Supplementary Material for RAID-G: Robust Estimation of Approximate Infinite Dimensional Gaussian with Application to Material Recognition

Qilong Wang[1], Peihua Li[1], Wangmeng Zuo[2], Lei Zhang[3]
[1]Dalian University of Technology, [2]Harbin Institute of Technology, [3]Hong Kong Polytechnic University
qlwang@mail.dlut.edu.cn, peihuali@dlut.edu.cn, wmzuo@hit.edu.cn, cslzhang@comp.polyu.edu.hk

Appendix A. Proof of Theorem 1

Let

$$\begin{aligned}\mathbf{f}(\widehat{\mathbf{\Sigma}}) &= \log|\widehat{\mathbf{\Sigma}}| + \mathrm{tr}(\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{S}}) + \alpha D_{\mathrm{vN}}(\mathbf{I},\widehat{\mathbf{\Sigma}}), \\ &= \log|\widehat{\mathbf{\Sigma}}| + \mathrm{tr}(\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{S}}) + \alpha\mathrm{tr}(\log(\mathbf{I}) - \log(\widehat{\mathbf{\Sigma}}) - \mathbf{I} + \widehat{\mathbf{\Sigma}}), \\ &= (1-\alpha)\log|\widehat{\mathbf{\Sigma}}| + \mathrm{tr}(\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{S}}) + \alpha\mathrm{tr}(\widehat{\mathbf{\Sigma}}) - \mathrm{tr}(\mathbf{I}).\end{aligned} \tag{1}$$

We discard $\mathrm{tr}(\mathbf{I})$ in Eq. (1), which is irrelevant to $\widehat{\mathbf{\Sigma}}$. Hence, $\mathbf{f}(\widehat{\mathbf{\Sigma}})$ can be written as

$$\mathbf{f}(\widehat{\mathbf{\Sigma}}) = (1-\alpha)\log|\widehat{\mathbf{\Sigma}}| + \mathrm{tr}(\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{S}}) + \alpha\mathrm{tr}(\widehat{\mathbf{\Sigma}}). \tag{2}$$

In the following, we first transform the minimization of $f(\widehat{\mathbf{\Sigma}})$ to that of its lower bound function, and then prove that $f(\widehat{\mathbf{\Sigma}})$ has a unique nonnegative solution and the solution we obtained in the first step is the global optimization of $f(\widehat{\mathbf{\Sigma}})$ (Eq. (2)).Let the SVD of $\widehat{\mathbf{S}} = \widehat{\mathbf{U}}\mathbf{\Delta}\widehat{\mathbf{U}}^T$ and $\widehat{\mathbf{\Sigma}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{\Delta}$ and $\mathbf{\Lambda}$ are diagonal matrices with diagonal entries being $\{\delta_1 \geq \delta_2 \ldots \geq \delta_d > 0\}$ and $\{\lambda_1 \geq \lambda_2 \ldots \geq \lambda_d > 0\}$, respectively. Then Eq. (2) becomes

$$\begin{aligned}\mathbf{f}(\widehat{\mathbf{\Sigma}}) &= \gamma\log|\mathbf{\Lambda}| + \mathrm{tr}(\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T\widehat{\mathbf{U}}\mathbf{\Delta}\widehat{\mathbf{U}}^T) + \alpha\mathrm{tr}(\mathbf{\Lambda}). \\ &\geq \gamma\log|\mathbf{\Lambda}| + \mathrm{tr}(\mathbf{\Lambda}^{-1}\mathbf{\Delta}) + \alpha\mathrm{tr}(\mathbf{\Lambda}),\end{aligned} \tag{3}$$

where $\gamma = 1-\alpha$. Note that $\mathrm{tr}(\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T\widehat{\mathbf{U}}\mathbf{\Delta}\widehat{\mathbf{U}}^T) \geq \mathrm{tr}(\mathbf{\Lambda}^{-1}\mathbf{\Delta})$ [1, Fact 5.12.4], and the equality holds if $\widehat{\mathbf{U}} = \mathbf{U}$. Therefore, instead of solving the objective function in $\mathbf{f}(\widehat{\mathbf{\Sigma}})$, we solve the following optimization problem:

$$\min_{\mathbf{\Lambda}} \sum_i \gamma\log\lambda_i + \lambda_i^{-1}\delta_i + \alpha\lambda_i, \tag{4}$$

where γ is nonnegative and $\log\lambda_i$ is concave. We let $\omega_i = \lambda_i^{-1}$, and the problem in Eq. (4) can be reformulated as:

$$\min_{\boldsymbol{\omega}} \sum_i -\gamma\log\omega_i + \omega_i\delta_i + \alpha\omega_i^{-1}. \tag{5}$$

The three components of Eq. (5) are convex, and thus the problem in Eq. (5) is convex and has global optimal solution. Moreover, the problem in Eq. (5) can be further decomposed into d independent subproblems:

$$\min_{\omega_i} -\gamma\log\omega_i + \omega_i\delta_i + \alpha\omega_i^{-1}. \tag{6}$$

It is easy to see that the global optimum of ω_i can be obtained at $\frac{\partial\mathbf{c}_i(\omega_i)}{\partial\omega_i} = 0 \ \ (\omega_i > 0)$, i.e.,

$$\begin{aligned}&\frac{\partial\mathbf{c}_i(\omega_i)}{\partial\omega_i} = 0 \Rightarrow -\gamma\omega_i^{-1} + \delta_i - \alpha\omega_i^{-2} = 0 \\ &\Rightarrow \alpha\lambda_i^2 + (1-\alpha)\lambda_i - \delta_i = 0.\end{aligned} \tag{7}$$

It is clear that Eq. (7) is a quadratic equation with one unknown, and its unique positive solution is

$$\lambda_i = \sqrt{\left(\frac{1-\alpha}{2\alpha}\right)^2 + \frac{\delta_i}{\alpha}} - \frac{1-\alpha}{2\alpha}. \tag{8}$$

By computing the derivative of Eq. (2) with respect to $\widehat{\mathbf{\Sigma}}$ and setting it to zero, we can obtain

$$\partial\mathbf{f}/\partial\widehat{\mathbf{\Sigma}} = (1-\alpha)\widehat{\mathbf{\Sigma}}^{-1} - \widehat{\mathbf{\Sigma}}^{-1}\mathbf{S}\widehat{\mathbf{\Sigma}}^{-1} + \alpha\mathbf{I} = \mathbf{0}, \tag{9}$$

which is an Algebraic Riccati equation (ARE) in the unknown $\widehat{\mathbf{\Sigma}}^{-1}$ [1]. According to [2, Sec. II] and control theory, we know that this ARE has an unique nonnegative solution. Moreover, it is not difficult to know that

$$\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{U}}\mathrm{diag}(\lambda_i)\widehat{\mathbf{U}}^T, \tag{10}$$

where λ_i is given in Eq. (8), satisfies the ARE (Eq. (9)),i.e.,

$$\begin{aligned}\partial\mathbf{f}/\partial\widehat{\mathbf{\Sigma}} &= (1-\alpha)\widehat{\mathbf{U}}\mathbf{\Lambda}^{-1}\widehat{\mathbf{U}}^T + \alpha\mathbf{I} \\ &- \widehat{\mathbf{U}}\mathbf{\Lambda}^{-1}\widehat{\mathbf{U}}^T\widehat{\mathbf{U}}\mathbf{\Delta}\widehat{\mathbf{U}}^T\widehat{\mathbf{U}}\mathbf{\Lambda}^{-1}\widehat{\mathbf{U}}^T \\ &= \widehat{\mathbf{U}}\big((1-\alpha)\mathbf{\Lambda}^{-1} - \mathbf{\Lambda}^{-1}\mathbf{\Delta}\mathbf{\Lambda}^{-1} + \alpha\mathbf{I}\big)\widehat{\mathbf{U}}^T \\ &= \mathbf{0}.\end{aligned} \tag{11}$$

Hence, we can conclude that $\widehat{\mathbf{\Sigma}}$ (10) is the unique optimal solution of objective function (2).

[1]We thank the stimulating comment of the anonymous reviewer who pointed out Eq. (9) may be a Riccati equation.

References

[1] D. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas (2nd ed.)*. Princeton University Press, 2009.

[2] A. Laub. A Schur method for solving algebraic Riccati equations. *IEEE Trans. on Automat. Contr.*, 24:913–921, 1979.