Errata for **Learning output kernels with block coordinate descent**

Differently from what claimed in Lemma 4.1 of [1], the update for $\mathbf{L}$ does not preserve symmetry. However, this doesn't compromise the validity of Algorithm 1 as a way to solve the original optimization problem. The analysis can be fixed as follows:

- Equation (7) is replaced by

$$\arg\min_{\mathbf{L}\in\mathbb{S}_+^m} Q\left(\mathbf{L},\mathbf{C}\right) = \arg\min_{\mathbf{L}\in\mathbb{S}^m} Q\left(\mathbf{L},\mathbf{C}\right),$$

- Equation (8) (and line 6 of Algorithm 1) are replaced by the Lyapunov equation:

$$\left(\frac{\mathbf{E}^T\mathbf{E}+\lambda\mathbf{I}}{2}\right)\mathbf{Q}+\mathbf{Q}^T\left(\frac{\mathbf{E}^T\mathbf{E}+\lambda\mathbf{I}}{2}\right)=\mathbf{P},\quad \mathbf{P}:=\frac{1}{2}\mathbf{E}^T\mathbf{C}-\mathbf{L}_p.$$

- Finally, the proof of Lemma 4.1 is modified as follows:

  Proof: Observe that matrix $\mathbf{L}$ is optimal for the problem on the right hand side of (7) if and only if

$$\mathbf{0}=\frac{\partial Q}{\partial \mathbf{L}}\left(\frac{\mathbf{L}+\mathbf{L}^T}{2},\mathbf{C}\right)=-\frac{\mathbf{C}^T\mathbf{K}\left(\mathbf{Y}-\lambda\mathbf{C}/2-\mathbf{KCL}\right)}{2\lambda}+\frac{\mathbf{L}}{2},$$
$$+-\frac{\left(\mathbf{Y}-\lambda\mathbf{C}/2-\mathbf{KCL}\right)^T\mathbf{KC}}{2\lambda}+\frac{\mathbf{L}^T}{2}.$$

  Now, recall that $\mathbf{C}$ satisfies

$$\mathbf{Y}-\frac{\lambda}{2}\mathbf{C}=\frac{\lambda}{2}\mathbf{C}+\mathbf{KCL}_p,$$

  where $\mathbf{L}_p$ denote the previous $\mathbf{L}$, which is positive semidefinite. Hence, we obtain the following Lyapunov equation:

$$\left(\mathbf{E}^T\mathbf{E}+\lambda\mathbf{I}\right)\mathbf{L}+\mathbf{L}\left(\mathbf{E}^T\mathbf{E}+\lambda\mathbf{I}\right)=\lambda\mathbf{C}^T\mathbf{KC}+\mathbf{E}^T\mathbf{EL}_p+\mathbf{L}_p\mathbf{E}^T\mathbf{E}.$$

  Since the right hand side is a symmetric positive semidefinite matrix, and $\left(\mathbf{E}^T\mathbf{E}+\lambda\mathbf{I}\right)$ is positive semidefinite, standard theory for stable Lyapunov equations [2] ensure that there exists a unique symmetric and positive semidefinite solution $\mathbf{L}$. Now, let's rewrite the equation in the form:

$$\mathbf{E}^T\mathbf{E}\left(\frac{\mathbf{L}-\mathbf{L}_p}{\lambda}\right)+\left(\frac{\mathbf{L}-\mathbf{L}_p}{\lambda}\right)^T\mathbf{E}^T\mathbf{E}+\left(\mathbf{L}+\mathbf{L}^T\right)=\mathbf{C}^T\mathbf{KC}.$$

  Letting $\mathbf{L}=\mathbf{L}_p+\lambda\mathbf{Q}$, we obtain (8).

Remark 0.1 *Observe that the original version of Algorithm 1 correctly solves the problem, even if the update for* $\mathbf{L}$ *doesn't preserve symmetry. The reason is that the original Algorithm 1 minimizes the unconstrained functional* Q*, and all the stationary points of* Q *satisfy*

$$\mathbf{L} = \frac{1}{2}\mathbf{C}^T\mathbf{K}\mathbf{C}.$$

Therefore, the sequence of matrices $\mathbf{L}$ *generated by the original version of Algorithm 1 asymptotically approaches a feasible matrix for the constrained problem.*

References

[1] F. Dinuzzo, C. S. Ong, P. Gehler, and G. Pillonetto. Learning output kernels with block coordinate descent. In *Proceedings of the 28th Annual International Conference on Machine Learning*, Bellevue, WA, USA, 2011.

[2] V. Sima. *Algorithms for Linear-quadratic Optimization.* Marcel Dekker, New York, 1996.