Supplementary material

Proof of the convergence of the two-tier iterative algorithm

When all binary variables become constants, the two-tier iteration process is simplified to **Algorithm 2.** $P_{i,t}^{\mathrm{P2P}}$, $P_{ij,t}^{\mathrm{P2P}}$ and $p_{i,t}$ are simplified as x_i, x_{ij} and y for brevity, respectively. The augmented Lagrangian functions L_ρ^x and L_ρ^y are simplified as

$$L_\rho^x=\sum_{i\in\mathcal{I}}\left(f_i(x_i)+\sum_{j\in\mathcal{I}\backslash\{i\}}\lambda_{ij}(x_{ij}+x_{ji})+\mu_{i,t}(x_i-y)+\frac{\rho}{2}\left(\sum_{j\in\mathcal{I}\backslash\{i\}}\left\|x_{ij}+x_{ji}\right\|_2^2+\left\|x_i-y\right\|_2^2\right)\right),x_i\in\mathcal{M}_i \tag{1}$$

$$L_\rho^y=g(y)+\sum_{i\in\mathcal{I}}\left(\mu_i(x_i-y)+\frac{\rho}{2}\left\|x_i-y\right\|_2^2\right),y\in\mathcal{N} \tag{2}$$

Algorithm 2 Simplified two-tier iterative process

1	**While** $\left\|r^h\right\|\ge\varepsilon_h^{\mathrm{pri}}$ or $\left\|s^h\right\|\ge\varepsilon_h^{\mathrm{dual}}$ **do**
2	Set k=0
3	**While** $\left\|r^k\right\|\ge\varepsilon_k^{\mathrm{pri}}$ or $\left\|s^k\right\|\ge\varepsilon_k^{\mathrm{dual}}$ **do**
4	$(x_i^{(h+1,k+1)},x_{ij}^{(h+1,k+1)})=\arg\min L_\rho^x(x_i,x_{ij},\lambda_{ij}^{(h,k)},y^{(h)},\mu_i^{(h)})$
5	$\lambda_{ij}^{(h+1,k+1)}=\lambda_{ij}^{(h+1,k)}+\rho(x_{ij}^{(h+1,k+1)}+x_{ji}^{(h+1,k+1)})$
6	k=k+1
7	**End**
8	$x_i^{(h+1,*)}=x_i^{(h+1,k)}$
9	$y^{(h+1)}=\arg\min L_\rho^y(x_i^{(h+1,*)},y,\mu_i^{(h)})$
10	$\mu_i^{(h+1)}=\mu_i^{(h)}+\rho(x_i^{(h+1,*)}-y^{(h+1)})$
11	h=h+1
12	**End**
13	**Output** $x_i^{(*,*)}=x_i^{(h,*)}$, $y^{(*)}=y^{(h)}$

Assume $f_i(x_i)$ and $g(y)$ are closed convex functions, and the feasible domains of the optimization problem are not empty. **Lines 3-7** constitute typical ADMM algorithm iteration steps, and the iteration sequence { $x_i^{(h+1,k)}$, $x_{ij}^{(h+1,k)}$, $\lambda_{ij}^{(h+1,k)}$ } will converge to a certain fixed sequence { $x_i^{(h+1,*)}$, $x_{ij}^{(h+1,*)}$, $\lambda_{ij}^{(h+1,*)}$ }. According to the KKT conditions, it follows that

$$0\in\left.\frac{\partial f_i(x_i)}{\partial x_i}\right|_{x_i=x_i^{(h+1,*)}}+\mu_i^{(h)}+\rho(x_i^{(h+1,*)}-y^{(h)}),\forall i\in\mathcal{I} \tag{3}$$

Due to that $\mu_i^{(h+1)}=\mu_i^{(h)}+\rho(x_i^{(h+1,*)}-y^{(h+1)})$, we have

$$0\in\left.\frac{\partial f_i(x_i)}{\partial x_i}\right|_{x_i=x_i^{(h+1,*)}}+\mu_i^{(h+1)}+\rho(y^{(h+1)}-y^{(h)}),\forall i\in\mathcal{I} \tag{4}$$

so that

$$x_i^{(h+1,*)} = \arg\min f_i(\boldsymbol{x}_i) + \left[\mu_i^{(h+1)} + \rho(y^{(h+1)} - y^{(h)})\right] x_i, \forall i \in \mathcal{I} \tag{5}$$

Similarly,

$$y^{(h+1)} = \arg\min g(y) - \sum_i (\mu_i^{(h+1)}) y \tag{6}$$

According to the optimality theory, the following inequality can be obtained:

$$\sum_i f_i(x_i^{(h+1,*)}) + \sum_i \left[\mu_i^{(h+1)} + \rho(y^{(h+1)} - y^{(h)})\right] x_i^{(h+1,*)} \le \sum_i f(x_i^{(*,*)}) + \sum_i \left[\mu_i^{h+1} + \rho(y^{h+1} - y^h)\right] x_i^{(*,*)} \tag{7}$$

$$g(y^{(h+1)}) - \sum_i (\mu_i^{(h+1)}) y^{(h+1)} \le g(y^{(*)}) - \sum_i (\mu_i^{(h+1)}) y^{(*)} \tag{8}$$

Add (7) and (8), define $p^{(h+1)} = \sum_i f_i(x_i^{(h+1,*)}) + g(y^{h+1})$, and according to the KKT conditions, we obtain

$$p^{(h+1)} - p^{(*)} + \sum_i \mu_i^{(h+1)}(x_i^{(h+1)} - y^{(h+1)}) + \sum_i \rho(y^{(h+1)} - y^{(h)}) x_i^{(h+1)} \le \sum_i \rho(y^{(h+1)} - y^{(h)}) x_i^{(*)} \tag{9}$$

Let $r_i^{(h)} = x_i^{(h)} - y^{(h)}$, (9) can be expressed as

$$p^{(h+1)} - p^{(*)} \le -\sum_i \mu_i^{(h+1)} r_i^{(h+1)} + \sum_i \rho(y^{(h+1)} - y^{(h)})(-r_i^{(h+1)} - (y^{(h+1)} - y^{(*)})) \tag{10}$$

Further, analogously to the derivation of [1], we can conclude that

$$p^{(*)} - p^{(h+1)} \le \sum_i \mu_i^{(*)} r_i^{(h+1)} \tag{11}$$

$$V^{(h+1)} \le V^{(h)} - \rho \sum_i \left\| r_i^{(h+1)} \right\|_2^2 - \rho \left\| y^{(h+1)} - y^{(h)} \right\|_2^2 \tag{12}$$

where $V^{(h)} = \frac{1}{\rho} \sum_i \left\| \mu_i^{(h+1)} - \mu_i^{(*)} \right\|_2^2 + \rho \left\| y^{(h)} - y^{(*)} \right\|_2^2$. Summing both sides of (10) in terms of *h* yields

$$\rho \sum_{h=0}^{\infty} \left(\sum_i \left\| r_i^{(h+1)} \right\|_2^2 + \left\| y^{(h+1)} - y^{(h)} \right\|_2^2 \right) \le V^0 \tag{13}$$

Hence, $r_i^{(h)} \to 0$, i.e. $\mu_i^{(h+1)} - \mu_i^{(h)} \to 0$. And $y^{(h+1)} - y^{(h)} \to 0$. This implies that the iterations for both the primal and dual variables converge. For (10) and (11), when $h \to 0$, the right-hand side of the equation both converges to 0 and hence $p^{(h+1)}$ converges to $p^{(*)}$. The proof is complete. ■

[1] S. Boyd, N. Parikh, and E. Chu, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning 3.1*, pp 1-122, 2011.