Supplementary Material: Non-Linear Stationary Subspace Analysis with Application to Video Classification

Mahsa Baktashmotlagh[†‡] MAHSA.BAKTASHMOTLAGH@NICTA.COM.AU

Mehrtash T. Harandi[‡] MEHRTASH.HARANDI@NICTA.COM.AU

Abbas Bigdeli[‡] ABBAS.BIGDELI@NICTA.COM.AU

Brian C. Lovell[†] LOVELL@ITEE.UQ.EDU.AU

Mathieu Salzmann[‡] MATHIEU.SALZMANN@NICTA.COM.AU

[†]University of Queensland, School of ITEE, QLD 4072, Australia
[‡]NICTA*, Locked Bag 8001, Canberra, ACT 2601, Australia

1. Bounds for KSSA and NLSSA

In Analytic SSA, an approximate upper bound to the logdet term in Eq. 4 is computed (Hara et al., 2012). Without re-deriving this bound entirely, we present its final steps, which we use to obtain the bounds for KSSA and NLSSA.

Following (Hara et al., 2012), let us denote by $\tilde{f}(\boldsymbol{W}, \boldsymbol{W}_*)$ the second order Taylor approximation of the logdet term around the optimal solution $\boldsymbol{W}_*$. It can be shown that

$$\tilde{f}(\boldsymbol{W}, \boldsymbol{W}^*) \leq \frac{2}{N}\sum_{i=1}^{N} \mathrm{Tr}\left(\boldsymbol{W}(\boldsymbol{\Sigma}_i - \bar{\boldsymbol{\Sigma}})\boldsymbol{W}_*^T\boldsymbol{W}_*(\boldsymbol{\Sigma}_i - \bar{\boldsymbol{\Sigma}})\boldsymbol{W}^T\right).$$

Let $\boldsymbol{B}_*^T = \left[\boldsymbol{W}_*^T, \boldsymbol{B}_*^{nT}\right]$, where $\boldsymbol{B}_*^n$ is the orthogonal complement of $\boldsymbol{W}_*$. Since

$$\mathrm{Tr}\left(\boldsymbol{W}(\boldsymbol{\Sigma}_i - \bar{\boldsymbol{\Sigma}})\boldsymbol{B}_*^{nT}\boldsymbol{B}_*^n(\boldsymbol{\Sigma}_i - \bar{\boldsymbol{\Sigma}})\boldsymbol{W}^T\right) \geq 0\,,$$

the bound can be re-written as

$$\tilde{f}(\boldsymbol{W}, \boldsymbol{W}_*) \leq \frac{2}{N}\sum_{i=1}^{N} \mathrm{Tr}\left(\boldsymbol{W}(\boldsymbol{\Sigma}_i - \bar{\boldsymbol{\Sigma}})\boldsymbol{B}_*^T\boldsymbol{B}_*(\boldsymbol{\Sigma}_i - \bar{\boldsymbol{\Sigma}})\boldsymbol{W}^T\right). \tag{1}$$

The constraint $\boldsymbol{B}_*\bar{\boldsymbol{\Sigma}}\boldsymbol{B}_*^T = \boldsymbol{I}$ yields $\boldsymbol{B}_*^T\boldsymbol{B}_* = \bar{\boldsymbol{\Sigma}}^{-1}$, which results in the final bound

$$\tilde{f}(\boldsymbol{W}, \boldsymbol{W}_*) \leq \frac{2}{N}\sum_{i=1}^{N} \mathrm{Tr}\left(\boldsymbol{W}(\boldsymbol{\Sigma}_i - \bar{\boldsymbol{\Sigma}})\bar{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\Sigma}_i - \bar{\boldsymbol{\Sigma}})\boldsymbol{W}^T\right).$$

KSSA: In kernel space, we can write $\boldsymbol{W} = \boldsymbol{\alpha}\Phi(\boldsymbol{V})^T$, with $\boldsymbol{\alpha}$ unknown. Similarly, for the entire subspace at the optimal solution, we have $\boldsymbol{B}_* = \hat{\boldsymbol{\alpha}}_*\Phi(\boldsymbol{V})^T$, with $\hat{\boldsymbol{\alpha}}_* = [\boldsymbol{\alpha}_*^T, \boldsymbol{\alpha}_*^{nT}]^T$ and $\boldsymbol{\alpha}_*^n$ the coefficients of the orthogonal complement of $\boldsymbol{W}_*$. From Eq. 1, and by recalling that the covariances are now defined in kernel space, we have

$$\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\alpha}_*) \leq \frac{2}{N}\sum_{i=1}^{N} \mathrm{Tr}\left(\boldsymbol{\alpha}(\tilde{\boldsymbol{K}}_i - \bar{\boldsymbol{K}})\hat{\boldsymbol{\alpha}}_*^T\hat{\boldsymbol{\alpha}}_*(\tilde{\boldsymbol{K}}_i - \bar{\boldsymbol{K}})\boldsymbol{\alpha}^T\right).$$

With the constraint $\boldsymbol{B}_*\bar{\boldsymbol{\Sigma}}\boldsymbol{B}_*^T = \hat{\boldsymbol{\alpha}}_*\bar{\boldsymbol{K}}\hat{\boldsymbol{\alpha}}_*^T = \boldsymbol{I}$, we can see that $\hat{\boldsymbol{\alpha}}_*^T\hat{\boldsymbol{\alpha}}_* = \bar{\boldsymbol{K}}^{-1}$, which yields the final bound

$$\tilde{f}(\boldsymbol{\alpha}, \boldsymbol{\alpha}_*) \leq \frac{2}{N}\sum_{i=1}^{N} \mathrm{Tr}\left(\boldsymbol{\alpha}(\tilde{\boldsymbol{K}}_i - \bar{\boldsymbol{K}})\bar{\boldsymbol{K}}^{-1}(\tilde{\boldsymbol{K}}_i - \bar{\boldsymbol{K}})\boldsymbol{\alpha}^T\right).$$

NLSSA: The derivation of the bound for NLSSA directly follows from the fact that $\bar{\boldsymbol{\Sigma}}^P = \boldsymbol{I}$, which thus yields

$$\tilde{f}(\boldsymbol{W}, \boldsymbol{W}_*) \leq \frac{2}{N}\sum_{i=1}^{N} \mathrm{Tr}\left(\boldsymbol{W}(\boldsymbol{\Sigma}_i^P - \boldsymbol{I})(\boldsymbol{\Sigma}_i^P - \boldsymbol{I})\boldsymbol{W}^T\right)\,.$$

Moreover, in NLSSA, we assumed that the average of the means over the epochs is 0 in $\mathcal{H}$. This centering can be achieved by re-writing the kernel matrix as $\hat{\boldsymbol{K}} = \boldsymbol{K} - \boldsymbol{1K} - \boldsymbol{K1}^T + \boldsymbol{1K1}^T$, where $\boldsymbol{1}$ is a $\sum_{i=1}^N m_i \times \sum_{i=1}^N m_i$ block matrix with all entries in each block i (corresponding to one video in the class) set to $1/(N \times m_i)$.

References

Hara, S., Kawahara, Y., Washio, T., Bünau, P., Tokunaga, T., and Yumoto, K. Separation of stationary and non-stationary sources with a generalized eigenvalue problem. *Neural networks*, 2012.