Submitted to *INFORMS Journal on Computing*
manuscript (Please, provide the manuscript number!)

Online Supplement for Convex Optimization for Group Feature Selection of Networked Data

Daehan Won
Department of Systems Science & Industrial Engineering, Binghamton University, the State University of New York, NY, dhwon@binghamton.edu

Hasan Manzour
Department of Industrial & Systems Engineering, University of Washington, Seattle, WA, hmanzour@uw.edu

W. Art Chaovalitwongse
Institute for Advanced Data Analytics, Department of Industrial Engineering, University of Arkansas, Fayetteville, AR, artchao@uark.edu

Appendix

Analysis of Convergence

In this section, we analyze the convergence property of Algorithm 1. Assume the current iteration step is l. At the l-th iteration, a new constraint $\mathbf{s}_l$ below can be found via the CGP derived by $\boldsymbol{\alpha}^{l-1}$

$$h(\boldsymbol{\alpha}^{l-1}, \mathbf{s}_l) = \max_{\mathbf{s} \in \mathcal{E}(\mathcal{V})} h(\boldsymbol{\alpha}^{l-1}, \mathbf{s}). \tag{1}$$

As a result, the new constraint set F_l is given by $F_l = F_{l-1} \cup \{\mathbf{s}_l\}$.

Define $ub^l = \min_{1 \leq k \leq l} h(\boldsymbol{\alpha}^{k-1}, \mathbf{s}_k)$ and (η^l, $\boldsymbol{\alpha}^l$) as the intermediate solution pair to Problem (12), which is obtained by

$$\boldsymbol{\alpha}^l = \arg\min_{\boldsymbol{\alpha} \in A} \left(\max_{\mathbf{s}_t \in F_l} h(\boldsymbol{\alpha}, \mathbf{s}_t) \right), \tag{2}$$

$$\eta^l = \max_{1 \leq k \leq l} h(\boldsymbol{\alpha}^l, \mathbf{s}_k) = \min_{\alpha \in A} \left(\max_{1 \leq k \leq l} h(\boldsymbol{\alpha}, \mathbf{s}_k) \right). \tag{3}$$

For notational convenience, we introduce lb^l such that $\eta^l = lb^l$. We obtain

$$lb^l = \min_{\boldsymbol{\alpha}\in A}\max_{1\le k\le l} h(\boldsymbol{\alpha}, \mathbf{s}_k), \tag{4}$$

$$ub^l = \min_{1\le k\le l} h(\boldsymbol{\alpha}^{k-1}, \mathbf{s}_k) = \min_{1\le k\le l}\Big(\max_{\mathbf{s}\in\mathcal{E}(\mathcal{V})} h(\boldsymbol{\alpha}^k, \mathbf{s})\Big). \tag{5}$$

Based on the definitions of lb^l and ub^l, we present the following two lemmas to prove that Algorithm 1 converges to the optimal solution.

LEMMA 1. *Suppose* (η^{opt}, $\boldsymbol{\alpha}^{opt}$) *is an optimal solution pair of Problem* (11). *Inequality* $lb^l \le \eta^{opt} \le ub^l$ *holds at the* l*-th iteration. As iteration step increases, the sequence* $\{lb^l\}$ *is monotonically increasing and* $\{ub^l\}$ *is monotonically decreasing.*

Proof. Firstly, we have the optimal solution $\eta^{opt} = \min_{\boldsymbol{\alpha}\in A}\max_{\mathbf{s}\in\mathcal{E}(\mathcal{V})} h(\alpha, \mathbf{s})$ from Problem (11). For any feasible $\boldsymbol{\alpha}$ in Problem (11), we have $\max_{\mathbf{s}\in F_l} h(\boldsymbol{\alpha}, \mathbf{s}) \le \max_{\mathbf{s}\in\mathcal{E}(\mathcal{V})} h(\boldsymbol{\alpha}, \mathbf{s})$ under $F_l \subseteq \mathcal{E}(\mathcal{V})$. With consideration of $\boldsymbol{\alpha}$, we obtain:

$$\min_{\boldsymbol{\alpha}\in A}\max_{\mathbf{s}\in F_l} h(\boldsymbol{\alpha}, \mathbf{s}) \le \min_{\boldsymbol{\alpha}\in A}\max_{\mathbf{s}\in\mathcal{E}(\mathcal{V})} h(\boldsymbol{\alpha}, \mathbf{s}).$$

By Eq. (4), we have $lb^l \le \eta^{opt}$.

On the other hand, from Problem (1), $\mathbf{s}_k = \arg\max_{\mathbf{s}\in\mathcal{E}(\mathcal{V})} h(\boldsymbol{\alpha}^{k-1}, \mathbf{s})$ for $k = 1, ..., l$. Thus, $\{(\boldsymbol{\alpha}^{k-1}, h(\boldsymbol{\alpha}^{k-1}, \mathbf{s}_k))\}_{k=1}^{l}$ is a set of feasible solutions to Problem (11). Since ($\boldsymbol{\alpha}^{opt}, \eta^{opt}$) is the optimal solution to Problem (11), any feasible solution $h(\boldsymbol{\alpha}^{k-1}, \mathbf{s}_k))$ satisfies $\eta^{opt} \le h(\boldsymbol{\alpha}^{k-1}, \mathbf{s}_k)$, for $k = 1, ..., l$. This follows from Eq. (5) and finally we have $\eta^{opt} \le ub^l$.

Localized set F_l is monotonically expanding with increasing iteration step l. Therefore, the sequence $\{lb^l\} = \{\eta^l\}$ is also monotonically increasing. Along the same line, the sequence $\{ub_l\}$ is monotonically decreasing. ■

According to the Lemma 1, the convergence can be traced by the gap between lb^l and ub^l. For example, if this gap is less than a predetermined tolerance, the Algorithm 1 is terminated.

LEMMA 2. *Assume that our algorithm converges at the* l*-th step. The intermediate solution* ($\eta^{l-1}, \boldsymbol{\alpha}^{l-1}$) *is the global optimal solution to Problem* (11).

Proof. Once the algorithm is terminated at the l-th iteration, there is no more update from the $(l-1)$-th step to l-th step. Thus, $F_l = F_{l-1}$ and $\boldsymbol{\alpha}^l = \alpha^{l-1}$. A new constraint $\mathbf{s}_l$ at the l-th iteration is given by $\mathbf{s}_l = \arg\max_{\mathbf{s}\in\mathcal{E}(\mathcal{V})} h(\boldsymbol{\alpha}^l)$. From Problem (1), we have

$h(\boldsymbol{\alpha}^{l-1}, \mathbf{s}_l) = \max_{\mathbf{s}\in\mathcal{E}(\mathcal{V})} h(\boldsymbol{\alpha}^{l-1}, \mathbf{s})$ and it is equivalent to $h(\boldsymbol{\alpha}^{l-1}, \mathbf{s}_l) = \max_{\mathbf{s}\in F_l} h(\boldsymbol{\alpha}^{l-1}, \mathbf{s})$ as the new constraint $\mathbf{s}_l \in F_l = F_{l-1}$. Following from Eqs. (4) and (5), we obtain

$$\begin{aligned} h(\boldsymbol{\alpha}^{l-1}, \mathbf{s}_l) &= \max_{\mathbf{s}\in\mathcal{E}(\mathcal{V})} h(\boldsymbol{\alpha}^{l-1}, \mathbf{s}) = \max_{\mathbf{s}\in F_{l-1}} h(\boldsymbol{\alpha}^{l-1}, \mathbf{s}) = \eta^{l-1} \\ ub^{l-1} &= \min_{1\leq k\leq l} h(\boldsymbol{\alpha}^{k-1}, \mathbf{s}_k) \leq \eta^{l-1} \end{aligned} \tag{6}$$

Since we already proved that $lb^l \leq \eta^{opt} \leq ub^l$ in Lemma 1, we obtain $\eta^l = lb^l$. Consequently, we conclude that $lb^{l-1} = ub^{l-1} = \eta^{l-1} = \eta^{opt}$ and $(\eta^{l-1}, \boldsymbol{\alpha}^{l-1})$ is the optimal solution to Problem (11). ■