A. Proofs of the Theorems

Asymptotic Consistency of ADMM

Theorem 3.1. *If we set $\bar{\theta}$ to be asymptotically consistent and $\lambda^i_{\beta_i} = 0$ for all i in the initial step of ADMM, then $\bar{\theta}$ remains asymptotically consistent at every iteration.*

Proof. To avoid overloading the notation, we consider the following consensus problem with dropped index β_i,

$$\min_{\theta^i, \bar{\theta}} f^i(\theta^i) \quad \text{s.t.} \quad \theta^i = \bar{\theta}, \tag{1}$$

where each $f^i_t(\theta) \overset{def}{=} -\mathbb{P}_n \ell^i(\theta)$ defines an asymptotically consistent M-estimator of parameter θ, whose true value is θ^*; $\mathbb{P}_n$ denotes the empirical average on n samples. The corresponding ADMM algorithm is

$$\theta^i_{t+1} = \arg\min_{\theta^i} \{f^i(\theta^i) + {\lambda^i_t}^T(\theta^i - \bar{\theta}_t) + \frac{1}{2}\rho^i||\theta^i - \bar{\theta}_t||^2\},$$
$$\bar{\theta}_{t+1} = \sum_i \rho^i \theta^i_t / \sum_i \rho^i,$$
$$\lambda^i_{t+1} = \lambda^i_t + \rho^i(\theta^i_t - \bar{\theta}_t),$$

where t represents the iteration. We prove the result by induction on iteration t. Assume that $\lambda^i_t \xrightarrow{p} 0$ and $\theta^i_t \xrightarrow{p} \theta^*$ at the t-th iteration; we want to prove $\lambda^i_{t+1} \xrightarrow{p} 0$ and $\theta^i_{t+1} \xrightarrow{p} \theta^*$ at the $(t+1)$-th iteration.

First, note that $\lambda^i_{t+1} = \lambda^i_t + \rho^i(\theta^i_t - \bar{\theta}_t)$, and by the induction assumption, $\lambda^i_t \xrightarrow{p} 0$, $\theta^i_t - \bar{\theta}_t \xrightarrow{p} 0$. Therefore $\lambda^i_{t+1} \xrightarrow{p} 0$ follows immediately by the continuous mapping theorem.

To show $\theta^i_{t+1} \xrightarrow{p} \theta^*$, note that θ^i_{t+1} minimizes

$$f^i_t(\theta^i) = f^i(\theta^i) + {\lambda^i_t}^T(\theta^i - \bar{\theta}_t) + \frac{1}{2}\rho^i||\theta^i - \bar{\theta}_t||^2.$$

Treating $f^i_t(\cdot)$ as an M-estimator, we have (van der Vaart, 1998)

$$(\theta^i_{\beta_i} - \theta^*_{\beta_i}) = -(H^i + \rho^i \mathbf{1})^{-1}(\nabla f^i_t(\theta^*)) + o_p(\frac{1}{\sqrt{n}}),$$

where $H^i = \mathbb{E}(\nabla^2 f^i_t(\theta^*))$ and $\nabla f^i_t(\theta^*) = \nabla f^i(\theta^i) + \lambda^i_t + \rho^i(\theta^* - \bar{\theta}^i_t)$; $\mathbf{1}$ denotes the identity matrix.

Note that $\nabla f^i(\theta^*) = -\mathbb{P}_n \nabla \ell^i(\theta^*) \xrightarrow{p} -\mathbb{E}\nabla\ell^i(\theta^*) = 0$, and $\lambda^i_t \xrightarrow{p} 0$, $\theta^* - \bar{\theta}^i_t \xrightarrow{p} 0$ (by the induction assumption), we have $\nabla f^i_t(\theta^*) \xrightarrow{p} 0$. By Slutsky's theorem, we establish that $\theta^i_{\beta_i} \xrightarrow{p} \theta^*_{\beta_i}$.

The proof is completed by induction. □

Asymptotic Results of Linear Consensus

Theorem 4.1. *Assume $\hat{W}^i \xrightarrow{p} W^i$ and $\sum_i W^i$ is an invertible matrix. Let $\hat{\theta}^{\text{linear}} = (\sum_i \hat{W}^i)^{-1} \sum_i \hat{W}^i \hat{\theta}^i$, then $\hat{\theta}^{\text{linear}}$ is asymptotically consistent and asymptotically normal, with an asymptotic variance of* $\text{var}\big[(\sum_i W^i)^{-1} \sum_i W^i s^i\big]$.

Proof. For notation, let $W = \sum_i W^i$ and $\hat{W} = \sum_i \hat{W}^i$, hence $\hat{W} \xrightarrow{p} W$. Since W is invertible, we have $\hat{W}^{-1} \xrightarrow{p} W^{-1}$ and $\hat{W}^{-1}$ is invertible as $n \to +\infty$.

First, recall that we require that the weights $\hat{W}^i$ has non-zero elements only on its $\beta_i \times \beta_i$ sub-matrix; one can show that this implies $\theta^* = (\hat{W})^{-1} \sum_i \hat{W}^i \theta^*$. Since $\hat{\theta}^i \xrightarrow{p} \theta^*$, by the continuous mapping theorem, we have $\hat{\theta}^{\text{linear}} \xrightarrow{p} \theta^*$.

To establish the asymptotic normality, note that by standard asymptotic results, we have

$$\sqrt{n}(\hat{\theta}_i - \theta^*) = \hat{S}^i + o_p(1). \tag{2}$$

where $\hat{S}^i = \sqrt{n}\mathbb{P}_n s^i$, and $\mathbb{P}_n$ denotes the empirical average on n samples. Since $\hat{\theta}^{\text{linear}} = W^{-1}\sum_i \hat{W}^i \hat{\theta}^i$ and $\theta^* = (\hat{W})^{-1} \sum_i \hat{W}^i \theta^*$, taking a linear combination of estimators in (2) yields

$$\sqrt{n}(\hat{\theta}^{\text{linear}} - \theta^*) = \hat{S}^0 + o_p(1), \tag{3}$$

where $\hat{S}^0 = \hat{W}^{-1} \sum_i \hat{W}^i \hat{S}^i$. Since $\hat{W}^i \xrightarrow{p} W^i$ and $\sum_i W^i$ is invertible, by Slutsky's theorem, we have $\hat{S}^0 \rightsquigarrow W^{-1}\sum_i W^i \hat{S}^i = \sqrt{n}\mathbb{P}_n(W^{-1}\sum_i W^i s^i)$. Denote $s^0 = W^{-1}\sum_i W^i s^i$; then we have $\mathbb{E}(s^0) = 0$ and $\text{var}(s^0) = \text{var}[W^{-1}\sum_i W^i s^i] \overset{def}{=} V$. Therefore, by the central limit theorem, we have $\hat{S}^0 \rightsquigarrow \mathcal{N}(0, V)$. Finally, we have $\sqrt{n}(\hat{\theta}^{\text{linear}} - \theta^*) \rightsquigarrow \mathcal{N}(0, V)$ by applying the Slutsky's theorem on (3). □

Asymptotic Results of Max Consensus

Theorem 4.3. *The max consensus estimator $\hat{\theta}^{\max}$ as defined in (5) is asymptotically consistent. Further, for any $\alpha \in \mathcal{I}$, if $\hat{w}^i_\alpha \xrightarrow{p} w^i_\alpha$ and $w^{i_0}_\alpha > \max_{i \in \alpha, i \neq i_0} w^i_\alpha$, then the α-th element of $\hat{\theta}^{\max}$ is asymptotically normal, with the asymptotic variance equal to the α-th diagonal element of V^{i_0}.*

Proof. We start by proving asymptotic consistency. Note that for any $\alpha \in \mathcal{I}$ and any $\epsilon > 0$,

$$\begin{aligned} &\Pr(||\hat{\theta}^{\max}_\alpha - \theta^*_\alpha|| > \epsilon) \\ &\le \Pr(\cup_i |\hat{\theta}^i_\alpha - \theta^*_\alpha| > \epsilon) \\ &\le \sum_i \Pr(|\hat{\theta}^i_\alpha - \theta^*_\alpha| > \epsilon) \to 0, \end{aligned}$$

where the last step follows by the asymptotic consistency of $\hat{\theta}^i$. Therefore, we have $\hat{\theta}^{\max}_\alpha \xrightarrow{p} \theta^*_\alpha$.

To establish the asymptotic normality of $\hat{\theta}^{\max}_\alpha$, we just need to prove that $\hat{\theta}^{\max}_\alpha \xrightarrow{p} \hat{\theta}^i_\alpha$. To see this, note that for any $\epsilon > 0$

$$\begin{aligned}
&\Pr(|\hat{\theta}^{\max}_\alpha - \theta^{i_0}_\alpha| > \epsilon) \\
&\le \Pr(\cup_{j \ne i_0} \{\hat{w}^{i_0}_\alpha \le \hat{w}^j_\alpha\}) \\
&\le \sum_{j \ne i_0} \Pr(\hat{w}^{i_0}_\alpha \le \hat{w}^j_\alpha).
\end{aligned}$$

Recall that $\hat{w}^i_\alpha \xrightarrow{p} w^i_\alpha$ and that $w^{i_0}_\alpha$ is strictly larger than the others. We have $(\hat{w}^{i_0}_\alpha - \hat{w}^j_\alpha) \xrightarrow{p} (w^{i_0}{}_\alpha - w^j{}_\alpha) > 0$, and hence $\Pr(\hat{w}^{i_0}_\alpha \le \hat{w}^j_\alpha) \to 0$. This implies

$$\Pr(|\hat{\theta}^{\max}_\alpha - \theta^{i_0}_\alpha| > \epsilon) \to 0,$$

that is, $\hat{\theta}^{\max}_\alpha \xrightarrow{p} \theta^{i_0}_\alpha$. Since $\theta^{i_0}_\alpha \rightsquigarrow \mathcal{N}(0, V^i_{\alpha,\alpha})$, we establish that $\hat{\theta}^{\max}_\alpha \rightsquigarrow \mathcal{N}(0, V^i_{\alpha,\alpha})$ by the chain rule (van der Vaart, 1998, Theorem 2.7 in page 10). □

Weights for Max Consensus

Proposition 4.4. *For the max consensus estimator $\hat{\theta}^{\max}$ as defined in (5), the weight $w^i_\alpha = 1/V^i_{\alpha,\alpha}$ achieves minimum least square error asymptotically.*

Proof. The result is straightforward, by noting that the optimization is de-coupled in terms of the parameter components. For each parameter component θ_α, the smallest possible asymptotic variance of $\hat{\theta}^{\max}_\alpha$ is $\min_{i \in \alpha}\{V^i_{\alpha,\alpha}\}$, which is achieved by max consensus if the weights are set to be $w^i_\alpha = 1/V^i_{\alpha,\alpha}$. □

Interestingly, if the weights are constrained to be a single scalar for each local estimator, i.e., $w^i = w^i_\alpha$ for all $\alpha \in \beta_i$, finding optimal weights becomes significantly more difficult, and can be framed as a quadratic assignment problem.

Matrix Weights for Linear Consensus

By Theorem 4.1, the optimal weight problem for matrix linear consensus is formulated as

$$\min_{W^i} \operatorname{tr}[\operatorname{var}(\sum_i W^i s^i)] \quad \text{s.t.} \quad \sum_i W^i = \mathbf{1}, \qquad (4)$$

where $\mathbf{1}$ denotes the identity matrix of the same size as W^i. If each local estimator $\hat{\theta}^i$ is only non-degenerate on θ_{β_i} (as is the case in our paper), the optimization should be subject to the additional constraint that W^i has non-zero elements only on $\beta_i \times \beta_i$ sub-matrix.

Proposition 4.5. *Suppose $\hat{\theta}^i$ are information unbiased. If $\operatorname{cov}(s^i, s^j) = 0$ for all $i \ne j$, then $W^i = (\sum_i H^i)^{-1} H^i$ is the solution of (4).*

Proof. When the $\operatorname{cov}(s^i, s^j) = 0$, (4) reduces to

$$\min_{W^i} \sum_i \operatorname{tr}(W^i V^i {W^i}^T) \quad \text{s.t.} \quad \sum_i W^i = \mathbf{1}, \qquad (5)$$

where $V^i = \operatorname{var}(s^i)$. Solving this quadratic program (with the additional constraint that W^i is non-zero only on the $\beta_i \times \beta_i$ sub-matrix), we get a closed form solution, $W^i = (\sum_i \Lambda^i)^{-1} \Lambda^i$, where Λ^i is defined by $\Lambda^i_{\beta_i,\beta_i} = (V^i_{\beta_i,\beta_i})^{-1}$ and zero on the other elements. Since for information unbiased estimators, we have $H^i = \Lambda^i$, the result follows. □

Vector Weights for Linear Consensus

Proposition 4.6. *For the linear consensus estimator $\hat{\theta}^{\text{linear}}$ as defined in (4), the weights $w_\alpha = V_\alpha^{-1} e$, where e is a column vector of all ones, achieves the minimum asymptotic least square error.*

Proof. First, note that the optimization is decoupled w.r.t. to the parameter components. Consider the component θ_α; the relevant optimization problem is

$$\min_{w_\alpha} \{\operatorname{var}(\sum_{i \in \alpha} w^i_\alpha s^i_\alpha)\} \quad \text{s.t.} \quad \sum_{i \in \alpha} w^i_\alpha = 1.$$

which is equivalent to

$$\min_{w_\alpha} w^T_\alpha V_\alpha w_\alpha \quad \text{s.t.} \quad \sum_{i \in \alpha} w^i_\alpha = 1. \qquad (6)$$

The conclusion follows by solving this simple quadratic optimization. □

Proposition 4.7. *If $\operatorname{cov}(s^i, s^j) = 0$, $\forall i \ne j$, then the linear consensus estimator $\hat{\theta}^{\text{linear}}$ as defined in (4), achieves the lowest asymptotic MSE when taking weights $w^i_\alpha = 1/V^i_{\alpha,\alpha}$.*

Proof. When $\operatorname{cov}(s^i, s^j) = 0$, V_α is a diagonal matrix. The optimization in (6) reduces to

$$\min \sum_{i \in \alpha} V^i_{\alpha,\alpha} (w^i_\alpha)^2 \quad \text{s.t.} \quad \sum_{i \in \alpha} w^i_\alpha = 1.$$

Solving this quadratic yields the conclusion. □

Proposition 4.8. *If s^i $(i = 1, \ldots, p)$ are deterministically positively correlated, i.e., there exists a random vector s^0, and constants $v^i_\alpha \ge 0$, such that $s^i_\alpha = v^i_\alpha s^0_\alpha$, then the optimal vector weights $\{w^i_\alpha\}$ for linear consensus, under the constraint $w^i_\alpha \ge 0$, are $w^i_\alpha = 1$ if $v^i_\alpha \le v^j_\alpha$ for any $j \in \alpha$ and $w^i_\alpha = 0$ if otherwise.*

Proof. By the assumption of deterministic correlation, we have $V_\alpha^{ij} = v_\alpha^i v_\alpha^j \mathrm{var}(s_\alpha^0)$. The optimization in (6) reduces to

$$\min(\sum_{i\in\alpha} w_\alpha^i v_\alpha^i)^2, \quad \text{s.t.} \quad \sum_{i\in\alpha} w_\alpha^i = 1, w_\alpha^i \geq 0.$$

The conclusion follows by solving the quadratic form. □

Scalar Weights for Linear Consensus

Here, we consider a more restricted choice of weights, in which the performance of the i-th sensor for estimating all its related parameters θ_{β_i} is quantified by a single scalar w^i, that is, $w^i = w_\alpha^i$ for all $\alpha \in \beta_i$. It turns out that it is more difficult to find the optimal weights in this case, because the equality constraints induce global constraints on the overall problem.

In the case of linear consensus, we show in the following that the optimum scalar weights have a closed form solution, which involves a global matrix inversion. Let $V_\mathcal{I}$ be a $p \times p$ matrix with $V_\mathcal{I}^{ij} = \sum_{\alpha\in\beta_i} \mathrm{cov}(s_\alpha^i, s_\alpha^j)$, that is, $V_\mathcal{I}$ is the sum of covariance between estimator i and j over all the parameters that they share. We have

Proposition A.1. *Among the scalar weights w^i for linear consensus estimators $\hat{\theta}^{\mathrm{linear}}$, the weight setting $w^i = (V_\mathcal{I})^{-1}e$, where e is a column vector of all ones, achieves the lowest asymptotic mean square error.*

Proof. With the set of scalar weights w^i, the problem of minimizing the trace of the asymptotic variance reduces to

$$\min_w w^T V_\mathcal{I} w \quad \text{s.t.} \quad \sum_{i=1}^p w^i = 1.$$

Solving it leads to $w = V_\mathcal{I}^{-1} e$ (up to a constant). Note that w is the sum of columns of $V_\mathcal{I}^{-1}$. □

Analogous to Proposition 4.7, we have

Proposition A.2. *If $\mathrm{cov}(s^i, s^j) = 0$, $\forall i \neq j$, then among the scalar weights w^i for linear consensus estimators $\hat{\theta}^{\mathrm{linear}}$, the weight setting $w^i = 1/\mathrm{tr}(V^i)$ achieves the lowest asymptotic mean square error.*

Proof. Similar to the proof of Proposition 4.7. □