Passivity-based Analysis of the ADMM Algorithm for Constraint-Coupled Optimization

Ivano Notarnicola[a], Alessandro Falsone[b]

[a] Department of Electrical, Electronic and Information Engineering, Alma Mater Studiorum Università di Bologna, Bologna, Italy.
[b] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy.

Abstract

We propose a novel, system theoretic analysis of the Alternating Direction Method of Multipliers (ADMM) applied to a convex constraint-coupled optimization problem. The resulting algorithm can be interpreted as a linear, discrete-time dynamical system (modeling the multiplier ascent update) in closed loop with a static nonlinearity (representing the minimization of the augmented Lagrangian). When expressed in suitable coordinates, we prove that the discrete-time linear dynamical system has a discrete positive-real transfer function and is interconnected in closed loop with a static, passive nonlinearity. This readily shows that the origin is a stable equilibrium for the feedback interconnection. Finally, we also show global asymptotic stability of the origin for the closed-loop system and, thus, global asymptotic convergence of ADMM to the optimal solution of the optimization problem.

Keywords: ADMM, Constraint-Coupled Optimization, Passivity theory, Nonsmooth optimization, Control for optimization

1. Introduction

Many relevant applications in control engineering can be posed as constraint-coupled optimization problems. They usually involve the minimization of the sum of N independent, nonsmooth, convex cost functions whose decision variables must satisfy individual, convex constraints. On top of these constraints, all the variables are linearly coupled through equality constraints making the problem meaningful and its solution nontrivial. Formally, we focus on optimization problems in the form

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} f_i(x_i) \tag{1a}$$

$$\text{subj. to} \quad x_i \in X_i \qquad \forall\, i = 1,\dots,N \tag{1b}$$

$$\sum_{i=1}^{N} H_i x_i = b \tag{1c}$$

where $x_i \in \mathbb{R}^{n_i}$ are the decision variables, with $n_i \in \mathbb{N}$, each cost function $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ is convex (possibly nonsmooth), and each decision variable must belong to a convex and nonempty compact set $X_i \subset \mathbb{R}^{n_i}$, for all $i = 1,\dots,N$. Finally, all the decision variables must satisfy a set of $p \in \mathbb{N}$ linear, coupling constraints described by matrices $H_i \in \mathbb{R}^{p\times n_i}$, $i = 1,\dots,N$, and vector $b \in \mathbb{R}^p$.

This structure is often encountered in resource allocation problems where the coupling constraint is given by the resource budget. To ease the notation, we impose the following assumption.

Assumption 1.1. *Problem* (1) *admits a unique optimal solution* $\boldsymbol{x}_\star = [x_{1,\star}^\top \cdots x_{N,\star}^\top]^\top$ *and its dual problem admits a unique optimal vector of Lagrange multipliers* $\lambda_\star$. *Moreover, strong duality holds.*

The proposed analysis would carry over to multiple primal-dual solutions at the expense of a heavier notation.

An effective and widely adopted algorithm to address this class of problems is the Alternating Direction Method of Multipliers (ADMM). When applied to problem (1), ADMM takes the form (see, e.g., [1, §3.4] and [2, §7.3])

$$x_{i,k+1} \in \operatorname*{argmin}_{x_i\in X_i} f_i(x_i) + \tfrac{1}{2c}\left\|\lambda_k + c(H_i x_i - H_i x_{i,k} + d_k)\right\|^2 \qquad \forall\, i = 1,\dots,N \tag{2a}$$

$$d_{k+1} = \frac{1}{N}\Big(\sum_{i=1}^{N} H_i x_{i,k+1} - b\Big) \tag{2b}$$

$$\lambda_{k+1} = \lambda_k + c\, d_{k+1}, \tag{2c}$$

with $x_{i,0} \in X_i$, $d_0 \in \mathbb{R}^p$, $\lambda_0 \in \mathbb{R}^p$, and $c > 0$. Vector $d_k \in \mathbb{R}^p$ represents the current violation of the coupling constraints (1c) and $\lambda_k \in \mathbb{R}^p$ represents the current value of the Lagrange multiplier vector associated to those constraints.

In this note, we analyze (2) from a system theoretic perspective. The growing interest in taking such a perspective is testified by the numerous contributions recently appeared in the literature. In [3] the authors reinterpret different gradient-based optimization algorithms in terms of dynamical systems, without however giving proofs for their convergence. The authors of [4] study the Nesterov accelerated version of a continuous-time gradient algorithm and its discretization. In [5] the dynamical system perspective and dissipativity theory are leveraged to design Nesterov-like acceleration schemes for discrete-time gradient-based optimization algorithms. Similarly, the work in [6] leverages a control system perspective to show that gradient algorithms with Nesterov acceleration are robust to noise in the gradient evaluation. The authors in [7, 8] use Lyapunov stability and linear matrix inequalities to design first-order algorithms with robustness guarantees and acceleration schemes. In [9, 10] linear matrix inequalities and integral quadratic constraints are used to analyze and design optimization algorithms. Finally, [11] leverages linear system theory to analyze the convergence properties of the (distributed) gradient tracking algorithm. Notably, all mentioned approaches focus on first-order methods such as the gradient method and its variants and their line of analysis cannot be carried over to ADMM.

The literature concerning ADMM is quite vast and we here provide, in the interest of space, a (necessarily incomplete) summary of recent advances. It is well known that ADMM can be seen as an application of the Douglas-Rachford splitting and enjoys favorable convergence properties of monotone operators, [2]. Along this operator theory perspective, [12] proposes several acceleration schemes (without convergence proofs), while [13] studies the linear convergence of the ADMM operator to derive better convergence rates. Similarly, [14] studies the ADMM operator to find the penalty coefficient achieving the best convergence rate, but focusing of quadratic objective functions only. Convergence properties of ADMM have also been studied from a system-theory perspective. In [15], a continuous-time (accelerated) version of ADMM is interpreted as a dynamical system and stability and convergence rates are analyzed. Similarly, in [16] a continuous-time version of ADMM is interpreted as a dynamical system and analyzed by means of integral quadratic constraints. In [17], a continuous-time version of ADMM is robustified against delays using passivity theory. Finally, in [18] the convergence rate of ADMM is studied under strong convexity assumption, while in [19] a distributed version of ADMM is studied using a Lyapunov approach.

In this note we follow this latter perspective: we study ADMM for constraint-coupled optimization problems by interpreting (2) as the feedback interconnection of two discrete-time dynamical systems and we analyze the stability properties of the feedback loop leveraging passivity theory. Differently from [15–17] our analysis is carried out directly in discrete-time and, in contrast to [18] we do not require the objective function to be strongly convex. Besides being interesting per se, this approach can shed light and give insights about the inner working of ADMM, possibly leading to the synthesis of some variants or even novel optimization strategies resulting from the control-oriented interpretation.

2. System Theoretic Reformulation

In this section we rephrase the ADMM algorithm for constraint-coupled problems (2) adopting a system theoretic perspective so that its convergence analysis will be posed as an asymptotic stability problem.

2.1. Aggregate Reformulation

We start by introducing a compact formulation of (2). Collecting the primal iterates into $\boldsymbol{x}_k = [x_{1,k}^\top \cdots x_{N,k}^\top]^\top$, the N minimization steps described by (2a) can be arranged as a single optimization given by

$$\boldsymbol{x}_{k+1} \in \underset{\boldsymbol{x}\in X}{\operatorname{argmin}}\ f(\boldsymbol{x}) + \tfrac{1}{2c}\left\|\mathbf{1}\lambda_k + c\left(H_\mathrm{d}\boldsymbol{x} - H_\mathrm{d}\boldsymbol{x}_k + \mathbf{1}d_k\right)\right\|^2 \tag{3a}$$

where $H_\mathrm{d} = \mathrm{blkdiag}(H_1, \ldots, H_N)$, $\mathbf{1} = 1_N \otimes I_p$ with 1_N being the all-one vector of dimension N, $\boldsymbol{x} = [x_1^\top \cdots x_N^\top]^\top$, and $f(\boldsymbol{x}) = \sum_{i=1}^N f_i(x_i)$. Also the coupling constraints can be compactly written as $\sum_{i=1}^N H_i x_i - b = \mathbf{1}^\top H_\mathrm{d}\boldsymbol{x} - b = \mathbf{1}^\top(H_\mathrm{d}\boldsymbol{x} - H_\mathrm{d}\boldsymbol{x}_\star)$, where we used the fact that $b = \mathbf{1}^\top H_\mathrm{d}\boldsymbol{x}_\star$ for the optimal (thus feasible) solution $\boldsymbol{x}_\star$ of problem (1). Therefore, updates (2b) and (2c) can be rephrased as

$$d_{k+1} = \tfrac{1}{N}\mathbf{1}^\top\left(H_\mathrm{d}\boldsymbol{x}_{k+1} - H_\mathrm{d}\boldsymbol{x}_\star\right) \tag{3b}$$

$$\lambda_{k+1} = \lambda_k + c\, d_{k+1}. \tag{3c}$$

Exploiting the definition (3b) the term $\mathbf{1}d_k$ appearing in (3a) can be substituted by $J(H_\mathrm{d}\boldsymbol{x}_k - H_\mathrm{d}\boldsymbol{x}_\star)$ with $J = \frac{1}{N}1_N 1_N^\top \otimes I_p$, so that (3) further simplifies into

$$\boldsymbol{x}_{k+1} \in \underset{\boldsymbol{x}\in X}{\operatorname{argmin}}\ f(\boldsymbol{x}) + \tfrac{1}{2c}\|\mathbf{1}\lambda_k + c\,(-(I-J)H_\mathrm{d}\boldsymbol{x}_k + H_\mathrm{d}\boldsymbol{x} - JH_\mathrm{d}\boldsymbol{x}_\star)\|^2 \tag{4a}$$

$$\lambda_{k+1} = \lambda_k + c\,\tfrac{1}{N}\mathbf{1}^\top(H_\mathrm{d}\boldsymbol{x}_{k+1} - H_\mathrm{d}\boldsymbol{x}_\star). \tag{4b}$$

Two remarks are in order. First, notice that the quantities passed on from one iteration to the next in (4) are $(I-J)H_\mathrm{d}\boldsymbol{x}_k$ and λ_k only. Secondly, let us point out that, despite the optimization in (4a) admits multiple minimizers, it has a strictly convex objective function in $H_\mathrm{d}\boldsymbol{x}$. Hence, the quantity $H_\mathrm{d}\boldsymbol{x}_{k+1}$ is uniquely defined for all k, no matter which $\boldsymbol{x}_{k+1}$ results from (4a). In light of these two considerations, it is sensible to interpret (4) as a discrete-time (nonlinear) dynamical system with *state variables* $\lambda_k \in \mathbb{R}^p$ and $(I-J)H_\mathrm{d}\boldsymbol{x}_k \in \mathbb{R}^{pN}$.

Next we study the stability properties related to (4).

2.2. Steady-state Analysis

Let us study the equilibria of (4) by investigating its steady-state. By setting

$$\lambda_{k+1} = \lambda_k = \lambda_\mathrm{ss} \tag{5a}$$

$$(I-J)H_\mathrm{d}\boldsymbol{x}_{k+1} = (I-J)H_\mathrm{d}\boldsymbol{x}_k = \xi_\mathrm{ss} \tag{5b}$$

in (4) it gives

$$\xi_\mathrm{ss} = (I-J)H_\mathrm{d}\boldsymbol{x}^+ \tag{6a}$$

$$\boldsymbol{x}^+ \in \operatorname*{argmin}_{\boldsymbol{x}\in X} \; f(\boldsymbol{x}) + \tfrac{1}{2c}\|\mathbf{1}\lambda_{\rm ss} + c\,(H_{\rm d}\boldsymbol{x} - \xi_{\rm ss} - JH_{\rm d}\boldsymbol{x}_\star)\|^2 \tag{6b}$$

$$\lambda_{\rm ss} = \lambda_{\rm ss} + c\,\tfrac{1}{N}\mathbf{1}^\top\left(H_{\rm d}\boldsymbol{x}^+ - H_{\rm d}\boldsymbol{x}_\star\right). \tag{6c}$$

We used the symbol $\boldsymbol{x}^+$ in (6b) to stress that dependence on k is irrelevant. We shall show that the (unique) pair $(\boldsymbol{x}^+, \lambda_{\rm ss})$, associated to the (unique) pair $(\xi_{\rm ss}, \lambda_{\rm ss})$ solving (6), is the optimal primal-dual solution of problem (1).

By simplifying $\lambda_{\rm ss}$ in (6c), we have

$$0 = \mathbf{1}^\top(H_{\rm d}\boldsymbol{x}^+ - H_{\rm d}\boldsymbol{x}_\star) \Longrightarrow \mathbf{1}^\top H_{\rm d}\boldsymbol{x}^+ = \mathbf{1}^\top H_{\rm d}\boldsymbol{x}_\star = b, \tag{7}$$

which shows that $\boldsymbol{x}^+$ satisfies the coupling constraints of (1). Moreover, left-multiplying (6c) by $\mathbf{1}$ yields

$$0 = JH_{\rm d}\boldsymbol{x}^+ - JH_{\rm d}\boldsymbol{x}_\star = H_{\rm d}\boldsymbol{x}^+ - \xi_{\rm ss} - JH_{\rm d}\boldsymbol{x}_\star, \tag{8}$$

where in the last equality we used (6a). By plugging (8) into (6b), it becomes

$$\boldsymbol{x}^+ \in \operatorname*{argmin}_{\boldsymbol{x}\in X} \; f(\boldsymbol{x}) + \tfrac{1}{2c}\|\mathbf{1}\lambda_{\rm ss} + c\,(H_{\rm d}\boldsymbol{x} - H_{\rm d}\boldsymbol{x}^+)\|^2.$$

By applying [1, Lemma 4.1, p. 257] to the previous optimization problem, it turns out that $\boldsymbol{x}^+$ must satisfy

$$f(\boldsymbol{x}^+) + (\mathbf{1}\lambda_{\rm ss})^\top H_{\rm d}\boldsymbol{x}^+ \leq f(\boldsymbol{x}) + (\mathbf{1}\lambda_{\rm ss})^\top H_{\rm d}\boldsymbol{x} \tag{9}$$

for any $\boldsymbol{x} \in X$. Evaluating the right-hand side of (9) at $\boldsymbol{x}$ equal to the optimal solution $\boldsymbol{x}_\star$ of (1) gives

$$f(\boldsymbol{x}^+) + \lambda_{\rm ss}^\top\mathbf{1}^\top H_{\rm d}\boldsymbol{x}^+ \leq f(\boldsymbol{x}_\star) + \lambda_{\rm ss}^\top\mathbf{1}^\top H_{\rm d}\boldsymbol{x}_\star. \tag{10}$$

Since $\mathbf{1}^\top H_{\rm d}\boldsymbol{x}^+ = \mathbf{1}^\top H_{\rm d}\boldsymbol{x}_\star$ by (7), then $f(\boldsymbol{x}^+) \leq f(\boldsymbol{x}_\star)$. Since $\boldsymbol{x}^+$ is in X and satisfies also the coupling constraints, then $\boldsymbol{x}^+$ is feasible for (1), hence $f(\boldsymbol{x}^+) = f(\boldsymbol{x}_\star)$ and $\boldsymbol{x}^+$ is the (unique) optimal solution of problem (1).

Finally, combining (7) and (9), we get for all $\boldsymbol{x} \in X$

$$\begin{aligned} f(\boldsymbol{x}^+) &\leq f(\boldsymbol{x}) + \lambda_{\rm ss}^\top(\mathbf{1}^\top H_{\rm d}\boldsymbol{x} - \mathbf{1}^\top H_{\rm d}\boldsymbol{x}^+) \\ &= f(\boldsymbol{x}) + \lambda_{\rm ss}^\top(\mathbf{1}^\top H_{\rm d}\boldsymbol{x} - b) = f(\boldsymbol{x}) + \lambda_{\rm ss}^\top(H\boldsymbol{x} - b). \end{aligned}$$

If we set $\boldsymbol{x}$ in the previous relation equal to any element of $\operatorname{argmin}_{\boldsymbol{x}\in X} f(\boldsymbol{x}) + \lambda_{\rm ss}^\top[H\boldsymbol{x} - b]$, we have

$$f(\boldsymbol{x}_\star) = f(\boldsymbol{x}^+) \leq \min_{\boldsymbol{x}\in X} f(\boldsymbol{x}) + \lambda_{\rm ss}^\top[H\boldsymbol{x} - b] = q(\lambda_{\rm ss}), \tag{11}$$

where $q(\lambda) = \sum_{i=1}^N \min_{x_i\in X_i}\left(f_i(x_i) + \lambda(H_ix_i - H_ix_{i,\star})\right)$ is the dual function of (1). Since $q(\lambda_{\rm ss}) \leq \max_\lambda q(\lambda) = q(\lambda_\star)$ and $q(\lambda_\star) = f(\boldsymbol{x}_\star)$ by strong duality, inequality (11) shows that $\lambda_{\rm ss}$ is the (unique) optimal dual solution.

The previous discussion shows that any equilibrium $(\xi_{\rm ss}, \lambda_{\rm ss})$ of (4) is associated with the unique primal-dual optimal pair $(\boldsymbol{x}_\star, \lambda_\star)$. We can thus safely take $\xi_{\rm ss} = (I-J)H_{\rm d}\boldsymbol{x}_\star$ and $\lambda_{\rm ss} = \lambda_\star$ and, since $\boldsymbol{x}_\star$ and $\lambda_\star$ are unique, then also the equilibrium $(\lambda_{\rm ss}, \xi_{\rm ss})$ is unique.

2.3. Feedback Interconnection

Let us perform a change of coordinates to shift the equilibrium to the origin by defining the state vector

$$\begin{bmatrix} e_k \\ v_k \end{bmatrix} = \begin{bmatrix} \lambda_k - \lambda_\star \\ c\,(I-J)(H_{\rm d}\boldsymbol{x}_k - H_{\rm d}\boldsymbol{x}_\star) \end{bmatrix},$$

with $e_k \in \mathbb{R}^p$ and $v_k \in \mathbb{R}^{pN}$. Then, in order to express (4a) in terms of (e_k, v_k), we add and subtract $H_{\rm d}\boldsymbol{x}_\star$ and $\mathbf{1}\lambda_\star$ inside the squared-norm term in (4a) to write

$$\boldsymbol{x}_{k+1} \in \operatorname*{argmin}_{\boldsymbol{x}\in X} f(\boldsymbol{x}) + \tfrac{1}{2c}\|\mathbf{1}\lambda_\star + c\,(H_{\rm d}\boldsymbol{x} - H_{\rm d}\boldsymbol{x}_\star) + \mathbf{1}e_k - v_k\|^2.$$

Introducing the following nonlinear map $\varphi(\cdot) : \mathbb{R}^{pN} \to \mathbb{R}^{pN}$

$$\varphi(y) = c\,(H_{\rm d}\boldsymbol{x}^+ - H_{\rm d}\boldsymbol{x}_\star) \tag{12a}$$

$$\boldsymbol{x}^+ \in \operatorname*{argmin}_{\boldsymbol{x}\in X} f(\boldsymbol{x}) + \tfrac{1}{2c}\left\|\mathbf{1}\lambda_\star + c\,(H_{\rm d}\boldsymbol{x} - H_{\rm d}\boldsymbol{x}_\star) + y\right\|^2, \tag{12b}$$

system (4) can be conveniently rewritten as

$$\begin{bmatrix} e_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} e_k + \frac{1}{N}\mathbf{1}^\top\varphi(\mathbf{1}e_k - v_k) \\ (I-J)\varphi(\mathbf{1}e_k - v_k) \end{bmatrix}. \tag{13}$$

Defining $y_k = \mathbf{1}e_k - v_k$ and $u_k = \varphi(\mathbf{1}e_k - v_k) = \varphi(y_k)$, we can rewrite (13) as

$$\Sigma : \begin{cases} \begin{bmatrix} e_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} e_k \\ v_k \end{bmatrix} + \begin{bmatrix} \frac{1}{N}\mathbf{1}^\top \\ I - J \end{bmatrix} u_k \\ \qquad y_k = \begin{bmatrix} \mathbf{1} & -I \end{bmatrix}\begin{bmatrix} e_k \\ v_k \end{bmatrix} \end{cases} \tag{14a}$$

$$u_k = \varphi(y_k), \tag{14b}$$

which highlights how the *autonomous* discrete-time dynamical system in (13) can be easily interpreted as the feedback interconnection of a linear dynamical system Σ with states e_k and v_k, input u_k and output y_k, and the static nonlinearity $\varphi(\cdot)$. In light of the reformulation above, the standard convergence result for (2) (see, e.g., [1, Prop. 4.2, p. 256]) can be restated as follows.

Theorem 2.1. *Under Assumption 1.1, the origin is a globally asymptotically stable equilibrium for the closed-loop system in* (14).

In the following Section 3 we will prove Theorem 2.1 leveraging a clever reformulation of (14) together with arguments from discrete-time passivity theory.

3. Passivity-based Stability Analysis

We investigate the two subsystems in (14) separately and, then, combine their properties to assert global asymptotic stability of the origin for their feedback interconnection.

3.1. Passivity of the Nonlinear Part

We first focus on the static, nonlinear part of the feedback loop in (14). By [1, Lemma 4.1, p. 257] applied to (12b), given any $y \in \mathbb{R}^{pN}$ we have that $\boldsymbol{x}^+$ satisfies

$$\begin{aligned} & f(\boldsymbol{x}^+) + [\mathbf{1}\lambda_\star + c\,(H_\mathrm{d}\boldsymbol{x}^+ - H_\mathrm{d}\boldsymbol{x}_\star) + y]^\top H_\mathrm{d}\boldsymbol{x}^+ \\ & \qquad \le f(\boldsymbol{x}) + [\mathbf{1}\lambda_\star + c\,(H_\mathrm{d}\boldsymbol{x}^+ - H_\mathrm{d}\boldsymbol{x}_\star) + y]^\top H_\mathrm{d}\boldsymbol{x}, \end{aligned}$$

for all $\boldsymbol{x} \in X$. Setting $\boldsymbol{x} = \boldsymbol{x}_\star$, scaling by c and using (12a), the previous inequality can be rearranged as

$$c\,f(\boldsymbol{x}^+) + [\mathbf{1}\lambda_\star + \varphi(y) + y]^\top \varphi(y) \le c\,f(\boldsymbol{x}_\star). \tag{15}$$

Optimality of $\boldsymbol{x}_\star$ and the Saddle-Point Theorem imply

$$\begin{aligned} f(\boldsymbol{x}_\star) &\le f(\boldsymbol{x}^+) + \lambda_\star^\top (H\boldsymbol{x}^+ - b) \\ &= f(\boldsymbol{x}^+) + \lambda_\star^\top \mathbf{1}^\top (H_\mathrm{d}\boldsymbol{x}^+ - H_\mathrm{d}\boldsymbol{x}_\star), \end{aligned}$$

with $H = \mathbf{1}^\top H_\mathrm{d}$. Using $\varphi(\cdot)$ as in (12a), we can combine (15) with the previous inequality to get

$$[\varphi(y) + y]^\top \varphi(y) \le 0. \tag{16}$$

Condition (16) proves that the nonlinear part of the feedback loop is an output strictly passive (static) system. This property is stronger than mere passivity and it means that $\varphi(\cdot)$ actually has an excess of passivity. Such excess can be exploited to manipulate the feedback interconnection if needed as discussed in the following subsection.

3.2. Loop Transformation

If also Σ had been passive, then the whole interconnection would have been. Unfortunately, this is not the case. However, as is customary in a passivity framework, one can "steal" the excess of passivity in one system to compensate the shortage in the other.

To this end, we define a new output $\tilde{y}_k = y_k + u_k$ and, consistently, a new static nonlinearity $\tilde{\varphi}(\cdot) : \mathbb{R}^{pN} \to \mathbb{R}^{pN}$ such that $\tilde{\varphi}(\tilde{y}_k) = \varphi(y_k)$. Accordingly, (14) becomes

$$\widetilde{\Sigma} : \begin{cases} \begin{bmatrix} e_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} e_k \\ v_k \end{bmatrix} + \begin{bmatrix} \frac{1}{N}\mathbf{1}^\top \\ I - J \end{bmatrix} u_k \\ \tilde{y}_k = \begin{bmatrix} \mathbf{1} & -I \end{bmatrix} \begin{bmatrix} e_k \\ v_k \end{bmatrix} + u_k \end{cases} \tag{17a}$$

$$u_k = \tilde{\varphi}(\tilde{y}_k). \tag{17b}$$

We shall emphasize that (17) and (14) are equivalent. However, in the former the input u_k directly affects the output $\tilde{y}_k$ resulting in $\tilde{\varphi}(\cdot)$ defined only implicitly as

$$\begin{aligned} & \tilde{\varphi}(\tilde{y}) = c\,(H_\mathrm{d}\boldsymbol{x}^+ - H_\mathrm{d}\boldsymbol{x}_\star) \\ & \boldsymbol{x}^+ \in \operatorname*{argmin}_{\boldsymbol{x}\in X}\; f(\boldsymbol{x}) + \tfrac{1}{2c}\left\|\mathbf{1}\lambda_\star + c\,(H_\mathrm{d}\boldsymbol{x} - H_\mathrm{d}\boldsymbol{x}^+) + \tilde{y}\right\|^2. \end{aligned}$$

On the other hand, from (16) one can see that $\tilde{\varphi}(\cdot)$ satisfies

$$\tilde{y}_k^\top \tilde{\varphi}(\tilde{y}_k) = \tilde{y}_k^\top u_k \le 0, \tag{18}$$

proving that passivity of the nonlinear part has been preserved and made tight.

By looking at the second state equation in (17a), we can see that the state v_k satisfies $\mathbf{1}^\top v_k = 0$ for all $k \ge 1$ regardless of the value of the input u_k. It means that the state-space representation of $\widetilde{\Sigma}$ is not minimal (i.e., $\widetilde{\Sigma}$ can be described with fewer states). Though not strictly necessary, let us consider a minimal realization of (17a). To this end, consider the orthogonal projection $\tilde{v}_k$ of v_k onto the orthogonal complement of the subspace spanned by $\mathbf{1}$. Let $S \in \mathbb{R}^{pN \times p(N-1)}$ be the matrix whose columns form an orthonormal basis of such orthogonal complement. Then $S^\top \mathbf{1} = 0$, $S^\top S = I_{p(N-1)}$ and $SS^\top = I_{pN} - J$. We thus have $\tilde{v}_k = S^\top v_k$ and we can rewrite (17a) as

$$\widehat{\Sigma} : \begin{cases} \begin{bmatrix} e_{k+1} \\ \tilde{v}_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} e_k \\ \tilde{v}_k \end{bmatrix} + \begin{bmatrix} \frac{1}{N}\mathbf{1}^\top \\ S^\top \end{bmatrix} u_k \\ \tilde{y}_k = \begin{bmatrix} \mathbf{1} & -S \end{bmatrix} \begin{bmatrix} e_k \\ \tilde{v}_k \end{bmatrix} + u_k \end{cases} \tag{19}$$

which is a *minimal* state-space realization of $\widetilde{\Sigma}$.

3.3. Passivity of the Linear Part

We next focus on the linear part $\widehat{\Sigma}$ and prove that it is a passive system by showing that its transfer matrix $G(z)$ is positive real, [20, Section 3]. Passivity will then follow by the KYP Lemma, [20, Lemma 3]. Let

$$A = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix},\; B = \begin{bmatrix} \frac{1}{N}\mathbf{1}^\top \\ S^\top \end{bmatrix},\; C = \begin{bmatrix} \mathbf{1} & -S \end{bmatrix},\; D = I. \tag{20}$$

The transfer matrix of the discrete-time system $\widehat{\Sigma}$ is

$$\begin{aligned} G(z) &= C(zI - A)^{-1}B + D \\ &= \begin{bmatrix} \mathbf{1} & -S \end{bmatrix} \begin{bmatrix} \frac{1}{z-1}I & \\ & \frac{1}{z}I \end{bmatrix} \begin{bmatrix} \frac{1}{N}\mathbf{1}^\top \\ S^\top \end{bmatrix} + I \\ &= \tfrac{1}{z-1}\tfrac{1}{N}\mathbf{1}\mathbf{1}^\top - \tfrac{1}{z}SS^\top + I \\ &= \left(\tfrac{1}{z-1} + \tfrac{1}{z}\right) J + \left(1 - \tfrac{1}{z}\right) I, \end{aligned} \tag{21}$$

where $z \in \mathbb{C}$ is the complex argument. The poles of $G(z)$ are in $z = 0$ and $z = 1$ and all elements of $G(z)$ are analytic for each z such that $|z| > 1$.

Recalling the properties of S, we can define the (orthonormal) transformation matrix $T = [\frac{1}{\sqrt{N}}\mathbf{1} \;\; S] \in \mathbb{R}^{pN \times pN}$ which jointly diagonalizes J and I_{pN} as

$$J = T \begin{bmatrix} I_p & \\ & 0_{p(N-1)} \end{bmatrix} T^\top, \quad I_{pN} = T \begin{bmatrix} I_p & \\ & I_{p(N-1)} \end{bmatrix} T^\top.$$

This fact can be used in (21) to get

$$G(z) = T \begin{bmatrix} \frac{z}{z-1} I_p & \\ & \frac{z-1}{z} I_{p(N-1)} \end{bmatrix} T^\top$$

and

$$\tfrac{1}{2}\big(G(z) + G(z)^\top\big)$$

$$= \tfrac{1}{2} T \begin{bmatrix} \left(\frac{z}{z-1} + \frac{\bar{z}}{\bar{z}-1}\right) I_p & \\ & \left(\frac{z-1}{z} + \frac{\bar{z}-1}{\bar{z}}\right) I_{p(N-1)} \end{bmatrix} T^\top$$

$$= \left(|z|^2 - \Re[z]\right) T \begin{bmatrix} \frac{1}{|z-1|^2} I_p & \\ & \frac{1}{|z|^2} I_{p(N-1)} \end{bmatrix} T^\top \succ 0$$

for all $z \in \mathbb{C}$ such that $|z| > 1$. This shows that $G(z)$ is discrete positive real and, hence, system $\widehat{\Sigma}$ is passive.

3.4. Proof of Theorem 2.1

Since the linear dynamical part $\widehat{\Sigma}$ and the nonlinear static part are both passive, it is well known that also their feedback interconnection is passive too. Moreover, by [20, Lemma 3], the system matrices of $\widehat{\Sigma}$ in (20) satisfy

$$P - A^\top P A = L^\top L \tag{22a}$$

$$A^\top P B = C^\top - L^\top U \tag{22b}$$

$$B^\top P B = D + D^\top - U^\top U, \tag{22c}$$

for some positive-definite matrix P and matrices L and U. This means that there exists a positive-definite Lyapunov function $V(s_k) = \frac{1}{2} s_k^\top P s_k$ with $s_k = [e_k^\top \ \tilde{v}_k^\top]^\top$ satisfying

$$V(s_{k+1}) - V(s_k) = \underbrace{\tilde{y}_k^\top u_k}_{\overset{(18)}{\leq}\, 0} - \tfrac{1}{2} \|L s_k + U u_k\|^2 . \tag{23}$$

Let $\Phi(s_k)$ denote the right-hand side of (23) (recall that both u_k and $\tilde{y}_k$ are functions of the state s_k). Clearly, it is negative semidefinite. To prove asymptotic stability, we shall show that $\Phi(\cdot)$ is also negative definite. To this end, we show that $\Phi(s_k) = 0$ implies $s_k = 0$.

If $\Phi(s_k) = 0$, then from (23) we can write

$$L s_k = -U u_k, \tag{24a}$$

$$\tilde{y}_k^\top u_k = 0. \tag{24b}$$

Let us focus on (24a) first. By definition of A and B in (20), one can easily check the following identity

$$\begin{bmatrix} I_p & 0_{p \times p(N-1)} \end{bmatrix} A^\top = \begin{bmatrix} I_p & 0_{p \times p(N-1)} \end{bmatrix} = \mathbf{1}^\top B^\top, \tag{25}$$

which can be used together with (22b) and (22c) to get

$$\begin{array}{ccc} \begin{bmatrix} I & 0 \end{bmatrix} A^\top P B & \overset{(22b)}{=} & \begin{bmatrix} I & 0 \end{bmatrix} \left(C^\top - L^\top U\right) \\ (25) \parallel & & \\ \mathbf{1}^\top B^\top P B & \overset{(22c)}{=} & \mathbf{1}^\top \left(D + D^\top - U^\top U\right). \end{array}$$

In light of the definition of C and D in (20), the latter chain of equalities translates into

$$\mathbf{1}^\top U^\top U = \mathbf{1}^\top + \begin{bmatrix} I & 0 \end{bmatrix} L^\top U. \tag{26}$$

Given the sparsity pattern imposed by (22a), we know that $L = [0_{pN \times p} \quad L_2]$, which implies $L s_k = L_2 \tilde{v}_k$ and $[I \quad 0] L^\top = 0_{p \times pN}$ which, in turn, can be used in (24a) and (26) to get

$$L_2 \tilde{v}_k = -U u_k, \tag{27a}$$

$$\mathbf{1}^\top U^\top U = \mathbf{1}^\top. \tag{27b}$$

From (22b) we also have that $L_2^\top U = -S^\top$, irrespectively of P. This fact can be used in (27a) to obtain

$$U^\top U u_k = S \tilde{v}_k \tag{28}$$

which, together with (27b), yields

$$\mathbf{1}^\top u_k = \mathbf{1}^\top U^\top U u_k = \mathbf{1}^\top S \tilde{v}_k = 0. \tag{29}$$

By definition of u_k in (14b), condition (29) implies

$$\mathbf{1}^\top u_k = c\, \mathbf{1}^\top (H_{\mathrm{d}} \boldsymbol{x}^+ - H_{\mathrm{d}} x_\star) = c\,(H \boldsymbol{x}^+ - b) = 0, \tag{30}$$

meaning that $\boldsymbol{x}^+$ satisfies the coupling constraints (1c).

We now add condition (24b) to show that $s_k = 0$. By [1, Lemma 4.1, p. 257], $\boldsymbol{x}^+$ satisfies

$$f(\boldsymbol{x}^+) + [\mathbf{1}\lambda_\star + \tilde{y}_k]^\top H_{\mathrm{d}} \boldsymbol{x}^+ \leq f(\boldsymbol{x}) + [\mathbf{1}\lambda_\star + \tilde{y}_k]^\top H_{\mathrm{d}} \boldsymbol{x}, \tag{31}$$

for all $\boldsymbol{x} \in X$. Setting $\boldsymbol{x} = \boldsymbol{x}_\star$ in (31) and using the definition of u_k, the previous relation can be rephrased as

$$f(\boldsymbol{x}^+) + \lambda_\star^\top \underbrace{\mathbf{1}^\top [H_{\mathrm{d}} \boldsymbol{x}^+ - H_{\mathrm{d}} \boldsymbol{x}_\star]}_{= H\boldsymbol{x}^+ - b} + \tfrac{1}{c} \tilde{y}_k^\top u_k \leq f(\boldsymbol{x}_\star).$$

Since $\tilde{y}_k^\top u_k = 0$ by (24b) and $H\boldsymbol{x}^+ - b = 0$ by (30), then $f(\boldsymbol{x}^+) \leq f(\boldsymbol{x}_\star)$, meaning that $\boldsymbol{x}^+$ is both feasible and achieves the optimal cost of problem (1). Hence, $\boldsymbol{x}^+ = \boldsymbol{x}_\star$ and $u_k = 0$. Left-multiplying (28) by $S^\top$ gives $0 = S^\top U^\top U u_k = S^\top S \tilde{v}_k = \tilde{v}_k$. Since $u_k = 0$ and $\tilde{v}_k = 0$, then $\tilde{y}_k = \mathbf{1} e_k = \mathbf{1}(\lambda_k - \lambda_\star)$, which can be used in (31) to obtain $f(\boldsymbol{x}^+) + \lambda_k^\top H \boldsymbol{x}^+ \leq f(\boldsymbol{x}) + \lambda_k^\top H \boldsymbol{x}$, for all $\boldsymbol{x} \in X$. Subtracting $\lambda_k^\top b$ on both sides of the previous inequality yields $f(\boldsymbol{x}^+) + \lambda_k^\top (H\boldsymbol{x}^+ - b) \leq f(\boldsymbol{x}) + \lambda_k^\top (H\boldsymbol{x} - b)$, for all $\boldsymbol{x} \in X$, meaning that $f(\boldsymbol{x}_\star) = f(\boldsymbol{x}^+) + \lambda_k^\top (H\boldsymbol{x}^+ - b) = \min_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \lambda_k) = q(\lambda_k)$, where the first equality is due to optimality of $\boldsymbol{x}^+$. Since, by strong duality (cf. Assumption 1.1), $q(\lambda_\star) = f(\boldsymbol{x}_\star)$, then λ_k achieves the optimal dual cost and is, therefore, the optimal dual solution. This means that we can set $\lambda_k = \lambda_\star$ and, thus, $e_k = 0$ and $s_k = 0$. Therefore $\Phi(\cdot)$ is negative definite and, thus, by the Lyapunov theorem, the equilibrium $s = 0$ is globally asymptotically stable for system (17), and, in light of their equivalence, also for system (14).

References

[1] D. P. Bertsekas, J. N. Tsitsiklis, Parallel and distributed computation: numerical methods, Prentice Hall, 1989.

[2] S. Boyd, N. Parikh, E. Chu, Distributed optimization and statistical learning via the alternating direction method of multipliers, Now Publishers Inc, 2011.

[3] J. Wang, N. Elia, A control perspective for centralized and distributed convex optimization, in: IEEE Conference on Decision and Control (CDC) and European Control Conference (ECC), 2011, pp. 3800–3805.

[4] M. Muehlebach, M. Jordan, A dynamical systems perspective on Nesterov acceleration, in: International Conference on Machine Learning (ICML), 2019, pp. 4656–4662.

[5] B. Hu, L. Lessard, Dissipativity theory for Nesterov's accelerated method, in: International Conference on Machine Learning (ICML), 2017, pp. 1549–1557.

[6] H. Mohammadi, M. Razaviyayn, M. R. Jovanović, Robustness of accelerated first-order algorithms for strongly convex optimization problems, IEEE Trans. on Automatic Control 66 (6) (2020) 2480–2495.

[7] S. Michalowsky, C. Scherer, C. Ebenbauer, Robust and structure exploiting optimization algorithms: an integral quadratic constraint approach, Intern. Journal of Control (2020) 1–24.

[8] D. Gramlich, C. Ebenbauer, C. W. Scherer, Convex synthesis of accelerated gradient algorithms for optimization and saddle point problems using Lyapunov functions, arXiv preprint arXiv:2006.09946 (2020).

[9] L. Lessard, B. Recht, A. Packard, Analysis and design of optimization algorithms via integral quadratic constraints, SIAM Journal on Optimization 26 (1) (2016) 57–95.

[10] M. Fazlyab, A. Ribeiro, M. Morari, V. M. Preciado, Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems, SIAM Journal on Optimization 28 (3) (2018) 2654–2689.

[11] M. Bin, I. Notarnicola, L. Marconi, G. Notarstefano, A system theoretical perspective to gradient-tracking algorithms for distributed quadratic optimization, in: IEEE Conference on Decision and Control (CDC), 2019, pp. 2994–2999.

[12] I. Pejcic, C. N. Jones, Accelerated ADMM based on accelerated Douglas-Rachford splitting, in: European Control Conference (ECC), 2016, pp. 1952–1957.

[13] P. Giselsson, S. Boyd, Linear convergence and metric selection for Douglas-Rachford splitting and ADMM, IEEE Trans. on Automatic Control 62 (2) (2016) 532–544.

[14] E. Ghadimi, A. Teixeira, I. Shames, M. Johansson, Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems, IEEE Trans. on Automatic Control 60 (3) (2014) 644–658.

[15] G. Franca, D. Robinson, R. Vidal, ADMM and accelerated ADMM as continuous dynamical systems, in: International Conference on Machine Learning (ICML), 2018, pp. 1559–1567.

[16] S. Hassan-Moghaddam, M. R. Jovanović, Proximal gradient flow and Douglas–Rachford splitting dynamics: global exponential stability via integral quadratic constraints, Automatica 123 (2021) 109311.

[17] T. Miyano, S. Yamashita, T. Hatanaka, K. Shibata, T. Jimbo, M. Fujita, Continuous-time optimization dynamics mirroring ADMM architecture and passivity-based robustification against delays, IEEE Trans. on Control of Network Systems 7 (3) (2020) 1296–1307.

[18] R. Nishihara, L. Lessard, B. Recht, A. Packard, M. Jordan, A general analysis of the convergence of ADMM, in: International Conference on Machine Learning (ICML), 2015, pp. 343–352.

[19] A. Falsone, I. Notarnicola, G. Notarstefano, M. Prandini, Tracking-ADMM for distributed constraint-coupled optimization, Automatica 117 (2020) 108962.

[20] L. Hitz, B. Anderson, Discrete positive-real functions and their application to system stability, in: Proceedings of the Institution of Electrical Engineers, Vol. 116, IET, 1969, pp. 153–155.