Comments and Corrections

Corrections to "Sparse-Group Lasso for Graph Learning From Multi-Attribute Data"

Jitendra K. Tugnait, *Life Fellow, IEEE*

We correct errors in step 6 of Algorithm 1 given in [1]. Other parts of the paper including numerical results are unchanged as the algorithm was implemented correctly.

Step 6 of Algorithm 1 in [1] appears as follows.

"Define soft thresholding scalar operator $S(a,\beta) := (1-\beta/|a|)_+ a$ where $(a)_+ := \max(0,a)$. The diagonal $m\times m$ subblocks of $\boldsymbol{W}$ are updated as

$$[(\boldsymbol{W}^{(jj)})^{(i+1)}]_{st} = \begin{cases} [(\boldsymbol{\Omega}^{(jj)})^{(i+1)}]_{ss} & \text{if } s=t \\ S([(\boldsymbol{\Omega}^{(jj)})^{(i+1)}]_{st}, \frac{\alpha\lambda}{\rho^{(i)}}) & \text{if } s\neq t \end{cases}$$

$j=1,2,\ldots,p,\quad s,t=1,2,\ldots,m$. The off-diagonal $m\times m$ subblocks of $\boldsymbol{W}$ are updated as (denote $\boldsymbol{A} = (\boldsymbol{\Omega}^{(jk)})^{(i+1)} - (\boldsymbol{U}^{(jk)})^{(i)}$)

$$[(\boldsymbol{W}^{(jk)})^{(i+1)}]_{st} = \begin{cases} [(\boldsymbol{\Omega}^{(jk)})^{(i+1)}]_{ss} & \text{if } s=t \\ S([\boldsymbol{A}]_{st}, \frac{\alpha\lambda}{\rho^{(i)}}) \left(1 - \frac{(1-\alpha)\lambda}{\rho \|\boldsymbol{S}(\boldsymbol{A}, \frac{\alpha\lambda}{\rho^{(i)}})\|_F}\right)_+ & \text{if } s\neq t \end{cases}$$

where $\boldsymbol{S}(\boldsymbol{A},\alpha)$ denotes elementwise matrix soft thresholding, specified by $[\boldsymbol{S}(\boldsymbol{A},\alpha)]_{st} := S([\boldsymbol{A}]_{st},\alpha)$, and $j\neq k = 1,2,\ldots,p,\quad s,t=1,2,\ldots,m$."

The corrected step 6 is as follows.

"Set $\boldsymbol{A}^{(jk)} = (\boldsymbol{\Omega}^{(jk)})^{(i+1)} + (\boldsymbol{U}^{(jk)})^{(i)}$. Define soft thresholding scalar operator $S(a,\beta) := (1-\beta/|a|)_+ a$ where $(a)_+ := \max(0,a)$. The diagonal $m\times m$ subblocks of $\boldsymbol{W}$ are updated as

$$[(\boldsymbol{W}^{(jj)})^{(i+1)}]_{st} = \begin{cases} [\boldsymbol{A}^{(jj)}]_{ss} & \text{if } s=t \\ S([\boldsymbol{A}^{(jj)}]_{st}, \frac{\alpha\lambda}{\rho^{(i)}}) & \text{if } s\neq t \end{cases}$$

$j=1,2,\ldots,p,\quad s,t=1,2,\ldots,m$. The off-diagonal $m\times m$ subblocks of $\boldsymbol{W}$ are updated as

$$(\boldsymbol{W}^{(jk)})^{(i+1)} = \boldsymbol{B}\left(1 - \frac{(1-\alpha)\lambda}{\rho^{(i)}\|\boldsymbol{B}\|_F}\right)_+$$

where $m\times m$ matrix $\boldsymbol{B} = \boldsymbol{S}(\boldsymbol{A}^{(jk)}, \alpha\lambda/\rho^{(i)})$, $\boldsymbol{S}(\boldsymbol{A},\alpha)$ denotes elementwise matrix soft thresholding, specified by $[\boldsymbol{S}(\boldsymbol{A},\alpha)]_{st} := S([\boldsymbol{A}]_{st},\alpha)$, and $j\neq k=1,2,\ldots,p$."

The errors are related to update (b) of the ADMM algorithm discussed on p. 1774 (first column) of the paper. We now explain the corrections. We need to solve $(\boldsymbol{W}^{(jk)})^{(i+1)} \leftarrow \arg\min_{\boldsymbol{W}^{(jk)}} J_{bjk}(\boldsymbol{W}^{(jk)})$, for subblock indexed by (j,k), where, for $j=k$,

$$J_{bjj}(\boldsymbol{W}^{(jj)}) := \alpha\lambda \|(\boldsymbol{W}^{(jj)})^-\|_1 + \frac{\rho}{2}\left\| \left(\boldsymbol{\Omega}^{(i+1)} - \boldsymbol{W} + \boldsymbol{U}^{(i)}\right)^{(jj)} \right\|_F^2$$

and for $j\neq k$,

$$J_{bjk}(\boldsymbol{W}^{(jk)}) := \alpha\lambda \|\boldsymbol{W}^{(jk)}\|_1 + (1-\alpha)\lambda \|\boldsymbol{W}^{(jk)}\|_F + \frac{\rho}{2}\left\| \left(\boldsymbol{\Omega}^{(i+1)} - \boldsymbol{W} + \boldsymbol{U}^{(i)}\right)^{(jk)} \right\|_F^2$$

For $j=k$, the solution is the standard lasso solution given by the first equation in the corrected step 6, where in the original version, the error is in using $\boldsymbol{\Omega}^{(jj)}$ instead of $\boldsymbol{A}^{(jj)} = (\boldsymbol{\Omega}^{(jj)})^{(i+1)} + (\boldsymbol{U}^{(jj)})^{(i)}$. For $j\neq k$, we have sparse-group lasso, and following [2], [3] (see also [4]), the solution to minimization of $J_{bjk}(\boldsymbol{W}^{(jk)})$ w.r.t. $m\times m$ $\boldsymbol{W}^{(jk)}$ is given by the second equation in the corrected step 6. In the original version of the paper, there are two errors: we incorrectly defined $\boldsymbol{A} = (\boldsymbol{\Omega}^{(jk)})^{(i+1)} - (\boldsymbol{U}^{(jk)})^{(i)}$ which should have been $\boldsymbol{A} = (\boldsymbol{\Omega}^{(jk)})^{(i+1)} + (\boldsymbol{U}^{(jk)})^{(i)}$ (we now denote this $\boldsymbol{A}$ as $\boldsymbol{A}^{(jk)}$), and the diagonal terms $s=t$ of the off-diagonal $m\times m$ subblocks $\boldsymbol{W}^{(jk)}$ were incorrectly taken to have no penalties. Only the diagonal terms $s=t$ of the diagonal $m\times m$ subblocks $\boldsymbol{W}^{(jj)}$ have no penalties (lasso or group-lasso).

We note that the other parts of the paper including numerical results are unchanged as the algorithm was implemented correctly (it was implemented and debugged well before the paper was written).

References

[1] J. K. Tugnait, "Sparse-group lasso for graph learning from multi-attribute data," *IEEE Trans. Signal Process.*, vol. 69, pp. 1771–1786, 2021, doi: 10.1109/TSP.2021.3057699.
[2] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," *J. Comput. Graphical Statist.*, vol. 22, pp. 231–245, 2013.
[3] J. Friedman, T. Hastie, and R. Tibshirani, "A note on the group lasso and a sparse group lasso," 2010, *arXiv:1001.0736v1*.
[4] P. Danaher, P. Wang, and D. M. Witten, "The joint graphical lasso for inverse covariance estimation across multiple classes," *J. Roy. Stat. Society, Ser. B. (Methodological)*, vol. 76, pp. 373–397, 2014.

Manuscript received July 9, 2021; accepted August 10, 2021. Date of publication August 13, 2021; date of current version August 27, 2021. This work was supported by National Science Foundation under Grants CCF-1617610 and ECCS-2040536. The associate editor coordinating the review of this paper and approving it for publication was Prof. Justin Dauwels.

The author is with the Department of Electrical & Computer Engineering Auburn University, Auburn, AL 36849 USA (e-mail: tugnajk@eng.auburn.edu).

Digital Object Identifier 10.1109/TSP.2021.3104727