Proof of Robustness of the Relaxed-PRS: a Robust ADMM Approach

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

Appendix

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

A. Derivation of Algorithm 1

First of all we derive the augmented Lagrangian (9) for problem (24), and obtain

$$\mathcal{L}_\rho(x, y; w) = \sum_{i=1}^{N} f_i(x_i) + \iota_{(I-P)}(y) + \\ - w^\top (Ax + y) + \frac{\rho}{2}\|Ax + y\|^2, \tag{A1}$$

where $\|Ax + y\|^2 = \|Ax\|^2 + \|y\|^2 + 2\langle Ax, y\rangle$. We can now proceed to derive equations (19)–(21) for the problem at hand.

1) Equation (19): By (A1) and discarding the terms that do not depend on y we get

$$y(k+1) = \arg\min_y \Big\{ \iota_{(I-P)}(y) - w^\top(k)y + \frac{\rho}{2}\|y\|^2 \\ + 2\alpha\rho\langle Ax(k), y\rangle + \rho(2\alpha - 1)\langle y, y(k)\rangle \Big\}$$

where we summed the terms with the inner product $\langle Ax(k), y\rangle$. Therefore we need to solve the problem

$$y(k+1) = \arg\min_{y=Py} \Big\{ - w^\top(k)y + \frac{\rho}{2}\|y\|^2 \\ + 2\alpha\rho\langle Ax(k), y\rangle + \rho(2\alpha - 1)\langle y, y(k)\rangle \Big\}$$

that for simplicity we can write as

$$y(k+1) = \arg\min_{y=Py} \{ h_{\alpha,\rho}(y; x(k), w(k)) \}. \tag{A2}$$

We apply now the Karush-Kuhn-Tucker (KKT) conditions [2] to problem (A2) and obtain the system

$$\nabla \Big[h_{\alpha,\rho}(y; x(k), w(k)) - \nu^\top (I - P)y \Big|_{y(k+1),\nu^*} = 0 \tag{A3}$$

$$y(k+1) = Py(k+1) \tag{A4}$$

where ν^* is the optimal value of the Lagrange multipliers of the problem.
By computing the gradient in (A3) we obtain

$$y(k+1) = \frac{1}{\rho}\big[w(k) - 2\alpha\rho Ax(k) \\ - \rho(2\alpha - 1)y(k) + (I - P)\nu^* \big]. \tag{A5}$$

The authors are with the Department of Information Engineering, University of Padova, via Gradenigo 6/b 35131, Padova, Italy. nicola.bastianello.3@studenti.unipd.it, [todescat|carlirug|schenato]@dei.unipd.it

We substitute this formula for $y(k+1)$ in the right-hand side of (A4) which results in

$$y(k+1) = \frac{1}{\rho}\big[Pw(k) - 2\alpha\rho PAx(k) \\ - \rho(2\alpha - 1)Py(k) - (I - P)\nu^* \big] \tag{A6}$$

for the fact that $P^2 = I$ and hence $P(I - P) = -(I - P)$. We sum now equations (A5) and (A6) and obtain

$$y(k+1) = \frac{1}{2\rho}(I + P)\big[w(k) - 2\alpha\rho Ax(k) - \rho(2\alpha - 1)y(k) \big]. \tag{A7}$$

Finally noting that, given a vector t of dimension equal to that of y, the ij-th element of $(I + P)t$ is equal to $t_{ij} + t_{ji}$, then the update for $y_{ij}(k + 1)$ follows.

2) Equation (20): By equation (20) and (A7) we can write

$$\begin{aligned} w(k+1) =& w(k) - 2\alpha\rho Ax(k) - \rho(2\alpha - 1)y(k) + \\ & - \frac{1}{2}(I + P)[w(k) - 2\alpha\rho Ax(k) - \rho(2\alpha - 1)y(k)] \\ =& \frac{1}{2}(I - P)[w(k) - 2\alpha\rho Ax(k) - \rho(2\alpha - 1)y(k)] \end{aligned}$$

and by the definition of $I - P$ we get the update equation for $w_{ij}(k + 1)$ stated in Algorithm 1.

3) Equation (21): Finally we apply equation (21) to the problem at hand, which means that we need to solve

$$x(k+1) = \arg\min_x \Bigg\{ \sum_{i=1}^{N} f_i(x_i) + \\ - \Big(w(k+1) - \rho y(k+1) \Big)^\top Ax + \frac{\rho}{2}\|Ax\|^2 \Bigg\}.$$

We know that each variable x_i appears in $|\mathcal{N}_i|$ constraints and therefore $\|Ax\|^2 = \sum_{i=1}^{N} |\mathcal{N}_i| \|x_i\|^2$. Moreover, given a vector t with the same size as y, we have

$$\begin{aligned} t^\top Ax &= \begin{bmatrix} \cdots & t_{ji}^\top & \cdots & t_{ji}^\top & \cdots \end{bmatrix} \begin{bmatrix} \vdots \\ -x_i \\ \vdots \\ -x_j \\ \vdots \end{bmatrix} \\ &= \sum_{(i,j)\in\mathcal{E}} \left(t_{ji}^\top x_i + t_{ij}^\top x_j \right) \\ &= \sum_{i=1}^{N} \left(\sum_{j\in\mathcal{N}_i} t_{ji}^\top \right) x_i. \end{aligned}$$

and we get the update equation for $x_i(k+1)$ substituting $\Big(w(k+1) - \rho y(k+1)\Big)$ to t. Notice that by the results obtained above we have

$$\Big(w(k+1) - \rho y(k+1)\Big) = \\ = -P[w(k) - 2\alpha\rho A x(k) - \rho(2\alpha - 1)y(k)]$$

which means that $x(k+1)$ can be computed as a function of the x, y and w variables at time k only. ■

B. Proof of Proposition 1

1) Equations (14)*:* The following derivation shares some points with the derivation described in the section above. Indeed, applying the first equation of (14) to the problem at hand requires that we solve

$$y(k) = \arg\min_{y=Py} \left\{ -z^\top(k)y + \frac{\rho}{2}\|y\|^2 \right\},$$

which can be done by solving the system of KKT conditions of the problem as performed above. The result is

$$y(k) = \frac{1}{2\rho}(I + P)z(k). \tag{A8}$$

It easily follows from (A8) that $\psi(k) = \frac{1}{2}(I - P)z(k)$.

2) Equations (15)*:* First of all we have $(2\psi(k) - z(k)) = -Pz(k)$, hence according to the same reasoning employed above to derive the expression for $x(k+1)$ we find (25). Moreover, we have $\xi(k) = -Pz(k) - \rho A x(k)$.

3) Equation (7)*:* By the results derived above we can easily compute

$$z(k+1) = (1-\alpha)z(k) - \alpha P z(k) - 2\alpha\rho A x(k)$$

which gives equations (26).

Notice that to compute the variables $y(k)$, $\psi(k)$, $x(k)$ and $\xi(k)$ we need only the variables $z(k)$. Moreover, to update z we require only $z(k)$ and $x(k)$. Hence the five update equations reduce to the updates for x and z only. ■

C. Proof of Proposition 2

To prove convergence of the R-ADMM in the two implementations of Algorithms 1 and 2, we resort to the following result, adapted from [3, Corollary 27.4].

Proposition 1 ([3, Corollary 27.4]): Consider problem (2) and assume that it has solution; let $\alpha \in (0, 1)$, $\rho > 0$, and $x(0) \in \mathcal{X}$. Assume to apply equations (5)–(7) to the problem. Then there exists z^* such that

- $x^* = \text{prox}_{\rho g}(z^*) \in \arg\min_x\{f(x) + g(x)\}$, and
- $\{z(k)\}_{k\in\mathbb{N}}$ converges weakly to z^*.

□

We need to show now that this result applies to the dual problem of problem (24). First of all, by formulation of the problem we have that f is convex and proper (and also closed). Moreover, by [3, Example 8.3] we know that the indicator function of a convex set is convex (and, by definition, proper). But the set of vectors y that satisfy $(I - P)y = 0$ is indeed convex, hence also g is convex and proper.

Now [4, Theorem 12.2] states that the convex conjugate of a convex and proper function is closed, convex and proper. Therefore both d_f and d_g are closed, convex and proper, which means that we can apply the convergence result in Proposition 1 to the dual problem of (24).

Therefore we have that $w^* = \text{prox}_{\rho d_g}(z^*)$ is indeed a solution of the dual problem and $\{z(k)\}_{k\in\mathbb{N}}$ converges to z^*. But since the duality gap is zero, then when we attain the optimum of the dual problem we have obtained that of the primal as well. ■

D. Proof of Proposition 3

In order to prove the convergence of Algorithm 3 we need to introduce a probabilistic framework in which to reformulate the KM update. For this stochastic version of the KM iteration we can state a convergence result adapted from [5, Theorem 3] and show that indeed Algorithm 3 is represented by this formulation.

We are therefore interested in altering the standard KM iteration (1) in order to include a stochastic selection of which coordinates in $\mathcal{I} = \{1, \dots, M\}$ to update at each instant. To do so we introduce the operator $\hat{T}^{(\xi)} : \mathcal{X} \to \mathcal{X}$ whose i-th coordinate is given by $\hat{T}_i^{(\xi)}x = T_i x$ if the coordinate is to be updated ($i \in \xi$), $\hat{T}_i^{(\xi)}x = x_i$ otherwise ($i \notin \xi$). In general the subset of coordinates to be updated changes from one instant to the next. Therefore, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define the random i.i.d. sequence $\{\xi_k\}_{k\in\mathbb{N}}$, with $\xi_k : \Omega \to 2^{\mathcal{I}}$, to keep track of which coordinates are updated at each instant. The stochastic KM iteration is finally defined as

$$x(k+1) = (1-\alpha)x(k) + \alpha\hat{T}^{(\xi_{k+1})}x(k) \tag{A9}$$

and consists of the α-averaging of a stochastic operator.

The stochastic iteration satisfies the following convergence result, which is particularized from [5] using the fact that a nonexpansive operator is 1-averaged, and a constant step size.

Proposition 2 ([5, Theorem 3]): Let T be a nonexpansive operator with at least a fixed point, and let the step size be $\alpha \in (0, 1)$. Let $\{\xi_k\}_{k\in\mathbb{N}}$ be a random i.i.d. sequence on $2^{\mathcal{I}}$ such that

$$\forall i \in \mathcal{I},\ \exists I \in 2^{\mathcal{I}} \text{ s.t. } i \in I \text{ and } \mathbb{P}[\xi_1 = I] > 0.$$

Then for any deterministic initial condition $x(0)$ the stochastic KM iteration (A9) converges almost surely to a random variable with support in the set of fixed points of T. □

We turn now to the distributed optimization problem, in which the stochastic KM iteration is performed on the auxiliary variables z. In particular we assume that the packet loss occurs with probability p, and that in the case of packet loss the relative variable is not updated. As shown in the main paper, this update rule can be compactly written as

$$\hat{T}^{(\xi_{k+1})}z(k) = L_k z(k) + (I - L_k)Tz(k) \tag{A10}$$

where L_k is the diagonal matrix with elements the realizations of the binary random variables that model the packet loss at time k. Recall that these variables take value 1 if the packet is lost.

Substituting now the operator (A10) into (A9) we get the update equation

$$z(k+1) = (1-\alpha)z(k) + \alpha \left[L_k z(k) + (I - L_k) T z(k) \right] \tag{A11}$$

which conforms to the stochastic KM iteration for which the convergence result is stated.

Finally, notice that in the main article the α-averaging is applied before the stochastic coordinate selection, that is the update is given by

$$z(k+1) = L_k z(k) + (I - L_k) \left[(1-\alpha) z(k) + \alpha T z(k) \right]. \tag{A12}$$

However it can be easily shown that (A11) and (A12) do indeed coincide, hence proving the convergence of our update scheme. ■

References

[1] N. Bastianello, M. Todescato, R. Carli, and L. Schenato, “Distributed optimization over lossy networks via relaxed peaceman-rachford splitting: a robust admm approach,” in *European Control Conference 2018 [submitted]*.

[2] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[3] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011, vol. 408.

[4] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.

[5] P. Bianchi, W. Hachem, and F. Iutzeler, “A coordinate descent primal-dual algorithm and application to distributed asynchronous optimization,” *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2947–2957, 2016.