Errata for Learning output kernels with block coordinate descent

Differently from what claimed in Lemma 4.1 of [1], the update for $L$ 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
  \[
  \arg\min_{L \in \mathbb{S}^m_+} Q(L, C) = \arg\min_{L \in \mathbb{S}^m} Q(L, C),
  \]

- Equation (8) (and line 6 of Algorithm 1) are replaced by the Lyapunov equation:
  \[
  \left( \frac{E^T E + \lambda I}{2} \right) Q + Q^T \left( \frac{E^T E + \lambda I}{2} \right) = P, \quad P := \frac{1}{2} E^T C - L_p.
  \]

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

  **Proof:** Observe that matrix $L$ is optimal for the problem on the right hand side of (7) if and only if
  \[
  0 = \frac{\partial Q}{\partial L} \left( \frac{L + L^T}{2}, C \right) = - \frac{C^T K (Y - \lambda C/2 - KCL)}{2\lambda} + \frac{L}{2}^T + \frac{L}{2} + \frac{E^T (Y - \lambda C/2 - L_p) L_p E}{2}. \]

  Now, recall that $C$ satisfies
  \[
  Y - \frac{\lambda}{2} C = \frac{\lambda}{2} C + KCL_p,
  \]
  where $L_p$ denote the previous $L$, which is positive semidefinite. Hence, we obtain the following Lyapunov equation:
  \[
  (E^T E + \lambda I) L + L (E^T E + \lambda I) = \lambda C^T K C + E^T E L_p + L_p E^T E.
  \]
  Since the right hand side is a symmetric positive semidefinite matrix, and $(E^T E + \lambda I)$ is positive semidefinite, standard theory for stable Lyapunov equations [2] ensure that there exists a unique symmetric and positive semidefinite solution $L$. Now, let’s rewrite the equation in the form:
  \[
  E^T E \left( \frac{L - L_p}{\lambda} \right) + \left( \frac{L - L_p}{\lambda} \right)^T E^T E + (L + L^T) = C^T K C.
  \]
  Letting $L = L_p + \lambda Q$, we obtain (8).
Remark 0.1 Observe that the original version of Algorithm 1 correctly solves the problem, even if the update for $L$ 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

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

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

References
