Input/Output Analysis of Primal-Dual Gradient Algorithms

John W. Simpson-Porco

***Abstract*— The primal-dual or saddle-point gradient algorithm has recently attracted interest as a systematic technique for solving distributed optimization problems. To examine the robustness of the algorithm, here we introduce exogenous disturbance inputs and quantify the performance of the algorithm in terms of the induced L_2-gain from the disturbance to deviations around the optimizer. For convex problems without inequality constraints, we find that the L_2-gain from a disturbance to the deviation of the primal state from the optimizer depends only on how strongly convex the agent objective functions are, and not on the equality constraints or on algorithm time constants. For primal-dual laws derived from an augmented Lagrangian, we show that the L_2-gain is a non-increasing function of the augmentation parameters, and therefore that augmentation may be beneficial for improving input/output performance.**

I. Introduction

A distributed optimization problem is one in which individual agents must cooperatively make decisions based on local information to minimize a global cost subject to global constraints. The defining aspects of the problem class are that (i) each agent possess a local cost function, this cost being known only to the respective agent, with the global cost being the sum of individual costs (ii) communication between agents is restricted, with allowable information flows being described by a graph, and (iii) no centralized coordination is permitted. As such, the approach is most well-motivated for applications where central coordination is undesirable or not possible; proposed applications include the fusion of measurements in sensor networks, resource allocation/congestion control in communication networks, and optimal distributed frequency control in power systems. The authors interest in distributed optimization stems from the last of these application areas; surveying the literature here is beyond our scope, but see for example [1]–[3].

A common continuous-time gradient-based algorithm for distributed convex optimization is the primal-dual or saddle-point algorithm, introduced in the early days of mathematical economics by Kose [4] and Arrow *et al.* [5]; see [6] for connections to Hamiltonian dynamics. The terminology saddle-point comes from the observation that the algorithm seeks the saddle points of the Lagrangian function of the problem. Various convergence results are available in [7]–[9].

A. Some Relevant Literature

In [10] Nedić and Ozdaglar posed a distributed optimization problem and studied a discrete-time algorithm for its solution. When translated to continuous-time for differentiable cost functions, the algorithm reads roughly as

$$\dot{x}_i = -k_i(t)\nabla_{x_i} f_i(x_i) - \sum\nolimits_{j=1}^{n} a_{ij}(t)(x_i - x_j) \qquad (1)$$

for each $i \in \{1, \ldots, n\}$, where $\{k_i(t)\}$ are specified time-varying gains and $\{a_{ij}(t)\}$ are the edge weights for a time-varying communication graph. The idea key idea is to simultaneously apply both gradient descent and consensus, adjusting the gain $K(t)$ as necessary to ensure convergence to the optimizer. Among many extensions, the framework has been further developed using push-sum protocols in [11]. Gharisifard and Cortés [12] studied the convergence of a modified primal-dual algorithm on strongly connected weight-balanced digraphs; extensions to time-varying strongly connected digraphs via push-sums were proposed in [13], with an alternative design proposed in [14]. The design of algorithms under more complex influence/information structures was studied by Kvaternik *et al.* in [15], [16]. Stegink *et al.* [17] take a port-Hamiltonian approach to study the stability of primal-dual algorithms, and apply their results to distributed frequency regulation in power networks.

Elia and Wang [8], [18] studied primal-dual algorithms from a control perspective, and observed that designs based on augmented Lagrangians can be interpreted as PI controllers. Moreover, these two-state augmented designs appear to possess superior noise rejection properties compared to the single-state design (1); see Figure 4 in [18]. Droge *et al.* [19] continue the discussion on control-engineering aspects of distributed optimization, suggesting that algorithm performance be quantified in terms of time-domain criteria such overshoot, settling time, and percent steady-state error from the optimizer. Algorithms subject to disturbances generated from a reference model have been studied in [20] via the internal model principle, with mean-square stability under stochastic disturbances studied in [21], [22].

B. Stability vs. Robustness

The above literature shares an emphasis on studying algorithm convergence, with the goal being to find the weakest convexity/connectivity conditions under which convergence to an optimizer is guaranteed. In most control problems however, a stable design is only the first step. It is well-understood that stable systems with good convergence rates can display poor transient behaviour when subjected to exogeneous disturbances [23, Section 3.3.3]. Rather than appealing to convergence rates alone, system norms are a useful tool for quantifying an algorithm's ability to tolerate

J. W. Simpson-Porco is with the Department of Electrical and Computer Engineering, University of Waterloo, ON, Canada. Email: jwsimpson@uwaterloo.ca.

and attenuate disturbances, with the L_2-norm being the most common metric for nonlinear systems.

Our goal here is to provide such a robustness analysis for the primal-dual algorithm. While it is generally interesting to understand how sensitive the algorithm is to disturbances, as a specific example consider the case of *online optimization*. In this scenario, we do not simply run an optimization algorithm *once*, but have it continuously run in real-time, taking in measurements and allowing the agents to cooperatively track the optimizer of the problem as problem parameters change. In other words, the optimization algorithm is now acting much like a controller, and like all controllers, becomes subject to measurement noise and unknown disturbances. We therefore arrive at the question of dynamic input/output performance or sensitivity: how much (or how little) does the primal-dual algorithm amplify exogenous noise and disturbances? In the power systems context, disturbances arise from noisy measurements or time-varying power injections. Understanding how well primal-dual frequency-control algorithms can attenuate these disturbances has immediate consequences for the viability of the approach in practice.

C. Contributions

Here we continue with the application of control-theoretic ideas to distributed optimization, much in the spirit of the recent works [8], [16]–[19], [24]. For problems without inequality constraints, we interpret the primal-dual algorithm as the interconnection of three distinct subsystems; a primal subsystem, a dual subsystem, and a (static) interconnection subsystem. By studying the input/output dissipativity properties of these three subsystems, we are able to bound the L_2-gain of a particular input/output map, providing one indication of how sensitive the algorithm is to external disturbances. We then study the more general situation where the primal-dual algorithm is derived from an augmented Lagrangian, and find that this may — or more interestingly, may not — lead to a decrease in the previously determined L_2-gain.

D. Notation

The set $\mathbb{R}$ (resp. $\mathbb{R}_{\geq 0}$) is the set of real (resp. nonnegative) numbers. The $n \times n$ identity matrix is I_n, $\mathbb{0}$ is a matrix of zeros of appropriate dimension, while $\mathbb{1}_n$ (resp. $\mathbb{0}_n$) are column n-vectors of all ones and all zeros, respectively. If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then $\nabla_x f : \mathbb{R}^n \to \mathbb{R}^n$ is its gradient; we will suppress the subscript x when no confusion can arise. A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if there exists a function $m : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ (the *modulus of convexity*) such that

$$(\nabla f(x_1)-\nabla f(x_2))^{\mathsf{T}}(x_1-x_2) \geq m(x_1,x_2)\|x_1-x_2\|_2^2\,, \quad (2)$$

for all $x_1, x_2 \in \mathbb{R}^n$. If $m(x,y) > 0$ for all $x, y \in \mathbb{R}^n$, then f is *strictly convex*, while if $m(x,y) \geq m > 0$ for all $x, y \in \mathbb{R}^n$, f is m*-strongly convex*.

II. Review of L_2-Stability and Passivity

While we assume some familiarity, we provide a brief refresher on these input-output system properties; see [25, Chapters 5 and 6] for details. We consider the square, continuous-time nonlinear control-affine system

$$\Sigma : \begin{cases} \dot{x} = F(x) + G(x)u \\ y = H(x) + J(x)u\,, \end{cases} \quad (3)$$

with state $x \in \mathbb{R}^n$, input $u \in \mathbb{R}^m$ and output $y \in \mathbb{R}^m$, and assume that F, G, H, J are of appropriate dimension and sufficiently smooth, with $F(0) = H(0) = 0$.

A. L_2-Stability of State-Space Control Systems

A signal $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$ is in $L_2[0,\infty)$ (or simply L_2) if its L_2-norm $\|u\|_{L_2}$ is finite, i.e., if

$$\|u\|_{L_2} = \left(\int_0^\infty u(t)^{\mathsf{T}} u(t)\,\mathrm{d}t\right)^{\frac{1}{2}} < \infty\,.$$

Signals in L_2 have finite energy, which is rather restrictive. To relax this, we say u is in the extended L_2-space L_{2e} if its truncation u_T to the interval $[0,T]$ is in L_2 for all finite $T \geq 0$. For example, step and ramp functions are in L_{2e} but not in L_2. The control system Σ in (3) is said to be *finite-gain L_2-stable with L_2-gain less than or equal to $\gamma \geq 0$* if for all initial conditions $x(0) \in \mathbb{R}^n$ there is a constant bias $b_{x(0)} \geq 0$ such that

$$\|y_T\|_{L_2} \leq \gamma\|u_T\|_{L_2} + b_{x(0)} \quad (4)$$

for all $T \geq 0$ and for all $u \in L_{2e}$. The L_2-gain $\|\Sigma\|_{L_2}$ of Σ is then defined as the smallest γ satisfying the above:

$$\|\Sigma\|_{L_2} \triangleq \inf\{\gamma \mid \exists\, b_{x(0)} \text{ such that (4) holds}\}\,. \quad (5)$$

In other words, the L_2-gain of a system quantifies how much input signals are amplified by the system, with both input and output measured in terms of the L_2 norm. In general the L_2-gain of a system is difficult to calculate; even verifying a given upper bound requires the solution of a PDE [26, Equation 3.49]. However, dissipation inequalities – and here in particular, passivity – provide a simpler verification tool.

B. Passivity of State-Space Control Systems

The system Σ in (3) is called *passive* if there exists a continuously differentiable *storage function* $V : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ satisfying $V(0) = 0$ and a scalar $\rho \geq 0$ such that along trajectories of (3) we have*

$$\dot{V} = \nabla_x V(x)^{\mathsf{T}}(F(x) + G(x)u) \leq -\rho y^{\mathsf{T}} y + y^{\mathsf{T}} u\,.$$

If $\rho > 0$, then Σ is called *output-strictly passive*. It can be shown that if Σ is output-strictly passive, then Σ is finite L_2-gain stable with L_2-gain less than or equal to $1/\rho$. Thus, establishing output strict passivity establishes an upper bound on the L_2-gain of the system.

*Since the $(x,u,y) = (0,0,0)$ is an equilibrium configuration of the system Σ, one could call this *passive around the origin*. Later, we will establish passivity of the primal-dual dynamics around its (typically non-zero) equilibrium configuration.

III. Primal-Dual Algorithm for Distributed Optimization

Consider the optimization problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) = \sum\nolimits_{i=1}^{n} f_i(x_i) \\ \text{subject to} \quad & Sx = b\,, \end{aligned} \tag{6}$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is the agent state vector, each $f_i : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and strictly convex, $b \in \mathbb{R}^r$, and $S \in \mathbb{R}^{r\times n}$ where $r < n$. We assume that S has full row rank ($\mathrm{rank}(S) = r$), which simply means that the constraints $Sx = b$ are not redundant. Relaxing the problem setup to include inequality constraints, vector local variables $x_i \in \mathbb{R}^{n_i}$, to the case of $r \geq n$, to the case where $\mathrm{rank}(S) < r$, and weakening the differentiability/convexity requirements on f are topics for future work. One should think of the matrix S as encoding sparse constraints between agents, and hence (6) has a naturally distributed structure. For example, if S^{T} is the incidence matrix of a graph [27, Chapter 8], then $Sx = \mathbb{0}$ simply says that $x_i = x_j$ for all $i, j \in \{1, \ldots, n\}$, implying that agents agree on their states at any optimizer.

Under the above assumptions the problem (6) is a convex optimization problem; it has a finite optimum, the equality constraints are strictly feasible, and (6) may be equivalently studied through its Lagrange dual with zero duality gap [28]. The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ of the problem (6) is

$$L(x,\nu) = f(x) + \nu^{\mathsf{T}}(Sx - b)\,, \tag{7}$$

where $\nu \in \mathbb{R}^r$ is a vector of Lagrange multipliers. By strong duality, the KKT conditions

$$\begin{aligned} \nabla_x L(x^*,\nu^*) = \mathbb{0}_n \quad &\iff \quad \mathbb{0}_n = \nabla f(x^*) + S^{\mathsf{T}}\nu^*\,, \\ \nabla_\nu L(x^*,\nu^*) = \mathbb{0}_r \quad &\iff \quad \mathbb{0}_r = Sx^* - b\,, \end{aligned} \tag{8}$$

are necessary and sufficient for optimality, and (8) determines the *unique* primal-dual optimizer (x^*, ν^*).

To calculate this optimizer, the primal-dual algorithm [7]–[9], [29], [30] says that we should perform gradient descent on the primal variables and gradient ascent on the dual variables

$$\tau_x\dot{x} = -\nabla_x L(x,\nu)\,, \qquad \tau_\nu\dot{\nu} = \nabla_\nu L(x,\nu)\,, \tag{9}$$

which for the Lagrangian (7) reduces to

$$\begin{aligned} \tau_x\dot{x} &= -\nabla f(x) - S^{\mathsf{T}}\nu \\ \tau_\nu\dot{\nu} &= Sx - b\,, \end{aligned} \tag{10}$$

where τ_x, τ_ν are positive definite diagonal matrices of time constants. By construction, the equilibrium points of (10) are in one-to-one correspondence with the solutions of the KKT conditions (8). One may show [7]–[9] that if each local cost f_i is strictly convex, then (10) converges from every initial condition to the global optimizer of (6).

IV. L_2-Gain Analysis of Primal-Dual Algorithm

We interpret the dynamics (10) as follows: the job of the primal subsystem with state x is to solve the unconstrained optimization problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) = \sum\nolimits_{i=1}^{n} f_i(x_i)\,. \tag{11}$$

Since f_i depends only on x_i, each agent locally implements a first-order gradient descent to reach the minimum of its cost function, and outputs the local state. The job of the dual subsystem is to act as a feedback controller for the error signal $Sx-b$, here an integral controller. The interconnection of these two subsystems then yields the dynamics (10), and the points where the two systems interconnect are a natural place to insert exogenous disturbances and study their effects. Figure 1 depicts the described interconnection.

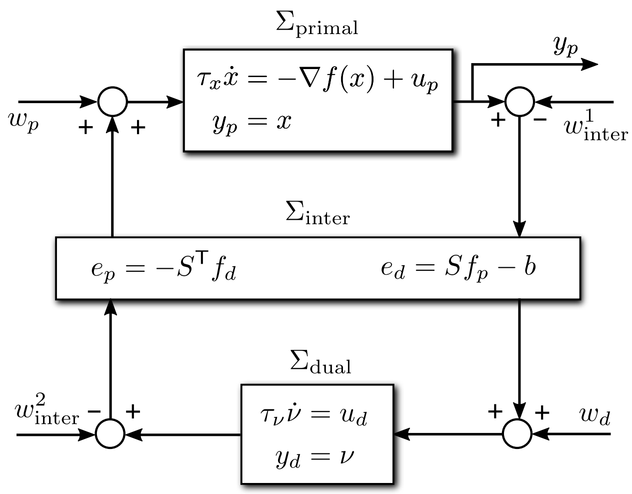

Fig. 1. Block-diagram of input/output primal-dual dynamics, with exogenous disturbances w.

To study input/output performance, we augment the dynamics (10) with primal/dual disturbance inputs $w = (w_p, w_d)$ and performance outputs $y = (y_p, y_d)$ as

$$\Sigma_{\text{pd}} : \begin{cases} \tau_x\dot{x} = -\nabla f(x) - S^{\mathsf{T}}\nu + w_p \\ \tau_\nu\dot{\nu} = Sx - b + w_d \\ \;y_p = x \\ \;y_d = \nu \end{cases} \tag{12}$$

Our first result shows that the system in Figure 1 is passive, and quantifies the L_2-gain from primal disturbances w_p to primal outputs y_p.

*Theorem 4.1 (***Primal-Dual I/O Performance***):* Consider the input-output primal-dual dynamics Σ_{pd} in (10), and let (x^*, ν^*) be the unique equilibrium point as determined by the KKT conditions (8). If for each agent $i \in \{1, \ldots, n\}$,

(i) $f_i(x_i)$ is strictly convex, then Σ_{pd} is passive around (x^*, ν^*);

(ii) $f_i(x_i)$ is m_i-strongly convex, then the map from w_p to y_p is finite L_2-gain stable around the equilibrium point (x^*, ν^*), with L_2-gain satisfying

$$\|\Sigma_{\rm pd}\|_{L_2} \leq \frac{1}{m_{\min}}\,,$$

where $m_{\min} = \min_{i\in\{1,\ldots,n\}} m_i$ is the minimum modulus of convexity among the agent cost functions.

Theorem 4.1(ii) shows that the ability of the system to attenuate primal disturbances depends only on how strongly convex the objective functions are, and it is the *least* strongly convex cost which determines the L_2-gain. Surprisingly, the bound does not depend in any way on the constraint matrix S or on the time-constants τ_x, τ_ν of the system. Moreover, in the limit where $m_{\min} \to 0$, our finite L_2-gain bound becomes arbitrarily large. This suggests that — while strict convexity is sufficient to guarantee asymptotic convergence — the system may be quite sensitive to external disturbances.

Proof: We begin by decomposing (10) into three subsystems:

$$\Sigma_{\rm primal}: \quad \begin{cases} \tau_x\dot{x} = -\nabla f(x) + u_p \\ y_p = x \end{cases}$$

$$\Sigma_{\rm dual}: \quad \begin{cases} \tau_\nu\dot{\nu} = u_d \\ y_d = \nu \end{cases} \qquad \Sigma_{\rm inter}: \quad \begin{cases} e_p = -S^{\mathsf T} f_d \\ e_d = Sf_p - b \end{cases}$$

subject to the feedback interconnection

$$\begin{aligned} u_p &= e_p + w_p\,, & f_p &= y_p - w_{\rm inter}^1\,,\\ u_d &= e_d + w_d\,, & f_d &= y_d - w_{\rm inter}^2\,, \end{aligned}$$

The subsystems and the interconnection structure is shown in Figure 1, where for the sake of generality we have included additional inputs to the interconnection $w_{\rm inter}^1$ and $w_{\rm inter}^2$. The equilibrium values associated with these variables are

$$\begin{aligned} &x^*, && u_p^* = e_p^* = -S^{\mathsf T}\nu^*, && y_p^* = f_p^* = x^*\\ &\nu^*, && u_d^* = e_d^* = \mathbb{0}_r, && y_d^* = f_d^* = \nu^*\,. \end{aligned}$$

Beginning with $\Sigma_{\rm primal}$, consider the storage function candidate $V_{\rm primal}(x) = \frac{1}{2}(x-x^*)^{\mathsf T}\tau_x(x-x^*)$. Along trajectories of $\Sigma_{\rm primal}$, we calculate that

$$\begin{aligned} \dot{V}_{\rm primal} &= (x-x^*)^{\mathsf T}(-\nabla f(x) + u_p)\\ &= -(\nabla f(x))^{\mathsf T}(x-x^*) + (x-x^*)^{\mathsf T}u_p \end{aligned}$$

Since from (8) it holds that $-\nabla f(x^*) + u_p^* = \mathbb{0}_n$, we may subtract $(x-x^*)^{\mathsf T}(-\nabla f(x^*)+u_p^*)$ from the right-hand side, obtaining

$$\dot{V}_{\rm primal} = -(x-x^*)^{\mathsf T}((\nabla f(x) - \nabla f(x^*)) \tag{13}$$
$$+ (x-x^*)^{\mathsf T}(u_p - u_p^*)\,. \tag{14}$$

For the dual system $\Sigma_{\rm dual}$ consider the storage function candidate $V_{\rm dual}(\nu) = \frac{1}{2}(\nu-\nu^*)^{\mathsf T}\tau_\nu(\nu-\nu^*)$. Along trajectories of $\Sigma_{\rm dual}$,

$$\dot{V}_{\rm dual} = (\nu-\nu^*)^{\mathsf T}u_d = (y_d - y_d^*)^{\mathsf T}(u_d - u_d^*)\,, \tag{15}$$

where we have used that $u_d^* = \mathbb{0}$. Finally, for the interconnection $\Sigma_{\rm inter}$, using the fact that $Sf_p^* - b = \mathbb{0}_r$, one may compute that

$$\begin{bmatrix} e_p - e_p^* \\ e_d - e_d^* \end{bmatrix} = \begin{bmatrix} \mathbb{0} & -S^{\mathsf T} \\ S & \mathbb{0} \end{bmatrix} \begin{bmatrix} f_p - f_p^* \\ f_d - f_d^* \end{bmatrix}.$$

Since the matrix above is skew-symmetric, it follows that

$$(e_p - e_p^*)^{\mathsf T}(f_p - f_p^*) + (e_d - e_d^*)^{\mathsf T}(f_d - f_d^*) = 0\,. \tag{16}$$

Setting $V(x,\nu) = V_{\rm primal}(x) + V_{\rm dual}(\nu)$, we find by adding the dissipation inequalities (13) and (15) and subtracting (16) that

$$\begin{aligned} \dot{V} &\leq -(y_p - y_p^*)^{\mathsf T}(\nabla f(y_p) - \nabla f(y_p^*))\\ &\quad + \begin{bmatrix} y_p - y_p^* \\ y_d - y_d^* \\ e_p - e_p^* \\ e_d - e_d^* \end{bmatrix}^{\mathsf T} \begin{bmatrix} w_p \\ w_d \\ w_{\rm inter}^1 \\ w_{\rm inter}^2 \end{bmatrix}. \end{aligned} \tag{17}$$

Setting $w_{\rm inter}^1 = w_{\rm inter}^2 = \mathbb{0}$, and using the fact that $f(x)$ is strictly convex, we immediately obtain statement (i) of the theorem. For statement (ii), Since each function $f_i(x_i)$ is m_i-strongly convex, $\nabla f(x)$ satisfies

$$(y_p - y_p^*)^{\mathsf T}(\nabla f(y_p) - \nabla f(y_p^*)) \geq (y_p - y_p^*)^{\mathsf T}M(y_p - y_p^*)\,,$$

for any $x, y \in \mathbb{R}^n$, where $M = \mathrm{diag}(m_1, \ldots, m_n)$. Inserting this bound into (17) and setting $w_d = w_{\rm inter}^1 = w_{\rm inter}^2 = \mathbb{0}$, we obtain

$$\dot{V} \leq -(y_p - y_p^*)^{\mathsf T}M(y_p - y_p^*) + (y_p - y_p^*)^{\mathsf T}w_p\,.$$

From this we see that the map from w_p to $y_p - y_p^*$ is output strictly passive with $\rho = \lambda_{\min}(M) = m_{\min}$. It follows that the L_2-gain from w_p to $y_p - y_p^*$ is less than or equal to $1/\rho = \frac{1}{m_{\min}}$, which shows the result. □

V. L_2-Gain Analysis of Augmented Lagrangian Primal-Dual Algorithms

Consider now the augmented optimization problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) + \frac{1}{2}(Sx-b)^{\mathsf T}K(Sx-b)\\ \text{subject to} \quad & Sx = b\,, \end{aligned} \tag{18}$$

where we have incorporated the squared constraint $Sx - b = \mathbb{0}_r$ into the cost function, with $K = \mathrm{diag}(k_1, \ldots, k_r)$ being a positive diagonal matrix of gains. Letting $L_k(x,\nu)$ denote the Lagrangian function of (18), it follows that (x,ν) is a saddle point of $L_k(x,\nu)$ if and only if it is a saddle point of $L(x,\nu)$, and hence the (unique) optimizer is unchanged. Taking gradients of $L_k(x,\nu)$, we obtain the *augmented primal-dual algorithm*

$$\tau_x\dot{x} = -\nabla f(x) - S^{\mathsf T}\nu - S^{\mathsf T}K(Sx-b) \tag{19a}$$
$$\tau_\nu\dot{\nu} = Sx - b\,. \tag{19b}$$

From a static optimization point of view, we have incorporated additional convexity into the cost function. As noted by [8], [18], this leads to a proportional control action in the primal-dual system. While it is well understood that such augmentation tends to improve the convergence rate of the algorithm, the following result shows that an improvement in I/O performance may (or may not!) also occur.

*Theorem 5.1 (***Aug. Primal-Dual I/O Performance):** Consider the input-output augmented primal-dual dynamics

$$\Sigma_{\mathrm{apd}}: \begin{cases} \tau_x \dot{x} = -\nabla f(x) - S^{\mathsf{T}}\nu - S^{\mathsf{T}}K(Sx-b) + w_p \\ \tau_\nu \dot{\nu} = Sx - b \\ y_p = x \end{cases}$$

with disturbance input w_p and performance output y_p, and let (x^*, ν^*) be the unique equilibrium point of (18) as determined by the KKT conditions (8). If for each agent $i \in \{1, \ldots, n\}$ the cost $f_i(x_i)$ is m_i-strongly convex, then the L_2-gain from w_p to y_p around the equilibrium point (x^*, ν^*) satisfies

$$\|\Sigma_{\mathrm{apd}}\|_{L_2} \leq \frac{1}{\lambda_{\min}(M + S^{\mathsf{T}}KS)}\,. \tag{20}$$

Proof: The decomposition of the algorithm into subsystems now reads as , as can be seen through the following decomposition:

$$\Sigma_{\mathrm{primal}}: \quad \begin{cases} \tau_x \dot{x} = -\nabla f(x) + u_p \\ y_p = x \end{cases}$$

$$\Sigma_{\mathrm{dual}}: \quad \begin{cases} \tau_\nu \dot{\nu} = u_d \\ y_d = \nu + K u_d \end{cases} \qquad \Sigma_{\mathrm{inter}}: \quad \begin{cases} e_p = -S^{\mathsf{T}} f_d \\ e_d = S f_p - b \end{cases}$$

subject to the feedback interconnection

$$u_p = e_p + w_p\,, \qquad f_p = y_p\,,$$
$$u_d = e_d\,, \qquad f_d = y_d\,.$$

Note the change in Σ_{dual}. Using tildes to denote a deviation

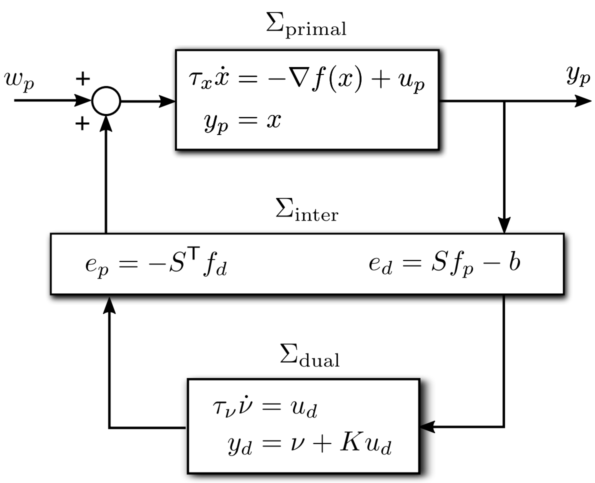

Fig. 2. Block-diagram of input/output augmented primal-dual dynamics for Theorem 5.1.

from the equilibrium value (e.g., $\tilde{y}_p = y_p - y_p^*$), we may proceed as before to arrive at the dissipation inequalities

$$\dot{V}_{\mathrm{primal}} \leq -\tilde{y}_p^{\mathsf{T}} M \tilde{y}_p + \tilde{y}_p^{\mathsf{T}} \tilde{u}_p\,, \tag{21a}$$
$$0 = -\tilde{e}_p^{\mathsf{T}} \tilde{f}_p - \tilde{e}_d^{\mathsf{T}} \tilde{f}_d\,, \tag{21b}$$

for the primal subsystem Σ_{primal} and the interconnection Σ_{inter}, respectively. For the dual system Σ_{dual}, we use the same storage function as before and compute that

$$\begin{aligned} \dot{V}_{\mathrm{dual}} &= (\nu - \nu^*)^{\mathsf{T}} u_d \\ &= (y_d - K u_d - y_d^*)^{\mathsf{T}} (u_d - u_d^*) \\ &= -(u_d - u_d^*)^{\mathsf{T}} K (u_d - u_d^*) + (y_d - y_d^*)^{\mathsf{T}} (u_d - u_d^*) \\ &= -\tilde{u}_d^{\mathsf{T}} K \tilde{u}_d + \tilde{y}_d^{\mathsf{T}} \tilde{u}_d\,. \end{aligned} \tag{22}$$

Using the interconnection equations and Σ_{inter} to eliminate everything other than $\tilde{y}_p$ and w_p, the dissipation inequality for $V = V_{\mathrm{primal}} + V_{\mathrm{dual}}$ reduces to

$$\begin{aligned} \dot{V} &\leq -\tilde{y}_p^{\mathsf{T}} M \tilde{y}_p - (S\tilde{y}_p)^{\mathsf{T}} K (S\tilde{y}_p) + \tilde{y}_p^{\mathsf{T}} w_p \\ &= -\tilde{y}_p^{\mathsf{T}} (M + S^{\mathsf{T}}KS) \tilde{y}_p^{\mathsf{T}} + \tilde{y}_p^{\mathsf{T}} w_p\,. \end{aligned}$$

Since $\mathrm{rank}(S) = r < n$ and $K > 0$, $S^{\mathsf{T}}KS \in \mathbb{R}^{n \times n}$ is positive semidefinite with rank r. Since $M > 0$, it follows that $M + S^{\mathsf{T}}KS$ is positive definite, and therefore

$$\dot{V} \leq -\lambda_{\min}(M + S^{\mathsf{T}}KS)\|\tilde{y}_p\|_2^2 + \tilde{y}_p^{\mathsf{T}} w_p\,,$$

with $\lambda_{\min}(M + S^{\mathsf{T}}KS) > 0$. The desired result now follows as in Theorem 4.1. □

By standard eigenvalue inequalities [31, Theorem 3.3.16], the L_2-gain bound of Theorem 5.1 is less than or equal to the bound obtained in Theorem 4.1 for every $K \geq 0$. Whether this new bound is *strictly* less than the previous bound depends on the particulars of K abd M, and on the the image S^{T}. For simplicity, assume that $K = kI_r$ for some $k > 0$, and without loss of generality assume that the minimum convexity parameter $m_{\min}$ is achieved by agent 1. W consider two examples. In the first example, augmentation lowers the L_2-gain, while in the second example it does not.

(i) **Agreement Constraints:** Consider the case where $r = n-1$ and $S = E^{\mathsf{T}}$, where E is the incidence matrix of an acyclic† graph. In this case, the constraint that $Sx = E^{\mathsf{T}}x = \mathbb{0}_{n-1}$ says that $x \in \mathrm{span}(\mathbb{1}_n)$, meaning that $x_i = x_j$ for all $i, j \in \{1, \ldots, n\}$. It follows that $S^{\mathsf{T}}KS = kEE^{\mathsf{T}} = k\mathsf{L}$, where $\mathsf{L} = EE^{\mathsf{T}}$ is an undirected unweighted Laplacian matrix for the graph described by E. A straightforward argument shows that

$$\lim_{k \to \infty} \lambda_{\min}(M + k\mathsf{L}) = m_{\mathrm{avg}} \triangleq \frac{1}{n}\sum\nolimits_{i=1}^{n} m_i\,.$$

In other words, by using high-gain Laplacian proportional control, the L_2-gain bound is improved from $\|\Sigma_{\mathrm{pd}}\|_{L_2} \leq m_{\min}^{-1}$ to $\|\Sigma_{\mathrm{apd}}\|_{L_2} \leq m_{\mathrm{avg}}^{-1}$ in the limit of large k. If all m_i are not equal, then this yields a strict decrease in the bound on the L_2-gain.

(ii) **Unconstrained Variable:** Consider the case where variable x_1 is not involved in any of the equality constraints $Sx = b$. In other words, the first column of S contains only zero elements, which implies that $S^{\mathsf{T}}KS$ has a zero eigenvalue with eigenvector $(1, 0, \ldots, 0)^{\mathsf{T}}$.

†For any connected graph it holds that $\mathrm{rank}(E) = n - 1$. The acyclic assumption ensures that $E \in \mathbb{R}^{n \times (n-1)}$, for consistency with our standing assumptions that $S \in \mathbb{R}^{r \times n}$ and $\mathrm{rank}(S) = r$.

Then $\lambda_{\min}(M + S^{\mathsf{T}}KS) = \lambda_{\min}(M) = m_1$ for all choices of diagonal gain matrices $K > 0$. Therefore, in this case augmentation is unable to reduce the L_2-gain. The problem here is that $(1, 0, \ldots, 0)^{\mathsf{T}}$ is not in the image of S^{T}. The proportional action therefore cannot influence the first state, which is the state limiting the L_2 gain of the system from being lowered.

VI. Conclusions

Here we have proposed a passivity and L_2-gain framework for studying the robustness of primal-dual gradient dynamics to unmodeled external disturbances. The L_2-gain from disturbances entering the primal dynamics to the primal variable seems to be limited by the *least* strongly convex agent cost function $f_i(x_i)$, which is rather intuitive. We found that using an augmented Lagrangian may or may not improve this performance depending on the particulars of the cost functions $\{f_i(x_i)\}_{i=1}^n$ and the constraint matrix S.

The decomposition of the dynamics in Figure 2 suggests the exploration of a dissipativity-based framework for optimizing and interconnecting optimization algorithms, in which input/output performance robustness becomes a focal point of design; the dual system Σ_{dual} arising from the augmented Lagrangian is simply one possibly dissipative (in this case, input-strictly passive) system. This broad framework will be pursued further in a future publication, and may be able to accommodate inequality constraints, directed communication, and delay, along with other types of distributed optimization algorithms. Future work will also attempt to connect these results to those in [24], and equilibrium-independent passivity or incremental passivity may allow for stronger conclusions to be drawn regarding system stability under interconnection.

References

[1] F. Dörfler, J. W. Simpson-Porco, and F. Bullo, "Breaking the hierarchy: Distributed control & economic optimality in microgrids," *IEEE Transactions on Control of Network Systems*, 2016, to appear.

[2] N. Li, L. Chen, C. Zhao, and S. H. Low, "Connecting automatic generation control and economic dispatch from an optimization view," in *American Control Conference*, Portland, OR, USA, Jun. 2014, pp. 735–740.

[3] A. Cherukuri and J. Cortés, "Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment," *Automatica*, 2016, to appear.

[4] T. Kose, "Solutions of saddle value problems by differential equations," *Econometrica*, vol. 24, no. 1, pp. 59–70, 1956.

[5] K. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*. Stanford University Press, 2006.

[6] A. M. Bloch, R. W. Brockett, and T. S. Ratui, "On the geometry of saddle point algorithms," in *IEEE Conf. on Decision and Control*, Tuscon, AZ, USA, 1992, pp. 1482–1487.

[7] D. Feijer and F. Paganini, "Stability of primal–dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.

[8] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *IEEE Conf. on Decision and Control and European Control Conference*, Orlando, FL, USA, Dec. 2011, pp. 3800–3805.

[9] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal–dual dynamics," *Systems & Control Letters*, vol. 87, pp. 10 – 15, 2016.

[10] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[11] A. Nedić and A. Olshevsky, "Distributed optimization of strongly convex functions on directed time-varying graphs," in *IEEE Global Conference on Signal and Information Processing*, Austin, TX, USA, Dec 2013, pp. 329–332.

[12] B. Gharesifard and J. Cortes, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.

[13] B. Touri and B. Gharesifard, "Continuous-time distributed convex optimization on time-varying directed networks," in *IEEE Conf. on Decision and Control*, Osaka, Japan, Dec 2015, pp. 724–729.

[14] S. S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, no. C, pp. 254–264, May 2015.

[15] K. Kvaternik, J. Llorca, D. Kilper, and L. Pavel, "A decentralized coordination strategy for networked multiagent systems," in *Allerton Conf. on Communications, Control and Computing*, Monticello, IL, USA, Oct 2012, pp. 41–47.

[16] K. Kvaternik and L. Pavel, "A continuous-time decentralized optimization scheme with positivity constraints," in *IEEE Conf. on Decision and Control*, Maui, HI, USA, Dec 2012, pp. 6801–6807.

[17] T. Stegink, C. D. Persis, and A. van der Schaft, "A unifying energy-based approach to stability of power grids with market dynamics," *IEEE Transactions on Automatic Control*, 2016, to appear.

[18] J. Wang and N. Elia, "Control approach to distributed optimization," in *Allerton Conf. on Communications, Control and Computing*, Monticello, IL, USA, 2010, pp. 557–561.

[19] G. Droge, H. Kawashima, and M. B. Egerstedt, "Continuous-time proportional-integral distributed optimisation for networked systems," *Journal of Control and Decision*, vol. 1, no. 3, pp. 191–213, 2014.

[20] X. Wang, P. Yi, and Y. Hong, "Dynamic optimization for multi-agent systems with external disturbances," *Control Theory and Technology*, vol. 12, no. 2, pp. 132–138, 2014.

[21] J. Wang and N. Elia, "Distributed least square with intermittent communications," in *American Control Conference*, Montreal, QC, Canada, June 2012, pp. 6479–6484.

[22] D. Mateos-Núñez and J. Cortés, "Noise-to-state exponentially stable distributed convex optimization on weight-balanced digraphs," *SIAM Journal on Control and Optimization*, vol. 54, no. 1, pp. 266–290, 2016.

[23] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control Analysis and Design*, 2nd ed. John Wiley & Sons, 2005.

[24] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.

[25] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2002.

[26] A. J. van der Schaft, *L2-Gain and Passivity Techniques in Nonlinear Control*, 2nd ed. Springer, 1999.

[27] F. Bullo, *Lectures on Network Systems*. Version 0.85, May 2016, with contributions by J. Cortés, F. Dörfler, and S. Martínez. [Online]. Available: http://motion.me.ucsb.edu/book-lns

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[29] M. Benzi, G. H. Golub, and J. Liesen, "Numerical solution of saddle point problems," *Acta Numerica*, vol. 14, pp. 1–137, 2005.

[30] A. Nedic and A. Ozdaglar, "Subgradient methods for saddle-point problems," *Journal of Optimization Theory and Applications*, vol. 142, no. 1, pp. 205–228, 2009.

[31] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1994.