Supplementary Materials

Label co-occurrence guided nonlinear disambiguation for partial multi-label learning

Yining Song[1,#], Fuyu Qi[1,#], Jinfu Fan[1], Jian Feng[1], Zhiyong Li[1], Qingkai Bu[1], Wenpeng Lu[2,3], Linqing Huang[4]

[1]School of Computer Science and Technology, Qingdao University, Qingdao 266071, Shandong, China.
[2]Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250000, Shandong, China.
[3]Shandong Provincial Key Laboratory of Computing Power Internet and Service Computing, Shandong Fundamental Research Center for Computer Science, Jinan 250000, Shandong, China.
[4]School of Computer Science, Shanghai Jiao Tong University, Shanghai 200240, China.
[#]These authors contributed equally to this work.

Correspondence to: Prof. Linqing Huang, School of Computer Science, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: huanglinqing@sjtu.edu.cn

Demonstration of Lemma 1

$$\min_{\{\mathbf{p_i}\}_{i=1}^{n}} \sum_{i=1}^{n} \sqrt{\mathbf{p_i}'\mathbf{K}\mathbf{p}_i} + \frac{\tau}{2}\sum_{i=1}^{n} \|\mathbf{p}_i - \mathbf{c}_i\|^2 \tag{1}$$

The problem can be broken down into n distinct subtasks, with each subtask taking the subsequent structure:

$$\min_{\mathbf{p}_i} \sqrt{\mathbf{p}_i'\mathbf{K}\mathbf{p}_i} + \frac{\tau}{2}\|\mathbf{p}_i - \mathbf{c}_i\|^2 \tag{2}$$

Lemma 1: Considering a positive scalar τ, a constant vector $\mathbf{b} \in \mathbb{R}^q$ (where q is a positive integer), and a diagonal matrix $\mathbf{S} = \mathrm{diag}(\{s_i\}_{1\le i\le q}) \in \mathbb{R}^{q\times q}$ with $\{s_i\}_{i=1}^{q}$ being positive scalars in descending order, the optimal solution $\mathbf{p}^*$ for the problem below can be derived as:

$$\min_{\mathbf{p}\in\mathbb{R}^q} \sqrt{\mathbf{p}^T\mathbf{S}^2\mathbf{p}} + \frac{\tau}{2}\|\mathbf{p} - \mathbf{b}\|^2 \tag{3}$$

is given by

$$\mathbf{p}^* = \begin{cases} \left(\frac{\mathbf{S}^2}{\tau\delta} + \mathbf{I}\right)^{-1}\mathbf{b}, & \text{if} \|\mathbf{S}^{-1}\mathbf{b}\| > \frac{1}{\tau} \\ \mathbf{0}_q, & \text{otherwise} \end{cases} \tag{4}$$

where the inverse of $\mathbf{S}$, denoted as $\mathbf{S}^{-1}$, is given by the diagonal matrix $\mathrm{diag}(s_i^{-1}{}_{i=1}^{q})$, and the positive scalar δ meets the following condition:

$$\mathbf{b}^T\mathrm{diag}\left(\left\{\frac{s_i^2}{(\tau\delta + s_i^2)^2}\right\}_{1\le i\le q}\right)\mathbf{b} = \frac{1}{\tau^2} \tag{5}$$

We refer to the goal of Eq. 3 as:

$$F(\mathbf{p}) = f(\mathbf{p}) + \psi(\mathbf{p}) \tag{6}$$

where $f(\mathbf{p}) = (\mathbf{p}^T\mathbf{S}^2\mathbf{p})^{1/2}$ and $\psi(\mathbf{p}) = \tau/2\|\mathbf{p} - \mathbf{b}\|^2$.

The function $F(\mathbf{p})$ exhibits convexity, and an optimal solution to Eq. 3 aligns with a stationary point of $F(\mathbf{p})$. Consequently, in order to ascertain the optimal resolution for the quandary presented in Eq. 3, we initially compute the subgradient of $F(\mathbf{p})$, followed by the pursuit of its stationary point.

We proceed to compute the subgradients of $f(\mathbf{p})$ and $\psi(\mathbf{p})$ individually. The subgradient of $\psi(\mathbf{p})$ with respect to $\mathbf{p}$ is given by:

$$\partial\psi(\mathbf{p}) = \tau(\mathbf{p} - \mathbf{b}) \tag{7}$$

Regarding $f(\mathbf{p})$, observe that it admits an alternative expression as $f(\mathbf{p}) = \|\mathbf{S}\mathbf{p}\|$. The subgradient of $f(\mathbf{p})$ concerning $\mathbf{p}$ is presented as:

$$\partial f(\mathbf{p}) = \begin{cases} \{\mathbf{S}^T\mathbf{r} \mid \|\mathbf{r}\| \le 1\}, & \text{if } \mathbf{S}\mathbf{p} = \mathbf{0}_q, \\ \frac{\mathbf{S}^T\mathbf{p}}{\|\mathbf{S}\mathbf{p}\|}, & \text{otherwise.} \end{cases} \tag{8}$$

We are aware that the diagonal matrix $\mathbf{S}$, where all diagonal elements are positive, constitutes a positive definite matrix. Consequently, $\mathbf{S}\mathbf{p} = \mathbf{0}_q$ is tantamount to $\mathbf{p} = \mathbf{0}_q$.

In conclusion, the subgradient of $F(\mathbf{p})$ can be articulated as:

$$\partial F(\mathbf{p}) = \begin{cases} \{\mathbf{S}^T\mathbf{r} + \tau(\mathbf{p} - \mathbf{b}) \mid \|\mathbf{r}\| \leq 1\}, & \text{if } \mathbf{p} = \mathbf{0}_q \\ \frac{\mathbf{S}^2\mathbf{p}}{\|\mathbf{S}\mathbf{p}\|} + \tau(\mathbf{p} - \mathbf{b}), & \text{otherwise} \end{cases} \tag{9}$$

Considering the aforementioned scenarios, we shall examine the stationary point of $F(\mathbf{p})$ correspondingly in the two cases delineated below:

1)When $\mathbf{p} = \mathbf{0}_q$, it follows that

$$\partial F(\mathbf{p}) = \{\mathbf{S}^T\mathbf{r} + \tau(\mathbf{p} - \mathbf{b}) \mid \|\mathbf{r}\| \leq 1\} \tag{10}$$

$\mathbf{p}^* = \mathbf{0}_q$ constitutes a stationary point, if and only if

$$\mathbf{0}_q \in \{\mathbf{S}^T\mathbf{r} + \tau(\mathbf{p}^* - \mathbf{b}) \mid \|\mathbf{r}\| \leq 1\} \tag{11}$$

In other words, $\mathbf{p}^* = \mathbf{0}_q$ constitutes a stationary point, if and only if there exists a vector $\mathbf{r}$ that meets the following two conditions:

$$\mathbf{S}^T\mathbf{r} + \tau(\mathbf{p}^* - \mathbf{b}) = \mathbf{0}_q \tag{12}$$

$$\|\mathbf{r}\| \leq 1 \tag{13}$$

Given $\mathbf{p}^* = \mathbf{0}_q$ and $\mathbf{S}$ being positive definite, Eq. 12 equates to $\mathbf{r} = \tau\mathbf{S}^{-1}\mathbf{b}$. By integrating this with the inequality in Eq. 13, we deduce:

$$\|\mathbf{S}^{-1}\mathbf{b}\| \leq \frac{1}{\tau} \tag{14}$$

Consequently, $\mathbf{p}^* = \mathbf{0}_q$ constitutes a stationary point of $F(\mathbf{p})$, if and only if $\|\mathbf{S}^{-1}\mathbf{b}\| \leq (1/\tau)$. Specifically, when $\|\mathbf{S}^{-1}\mathbf{b}\| \leq (1/\tau)$, we find that $\mathbf{r} = \tau\mathbf{S}^{-1}\mathbf{b}$ fulfills Eq. 11.

2) When $\mathbf{p} \neq \mathbf{0}_q$, it follows that

$$\partial F(\mathbf{p}) = \frac{\mathbf{S}^2\mathbf{p}}{\|\mathbf{S}\mathbf{p}\|} + \tau(\mathbf{p} - \mathbf{b}) \tag{15}$$

$\mathbf{p}^* \neq \mathbf{0}_q$ constitutes a stationary point, if and only if $\mathbf{S}^2\mathbf{p}^*/\|\mathbf{S}\mathbf{p}^*\| + \tau(\mathbf{p}^* - \mathbf{b}) = \mathbf{0}_q$. This condition can be reformulated as:

$$\left(\frac{\mathbf{S}^2}{\|\mathbf{S}\mathbf{p}^*\|} + \tau\mathbf{I}\right)\mathbf{p}^* = \tau\mathbf{b} \tag{16}$$

By introducing a scalar $\delta \triangleq \|\mathbf{S}\mathbf{p}^*\|$, we reexpress $\mathbf{p}^*$ as:

$$\mathbf{p}^* = \left(\frac{\mathbf{S}^2}{\tau\delta} + \mathbf{I}\right)^{-1}\mathbf{b} \tag{17}$$

Given that $\mathbf{p}^* \neq \mathbf{0}_q$, it follows that $\mathbf{S}\mathbf{p}^* \neq \mathbf{0}_q$. Consequently, we deduce that $\delta = \|\mathbf{S}\mathbf{p}^*\| > 0$. Furthermore, considering $\delta = \|\mathbf{S}\mathbf{p}^*\|$ and $\delta > 0$, we ascertain that δ is the positive root of

$$\delta^2 = (\mathbf{p}^*)^T\mathbf{S}^2\mathbf{p}^* \tag{18}$$

By substituting Eq. 17 into Eq. 18,we obtain

$$\delta^2 = \tau^2\delta^2\mathbf{b}^T\text{diag}\left(\left\{\frac{s_i^2}{(s_i^2 + \tau\delta)^2}\right\}_{1\leq i\leq q}\right)\mathbf{b} \tag{19}$$

Dividing both sides of Eq. 19 by $\tau^2\delta^2$, we deduce Eq. 5. Now, if the positive root of the equation in Eq. 5 exists, we can initially determine δ, and subsequently determine $\mathbf{p}^*$ by inserting δ into Eq. 17. In the subsequent discussion, we demonstrate the existence condition of the positive root of Eq. 5. For ease of reference, we introduce the following function with respect to δ:

$$\ell(\delta) \triangleq \mathbf{b}^T \mathrm{diag}\left(\left\{ \frac{s_i^2}{(\tau\delta + s_i^2)^2} \right\}_{1\le i\le q}\right)\mathbf{b} - \frac{1}{\tau^2} \tag{20}$$

With this definition in place, Eq. 5 can be reformulated as $\ell(\delta) = 0$. We can ascertain the following characteristics of $\ell(\delta)$.

1) The function $\ell(\delta)$ with respect to α exhibits continuity, and it is strictly monotonically decreasing when δ lies within the interval $[0, +\infty)$. Indeed, the continuity of $\ell(\delta)$ is self-evident. Furthermore, it can be readily confirmed that the gradient of $\ell(\delta)$ is negative for any positive α, hence $\ell(\delta)$ is strictly monotonically decreasing within the interval $[0, +\infty)$.

2) $\lim_{\delta\to 0} \ell(\delta) = \|\mathbf{S}^{-1}\mathbf{b}\|^2 - 1/\tau^2$.

Indeed, owing to the continuity of $\ell(\delta)$, $\lim_{\delta\to 0}\ell(\delta)$ can be evaluated as $\ell(0) = \mathbf{b}^T\mathrm{diag}({s_i^{-2}}_{1\le i\le q})\mathbf{b} - 1/\tau^2$, where the initial term can be represented as $\|\mathbf{S}^{-1}\mathbf{b}\|^2$.

3) $\lim_{\delta\to+\infty}\ell(\delta) < 0$.

Observe that $\lim_{\delta\to+\infty}\ell(\delta) = -1/\tau^2$ and $\tau^2 > 0$, hence $\lim_{\delta\to+\infty}\ell(\delta) < 0$.

Considering the aforementioned properties and the intermediate value theorem, it can be readily confirmed that there exists a positive scalar δ which satisfies $\ell(\delta) = 0$ if and only if $\|\mathbf{S}^{-1}\mathbf{b}\|^2 - (1/\tau^2) > 0$, that is

$$\|\mathbf{S}^{-1}\mathbf{b}\| > \frac{1}{\tau}$$

Specifically, when $\|\mathbf{S}^{-1}\mathbf{b}\| > 1/\tau$, the positive root of the equation $\ell(\delta) = 0$ exists, and such positive root is unique due to the strictly monotonically decreasing property of $\ell(\delta)$. Moreover, let δ^* be the unique positive root of $\ell(\delta) = 0$, then we can establish the following inequality that determines the range of δ^*:

$$\max(0, \delta_l) \le \delta^* \le \delta_u \tag{21}$$

wherein $\delta_u = (\mathbf{b}^T\mathrm{diag}(\{s_i^2\}_{1\le i\le q})\mathbf{b})^{1/2} - s_q^2/\tau$ is the positive root of the equation $f_u(\delta) = 0$; $\delta_l = (\mathbf{b}^T\mathrm{diag}(\{s_i^2\}_{1\le i\le q})\mathbf{b})^{1/2} - s_1^2/\tau$ is the larger root of the equation $f_l(\delta) = 0$, with the two functions $f_u(\delta)$ and $f_l(\delta)$ defined as:

$$f_u(\delta) = \mathbf{b}^T\frac{\mathrm{diag}(\{s_i^2\}_{1\le i\le q})}{(\tau\delta + s_q^2)^2}\mathbf{b} - \frac{1}{\tau^2}$$

$$f_l(\delta) = \mathbf{b}^T\frac{\mathrm{diag}(\{s_i^2\}_{1\le i\le q})}{(\tau\delta + s_1^2)^2}\mathbf{b} - \frac{1}{\tau^2}$$

Specifically, $f_u(\delta)$ is derived by amplifying $\ell(\delta)$ through the substitution of $(\tau\delta + s_i^2)$ with $(\tau\delta + s_q^2)$, whereas $f_l(\delta)$ is obtained by diminishing $f_u(\delta)$ via the replacement of $(\tau\delta + s_i^2)$ with $(\tau\delta + s_1^2)$.

Indeed, the following assertions can be readily confirmed.

1) When $\delta \in [0, +\infty)$, $f_u(\delta)$ and $f_l(\delta)$ exhibit continuity, are strictly monotonically decreasing with respect to δ, and fulfill

$$f_u(\delta) \ge \ell(\delta) \ge f_l(\delta)$$

2) $\lim_{\delta\to 0} f_u(\delta) = f_u(0) \ge \ell(0) > 0$ holds, while it is unclear whether $\lim_{\delta\to 0} f_l(\delta) = f_l(0)$ is positive or not.

3)$\lim_{\delta\to+\infty} f_u(\delta) = \lim_{\delta\to+\infty} f_l(\delta) = -1/\tau^2 < 0$

Consequently, $f_u(\delta) = 0$ possesses a unique positive root, denoted as δ_u, which is not less than δ^*. Furthermore, δ_l [i.e., the larger root of $f_l(\delta) = 0$] does not exceed δ, yet it is not necessarily positive. Thus, the inequalities in

Eq. 21 are confirmed. Consequently, we can determine δ^* using the bisection search method within the search interval $[\max(0, \delta_l), \delta_u]$.

In conclusion, if and only if $\|\mathbf{S}^{-1}\mathbf{b}\| > (1/\tau)$ holds, there exists $\mathbf{p}^* \neq \mathbf{0}_q$, which is optimal for the problem in Eq. 3. Specifically, when $\|\mathbf{S}^{-1}\mathbf{b}\| > (1/\tau)$, p^* can be determined as in Eq. 17, where α is the unique positive root of the equation in Eq. 5 and can be found using the bisection search method.

This concludes the demonstration of Lemma 1.

Demonstration of Theorem 1

Theorem 1:The optimal solution $\mathbf{p}_i^*$ for problem Eq. 5 (with $\tau > 0$) can be expressed as

$$\mathbf{p}_i^* = \begin{cases} \hat{\mathbf{p}}, & \text{if } \left\| \left[\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_{r_K}}\right]' \circ \tilde{\mathbf{b}}_u \right\| > \frac{1}{\tau} \\ \mathbf{c}_i - \mathbf{G}_K \tilde{\mathbf{b}}_u, & \text{otherwise} \end{cases} \tag{22}$$

in which $\tilde{\mathbf{b}}_u = \mathbf{G}_K' \mathbf{c}_i \in \mathbb{R}^{r_K}$, where $\mathbf{G}_K \in \mathbb{R}^{n \times r_K}$ is formed by the first r_K columns of $\mathbf{G}$, and the vector $\hat{\mathbf{p}}$ is defined as

$$\hat{\mathbf{p}} = \mathbf{c}_i - \mathbf{G}_K \left(\left[\frac{\sigma_1^2}{\tau\delta + \sigma_1^2}, \dots, \frac{\sigma_{r_K}^2}{\tau\delta + \sigma_{r_K}^2} \right]' \circ \tilde{\mathbf{b}}_u \right) \tag{23}$$

where δ is a positive scalar, satisfying

$$\tilde{\mathbf{b}}_u' \operatorname{diag}\left(\left\{ \frac{\sigma_i^2}{(\tau\delta + \sigma_i^2)^2} \right\}_{1 \le i \le r_t} \right) \tilde{\mathbf{b}}_u = \frac{1}{\tau^2} \tag{24}$$

Specifically, if the norm $|[1/\sigma_1, \dots, 1/\sigma_{r_K}]^T \odot \tilde{\mathbf{b}}_u|$ surpasses $1/\tau$, then Eq. 24 (concerning δ) possesses a singular positive solution

By inserting $\mathbf{K} = \mathbf{G}\Sigma^2\mathbf{G}^T$ into $(\mathbf{p}_i^T \mathbf{K} \mathbf{p}_i)^{1/2}$ [i.e., the initial term in the objective function of Eq. 2], we deduce

$$\sqrt{\mathbf{p}_i^T \mathbf{K} \mathbf{p}_i} = \sqrt{\mathbf{p}_i^T \mathbf{G}\Sigma^2\mathbf{G}^T \mathbf{p}_i} = \sqrt{(\mathbf{G}^T\mathbf{p}_i)^T \Sigma^2 (\mathbf{G}^T\mathbf{p}_i)} \tag{25}$$

since $\mathbf{G}\mathbf{G}^T = \mathbf{I}$, we have

$$\|\mathbf{p}_i - \mathbf{c}_i\|^2 = \|\mathbf{G}^T(\mathbf{p}_i - \mathbf{c}_i)\|^2 = \|\mathbf{G}^T\mathbf{p}_i - \mathbf{G}^T\mathbf{c}_i\|^2 \tag{26}$$

In Eq. 25 and Eq. 26, the variable $\mathbf{p}_i$ solely appears in $\mathbf{G}^T\mathbf{p}_i$, and the term $\mathbf{G}^T\mathbf{c}_i$ in Eq. 26 remains constant with respect to $\mathbf{p}_i$. Consequently, by introducing a new variable $\tilde{\mathbf{p}} = \mathbf{G}^T\mathbf{p}_i$ and a constant vector

$$\tilde{\mathbf{b}} = \mathbf{G}^T\mathbf{c}_i \tag{27}$$

we reframe the optimization problem Eq. 2 as the following equivalent formulation:

$$\min_{\tilde{\mathbf{p}}} \sqrt{\tilde{\mathbf{p}}^T \Sigma^2 \tilde{\mathbf{p}}} + \frac{\tau}{2} \|\tilde{\mathbf{p}} - \tilde{\mathbf{b}}\|^2 \tag{28}$$

Let $\tilde{\mathbf{p}}^*$ represent the optimal solution of the problem in Eq. 28. Given that $\mathbf{G}\mathbf{G}^T = \mathbf{I}$, it follows that $\mathbf{p}_i = \mathbf{G}\tilde{\mathbf{p}}$ for any $\tilde{\mathbf{p}} = \mathbf{G}^T\mathbf{p}_i$. Consequently, once we determine $\tilde{\mathbf{p}}^*$, the optimal solution $\mathbf{p}_i^*$ of the problem in Eq. 2 can be derived by

$$\mathbf{p}_i^* = \mathbf{G}\tilde{\mathbf{p}}^* \tag{29}$$

Now, let us delve into obtaining $\tilde{\mathbf{p}}^*$. It is worth noting that Σ is a diagonal matrix, with the last $n - r_K$ diagonal elements being all zeros. Hence, we can articulate the variable $\tilde{\mathbf{p}} \in \mathbb{R}^n$ as $\tilde{\mathbf{p}} = [\tilde{\mathbf{p}}_u^T, \tilde{\mathbf{p}}_d^T]^T$, where $\tilde{\mathbf{p}}_u \in \mathbb{R}^{r_K}$ and $\tilde{\mathbf{p}}_d \in \mathbb{R}^{(n-r_K)}$. In a similar vein, we can partition $\mathbf{b} \in \mathbb{R}^n$ into two vectors $\tilde{\mathbf{b}}_u \in \mathbb{R}^{r_K}$ and $\tilde{\mathbf{b}}_d \in \mathbb{R}^{(n-r_K)}$, which means

$$\tilde{\mathbf{b}} = [\tilde{\mathbf{b}}_u^T, \tilde{\mathbf{b}}_d^T]^T \tag{30}$$

Based on Eq. 27, we can express $\tilde{\mathbf{b}}_u$ and $\tilde{\mathbf{b}}_d$ as follows:

$$\tilde{\mathbf{b}}_u = [\mathbf{I}_{r_K}, \mathbf{O}_{r_K \times (n-r_K)}]\mathbf{G}^T \mathbf{c}_i = \mathbf{G}_K^T \mathbf{c}_i \tag{31}$$

$$\tilde{\mathbf{b}}_d = [\mathbf{O}_{(n-r_K)\times r_K}, \mathbf{I}_{(n-r_K)}]\mathbf{G}^T \mathbf{c}_i \tag{32}$$

Furthermore, we introduce a diagonal matrix $\Sigma_u \in \mathbb{R}^{r_K \times r_K}$ defined as

$$\Sigma_u = \text{diag}(\{\sigma_i\}_{1\le i\le r_K}) \tag{33}$$

With these definitions in place, we can reformulate Eq. 28 equivalently as

$$\min_{\tilde{\mathbf{p}}_u, \tilde{\mathbf{p}}_d} \left(\sqrt{\tilde{\mathbf{p}}_u^T \Sigma_u^2 \tilde{\mathbf{p}}_u} + \frac{\tau}{2}\|\tilde{\mathbf{p}}_u - \tilde{\mathbf{b}}_u\|^2 \right) + \left(\frac{\tau}{2}\|\tilde{\mathbf{p}}_d - \tilde{\mathbf{b}}_d\|^2 \right) \tag{34}$$

where the objective function is decomposable with respect to $\tilde{\mathbf{p}}_u$ and $\tilde{\mathbf{p}}_d$. Let $(\tilde{\mathbf{p}}_u^*, \tilde{\mathbf{p}}_d^*)$ represent the optimal solution to the aforementioned optimization problem. Consequently, $\tilde{\mathbf{p}}^*$ can be expressed as

$$\tilde{\mathbf{p}}^* = [(\tilde{\mathbf{p}}_u^*)^T, (\tilde{\mathbf{p}}_d^*)^T]^T \tag{35}$$

It is obvious that

$$\tilde{\mathbf{p}}_d^* = \tilde{\mathbf{b}}_d \tag{36}$$

and $\tilde{\mathbf{p}}_u^*$ is an optimal resolution to the subsequent issue:

$$\min_{\tilde{\mathbf{p}}_u} \sqrt{\tilde{\mathbf{p}}_u^T \Sigma_u^2 \tilde{\mathbf{p}}_u} + \frac{\tau}{2}\|\tilde{\mathbf{p}}_u - \tilde{\mathbf{b}}_u\|^2 \tag{37}$$

Observe that Eq. 37 precisely matches the form of Eq. 3 in Lemma 1. Based on Lemma 1, $\tilde{\mathbf{p}}_u^*$ is determined by

$$\tilde{\mathbf{p}}_u^* = \begin{cases} \left(\frac{\Sigma_u^2}{\tau\alpha} + \mathbf{I}\right)^{-1} \tilde{\mathbf{b}}_u, & \text{if} \|\Sigma_u^{-1}\tilde{\mathbf{b}}_u\| > \frac{1}{\tau} \\ \mathbf{0}_{r_K}, & \text{otherwise} \end{cases} \tag{38}$$

where δ is a positive scalar, satisfying the equation in Eq. 5. Specifically, when $\|\Sigma_u^{-1}\tilde{\mathbf{b}}_u\| > 1/\tau$, the equation in Eq. 5 (with respect to δ) possesses exactly one positive root, which can be found using the bisection method. Given Eqs. 29, 35, and 36, we deduce $\mathbf{p}_i^* = \mathbf{G}[(\tilde{\mathbf{p}}_u^*)^T, \tilde{\mathbf{b}}_d^T]^T$. By inserting Eq. 38 into the equation, we deduce

$$\mathbf{p}_i^* = \begin{cases} \mathbf{G}\left[\left(\frac{\Sigma_u{}^2}{\tau\alpha} + \mathbf{I}\right)^{-1} \tilde{\mathbf{b}}_u\right], & \text{if} \|\Sigma_u^{-1}\tilde{\mathbf{b}}_u\| > \frac{1}{\tau} \\ \mathbf{G}\begin{bmatrix} \mathbf{0}_{r_K} \\ \tilde{\mathbf{b}}_d \end{bmatrix}, & \text{otherwise.} \end{cases} \tag{39}$$

To finalize the proof of Theorem 1, the remaining tasks involve confirming the following three equalities:

$$\Sigma_u^{-1}\tilde{\mathbf{b}}_u = \left[\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_{r_K}}\right]^T \circ \tilde{\mathbf{b}}_u \tag{40}$$

$$\mathbf{G}\begin{bmatrix}\left(\frac{\Sigma_u{}^2}{\tau\alpha}+\mathbf{I}\right)^{-1}\tilde{\mathbf{b}}_u\\ \tilde{\mathbf{b}}_d\end{bmatrix}=\hat{\mathbf{p}} \tag{41}$$

$$\mathbf{G}\begin{bmatrix}\mathbf{0}_{r_K}\\ \tilde{\mathbf{b}}_d\end{bmatrix}=\mathbf{c}_i-\mathbf{G}_K\tilde{\mathbf{b}}_u \tag{42}$$

Initially, the equality in Eq. 40 can be readily confirmed based on the definition of $\sum_u$ in Eq. 33.

Furthermore, we can establish Eq. 41 and Eq. 42 based on the following property. For any $\mathbf{a} = [a_1, \dots, a_{r_K}]^T \in \mathbb{R}^{r_K}$, it follows that

$$\mathbf{G}\text{diag}([\mathbf{a}^T, \mathbf{1}_{n-r_K}^T]^T)\mathbf{G}^T\mathbf{c}_i$$
$$= \mathbf{c}_i - \mathbf{G}_K\ \text{diag}(\{1-a_i\}_{1\le i\le r_K})\mathbf{G}_K^T\mathbf{c}_i \tag{43}$$
$$= \mathbf{c}_i - \mathbf{G}_K\text{diag}(\{1-a_i\}_{1\le i\le r_K})\tilde{\mathbf{b}}_u \tag{44}$$

where the equality in Eq. 43 can be confirmed, given that $\mathbf{G}\mathbf{G}^T = \mathbf{I}_n$ and $[\mathbf{a}^T, \mathbf{1}_{n-r_K}^T]^T = \mathbf{1}_n - [(\mathbf{1}_{r_K} - \mathbf{a})^T, \mathbf{0}_{n-r_K}^T]^T$, while the equality in Eq. 44 is valid based on Eq. 31 and the definition of $\mathbf{G}_K$. Eq. 41 can be obtained based on the above-mentioned property, by replacing $\mathbf{a}$ with $[\tau\alpha/(\tau\alpha+\sigma_1^2), \dots, \tau\alpha/(\tau\alpha+\sigma_{r_K}^2)]^T$, while the equality in Eq. 64 holds based on Eq. 31 and the definition of $\mathbf{G}_K$.

In particular, for the equality in Eq. 41, its left-hand side can be rewritten as

$$\mathbf{G}\text{diag}([[\tau\delta/(\tau\delta+\sigma_1^2), \dots, \tau\delta/(\tau\delta+\sigma_{r_K}^2)]^T, \mathbf{1}_{n-r_K}^T]^T)\mathbf{G}^T\mathbf{c}_i \tag{45}$$

Based on the equality $(\Sigma_u{}^2/(\tau\delta) + \mathbf{I})^{-1} = \text{diag}(\{\tau\alpha/(\tau\delta+\sigma_i^2)\}_{1\le i\le r_K})$ and the equalities in Eq. 31 and Eq. 32. Moreover, the right-hand side of the equality in Eq. 41, namely, $\hat{\mathbf{p}}$ defined in Eq. 23, can be written as

$$\mathbf{c}_i - \mathbf{G}_K\text{diag}(\{1-(\tau\delta)/(\tau\delta+\sigma_i^2)\}_{1\le i\le r_K})\tilde{\mathbf{b}}_u \tag{46}$$

Based on the above-mentioned property, we can conclude that $\mathbf{G}\text{diag}([[\tau\delta/(\tau\delta+\sigma_1^2), \dots, \tau\delta/(\tau\delta+\sigma_{r_K}^2)]^T, \mathbf{1}_{n-r_K}^T]^T)\mathbf{G}^T\mathbf{c}_i$ is equal to $\mathbf{c}_i - \mathbf{G}_K\text{diag}(\{1-(\tau\delta)/(\tau\delta+\sigma_i^2)\}_{1\le i\le r_K})\tilde{\mathbf{b}}_u$, so that Eq. 41 is obtained.

We can also derive Eq. 42 based on the aforementioned property, by substituting $\mathbf{a}$ with $\mathbf{0}_{r_K}$. Specifically, the left-hand side of Eq. 42 equals $\mathbf{G}\text{diag}([\mathbf{0}_{r_K}^T, \mathbf{1}_{n-r_K}^T]^T)\mathbf{G}^T\mathbf{c}_i$ according to Eq. 32, which can be reformulated as $\mathbf{G}\text{diag}([\mathbf{0}_{r_K}^T, \mathbf{1}_{n-r_K}^T]^T)\mathbf{G}^T\mathbf{c}_i$ leveraging the aforementioned property. Observe that we can readily reformulate $\mathbf{c}_i - \mathbf{G}_K\text{diag}(\mathbf{1}_{r_K} - \mathbf{0}_{r_K})\tilde{\mathbf{b}}_u$ as the right-hand side of Eq. 42, thus Eq. 42 is established.

This concludes the demonstration of Theorem 1.