Appendix

A. Proof of Proposition 3

Consider an arbitrary parameter $\tilde{\mu} = \sum_i \tilde{c}_i \tilde{\mathbf{a}}_i$ in the feasible set of parameters that satisfy the constraint of (6).

Let $\mathcal{P}_{\mathcal{M}}$ denote the matrix corresponding to the projection operator for the subspace $\mathcal{M}$.

Since $\tilde{\mu}$ should satisfy the constraint of $\mathcal{R}^*(\widehat{\mu}_n - \tilde{\mu}) \le \lambda_n$, for any fixed index i in the atoms of $\widehat{\mu}_n$,

$$\begin{aligned}&\mathcal{R}\Big(\mathcal{P}_{\mathbf{a}_i^*}(\widehat{\mu}_n - \tilde{\mu})\Big)\\ &= \mathcal{R}\Big(\sum_j \tilde{c}_j \mathcal{P}_{\mathbf{a}_i^*}(\tilde{\mathbf{a}}_j) - c_i^* \mathbf{a}_i^*\Big) \le \lambda_n,\end{aligned}$$

which implies $\max\{c_i^* - \lambda_n, 0\} \le \mathcal{R}\Big(\sum_j \tilde{c}_j \mathcal{P}_{\mathbf{a}_i^*}(\tilde{\mathbf{a}}_j)\Big)$. By summing over all i, we obtain

$$\begin{aligned}\mathcal{R}(\widehat{\mu}) &= \sum_i \max\{c_i^* - \lambda_n, 0\} \le \sum_i \mathcal{R}\Big(\sum_j \tilde{c}_j \mathcal{P}_{\mathbf{a}_i^*}(\tilde{\mathbf{a}}_j)\Big)\\ &\stackrel{(i)}{=} \mathcal{R}\Big(\sum_j \tilde{c}_j \sum_i \mathcal{P}_{\mathbf{a}_i^*}(\tilde{\mathbf{a}}_j)\Big) = \mathcal{R}\Big(\sum_j \tilde{c}_j \tilde{\mathbf{a}}_j\Big) = \mathcal{R}(\tilde{\mu}).\end{aligned}$$

Finally, since we can simply verify that $\widehat{\mu}$ is also feasible, we can conclude that $\widehat{\mu}$ is the optimal of (6).

B. Proof of Theorem 1

Let $\Delta := \widehat{\mu} - \mu^*$ be the error vector that we are interested in.

$$\begin{aligned}&\mathcal{R}^*(\widehat{\mu} - \mu^*) = \mathcal{R}^*(\widehat{\mu} - \widehat{\mu}_n + \widehat{\mu}_n - \mu^*)\\ &\le \mathcal{R}^*(\widehat{\mu}_n - \widehat{\mu}) + \mathcal{R}^*(\widehat{\mu}_n - \mu^*) \le 2\lambda_n\end{aligned}$$

By the fact that $\mu^*_{\mathcal{M}^\perp} = \mathbf{0}$, and the decomposability of $\mathcal{R}$ with respect to $(\mathcal{M}, \overline{\mathcal{M}}^\perp)$,

$$\begin{aligned}&\mathcal{R}(\mu^*)\\ &= \mathcal{R}(\mu^*) + \mathcal{R}\big[\Pi_{\bar{\mathcal{M}}^\perp}(\Delta)\big] - \mathcal{R}\big[\Pi_{\bar{\mathcal{M}}^\perp}(\Delta)\big]\\ &= \mathcal{R}\big[\mu^* + \Pi_{\bar{\mathcal{M}}^\perp}(\Delta)\big] - \mathcal{R}\big[\Pi_{\bar{\mathcal{M}}^\perp}(\Delta)\big]\\ &\stackrel{(i)}{\le} \mathcal{R}\big[\mu^* + \Pi_{\bar{\mathcal{M}}^\perp}(\Delta) + \Pi_{\bar{\mathcal{M}}}(\Delta)\big] + \mathcal{R}\big[\Pi_{\bar{\mathcal{M}}}(\Delta)\big]\\ &\qquad - \mathcal{R}\big[\Pi_{\bar{\mathcal{M}}^\perp}(\Delta)\big]\\ &= \mathcal{R}\big[\mu^* + \Delta\big] + \mathcal{R}\big[\Pi_{\bar{\mathcal{M}}}(\Delta)\big] - \mathcal{R}\big[\Pi_{\bar{\mathcal{M}}^\perp}(\Delta)\big] \qquad (18)\end{aligned}$$

where the equality (i) holds by the triangle inequality of norm. Since (6) minimizes $\mathcal{R}(\widehat{\mu})$, we have $\mathcal{R}(\mu^* + \Delta) = \mathcal{R}(\widehat{\mu}) \le \mathcal{R}(\mu^*)$. Combining this inequality with (18),

$$\mathcal{R}\big[\Pi_{\bar{\mathcal{M}}^\perp}(\Delta)\big] \le \mathcal{R}\big[\Pi_{\bar{\mathcal{M}}}(\Delta)\big]. \qquad (19)$$

Moreover, by Hölder's inequality and the decomposability of $\mathcal{R}(\cdot)$,

$$\begin{aligned}&\|\Delta\|_2^2 = \langle\!\langle \Delta, \Delta \rangle\!\rangle \le \mathcal{R}^*(\Delta)\,\mathcal{R}(\Delta) \le 2\lambda_n \mathcal{R}(\Delta)\\ &= 2\lambda_n\big[\mathcal{R}(\Pi_{\bar{\mathcal{M}}}(\Delta)) + \mathcal{R}(\Pi_{\bar{\mathcal{M}}^\perp}(\Delta))\big] \le 4\lambda_n \mathcal{R}(\Pi_{\bar{\mathcal{M}}}(\Delta))\\ &\le 4\lambda_n \Psi(\overline{\mathcal{M}})\|\Pi_{\bar{\mathcal{M}}}(\Delta)\|_2 \qquad (20)\end{aligned}$$

where $\Psi(\overline{\mathcal{M}})$ is a simple notation for $\Psi(\overline{\mathcal{M}}, \|\cdot\|_2)$.

Since the projection operator is defined in terms of $\|\cdot\|_2$ norm, it is non-expansive: $\|\Pi_{\bar{\mathcal{M}}}(\Delta)\|_2 \le \|\Delta\|_2$. Therefore, by (20), we have

$$\|\Pi_{\bar{\mathcal{M}}}(\Delta)\|_2 \le 4\lambda_n \Psi(\overline{\mathcal{M}}), \qquad (21)$$

and plugging it back into (20) yields the error bound (10).

Finally, (11) is straightforward from (19) and (21)

$$\begin{aligned}&\mathcal{R}(\Delta) \le 2\mathcal{R}(\Pi_{\bar{\mathcal{M}}}(\Delta))\\ &\le 2\Psi(\overline{\mathcal{M}})\|\Pi_{\bar{\mathcal{M}}}(\Delta)\|_2 \le 8\lambda_n \Psi(\overline{\mathcal{M}})^2.\end{aligned}$$

C. Proof of corollaries: Covariance Estimation in Section 4.1

In order to leverage Theorem 1, two ingredients need to be specified: (i) the convergence rate of $\mathcal{R}^*(\widehat{\mu}_n - \mu^*)$ for λ_n to satisfy $\lambda_n \ge \mathcal{R}^*(\widehat{\mu}_n - \mu^*)$, and (ii) the compatibility constant $\Psi(\overline{\mathcal{M}})$. In each corollary, we are going to show how these two components can be computed.

C.1. Proof of Corollary 1

For this case, we can directly appeal to the well known bound (e.g., the Lemma 1 of (Ravikumar et al., 2011)): Consider the following event:

$$\begin{aligned}&P(\|\widehat{\Sigma}_n - \Sigma\|_{\infty,\text{off}} > \delta)\\ &\le 4\exp\Big(-\frac{n\delta^2}{3200(\max_i \Sigma_{ii})^2} + \log p^2\Big).\end{aligned}$$

By setting $\delta = 40(\max_i \Sigma_{ii})\sqrt{\frac{2\tau \log p}{n}}$, we see that the choice of λ_n is valid with probability at least $1 - C_1\exp(-C_2 n\lambda_n^2)$.

For the second ingredient, let $\mathcal{M} = \overline{\mathcal{M}}$ correspond to the support of Σ^*. We have $\psi(\overline{\mathcal{M}}, \|\cdot\|_2) = \sqrt{s}$, where s is the cardinality of the support of Σ^*.

C.2. Proof of Corollary 2

$\mathcal{R}^*(\widehat{\Sigma}_n - \Sigma^*) = \max_{l=1,\ldots,L}\Big\|[\widehat{\Sigma}_n]_{G_l} - [\Sigma^*]_{G_l}\Big\|_{\nu^*}$. For a given entry (i,j) we have

$$P(|[\widehat{\Sigma}_n]_{ij} - \Sigma^*_{ij}| > t) \le C_1\exp(-C_2 n t^2).$$

For a given group G_l, by union bound over the group elements, we have

$$P(|[\widehat{\Sigma}_n]_{ij} - \Sigma^*_{ij}| > t \text{ for all } (i,j) \in G_l) \\ \leq C_1 \exp(-C_2 n t^2 + \log d).$$

This implies that

$$P(\|[\widehat{\Sigma}_n]_{G_l} - \Sigma^*_{G_l}\|_{\nu^\star} > t d^{1/\nu^\star}) \leq C_1 \exp(-C_2 n t^2 + \log d).$$

By a union bound over all groups we obtain

$$P\left(\max_{l=1,\ldots,L} \|[\widehat{\Sigma}_n]_{G_l} - \Sigma^*_{G_l}\|_2 > t d^{1/\nu^\star}\right) \\ \leq C_1 \exp(-C_2 n t^2 + \log d + \log L).$$

This yields

$$P\left(\max_{l=1,\ldots,L} \|[\widehat{\Sigma}_n]_{G_l} - \Sigma^*_{G_l}\|_2 > \delta\right) \\ \leq C_1 \exp(-C_2 n d^{-2/\nu^\star} \delta^2 + \log d + \log L).$$

We conclude by setting $\delta = d^{1/\nu^\star}\sqrt{(\log d + \log L)/n}$. Let $\mathcal{M} = \overline{\mathcal{M}}$ correspond to the support of Σ^*, which can be written as a union of groups in $\mathcal{G}$. Since $\nu \geq 2$, we have $\psi(\overline{\mathcal{M}}, \|\cdot\|_F) = \sqrt{k}$.

D. Proof of Theorem 2

In this proof, we consider the matrix parameter such as the covariance. Basically, the Frobenius norm can be simply replaced by ℓ_2 norm for the vector parameters. Let $\Delta_\alpha := \widehat{\mu}_\alpha - \mu^*_\alpha$, and $\Delta := \widehat{\mu} - \mu^* = \sum_{\alpha \in I} \Delta_\alpha$. The error bound (14) can be easily shown from the assumption in the statement with the constraint of (13). For every $\alpha \in I$,

$$\mathcal{R}^*_\alpha(\Delta) = \mathcal{R}^*_\alpha(\widehat{\mu} - \mu^*) = \mathcal{R}^*_\alpha(\widehat{\mu} - \widehat{\mu}_n + \widehat{\mu}_n - \mu^*) \\ \leq \mathcal{R}^*_\alpha(\widehat{\mu}_n - \widehat{\mu}) + \mathcal{R}^*_\alpha(\widehat{\mu}_n - \mu^*) \leq 2\lambda_\alpha. \tag{22}$$

By the similar reasoning as in (18) with the fact that $\Pi_{\mathcal{M}^\perp_\alpha}(\mu^*_\alpha) = \mathbf{0}$ in (C3), and the decomposability of $\mathcal{R}_\alpha$ with respect to $(\mathcal{M}_\alpha, \overline{\mathcal{M}}^\perp_\alpha)$, we have

$$\mathcal{R}_\alpha(\mu^*_\alpha) \leq \mathcal{R}_\alpha\big[\mu^*_\alpha + \Delta_\alpha\big] + \mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\big] \\ - \mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}^\perp_\alpha}(\Delta_\alpha)\big]. \tag{23}$$

Since $\{\widehat{\mu}_\alpha\}_{\alpha \in I}$ minimizes the objective function of (13),

$$\sum_{\alpha \in I} \lambda_\alpha \mathcal{R}_\alpha(\widehat{\mu}_\alpha) \leq \sum_{\alpha \in I} \lambda_\alpha \mathcal{R}_\alpha(\mu^*_\alpha).$$

Combining this inequality with (23), we have

$$\sum_{\alpha \in I} \lambda_\alpha \mathcal{R}_\alpha(\widehat{\mu}_\alpha) \leq \sum_{\alpha \in I} \lambda_\alpha \Big\{ \mathcal{R}_\alpha(\mu^*_\alpha + \Delta_\alpha) \\ + \mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\big] - \mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}^\perp_\alpha}(\Delta_\alpha)\big] \Big\},$$

which implies

$$\sum_{\alpha \in I} \lambda_\alpha \mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}^\perp_\alpha}(\Delta_\alpha)\big] \leq \sum_{\alpha \in I} \lambda_\alpha \mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\big], \tag{24}$$

Now, for each structure $\alpha \in I$, we have an application of Hölder's inequality; $|\langle\!\langle \Delta, \Delta_\alpha \rangle\!\rangle| \leq \mathcal{R}^*_\alpha(\Delta)\,\mathcal{R}_\alpha(\Delta_\alpha) \leq 2\lambda_\alpha \mathcal{R}_\alpha(\Delta_\alpha)$ where the notation $\langle\!\langle A, B \rangle\!\rangle$ denotes the trace inner product, $\text{trace}(A^\top B) = \sum_i \sum_j A_{ij} B_{ij}$, and we use the pre-computed bound in (22). Then, the Frobenius error $\|\Delta\|_F$ can be upper-bounded as follows:

$$\|\Delta\|_F^2 = \langle\!\langle \Delta, \Delta \rangle\!\rangle = \sum_{\alpha \in I} \langle\!\langle \Delta, \Delta_\alpha \rangle\!\rangle \leq \sum_{\alpha \in I} |\langle\!\langle \Delta, \Delta_\alpha \rangle\!\rangle| \\ \leq 2 \sum_{\alpha \in I} \lambda_\alpha \mathcal{R}_\alpha(\Delta_\alpha) \leq 2 \sum_{\alpha \in I} \Big\{ \lambda_\alpha \mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\big] + \\ \lambda_\alpha \mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}^\perp_\alpha}(\Delta_\alpha)\big] \Big\} \leq 4 \sum_{\alpha \in I} \lambda_\alpha \mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\big] \\ \leq 4 \sum_{\alpha \in I} \lambda_\alpha \Psi(\overline{\mathcal{M}}_\alpha)\, \|\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\|_F \tag{25}$$

where $\Psi(\overline{\mathcal{M}}_\alpha)$ denotes the compatibility constant of space $\overline{\mathcal{M}}_\alpha$ with respect to the Frobenius norm: $\Psi(\overline{\mathcal{M}}_\alpha, \|\cdot\|_F)$.

Here, we define a key notation in the error bound:

$$\Phi := \max_{\alpha \in I} \lambda_\alpha \Psi(\overline{\mathcal{M}}_\alpha).$$

Armed with this notation, (25) can be rewritten as

$$\|\Delta\|_F^2 \leq 4\Phi \sum_{\alpha \in I} \|\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\|_F \tag{26}$$

At this point, we directly appeal to the result in Proposition 2 of (Yang & Ravikumar, 2013) with a small modification:

Proposition 4. *Suppose that the structural incoherence condition (C4) as well as the condition (C3) hold. Then, we have*

$$2|\sum_{\alpha < \beta} \langle\!\langle \Delta_\alpha, \Delta_\beta \rangle\!\rangle| \leq \frac{1}{2} \sum_{\alpha \in I} \|\Delta_\alpha\|_F^2.$$

By this proposition, we have

$$\sum_{\alpha \in I} \|\Delta_\alpha\|_F^2 \leq \|\Delta\|_F^2 + 2|\sum_{\alpha < \beta} \langle\!\langle \Delta_\alpha, \Delta_\beta \rangle\!\rangle| \\ \leq \|\Delta\|_F^2 + \frac{1}{2} \sum_{\alpha \in I} \|\Delta_\alpha\|_F^2 \,,$$

which implies $\sum_{\alpha \in I} \|\Delta_\alpha\|_F^2 \leq 2\|\Delta\|_F^2$.

Moreover, since the projection operator is defined in terms of the Frobenius norm, it is non-expansive for all α:

$\|\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\|_F \leq \|\Delta_\alpha\|_F$. Hence, we finally obtain

$$\begin{aligned}\Big(\sum_{\alpha\in I}\|\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\|_F\Big)^2 &\leq \Big(\sum_{\alpha\in I}\|\Delta_\alpha\|_F\Big)^2\\ \leq |I|\sum_{\alpha\in I}\|\Delta_\alpha\|_F^2 &\leq 8|I|\Phi\sum_{\alpha\in I}\|\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\|_F,\end{aligned}$$

and therefore,

$$\sum_{\alpha\in I}\|\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\|_F \leq\ 8|I|\Phi \tag{27}$$

The Frobenius norm error bound (16) can be derived by plugging (27) back into (26):

$$\|\Delta\|_F \leq 32|I|\Phi^2.$$

The proof of the final error bound (15) is straightforward from (24) and (27) as follows: for each fixed $\alpha \in I$,

$$\begin{aligned}&\mathcal{R}_\alpha(\Delta_\alpha)\\ &\leq \frac{1}{\lambda_\alpha}\Big\{\lambda_\alpha\mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\big] + \lambda_\alpha\mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}_\alpha^\perp}(\Delta_\alpha)\big]\Big\}\\ &\leq \frac{1}{\lambda_\alpha}\Big\{\lambda_\alpha\mathcal{R}_\alpha\big[\Pi_{\bar{\mathcal{M}}_\alpha}(\Delta_\alpha)\big] + \sum_{\beta\in I}\lambda_\beta\mathcal{R}_\beta\big[\Pi_{\bar{\mathcal{M}}_\beta}(\Delta_\beta)\big]\Big\}\\ &\leq \frac{2}{\lambda_\alpha}\sum_{\beta\in I}\lambda_\beta\mathcal{R}_\beta\big[\Pi_{\bar{\mathcal{M}}_\beta}(\Delta_\beta)\big]\\ &\leq \frac{2}{\lambda_\alpha}\sum_{\beta\in I}\lambda_\beta\Psi(\overline{\mathcal{M}}_\beta)\|\Pi_{\bar{\mathcal{M}}_\beta}(\Delta_\beta)\|_F\\ &\leq \frac{2\Phi}{\lambda_\alpha}\sum_{\beta\in I}\|\Pi_{\bar{\mathcal{M}}_\beta}(\Delta_\beta)\|_F \leq \frac{16|I|\Phi^2}{\lambda_\alpha},\end{aligned}$$

which completes the proof.

D.1. Proof of Corollary 3

The proof for an element-wise sparse component is already proven in Section C.1. At the same time, for a low-rank component, we can directly appeal to the results for clean models (Agarwal et al., 2012):

$$\lambda_1 = 4\sqrt{\|\Sigma^*\|_2}\sqrt{\frac{p}{n}} \geq \|\Sigma^* - \widehat{\Sigma}_n\|_2$$

with probability at least $1-2\exp(-C_1 p)$. The subspace compatibility of any matrix A with rank k can be easily derived as

$$\sup_{A\neq 0}\frac{\|A\|_*}{\|A\|_F} \leq \sqrt{k_1}.$$

E. A Parallel Proximal algorithm for "Elem-Super-Moment" Estimation

The class of "Elem-Super-Moment" estimators solves

$$\begin{aligned}\underset{\mu_1,\mu_2,\ldots,\mu_{|I|}}{\text{minimize}}\ &\sum_{\alpha\in I}\lambda_\alpha\mathcal{R}_\alpha(\mu_\alpha)\\ \text{s. t. }&\mathcal{R}_\alpha^*(\widehat{\mu}_n-\mu)\leq\lambda_\alpha \quad \text{for } \forall\alpha\in I\\ &\mu=\sum_{\alpha\in I}\mu_\alpha.\end{aligned} \tag{28}$$

Let $\boldsymbol{\mu} = (\mu_1,\mu_2,\ldots,\mu_{|I|})$. Consider the operators $L_\alpha(\boldsymbol{\mu}) = \mu_\alpha$, for $\alpha\in I$ and $L_{\text{tot}}(\boldsymbol{\mu}) = \sum_{\alpha\in I}\mu_\alpha$. Then the problem can be rewritten as

$$\begin{aligned}\underset{\boldsymbol{\mu}}{\text{minimize}}\ &\sum_{\alpha\in I}\lambda_\alpha\mathcal{R}_\alpha(L_\alpha(\boldsymbol{\mu}))\\ \text{s. t. }&\mathcal{R}_\alpha^*(\widehat{\mu}_n - L_{\text{tot}}(\boldsymbol{\mu}))\leq\lambda_\alpha \quad \text{for } \forall\alpha\in I.\end{aligned} \tag{29}$$

For all $\alpha\in I$ let

$$f_\alpha(\cdot) = \lambda_\alpha\mathcal{R}_\alpha(L_\alpha(\cdot)).$$

Define the indicator function of a set C as

$$\imath_C : x\mapsto\begin{cases}0, & \text{if } x\in C\\ +\infty, & \text{if } x\notin C.\end{cases}$$

and let

$$g_\alpha(\cdot) = \imath_{(\mathcal{R}_\alpha^*(\widehat{\mu}_n - L_{\text{tot}}(\cdot))\leq\lambda_\alpha)}.$$

Then observe that (29) can be rewritten as

$$\begin{aligned}\underset{\bar{\boldsymbol{\mu}}_1,\ldots\bar{\boldsymbol{\mu}}_{|I|},\tilde{\boldsymbol{\mu}}_1,\ldots\tilde{\boldsymbol{\mu}}_{|I|}}{\text{minimize}}\ &\sum_{\alpha\in I}f_\alpha(\bar{\boldsymbol{\mu}}_\alpha)+\sum_{\alpha\in I}g_\alpha(\tilde{\boldsymbol{\mu}}_\alpha)\\ \text{s. t. }&\bar{\boldsymbol{\mu}}_1=\ldots=\bar{\boldsymbol{\mu}}_{|I|}=\tilde{\boldsymbol{\mu}}_1=\ldots=\tilde{\boldsymbol{\mu}}_{|I|}.\end{aligned} \tag{30}$$

We can then apply the parallel proximal method (Algorithm 3.1 of Combettes & Pesquet (2008)), which is derived from the classical Douglas-Rachford algorithm (Combettes & Pesquet, 2008), and obtain Algorithm 1. In this splitting algorithm, each function f_α is used separately via its own proximal operator. The same holds for each function g_α. Note that

$$\text{prox}_{2|I|\gamma f_\alpha} = \text{prox}_{2|I|\gamma\lambda_\alpha\mathcal{R}_\alpha\circ L_\alpha}$$

and

$$\text{prox}_{2|I|\gamma g_\alpha} = \text{prox}_{2|I|\gamma\imath_{(\mathcal{R}_\alpha^*(\widehat{\mu}_n-L_{\text{tot}}())\leq\lambda_\alpha)}}$$

For various popular choices of regularization R_α these proximal operators have simple closed-form formulas.

This can be seen by applying Lemma 2.4 of Combettes & Pesquet (2008) which states that if L is a bounded linear operator such that $L \circ L^* = \kappa \mathrm{Id}$ for some finite $\kappa > 0$ then

$$\mathrm{prox}_{h \circ L} = \mathrm{Id} + \frac{1}{\kappa} L^* \circ (\mathrm{prox}_{\kappa h} - \mathrm{Id}) \circ L.$$

and by noting that L_α and L_{tot} are such bounded linear operators.

Algorithm 1 Parallel proximal algorithm

Initialization: $\gamma > 0$, $(\bar{\boldsymbol{\mu}}^0_\alpha)_{\alpha \in I}$ and $(\tilde{\boldsymbol{\mu}}^0_\alpha)_{\alpha \in I}$
Set $\boldsymbol{\mu}^0 = \frac{1}{2|I|} \sum_{\alpha \in I} (\bar{\boldsymbol{\mu}}^0_\alpha + \tilde{\boldsymbol{\mu}}^0_\alpha)$.
for $i = 0, 1, \ldots$ **do**
  for $\alpha \in I$ **do**
    $\bar{\boldsymbol{p}}^i_\alpha = \mathrm{prox}_{2|I|\gamma f_\alpha} \bar{\boldsymbol{\mu}}^i_\alpha$ and $\tilde{\boldsymbol{p}}^i_\alpha = \mathrm{prox}_{2|I|\gamma g_\alpha} \tilde{\boldsymbol{\mu}}^i_\alpha$.
  end for
  $\boldsymbol{p}^i = \frac{1}{2|I|} \sum_{\alpha \in I} (\bar{\boldsymbol{p}}^i_\alpha + \tilde{\boldsymbol{p}}^i_\alpha)$.
  $0 < \rho_i < 2$
  for $\alpha \in I$ **do**
    $\bar{\boldsymbol{\mu}}^{i+1}_\alpha = \bar{\boldsymbol{\mu}}^i_\alpha + \rho_i (2\boldsymbol{p}^i - \boldsymbol{\mu}^i - \bar{\boldsymbol{p}}^i_\alpha)$.
    $\tilde{\boldsymbol{\mu}}^{i+1}_\alpha = \tilde{\boldsymbol{\mu}}^i_\alpha + \rho_i (2\boldsymbol{p}^i - \boldsymbol{\mu}^i - \tilde{\boldsymbol{p}}^i_\alpha)$.
  end for
  $\boldsymbol{\mu}^{i+1} = \boldsymbol{\mu}^i + \rho_i (\boldsymbol{p}^i - \boldsymbol{\mu}^i)$.
end for