A Partition-Based Relaxed ADMM for Distributed Convex Optimization over Lossy Networks: Technical Proofs

N. Bastianello, R. Carli, L. Schenato, M. Todescato

APPENDIX

In this paper we describe the technical proofs for the results presented in [1].

A. Proof of Proposition 1

As we showed in Section III-A of the main paper, it is possible to reformulate the partition-based problem (8) so that it conforms to problem

\[ \begin{align*}
\min_x & \quad \{ f(x) + \iota(I-P)(y) \} \\
\text{s.t.} & \quad Ax + y = 0
\end{align*} \tag{A1} \]

to which the R-ADMM can be applied. The three update equations (4), (5) and (6) that characterize the R-ADMM applied to problem (A1) yield

\[ y(k+1) = \arg \min_y \{ \mathcal{L}_\rho(x, y; w(k)) \} \tag{A2} \]
\[ w(k+1) = w(k) - \rho(\sum_{j \in \mathcal{N}_i} \xi(j) - \rho z(k)) \tag{A3} \]
\[ x(k+1) = \arg \min_x \mathcal{L}_\rho(x, y(k+1); w(k+1)) \tag{A4} \]

where \( w \) is the vector of Lagrange multipliers and the augmented Lagrangian is

\[ \mathcal{L}_\rho(x, y; w) = f(x) + \iota(I-P)(y) - w^\top (Ax + y) + \frac{\rho}{2} \|Ax + y\|^2. \]

However, as shown in [2], the R-ADMM for problem (A1) can be equivalently characterized with the set of four iterates

\[ y(k) = \arg \min_y \left\{ -z^\top(y) + \frac{\rho}{2} \|y\|^2 \right\} \tag{A5} \]
\[ w(k) = z(k) - \rho y(k) \tag{A6} \]
\[ x(k) = \arg \min_x \left\{ f(x) - (2w(k) - z(k))^\top Ax \right. \]
\[ \left. + \frac{\rho}{2} \|Ax\|^2 \right\} \tag{A7} \]
\[ z(k+1) = (1 - 2\alpha)z(k) + 2\alpha(w(k) - \rho Ax(k)). \tag{A8} \]

Similarly to what has been done in [3], it is now possible to simplify Equations (A5)–(A8).

First of all, solving the system of KKT conditions for (A5) yields \( y(k) = (I + P)z(k)/(2\rho) \), and therefore Equations (A5)–(A8) become

\[ \begin{align*}
y(k) &= (I + P)z(k)/(2\rho) \tag{A9} \\
w(k) &= (I - P)z(k)/2 \\
x(k) &= \arg \min_x \left\{ f(x) + (Pz(k))^\top Ax + \frac{\rho}{2} \|Ax\|^2 \right\} \tag{A10} \]
\[ z(k+1) = (1 - \alpha)z(k) - \alpha Pz(k) - 2\alpha \rho Ax(k). \tag{A12} \]

Since we are interested in the trajectory \( k \to x(k) \) and by the fact that the update (A11) depends only on the vector \( z(k) \), then the R-ADMM for problem (A1) can be described by Equations (A11) and (A12) only.

Notice now that the trajectory \( k \to x(k) \) generated by (A11) is equivalent to that generated by (A4) if the initial condition for \( x \) is the same and if \( z(0) = w(0) + \rho y(0) \) since Equation (A6) has to hold at time \( k = 0 \). Therefore Proposition 1 is proved if we can show that (A11) and (A12) can be rewritten as (11) and (12).

Recall that the permutation matrix \( P \) swaps the element \( z_i^{(j,i)} \) with the element \( z_i^{(j,i)} \) of vector \( z \), and that the row of Ax relative to the auxiliary variable \( z_i^{(j,i)} \) is \( -x_i^{(i)} \). Therefore it follows that

\[ (Pz)^\top Ax = \left[ \begin{array}{c} \cdots \ z_i^{(j,i)} \cdots \ z_i^{(j,i)} \cdots \end{array} \right] \begin{bmatrix} \vdots \ -x_i^{(i)} \ \cdots \ -x_i^{(i)} \ \cdots \end{bmatrix} = -\sum_{i=1}^N \left\{ \sum_{j \in \mathcal{N}_i} z_i^{(j,i)} x_i^{(i)} + \sum_{j \in \mathcal{N}_i} z_j^{(j,i)} x_j^{(i)} \right\}. \]

Moreover, for each node \( i \), \( x_i^{(i)} \) appears in \( |\mathcal{N}_i| \) constraints and \( \{x_j^{(i)} \}_{j \in \mathcal{N}_i} \), in one constraint each. Hence we have

\[ \|Ax\|^2 = |\mathcal{N}_i| \left\| x_i^{(i)} \right\|^2 + \sum_{j \in \mathcal{N}_i} \left\| x_j^{(i)} \right\|^2. \]

Therefore Equations (11) and (12) can be derived from (A11) and (A12) using the particular structure of the problem, proving Proposition 1.

B. Proof of Propositions 2 and 3

As was mentioned above, the partition-based problem can be reformulated as (A1) which can be solved by the application of the R-ADMM. Therefore both the convergence results
of Propositions 2 and 3 follow from those of Propositions 2 and 3 of [3].
Indeed the R-ADMM is guaranteed to converge in both the loss-less and lossy scenarios as long as the step-size and penalty parameters are such that $0 < \alpha < 1$ and $\rho > 0$. Moreover, the components of the primal variables vector, which in the partition-based case are the subvectors $x^{(i)}$, are guaranteed to converge to the optimum value, that is, each variable $x^{(i)}_i$ converges to the optimum $x^*_i$. □

REFERENCES