) = \frac{1}{2} \langle \Theta_{12} \rangle_\sigma Q_{11} - \langle h_1 \Theta_{12} \rangle_\sigma + \frac{1}{2} \langle \Theta_{21} \rangle_\sigma Q_{22} - \langle h_2 \Theta_{21} \rangle_\sigma \\
= - \langle h_1 \Theta_{12} \rangle_\sigma - \langle h_2 \Theta_{21} \rangle_\sigma + \frac{Q_{11}}{2} \langle \Theta_{12} \rangle_\sigma + \frac{Q_{22}}{2} \langle \Theta_{21} \rangle_\sigma \\
= - \frac{1}{\sqrt{2\pi} \tilde{\sigma}_{12,\sigma}} \exp \left( -\frac{Z^2}{2} \right) - \frac{C_{\tilde{\sigma}_{21,\sigma}}}{\sqrt{2\pi} \tilde{\sigma}_{21,\sigma}} \exp \left( -\frac{(-Z)^2}{2} \right) - (\mu_\sigma) \Phi(Z) \\
+ (\mu_\sigma) \Phi(-Z) + \frac{Q_{11}}{2} \Phi(Z) + \frac{Q_{22}}{2} \Phi(-Z) \\
= - \frac{\sqrt{\Delta Q}}{\sqrt{2\pi}} \exp \left( -\frac{Z^2}{2} \right) + \left( \frac{Q_{11}}{2} - \ell_\sigma R_{1\sigma} \right) \Phi(Z) + \left( \frac{Q_{22}}{2} - \ell_\sigma R_{2\sigma} \right) \Phi(-Z) ,
\]

with

\[
\Delta Q = \sqrt{Q_{11} - 2Q_{12} + Q_{22}} , \\
Z = (2\ell_\sigma (R_{1\sigma} - R_{2\sigma}) - Q_{11} + Q_{22}) / (2\sqrt{\Delta Q}) . \tag{B.55}
\]
Appendix C

Annealed approximation

Practical training procedures aim at an efficient minimization of the cost function. In the statistical physics interpretation of the learning process, this corresponds to low temperatures. While the correct treatment of finite $T$ requires sophisticated techniques such as the replica trick, a useful approximation method which is technically less difficult, can be employed to perform the quenched average $\langle \ln Z \rangle_I D$. In the annealed approximation, we approximate the function

$$\langle \ln Z \rangle_I D \approx \ln \langle Z \rangle_I D$$  \hspace{1cm} (C.1)

Equivalently,

$$\langle \exp (-\beta H(W)) / Z \rangle_I D \approx \langle \exp (-\beta H(W)) \rangle_I D / \langle Z \rangle_I D$$  \hspace{1cm} (C.2)

The annealed approximation becomes exact in the limit $\beta \to 0$ and coincides with the explicit treatment of this limit (Seung et al. 1992). At low temperatures the annealed free energy yields only an upper bound to the correct one, but the hope is that the position of minima in terms of the $\{R_{SS}, Q_{ST}\}$ is similar. The scheme has proven useful in predicting qualitative behavior of many learning systems, e.g. (Engel and van den Broeck 2001, Sompolinsky and Tishby 1990). The validity of the annealed approximation is discussed systematically in, for instance, (Solla and Levin 1992, Seung et al. 1992). The average partition function can be rewritten as

$$\langle Z \rangle_I D = \int d\mu(W) \left\langle \exp \left[ -\beta \sum_{\mu=1}^{P} e(\xi_{\mu}, W) \right] \right\rangle_I D$$ \hspace{1cm} (C.3)

$$= \int d\mu(W) \prod_{\mu=1}^{P} \langle \exp [-\beta e(\xi_{\mu}, W)] \rangle_{\xi}$$ \hspace{1cm} (C.4)

$$= \int d\mu(W) \exp \left[ P \ln \langle \exp [-\beta e(\xi_{\mu}, W)] \rangle_{\xi} \right]$$ \hspace{1cm} (C.5)

$$= \int d\mu(W) \exp [-\alpha N G_{A}(W)]$$

where $G_{A}$ involves an average over one random input only. Only for $\beta \to 0$ this average can be absorbed into the exponent and we recover the high temperature result.
The corresponding free energy function is

\[ f = \alpha G_A(\{R_{S\sigma}, Q_{ST}\}) - s(\{R_{S\sigma}, Q_{ST}\}) \]  

(C.7)

Unlike the high temperature limit, in the annealed approximation the empirical average over training set \( D \) is distinguished from the input density. The training and test set performances are given by

\[ \epsilon_{\text{train}} = \frac{1}{P} \sum_{\mu=1}^{P} e(W, \xi^\mu) = \alpha^{-1} \frac{\partial}{\partial f} f(\{R_{S\sigma}, Q_{ST}\}) \]

\[ \epsilon_{\text{test}} = \langle e \rangle_{\xi} \]  

(C.8)

which have to be evaluated in the minimum of \( f \).

The calculation of \( G_A \) can be done analytically for arbitrary \( \beta \) and two prototypes with the WTA cost function

\[ e(W, \xi) = \frac{1}{2} \sum_{i=1}^{K} d(w_i, \xi) \prod_{j \neq i} \Theta_{ij} - \frac{1}{2} \xi^2. \]  

(C.9)

we can obtain, for two prototypes,

\[ G_A = - \ln \int d\mu(\xi) \exp \left[ -\beta e(\xi^\mu, W) \right] \]

\[ = - \ln \left\langle \exp \beta \left( \frac{1}{2} \left( \Theta_{12} (\xi - w_1)^2 + \Theta_{21} (\xi - w_2)^2 \right) - \frac{1}{2} \xi^2 \right) \right\rangle_{\xi} \]  

(C.10)

\[ = - \ln \left\langle \exp \beta \left( \Theta_{12} \left( w_1 \cdot \xi - \frac{1}{2} w_1^2 \right) + \Theta_{21} \left( w_2 \cdot \xi - \frac{1}{2} w_2^2 \right) \right) \right\rangle_{\xi} \]  

(C.11)

Substituting

\[ \langle \exp(\Theta_{12} f + \Theta_{21} g) \rangle_{\xi} = \langle \Theta_{12} \exp(f) \rangle_{\xi} + \langle \Theta_{21} \exp(g) \rangle_{\xi}, \]  

(C.13)

we obtain

\[ G_A = - \ln \left\langle \Theta_{12} \exp \beta \left( w_1 \cdot \xi - \frac{1}{2} w_1^2 \right) \right\rangle_{\xi} \]

\[ + \left\langle \Theta_{21} \exp \beta \left( w_2 \cdot \xi - \frac{1}{2} w_2^2 \right) \right\rangle_{\xi} \]  

(C.14)

\[ = - \ln \left\{ \exp \left( -\frac{1}{2} \beta w_1^2 \right) \langle \Theta_{12} \exp(\beta w_1 \cdot \xi) \rangle_{\xi} \right\} \]

\[ + \exp \left( -\frac{1}{2} \beta w_2^2 \right) \langle \Theta_{21} \exp(\beta w_2 \cdot \xi) \rangle_{\xi} \]  

(C.15)
Converting into order parameters and projections,

\[
G_A = -\ln \left\{ \exp \left( -\frac{1}{2} \beta Q_{11} \right) \langle \Theta_{12} \exp (\beta h_1) \rangle \xi + \exp \left( -\frac{1}{2} \beta Q_{22} \right) \langle \Theta_{21} \exp (\beta h_2) \rangle \xi \right\}
\]

(C.16)

For the annealed approximation, the quantity of importance is

\[
\langle \Theta_{ab} \exp (\beta x_n) \rangle_k = \frac{1}{(2\pi)^{D/2}(\det C_\sigma)^{1/2}} \int_{\mathbb{R}^D} \exp \left( -\frac{1}{2} (x - \mu_k)^T C_\sigma^{-1} (x - \mu_k) \right) dx
\]

(17)

\[
= \frac{1}{(2\pi)^{D/2}(\det C_\sigma)^{1/2}} \int_{\mathbb{R}^D} \exp \left( \beta (x_n' + \mu_k) \right) \Theta_{ab} \exp \left( -\frac{1}{2} x_n'^T C_\sigma^{-1} x_n' \right) dx'
\]

(18)

We begin calculations by partitioning elements in the matrix \(C_\sigma^{-1}\) and vector \(x'\) associated with term \(n\) as follows

\[
x'^T C_\sigma^{-1} x' = \sum_j \sum_{i \neq n} x'_i (C_\sigma^{-1})_{ij} x'_j + \sum_{j \neq n} x'_n (C_\sigma^{-1})_{nj} x'_j + \sum_{i \neq n} x'_i (C_\sigma^{-1})_{in} x'_n + x_n (C_\sigma^{-1})_{nn} x_n
\]

(19)

where \(\bar{n}\) are elements complementary to \(n\). We use the shorthand

\[
\tilde{C}_k^{-1} = (C_\sigma^{-1})_{[\bar{n},\bar{n}]}.
\]

(20)

It can be easily shown that because \(C_\sigma\) is positive definite, \(\tilde{C}_k\) is also positive definite. Therefore \(\tilde{C}_k^{1/2}\) exists and we can define \(x'_n^T = \tilde{C}_k^{1/2} y\) to obtain

\[
x'^T C_\sigma^{-1} x' = y^2 + 2 \left( (C_\sigma^{-1})_{[\bar{n},\bar{n}]|\bar{n}} \tilde{C}_k^{1/2} y \right) x'_n + (C_\sigma^{-1})_{[n,n]} (x'_n)^2
\]

(21)
and from \( d\tilde{x} = \det(\tilde{C}_k^{1/2}) \, dy = (\det \tilde{C}_k)^{1/2} \, dy \),
\[
\langle \Theta_{ab} \exp(\beta x_n) \rangle_k
= \frac{\exp(\beta(\mu_{ab})_n) \det(\tilde{C}_k)^{1/2}}{(2\pi)^{D/2} \det(C_{\sigma})^{1/2}} \int_{\mathbb{R}^D} \exp(\beta x_n') \Theta_{ab} \exp\left(-\frac{1}{2} \left( y^2 + 2(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k^{1/2} y \cdot x_n' + (C_{\sigma}^{-1})_{nn}(x_n')^2 \right) \right) \, dy \, dx_n'
\]
\[
= \frac{\exp(\beta(\mu_{ab})_n) \sqrt{\det \tilde{C}_k}}{(2\pi)^{D/2} \det C_{\sigma}} \int_{\mathbb{R}} \exp\left(\beta x_n' - \frac{1}{2}(C_{\sigma}^{-1})_{nn}(x_n')^2\right) \left\{ \int_{\mathbb{R}^{D-1}} \Theta_{ab} \exp\left(-\frac{1}{2} \left( y^2 + 2(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k^{1/2} y \cdot x_n' \right) \right) \, dy \right\} \, dx_n'
\]
\[
= \int_{\mathbb{R}^{D-1}} \Theta_{ab} \exp\left(-\frac{1}{2} \left( y^2 + 2(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k^{1/2} y \cdot x_n' \right) \right) \, dy
\]
\[
= \exp\left(\frac{1}{2}(x_n')^2(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k(C_{\sigma}^{-1})_{[n,n]} \right) \times \int_{\mathbb{R}^{D-1}} \Theta_{ab} \exp\left(-\frac{1}{2} \left( y + x_n'(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k^{1/2} \right)^2 \right) \, dy
\]
\[
= \exp\left(\frac{1}{2}(x_n')^2(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k(C_{\sigma}^{-1})_{[n,n]} \right) \int_{\mathbb{R}^{D-1}} \Theta_{ab} \exp\left(-\frac{1}{2} z^2 \right) \, dz
\]
(22)

Calculating the integral \( I \) by completing the square,
\[
I = \int_{\mathbb{R}^{D-1}} \Theta_{ab} \exp\left(-\frac{1}{2} \left( y^2 + 2(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k^{1/2} y \cdot x_n' \right) \right) \, dy
\]
\[
= \int_{\mathbb{R}^{D-1}} \Theta_{ab} \exp\left(-\frac{1}{2} \left( y + x_n'(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k^{1/2} \right)^2 + \frac{1}{2}(x_n')^2(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k(C_{\sigma}^{-1})_{[n,n]} \right) \, dy
\]
\[
= \exp\left(\frac{1}{2}(x_n')^2(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k(C_{\sigma}^{-1})_{[n,n]} \right) \int_{\mathbb{R}^{D-1}} \Theta_{ab} \exp\left(-\frac{1}{2} \left( y + x_n'(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k^{1/2} \right)^2 \right) \, dy
\]
\[
= \exp\left(\frac{1}{2}(x_n')^2(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k(C_{\sigma}^{-1})_{[n,n]} \right) \int_{\mathbb{R}^{D-1}} \Theta_{ab} \exp\left(-\frac{1}{2} z^2 \right) \, dz
\]
(23)

where we defined \( z = y + x_n'(C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k^{1/2} \). The corresponding coordinate transforms for the Heaviside function is as follows
\[
\Theta_{ab} = \Theta \left( (\alpha_{ab})_n \cdot \tilde{x} + \beta_{ab,k} \right)
= \Theta \left( (\alpha_{ab})_n \cdot \tilde{x} + (\alpha_{ab})_n x_n' + \beta_{ab,k} \right)
= \Theta \left( (\alpha_{ab})_n \tilde{C}_k^{1/2} \tilde{y} + (\alpha_{ab})_n x_n' + \beta_{ab,k} \right)
= \Theta \left( (\alpha_{ab})_n \tilde{C}_k^{1/2} \tilde{z} + (\alpha_{ab})_n x_n' + \beta_{ab,k} \right)
= \Theta \left( (\alpha_{ab})_n \tilde{C}_k^{1/2} \tilde{z} + \tilde{\beta}_{ab,k} \right)
\]
(24)
where
\[ \tilde{\beta}_{ab,k} = \tilde{\beta}_{ab,k} + (\alpha_{ab})_n x'_n - (\alpha_{ab})_n \tilde{C}_k (C_{\sigma}^{-1})_{[n,n]} x'_n. \] (25)

Rotating the coordinate system so that one of the axes \( \tilde{z} \) coincides with \((\alpha_{ab})_n \tilde{C}_k^{1/2}\) and substituting \( \int_{\mathbb{R}} \exp(-\frac{1}{2}z^2)dz = \sqrt{2\pi} \)

\[ I = \exp \left( \frac{1}{2} (x'_n)^2 (C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k (C_{\sigma}^{-1})_{[n,n]} \right) (2\pi)^{(D-2)/2} \]
\[ \times \int_{\mathbb{R}} \Theta \left( \left\| (\alpha_{ab})_n \tilde{C}_k^{1/2} \right\| \tilde{z} + \tilde{\beta}_{ab,k} \right) \exp \left( -\frac{1}{2} \tilde{z}^2 \right) d\tilde{z} \]
\[ = \exp \left( \frac{1}{2} (x'_n)^2 (C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k (C_{\sigma}^{-1})_{[n,n]} \right) (2\pi)^{(D-2)/2} \]
\[ \times \int_{\mathbb{R}} \tilde{\beta}_{ab,k} \exp \left( -\frac{1}{2} \tilde{z}^2 \right) d\tilde{z} \] (26)

where
\[ \tilde{\alpha}_{ab,k} = \left\| (\alpha_{ab})_n \tilde{C}_k^{1/2} \right\| \] (28)

Substituting \( \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left( -\frac{1}{2} t^2 \right) dt \)

\[ I = \exp \left( \frac{1}{2} (C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k (C_{\sigma}^{-1})_{[n,n]} (x'_n)^2 \right) (2\pi)^{(D-1)/2} \Phi \left( \frac{\tilde{\beta}_{ab,k}}{\tilde{\alpha}_{ab,k}} \right) \] (29)

Hence, we obtain the form
\[ \langle \Theta_{ab} \exp (\beta x_n) \rangle_k \]
\[ = \frac{\exp (\beta (\mu_k)_n)}{\sqrt{2\pi}} \sqrt{\frac{\det C_k}{\det C_{\sigma}}} \int_{\mathbb{R}} \exp \left( \beta x'_n - \frac{1}{2} (C_{\sigma}^{-1})_{nn} (x'_n)^2 \right) \]
\[ \times \left( \frac{1}{2} (C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k (C_{\sigma}^{-1})_{[n,n]} (x'_n)^2 \right) \Phi \left( \frac{\tilde{\beta}_{ab,k}}{\tilde{\alpha}_{ab,k}} \right) dx'_n \]
\[ = \frac{\exp (\beta (\mu_k)_n)}{\sqrt{2\pi}} \sqrt{\frac{\det C_k}{\det C_{\sigma}}} \int_{\mathbb{R}} \exp \left( -\frac{1}{2} (C_{\sigma}^{-1})_{nn} (C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k (C_{\sigma}^{-1})_{[n,n]} (x'_n)^2 \right) \]
\[ \times \left( \frac{\tilde{\beta}_{ab,k} + (\alpha_{ab})_n x'_n - (\alpha_{ab})_n \tilde{C}_k (C_{\sigma}^{-1})_{[n,n]} x'_n}{\tilde{\alpha}_{ab,k}} \right) dx'_n \]
\[ = \frac{\exp (\beta (\mu_k)_n)}{\sqrt{2\pi}} \sqrt{\frac{\det C_k}{\det C_{\sigma}}} \int_{\mathbb{R}} \exp \left( -\frac{1}{2} (C_{\sigma}^{-1})_{nn} (C_{\sigma}^{-1})^T_{[n,n]} \tilde{C}_k (C_{\sigma}^{-1})_{[n,n]} (x'_n)^2 \right) \]
\[ \times \left( \frac{(\alpha_{ab})_n (\alpha_{ab})_n \tilde{C}_k (C_{\sigma}^{-1})_{[n,n]} x'_n + \tilde{\beta}_{ab,k}}{\tilde{\alpha}_{ab,k}} \right) dx'_n \] (30)
We use the substitution
\[
\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} ax^2 + bx \right) \Phi(cx + d) = \frac{1}{\sqrt{a}} \exp \left( \frac{b^2}{2a} \right) \Phi \left( \frac{bc + ad}{\sqrt{a^2 + ac^2}} \right)
\]
(31)
and get the final form,
\[
\langle \Theta_{ab} \exp (\beta x_n) \rangle_k = \exp (\beta \mu_k) \sqrt{\frac{\det \tilde{C}_k}{\det C_\sigma}} \left\{ \frac{1}{\sqrt{p}} \exp \left( \frac{q^2}{2p} \right) \Phi \left( \frac{qr + ps}{\sqrt{p^2 + ps}} \right) \right\}
\]
(32)
with
\[
p = (C_\sigma^{-1})_{nn} - (C_\sigma^{-1})_{[n,n]}^T \tilde{C}_k (C_\sigma^{-1})_{[n,n]}, \quad q = \beta,
\]
\[
r = \frac{(\alpha_{ab})_n - (\alpha_{ab})_n \tilde{C}_k (C_\sigma^{-1})_{[n,n]}}{\tilde{\alpha}_{ab,k}}, \quad s = \frac{\tilde{\beta}_{ab,k}}{\tilde{\alpha}_{ab,k}}
\]
(33)