NOISE-TO-STATE EXPONENTIALLY STABLE DISTRIBUTED CONVEX OPTIMIZATION ON WEIGHT-BALANCED DIGRAPHS*

DAVID MATEOS-NÚÑEZ AND JORGE CORTÉS†

Abstract. This paper studies the robustness properties against additive persistent noise of a class of distributed continuous-time algorithms for convex optimization. A team of agents, each with its own private objective function, seeks to collectively determine the global decision vector that minimizes the sum of all the objectives. The team communicates over a weight-balanced, strongly connected digraph and both inter-agent communication and agent computation are corrupted by noise. Under the proposed distributed algorithm, each agent updates its estimate of the global solution using the gradient of its local objective function while, at the same time, seeking to agree with its neighbors' estimates via proportional-integral feedback on their disagreement. Under mild conditions on the local objective functions, we show that this strategy is noise-to-state exponentially stable in second moment with respect to the optimal solution. Our technical approach combines notions and tools from graph theory, stochastic differential equations, Lyapunov stability analysis, and co-coercivity of vector fields. Simulations illustrate our results.

1. Introduction. Coordination problems that involve the collective optimization of a sum of convex functions, one per agent, find applications in a wide range of networked scenarios. These include, for example, distributed estimation in wireless sensor networks, motion coordination in multi-agent robotic systems, and large-scale optimization in machine learning. This paper is a contribution to the growing body of work that designs distributed coordination algorithms to allow agents to collectively determine a global solution of the optimization problem. In particular, we consider scenarios where the presence of noise in the agent-to-agent communications and in the agent computations induces errors in the algorithm execution. We study a family of distributed, continuous-time algorithms that have each agent update its estimate of the global optimizer doing gradient descent on its local cost function while, at the same time, seeking to agree with its neighbors' estimates via proportional-integral feedback on their disagreement. Our aim is to characterize the algorithm robustness properties against the additive persistent noise resulting from the errors in communication and computation.

Literature review. Our work here on distributed coordination for convex optimization under noise has connections mainly with two areas of research in the literature: distributed convex optimization and stability of stochastic differential equations. The multiple applications of the distributed optimization of a sum of convex functions has motivated the development of an increasing body of work. The model itself for what it means for a coordination strategy to be distributed leads to different architectures [3, 26]. While some emphasize the iterative selection of the component functions [2, 9, 21], here we focus on the multiagent approach, where the component functions are distributed among a group of agents that share information across a network. Some algorithms evolve in discrete time with associated gradient stepsize that is vanishing [6, 23, 25, 32], nonvanishing [18, 22, 23], or might require the solution

*A preliminary version of this manuscript appeared as [16] at the 2013 IEEE Conference on Decision and Control, Florence, Italy.

†The authors are with the Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA, {dmateosn,cortes}@ucsd.edu.

of a local optimization at each iteration [6, 29, 25, 19]; others evolve in continuous time [28, 7, 12] and even use separation of time scales [30]; and some are hybrid [27]. Most algorithms converge asymptotically to the solution, while others converge to an arbitrarily good approximation [18, 22]. Some examples of convergence rates, or size of the cost error as a function of the number of iterations, are $1/\sqrt{k}$ [6, 25] and $1/k$ [29]. The communication topologies might be undirected [18, 29, 12, 19, 28, 30], directed and weight-balanced or with a doubly stochastic adjacency matrix [6, 7, 32, 22, 23], or just directed under some knowledge about the number of in-neighbors and out-neighbors [25]; also, they can be fixed [7, 29, 12, 19, 30], or change over time under joint connectivity [6, 18, 32, 25, 22, 19, 23]. On the other hand, the objective functions might be required to be twice continuously differentiable [12, 30] or once differentiable [7, Sec. V], [19], or just Lipschitz [6], [7, Sec. IV], [18, 29, 32, 25, 22, 23]; in addition, they might need to be strongly convex [12], strictly convex [29, 19, 30], or just convex [6, 7, 18, 32, 25, 22, 23]. Some algorithms use the Hessian of the objective functions in addition to the gradients [12, 19, 30]. Also, the agents might need to share their gradients or second derivatives [12, 30] or even their objectives [19]. Some incorporate a global constraint known to all the agents using a projection method [6, 32, 25, 22] or a dual method [19], and in same cases each agent has a different constraint [32, 23]. Some algorithms impose a constraint on the initial condition [12, 19] in order to guarantee convergence. The algorithm execution can be synchronous [7, 29, 12], allow gossip/randomized communication [13, 23], or use event-triggered communication [27, 11]. Of particular interest to the subject matter of this paper are the works that consider noise affecting the dynamics through stochastically perturbed gradients with associated vanishing stepsize [6] or nonvanishing stepsize [22], while [23] considers both noisy communication links and subgradient errors. The characterization of the (discrete-time) algorithm performance under noise provided in these works builds on the fact that the projection onto a compact constraint set at every iteration effectively provides a uniform bound on the subgradients of the component functions.

The present work generalizes the class of continuous-time algorithms studied in [28] for undirected graphs and in [7] for weight-balanced digraphs by accounting for the presence of noise in the communication channels and in the agent computations. Under this strategy, each agent updates its estimate of the global solution using the gradient of its local objective function while, at the same time, performing proportional-integral distributed feedback on the disagreement among neighboring agents. As a result, the set of equilibria is given by the solution of the optimization problem together with an affine subspace of the integrator variables. The introduction of noise makes the resulting dynamical system a stochastic differential equation [15, 20, 10], with the particular feature that the stochastic perturbations do not decay with time and are present even at the equilibria of the underlying deterministic dynamics. The persistent nature of the noise rules out many classical stochastic notions of stability [24, 14, 15]. Instead, the concept of noise-to-state stability (NSS) [5] with respect to an equilibrium of the underlying ordinary differential equation is a notion of stochastic convergence to a neighborhood of that point. More precisely, it provides an ultimate bound for the state of the stochastic system, in probability, that depends on the magnitude of the covariance of the noise. Asymptotic convergence to the equilibrium follows in the absence of noise. Here, we build on our extension [17] of this concept to NSS in pth moment with respect to subspaces to establish NSS in second moment with respect to the subspace of equilibria of the underlying ordinary differential equation.

Statement of contributions. We consider a scenario where a group of agents communicating over a weight-balanced, strongly connected digraph seeks to collectively solve a convex optimization problem defined by a sum of local functions, one per agent. Both inter-agent communications and agent computations are corrupted by Gaussian white noise of locally bounded covariance. We study a family of distributed continuous-time coordination algorithms where each agent keeps track and interchanges with its neighbors two variables: one corresponding to its current estimate of the global optimizer and the other one being an auxiliary variable to guide agents towards agreement. According to this coordination strategy, each agent updates its estimate using gradient information of its local cost function while, at the same time, seeking to agree with its neighbors' estimates via proportional-integral feedback on their disagreement. The presence of noise both in the communication channels and the agent computations introduces errors in the algorithm execution that do not decay with time and are present even at the equilibria.

Our main contribution establishes that the resulting stochastic dynamics is noise-to-state exponentially stable in second moment and, therefore, robust against additive persistent noise. Our technical approach relies on the construction of a suitable candidate noise-to-state Lyapunov function whose nullspace is the affine subspace corresponding to the solution of the convex optimization problem and a direction of the auxiliary variables that absorbs the variance of the noise. To verify that the candidate function is in fact an NSS Lyapunov function, we analyze the interaction between local optimization and local consensus through the co-coercivity of a family of vector fields that are the sum of a gradient of a convex function plus a nonsymmetric Laplacian. Specifically, we give sufficient conditions for this family of vector fields to be co-coercive. In the absence of noise, our NSS-Lyapunov function is a strict Lyapunov function in the sense that it decreases along the trajectories outside of its nullspace. When noise is present, the expected rate of change of the NSS Lyapunov function is proportional to the difference between the square Frobenius norm of the covariance of the noise and the distance to its nullspace. The technical approach allows us to overcome the challenges posed by directed communication topologies and the presence of additive persistent noise. In addition, we also characterize the exponential rate of convergence of the coordination algorithm and the functional dependence on the size of the disturbance in the notion of noise-to-state exponential stability in second moment. Various simulations illustrate our results.

Organization. Section 2 introduces notational conventions and preliminary notions on graph theory and stochastic differential equations. Section 3 presents the network model and formulates the problem of interest. Section 4 introduces the family of distributed coordination algorithms along with the main convergence result and illustrative simulations. Section 5 presents the technical analysis of the algorithm properties. Finally, Section 6 gathers our conclusions and ideas for future work and the appendix contains an auxiliary result employed in our technical discussion.

2. Preliminaries.

Here we introduce some notations and review basic notions on graph theory and stochastic differential equations.

2.1. Notational conventions.

We let $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ denote the sets of real and nonnegative real numbers, respectively. For convenience, we use the shorthand notation $\mathbb{1}_n := [1, \dots, 1]^\top \in \mathbb{R}^n$, $0_n := [0, \dots, 0]^\top \in \mathbb{R}^n$, and denote by $\mathrm{I}_n \in \mathbb{R}^{n \times n}$

(or simply I) the identity matrix in $\mathbb{R}^n$ and by $e_i \in \mathbb{R}^n$, the ith column of I_n. We denote by $\|.\|_2$ the Euclidean norm for vectors or the induced two-norm for matrices, and by $|x|_{\mathcal{U}} := \inf\{\|x-u\|_2 : u \in \mathcal{U}\}$, the Euclidean distance from x to a set $\mathcal{U} \subseteq \mathbb{R}^n$. The linear subspace generated by a set $\{u_1, ..., u_m\} \subseteq \mathbb{R}^n$ of vectors is denoted by $\mathrm{span}\{u_1, ..., u_n\}$. Given a vector v whose entries are matrices, $\mathrm{diag}(v)$ is a block-diagonal matrix whose blocks are the entries of v. Given $B \in \mathbb{R}^{n\times m}$, its Frobenius norm is $|B|_{\mathcal{F}} := \sqrt{\mathrm{trace}\,(B^\top B)} = \sqrt{\mathrm{trace}\,(BB^\top)}$. We also define the seminorm on $\mathbb{R}^m$ associated to B by $\|x\|_B := \|Bx\|_2$ (note that we depart here from the usual convention of defining $\|x\|_A := \sqrt{x^\top Ax}$, which has the inconvenience of requiring A to be symmetric and positive semidefinite). The nullset of the seminorm corresponds to the nullspace of B, $\mathcal{N}(B) = \{x \in \mathbb{R}^m : Bx = 0\}$. For a symmetric real matrix $A \in \mathbb{R}^{n\times n}$, $\mathrm{spec}(A)$ denotes its set of eigenvalues, which we order as $\lambda_{\max}(A) := \lambda_1(A) \geq \cdots \geq \lambda_n(A) := \lambda_{\min}(A)$. For convenience, we also use the notation $\lambda^{\varnothing}_{\max}(A)$ to denote the maximum nonzero eigenvalue of A. Given a subspace $\mathcal{U}$, we let $\lambda^{\mathcal{U}^\perp}_{\max}(A) := \max_{\{x^\top u = 0\,:\, u\in\mathcal{U},\, \|x\|_2=1\}} x^\top Ax$. Given a matrix $B \in \mathbb{R}^{n\times m}$, its singular values are the square roots of the eigenvalues of $B^\top B$). We order them according to $\sigma_{\max}(B) := \sigma_1(B) \geq \cdots \geq \sigma_r(B) := \sigma_{\min}(A)$, where $r = \mathrm{rank}(B)$ is the rank of B. We let $A \otimes B$ denote the Kronecker product of matrices A and B. Recall that $\mathrm{spec}(A \otimes B) = \mathrm{spec}(A) \times \mathrm{spec}(B)$.

Given a convex set X, a function $f : X \to \mathbb{R}$ is convex if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ for each $x, y \in X$ and any $\lambda \in [0, 1]$. The function f is concave if $-f$ is convex. Given normed vector spaces X_1, X_2, a function $f : X_1 \to X_2$ is Lipschitz with constant κ if $\|f(x) - f(y)\|_{X_1} \leq \kappa\|x - y\|_{X_2}$ for each $x, y \in X_1$, where $\|.\|_X$ denotes the norm of X. If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, we denote its gradient and Hessian by ∇f and $\nabla^2 f$, respectively. Given a differentiable vector field $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$, we let $\mathbf{DF} : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ denote its Jacobian, where $\mathbf{DF}(x)_{ij} = \frac{\partial \mathbf{F}_i(x)}{\partial x_j}$ for all $x \in \mathbb{R}^n$. A continuous function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ belongs to class $\mathcal{K}$ if it is strictly increasing and $\alpha(0) = 0$, and it belongs to class $\mathcal{K}_\infty$ if $\alpha \in \mathcal{K}$ and it is unbounded. Also, a continuous function $\mu : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is class $\mathcal{KL}$ if, for each fixed $s \geq 0$, the function $r \mapsto \mu(r, s)$ is class $\mathcal{K}$, and for each fixed $r \geq 0$, the function $s \mapsto \mu(r, s)$ is decreasing and $\lim_{s\to\infty} \mu(r, s) = 0$.

2.2. Graph theory. The following notions in graph theory follow the exposition in [4]. A weighted digraph $\mathcal{G} = (\mathcal{I}, \mathcal{E}, \mathsf{A})$ is a triplet where $\mathcal{I} = \{1, \ldots, N\}$ is the vertex set, $\mathcal{E} \subseteq \{(i, j) \in \mathcal{I} \times \mathcal{I} : i \neq j\}$ is the edge set, and $\mathsf{A} \in \mathbb{R}^{N\times N}_{\geq 0}$ is the weighted adjacency matrix with the property that $\mathsf{a}_{ij} > 0$ if and only if $(i, j) \in \mathcal{E}$. The (out-)Laplacian matrix $\mathsf{L} \in \mathbb{R}^{N\times N}$ of $\mathcal{G}$ is $\mathsf{L} := \mathrm{diag}(\mathsf{A}\mathbb{1}_N) - \mathsf{A}$, which satisfies $\mathsf{L}\mathbb{1}_N = 0_N$. The complete graph is the digraph with edge set $\mathcal{E}_{\mathcal{K}} = \{(i, j) \in \mathcal{I} \times \mathcal{I} : i \neq j\}$. For convenience, we let $\mathsf{L}_{\mathcal{K}}$ denote the Laplacian of the complete graph with edge weights $1/N$. Note that $\mathsf{L}_{\mathcal{K}} = \mathrm{I}_N - \mathrm{M}$, where $\mathrm{M} := \frac{1}{N}\mathbb{1}_N\mathbb{1}_N^\top$, and that $\mathsf{L}_{\mathcal{K}}$ is idempotent, i.e., $\mathsf{L}^2_{\mathcal{K}} = \mathsf{L}_{\mathcal{K}}$. The weighted out-degree and in-degree of $i \in \mathcal{I}$ are, respectively, $d_{\mathrm{out}}(i) = \sum_{j=1}^N \mathsf{a}_{ij}$ and $d_{\mathrm{in}}(i) = \sum_{j=1}^N \mathsf{a}_{ji}$. A digraph is weight-balanced if $d_{\mathrm{out}}(i) = d_{\mathrm{in}}(i)$ for all $i \in \mathcal{I}$, that is, $\mathbb{1}_N^\top\mathsf{L} = 0_N$, which is also equivalent to the condition of $\mathsf{L} + \mathsf{L}^\top$ being positive semidefinite. A path is an ordered sequence of vertices such that any pair of vertices appearing consecutively is an edge. A digraph is strongly connected if there is a path between any pair of distinct vertices. If $\mathcal{G}$ is weight-balanced and strongly connected, then $\mathsf{L} + \mathsf{L}^\top$ is positive semidefinite and $\mathcal{N}(\mathsf{L} + \mathsf{L}^\top) = \mathrm{span}\{\mathbb{1}_N\}$.

2.3. Stochastic differential equations. A stochastic differential equation (SDE) [15, 20] is, roughly speaking, an ordinary differential equation driven by a "random process" called Brownian motion, $\mathrm{B} : \Omega \times [t_0, \infty) \to \mathbb{R}^m$. Here, Ω is the outcome space and $\mathbb{P}$ is a probability measure defined on the sigma-algebra $\mathcal{F}$ of measurable events (subsets) of Ω. These elements together form the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For each outcome $\omega \in \Omega$, the mapping $\mathrm{B}(\omega, .) : [t_0, \infty) \to \mathbb{R}^m$ is a sample path of the Brownian motion and is continuous with probability 1 and with $\mathrm{B}(., t_0) = 0$; and for each time $t \in [t_0, \infty)$, the function $\mathrm{B}(t) := \mathrm{B}(., t) : \Omega \to \mathbb{R}^m$ is a random variable such that the increments $\mathrm{B}(t) - \mathrm{B}(s)$ have a multivariate Gaussian distribution of zero mean and covariance $(t - s)\mathrm{I}_m$ and are independent from $\mathrm{B}(s)$ for all $t_0 \leq s < t$. Formally, we consider the SDE

$$\mathrm{d}x(\omega, t) = g(x(\omega, t), t)\mathrm{d}t + G(x(\omega, t), t)\Sigma(t)\mathrm{dB}(\omega, t)\,, \tag{2.1}$$

where $x(\omega, t_0) = x_0$ with probability 1 for some $x_0 \in \mathbb{R}^n$. The vector field $g : \mathbb{R}^n \times [t_0, \infty) \to \mathbb{R}^n$ is the drift, the matrix valued function $G : \mathbb{R}^m \times [t_0, \infty) \to \mathbb{R}^{n \times q}$ is the diffusion term that models the way in which the noise enters the dynamics, and $\Sigma : [t_0, \infty) \to \mathbb{R}^{q \times m}$ determines the size of the noise. The matrix $\Sigma(t)\Sigma(t)^\top$ is called the infinitesimal covariance. The following result, from [15, Th. 3.6, p. 58], guarantees the existence and uniqueness of solutions.

LEMMA 2.1. (Existence and uniqueness). *Let Σ be essentially locally bounded and, for any $T > t_0$ and $n \geq 1$, let $K_{T,n} \in \mathbb{R}_{>0}$ be such that, for almost every $t \in [t_0, T]$ and all $x, y \in \mathbb{R}^n$ with $\max\{\|x\|_2, \|y\|_2\} \leq n$, it holds that*

$$\max\{\, \|f(x,t) - f(y,t)\|_2^2\,,\, |G(x,t) - G(y,t)|_{\mathcal{F}}^2 \,\} \leq K_{T,n}\|x - y\|_2^2.$$

Furthermore, assume that for any $T > t_0$, there exists $K_T > 0$ such that, for almost every $t \in [t_0, T]$ and all $x \in \mathbb{R}^n$,

$$x^\top f(x,t) + \tfrac{1}{2}|G(x,t)|_{\mathcal{F}}^2 \leq K_T(1 + \|x\|_2^2).$$

Then, the SDE (2.1) enjoys global existence and uniqueness of solutions for each initial condition $x_0 \in \mathbb{R}^n$.

In particular, under the hypotheses of Lemma 2.1, the solution inherits some properties of the Brownian motion. For instance, $x : \Omega \times [t_0, \infty) \to \mathbb{R}^n$ has continuous sample paths $x(\omega, .) : [t_0, \infty) \to \mathbb{R}^n$ with probability 1, and for each $t \geq t_0$, $x(t) := x(., t) : \Omega \to \mathbb{R}^n$ is a random variable with certain distribution (so that we are able to measure the probabilities of certain events that involve them). Looking at (2.1), during a small time interval δ, the random outcome $x(\omega, t)$ changes approximately its value by an amount that is normally distributed with expectation $g(x(\omega, t))\delta$ and covariance $G(x(\omega, t), t)\Sigma(t)\Sigma(t)^\top G(x(\omega, t), t)^\top \delta$, and this change is independent of the previous history of the solution $\{x(s)\}_{s \leq t}$.

Next we introduce an important operator in the stability analysis of stochastic differential equations. For any twice continuously differentiable function $\mathrm{V} : \mathbb{R}^n \to \mathbb{R}$, we denote the generator of the SDE (2.1) acting on the function V as the mapping $\mathcal{L}[\mathrm{V}] : \mathbb{R}^n \times [t_0, \infty) \to \mathbb{R}$ given by

$$\mathcal{L}[\mathrm{V}](x, t) := \nabla \mathrm{V}(x)^\top g(x) + \tfrac{1}{2}\,\mathrm{trace}\left(\Sigma(t)^\top G(x,t)^\top \nabla^2 \mathrm{V}(x) G(x,t)\Sigma(t)\right). \tag{2.2}$$

The above quantity is the expected rate of change of the function V along the solutions of the SDE (2.1) that take the value x at time t. It can be thought of as a generalization of Lie derivative to SDEs. This operator plays a key role in the following result, which provides a useful tool to study the stability properties of SDEs. The formulation presented here is a distilled version of the discussion in [17, Section 3].

THEOREM 2.2. (Exponential pth moment noise-to-state stability). *Under the hypotheses of Lemma 2.1, let* $\mathrm{V} \in \mathcal{C}^2(\mathbb{R}^n;\mathbb{R}_{\geq 0})$ *satisfy the following properties with respect to a closed set* $\mathcal{U} \subseteq \mathbb{R}^n$*: there exist* $p > 0$ *and class* $\mathcal{K}_\infty$ *functions* α_1 *and* α_2*, where* α_1 *is convex, such that*

$$\alpha_1(|x|_{\mathcal{U}}^p) \leq \mathrm{V}(x) \leq \alpha_2(|x|_{\mathcal{U}}^p),$$

for all $x \in \mathbb{R}^n$*, and there exist* $\mathrm{W} \in \mathcal{C}(\mathbb{R}^n;\mathbb{R}_{\geq 0})$*,* $\sigma \in \mathcal{K}$*, and concave* $\eta \in \mathcal{K}_\infty$ *such that*

$$\mathcal{L}[\mathrm{V}](x,t) \leq -\mathrm{W}(x) + \sigma\big(|\Sigma(t)|_{\mathcal{F}}\big),$$

for all $(x,t) \in \mathbb{R}^n \times [t_0,\infty)$*, where, in addition,* $\mathrm{V}(x) \leq \eta(\mathrm{W}(x))$*, for all* $x \in \mathbb{R}^n$*. Then the system* (2.1) *is* pth moment noise-to-state stable ($p$thNSS) with respect to $\mathcal{U}$*, i.e., there exist* $\mu \in \mathcal{KL}$ *and* $\theta \in \mathcal{K}$ *such that*

$$\mathbb{E}\big[|x(t)|_{\mathcal{U}}^p\big] \leq \mu\big(|x_0|_{\mathcal{U}}, t-t_0\big) + \theta\big(\operatorname*{ess\,sup}_{t_0\leq s\leq t} |\Sigma(s)|_{\mathcal{F}}\big),$$

for all $t \geq t_0$ *and any* $x_0 \in \mathbb{R}^n$*. Specifically,* $\mu(r,s) := \alpha_1^{-1}\big(2\tilde{\mu}(\alpha_2(r^p),s)\big)$ *and* $\theta(r) := \alpha_1^{-1}\big(2\eta(2\sigma(r))\big)$*, where the class* $\mathcal{KL}$ *function* $(r,s) \mapsto \tilde{\mu}(r,s)$ *is well defined as the solution* $y(s)$ *to the initial value problem*

$$\dot{y}(s) = -\tfrac{1}{2}\eta^{-1}(y(s)), \quad y(0) = r.$$

We refer to the function V satisfying the hypotheses of this result as a *pth moment NSS-Lyapunov function with respect to* $\mathcal{U}$ for the system (2.1). If the functions α_1 and η are linear, then we refer to the above property as *pth moment noise-to-state exponential stability*.

3. Network model and problem statement. This section describes the model for the network of agents and the optimization problem we set out to solve in a distributed way. Consider a group of N agents with identities $\{1,\dots,N\}$ whose communication topology is modeled by a strongly connected and weight-balanced digraph $\mathcal{G}$. An edge $(i,j) \in \mathcal{E}$ represents the ability of agent i to receive information sent from agent j. We consider scenarios where the inter-agent communication is corrupted by Gaussian white noise. Specifically, if agent j sends the signal $x(t) \in \mathbb{R}^d$ to agent i at time $t \geq t_0$, agent i receives the corrupted signal

$$x(t) + \mathbf{J}^{ij}(t)\, W_{\mathrm{cmm}}^{(i,j)}(\omega,t), \tag{3.1}$$

where $W_{\mathrm{cmm}}^{(i,j)}(\omega,t) \in \mathbb{R}^d$ is a random variable representing Gaussian white noise, and $\mathbf{J}^{ij}(t) \in \mathbb{R}^{d\times d}$ is a weighting matrix. The noise we consider is additive, might be always present no matter what the value of the transmitted signal is, and is persistent because it might not decay with time. Our forthcoming algorithm design does not require that

agent $i \in \{1, \dots, N\}$ knows the weighting matrices $\mathbf{J}^{ij}$ for any $(i, j) \in \mathcal{E}$. Similarly, we also consider the possibility of the information available to any given agent being corrupted by noise when incorporated into its computations. Specifically, if agent i incorporates the quantity $q_i(t) \in \mathbb{R}^d$ into its computations at time instant $t \geq t_0$, it instead gets

$$q_i(t) + \mathbf{J}^{ii}(t)\, W^i_{\mathrm{cmp}}(\omega, t). \tag{3.2}$$

As before, $W^i_{\mathrm{cmp}}(\omega, t) \in \mathbb{R}^d$ is a random variable representing Gaussian white noise, and $\mathbf{J}^{ii}(t) \in \mathbb{R}^{d\times d}$ is a weighting matrix. Again, our algorithm design does not require that agent $i \in \{1, \dots, N\}$ knows the weighting matrix $\mathbf{J}^{ii}$.

With the model for the network in place, we next define the network objective. Consider a function $f : \mathbb{R}^d \to \mathbb{R}$ of the form

$$f(x) = \sum_{i=1}^{N} f_i(x), \tag{3.3}$$

where the local function $f_i : \mathbb{R}^d \to \mathbb{R}$ is only known to agent $i \in \{1, \dots, N\}$. We assume each f_i is convex and that at least one of them is strongly convex, so that the function f has a unique minimizer, which we denote by $x_{\min} \in \mathbb{R}^d$. Our goal is to design a distributed continuous-time coordination algorithm that helps the network collectively find the minimizer $x_{\min}$ in the presence of noise both in the communication channels and in the agent computations.

4. Robust distributed optimization. This section introduces a distributed coordination algorithm that allows the network of agents to solve the optimization problem as described in Section 3. Our study here generalizes the work in [7] to scenarios where the communication channels and the computations performed by the agents are subject to noise. In order to synthesize a strategy that allows the network to agree on the solution of the optimization problem, we have each agent $i \in \{1, \dots, N\}$ keep an estimate $x^i \in \mathbb{R}^d$ about the minimizer of the function f in (3.3). For convenience, we denote by $\boldsymbol{x} := [(x^1)^\top, \dots, (x^N)^\top]^\top \in (\mathbb{R}^d)^N$ the collection of estimates across the network and consider the function $\tilde{f} : (\mathbb{R}^d)^N \to \mathbb{R}$ defined by

$$\tilde{f}(\boldsymbol{x}) := \sum_{i=1}^{N} f_i(x^i). \tag{4.1}$$

In this computation, each agent can evaluate f_i at its own estimate x^i and the network objective function in (3.3) can be evaluated when agreement holds, $\tilde{f}(\mathbb{1} \otimes x) = f(x)$.

The continuous-time algorithm we consider is then given by the following system of stochastic differential equations,

$$\mathrm{d}\boldsymbol{x} = -(\nabla \tilde{f}(\boldsymbol{x}) + \tilde{\gamma}\, \mathbf{L}\boldsymbol{x} + \mathbf{L}\boldsymbol{z})\mathrm{d}t + G^1(\boldsymbol{x}, \boldsymbol{z}, t)\Sigma^1(t)\mathrm{dB}(t), \tag{4.2a}$$

$$\mathrm{d}\boldsymbol{z} = \mathbf{L}\boldsymbol{x}\mathrm{d}t + G^2(\boldsymbol{x}, \boldsymbol{z}, t)\Sigma^2(t)\mathrm{dB}(t), \tag{4.2b}$$

where we use the shorthand notation $\mathbf{L} := \mathsf{L} \otimes \mathrm{I}_d$ and L is the Laplacian of the digraph $\mathcal{G}$ modeling inter-agent communication. We assume that the matrix-valued functions $G^1, G^2 : \mathbb{R}^{2Nd} \times [t_0, \infty) \to \mathbb{R}^{Nd\times q}$ are uniformly bounded and uniformly

globally Lipschitz in the first two arguments, and measurable and essentially bounded in time. Also, we assume that the matrix-valued functions $\Sigma^1, \Sigma^2 : [t_0, \infty) \to \mathbb{R}^{q\times m}$, with $m \geq 1$, are measurable and essentially locally bounded, and that $\{\mathrm{B}(t)\}_{t\geq t_0}$ is an m-dimensional Brownian motion defined in the probability space.

We next provide some intuition behind the algorithm design in (4.2) and properly justify its distributed character. The deterministic part of the dynamics prescribes that each agent updates its estimate by following the gradient of its local cost function while, at the same time, seeking to agree with its neighbors' estimates. The latter is implemented through a second-order process that involves the auxiliary variables $\boldsymbol{z} := [(z^1)^\top, \dots, (z^N)^\top]^\top \in (\mathbb{R}^d)^N$ and employs proportional-integral feedback on the disagreement. When the graph $\mathcal{G}$ is undirected, one can in fact see [7] that the deterministic part corresponds exactly to the saddle-point dynamics associated with the augmented Lagrangian

$$L(\boldsymbol{x}, \boldsymbol{z}) = \tilde{f}(\boldsymbol{x}) + \tilde{\gamma}\, \boldsymbol{x}^\top \mathbf{L}\boldsymbol{x} + \boldsymbol{z}^\top \mathbf{L}\boldsymbol{x},$$

corresponding to the minimization of $\tilde{f}$ under the constraints $\mathbf{L}\boldsymbol{x} = 0$. The stochastic part of the dynamics (4.2) is motivated by the desire to capture the presence of noise affecting the execution of the coordination algorithm. In particular, Remark 4.1 below discusses how the noise model described in Section 3 affecting the communication channels and the agent computations is captured by the stochastic differential equation (4.2). Finally, the dynamics is distributed over the digraph $\mathcal{G}$ because each agent $i \in \{1, \dots, N\}$ can update its variables x^i and z^i using only the information sent from its neighbors and its knowledge of its local function f_i. This is not difficult to see from the observation that the gradient of $\tilde{f}$ takes the form $\nabla\tilde{f}(\boldsymbol{x}) = [\nabla\tilde{f}_1(x^1)^\top, \dots, \nabla\tilde{f}_N(x^N)^\top]^\top$ and that the agent i can compute the ith d-dimensional block $(\mathbf{L}\boldsymbol{x})^i \in \mathbb{R}^d$.

Remark 4.1. (Noise model for communication and computation is captured by the dynamics (4.2)). When communication along an edge $(i, j) \in \mathcal{E}$ occurs continuously over time, the model (3.1) gives rise to functions $\mathbf{J}^{ij} : [t_0, \infty) \to \mathbb{R}^{d\times d}$, which we assume measurable and essentially locally bounded, and $W^{(i,j)}_{\mathrm{cmm}} : \Omega \times [t_0, \infty) \to \mathbb{R}^d$, which essentially correspond to the derivative of Brownian motion. Similarly, when considering continuous-time dynamics, the computation model (3.2) gives rise to functions $\mathbf{J}^{ii} : [t_0, \infty) \to \mathbb{R}^{d\times d}$, which we also assume measurable and essentially locally bounded, and $W^i_{\mathrm{cmp}} : \Omega \times [t_0, \infty) \to \mathbb{R}^d$. Under this noise model, the implementation of the dynamics $\dot{\boldsymbol{x}} = -(\nabla\tilde{f}(\boldsymbol{x}) + \tilde{\gamma}\,\mathbf{L}\boldsymbol{x} + \mathbf{L}\boldsymbol{z})$ and $\dot{\boldsymbol{z}} = \mathbf{L}\boldsymbol{x}$ by the agent i actually results in the dynamics,

$$\begin{aligned}
\mathrm{d}x^i(t) = {} & \tilde{\gamma}\sum_{j=1}^N \mathsf{a}_{ij}\big(\big(x^j(t) - x^i(t)\big)\mathrm{d}t + \mathbf{J}^{ij}(t)\mathrm{dB}^{1,(i,j)}(t)\big) \\
& + \sum_{j=1}^N \mathsf{a}_{ij}\big(\big(z^j(t) - z^i(t)\big)\mathrm{d}t + \mathbf{J}^{ij}(t)\mathrm{dB}^{2,(i,j)}(t)\big) \\
& - \nabla f_i(x^i(t))\mathrm{d}t - \mathbf{J}^{ii}(t)\mathrm{dB}^{3,i}(t), \qquad \text{(4.3a)} \\
\mathrm{d}z^i(t) = {} & -\sum_{j=1}^N \mathsf{a}_{ij}\big(\big(x^j(t) - x^i(t)\big)\mathrm{d}t + \mathbf{J}^{ij}(t)\mathrm{dB}^{1,(i,j)}(t)\big), \qquad \text{(4.3b)}
\end{aligned}$$

where $\mathrm{B}^{1,(i,j)}$, $\mathrm{B}^{2,(i,j)}$ and $\mathrm{B}^{3,i}$ are independent d-dimensional Brownian motions for each edge $(i,j) \in \mathcal{E}$ and each agent $i \in \{1,\dots,N\}$, respectively. We next show how this dynamics is captured by (4.2). First, we set $G^1(\boldsymbol{x},\boldsymbol{z},t) = G^2(\boldsymbol{x},\boldsymbol{z},t) = \mathrm{I}_{Nd}$ for all $\boldsymbol{x}$, $\boldsymbol{z}$, t. Second, let $\mathbf{J}(t) \in \mathbb{R}^{Nd\times Nd}$ be the matrix whose (i,j)th d-dimensional block is $\mathsf{a}_{ij}\mathbf{J}^{ij}(t)$ and $\tilde{\mathbf{J}}(t) = \operatorname{diag}\big(\tilde{\mathbf{J}}_{11}(t),\dots,\tilde{\mathbf{J}}_{NN}(t)\big) \in \mathbb{R}^{Nd\times Nd}$. Define $\hat{\Sigma}^1(t) := \begin{bmatrix}\tilde{\gamma}\mathbf{J}(t) & \mathbf{J}(t) & -\tilde{\mathbf{J}}(t)\end{bmatrix} \in \mathbb{R}^{Nd\times 3Nd}$ and $\hat{\Sigma}^2(t) := \begin{bmatrix}-\mathbf{J}(t) & 0 & 0\end{bmatrix} \in \mathbb{R}^{Nd\times 3Nd}$, and set

$$\begin{bmatrix}\Sigma^1(t)\\ \Sigma^2(t)\end{bmatrix} := \begin{bmatrix}\big((e_1e_1^\top)\otimes \mathrm{I}_d\big)\hat{\Sigma}^1(t) & \cdots & \big((e_Ne_N^\top)\otimes \mathrm{I}_d\big)\hat{\Sigma}^1(t)\\ \big((e_1e_1^\top)\otimes \mathrm{I}_d\big)\hat{\Sigma}^2(t) & \cdots & \big((e_Ne_N^\top)\otimes \mathrm{I}_d\big)\hat{\Sigma}^2(t)\end{bmatrix} \in \mathbb{R}^{2Nd\times 3N^2d}.$$

Then, the dynamics (4.2) with this selection of functions G^1, G^2, Σ^1, and Σ^2 corresponds to (4.3). •

The main result of the paper is the characterization of the asymptotic stability properties of the stochastic differential equation (4.2) with respect to the solution of the optimization problem. In particular, the following result shows that the dynamics of $\boldsymbol{x}(t)$ is noise-to-state exponentially stable in second moment with respect to $\mathbb{1}\otimes x_{\min}$.

THEOREM 4.2. (Exponential noise-to-state stability of the dynamics (4.2)). *Assume the functions $\{f_i\}_{i=1}^N$ are convex and twice continuously differentiable with uniformly upper-bounded Hessians, i.e., there exists $R > 0$ such that $0 \preccurlyeq \nabla^2 f_i \preccurlyeq R\,\mathrm{I}_d$, for $i \in \{1,\dots,N\}$. Further assume that at least one of the functions is strongly convex, i.e., there exists $r > 0$ such that $r\,\mathrm{I}_d \preccurlyeq \nabla^2 f_{i_0}$ for some $i_0 \in \{1,\dots,N\}$. In addition, let the design parameter $\tilde{\gamma}$ be selected as follows. Given $\epsilon > 0$, let $K_1 := \lambda_{\min}\big(r\,e_{i_0}e_{i_0}^\top + \epsilon\,(\mathsf{L}+\mathsf{L}^\top)\big)$ and $K_2 := R+2\epsilon\,\sigma_{\max}(\mathsf{L})$, and, for any $\delta \in (0,\, K_1K_2^{-2})$, let $\beta_1^* \equiv \beta_1^*(\delta,\epsilon) := \sqrt{K_1^2K_2^{-2} - K_1\delta}$ and $\beta_2^* \equiv \beta_2^*(\delta)$ be such that $h(\beta,\delta) < 0$ for $\beta \in (0,\beta_2^*(\delta))$, where*

$$h(\beta,\delta) := \Big(-\tfrac{\beta^4+3\beta^2+2}{2\beta} + \sqrt{\big(\tfrac{\beta^4+3\beta^2+2}{2\beta}\big)^2 - 1}\,\Big)\lambda_2(\mathsf{L}+\mathsf{L}^\top) + \tfrac{\beta^2}{2\delta}\,. \tag{4.4}$$

Under the above selections, define $\tilde{\gamma}$ as

$$\tilde{\gamma}(\epsilon,\delta) := \tfrac{2+\beta^2}{\beta} + 2\epsilon, \qquad \beta \in (0, \min\{\beta_1^*(\delta,\epsilon), \beta_2^*(\delta)\}).$$

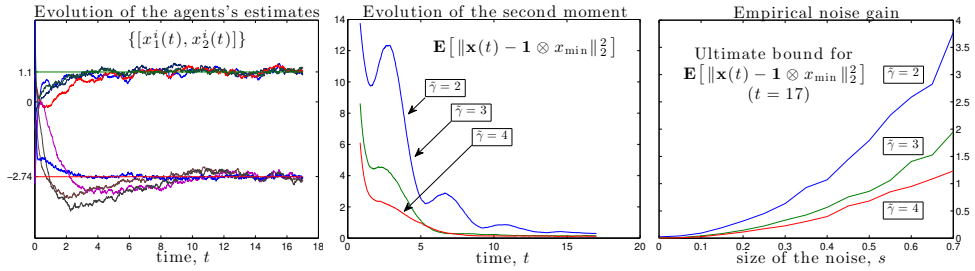
Then, the dynamics (4.2) executed over a strongly connected and weight-balanced digraph has the following stability property: there exist constants C_μ, D_μ, $C_\theta > 0$ such that, for any initial condition $(\boldsymbol{x}_0,\boldsymbol{z}_0) \in (\mathbb{R}^d)^N \times (\mathbb{R}^d)^N$ and all $t \ge t_0$, it holds that

$$\begin{aligned}\mathbb{E}\big[\|\boldsymbol{x}(t) - \mathbb{1}\otimes x_{\min}\|_2^2\big] &\le \mathbb{E}\big[\|\boldsymbol{x}(t) - \mathbb{1}\otimes x_{\min}\|_2^2 + \|\boldsymbol{z}(t) - \boldsymbol{z}^*\|_{\mathbf{L}_\mathcal{K}}^2\big]\\ &\le C_\mu(\|\boldsymbol{x}_0 - \mathbb{1}\otimes x_{\min}\|_2^2 + \|\boldsymbol{z}_0 - \boldsymbol{z}^*\|_{\mathbf{L}_\mathcal{K}}^2)\,e^{-D_\mu(t-t_0)} + C_\theta\big(\operatorname*{ess\,sup}_{t_0\le\tau\le t}|\Sigma(\tau)|_\mathcal{F}\big)^2,\end{aligned} \tag{4.5}$$

where $\mathbf{L}_\mathcal{K} := \mathsf{L}_\mathcal{K}\otimes \mathrm{I}_d$, $\Sigma(t) := [\Sigma^1(t)^\top, \Sigma^2(t)^\top]^\top$, $x_{\min} \in \mathbb{R}^d$ is the unique minimizer of (3.3), and $\boldsymbol{z}^* \in \mathbb{R}^d$ is any point satisfying $\mathbf{L}\boldsymbol{z}^* = -\nabla\tilde{f}(\mathbb{1}\otimes x_{\min})$.

Regarding Theorem 4.2, it is worth noting that the range of values of the design parameter $\tilde{\gamma}$ under which convergence is guaranteed depends on the network topology and has to be known by the agents. The expression (4.5) states that the dynamics (4.2) is noise-to-state stable in second moment with respect to the affine subspace of equilibria. In other words, the agreement direction of the agents' auxiliary states in $\boldsymbol{z}$ absorbs the cumulative variance of the noise while the estimates in $\boldsymbol{x}$ converge asymptotically,

in second moment, to a neighborhood of the minimizer of (3.3). The size of this neighborhood depends on the size of the noise, measured by $|\Sigma(t)|_{\mathcal{F}} := \sqrt{\operatorname{trace}(\Sigma(t)\Sigma(t)^\top)}$, which depends on the infinitesimal covariance $\Sigma(t)\Sigma(t)^\top$. Figure 4.1 illustrates empirically the evolution of the second moment using several realizations of the noise and also shows the values of the ultimate bounds as a function of the size of the noise.



(a) (b) (c)

Fig. 4.1: Evolution of the distributed optimization algorithm (4.2) with noise over a group of $N = 4$ agents communicating over the directed ring ($\mathcal{E} = \{(1,3),(3,2),(2,4),(4,1)\}$). The local objective functions are defined on $\mathbb{R}^2$ and given by $f_1(x_1,x_2) = \frac{1}{2}((x_1-4)^2 + (x_2-3)^2)$, $f_2(x_1,x_2) = x_1 + 3x_2 - 2$, $f_3(x_1,x_2) = \log(e^{x_1+3}+e^{x_2+1})$, and $f_4(x_1,x_2) = (x_1+2x_2+5)^2+(x_1-x_2-4)^2$. In all plots, the initial conditions are $\boldsymbol{x}_0 = (-3,-3,-1,-1,1,1,3,3)$, and $\boldsymbol{z}_0 = \mathbb{1}_8$. The dynamics is simulated using the Euler discretization with step size 0.01. Plot (a) shows the evolution of the first and second coordinates of the agents' estimates with $\tilde{\gamma} = 3$, $G^1 = G^2 = \mathrm{I}_8$, and $\Sigma^1 = \Sigma^2 = 0.2\,\mathrm{I}_8$. Despite the additive persistent noise, the estimates converge, in probability, to a neighborhood of the minimizer $x_{\min} = (1.10, -2.74)$. For three different values of the design parameter $\tilde{\gamma}$, plot (b) shows the asymptotic convergence in second moment to a neighborhood of the solution and plot (c) depicts the ultimate bound for the second moment when $\Sigma^1 = \Sigma^2 = s\,\mathrm{I}_8$ for increasing values of s. Observe that, as the design parameter gets larger (putting more emphasis on consensus among the agents) the noise gain gets smaller. Here, the expectations are computed averaging over 100 realizations of the noise.

We devote Section 5 to prove Theorem 4.2, where we provide explicit characterizations of the class $\mathcal{KL}$ function $\mu(r,s) := C_\mu r^2\, e^{-D_\mu\, s}$ and the class $\mathcal{K}_\infty$ function $\theta(r) := C_\theta\, r^2$. We end this section by noting that, in the noiseless case, a byproduct of Theorem 4.2 is a refinement of the result in [7], showing exponential convergence to the solution.

COROLLARY 4.3. (Global exponential stability in the noiseless case). *In the noiseless case (i.e., $\Sigma^1 = \Sigma^2 = 0$), and under the hypotheses of Theorem 4.2, the trajectory of the dynamics* (4.2) *starting from an arbitrary initial condition $(\boldsymbol{x}_0, \boldsymbol{z}_0) \in (\mathbb{R}^d)^N \times (\mathbb{R}^d)^N$ satisfies, for all $t \geq t_0$,*

$$\begin{aligned}\|\boldsymbol{x}(t) - \mathbb{1}\otimes x_{\min}\|_2^2 &\leq \|\boldsymbol{x}(t) - \mathbb{1}\otimes x_{\min}\|_2^2 + \|\boldsymbol{z}(t) - \boldsymbol{z}^*\|_2^2 \\ &\leq C_\mu(\|\boldsymbol{x}_0 - \mathbb{1}\otimes x_{\min}\|_2^2 + \|\boldsymbol{z}_0 - \boldsymbol{z}^*\|_{\mathbf{L}_\mathcal{K}}^2)e^{-D_\mu(t-t_0)} + \|\boldsymbol{z}_0 - \boldsymbol{z}^*\|_{\mathbf{M}}^2, \qquad (4.6)\end{aligned}$$

where $\mathbf{M} = \frac{1}{N}\mathbb{1}\mathbb{1}^\top \otimes \mathrm{I}_d$, and $\boldsymbol{z}^* \in \mathbb{R}^d$ is any point satisfying $\mathbf{L}\boldsymbol{z}^* = -\nabla\tilde{f}(\mathbb{1}\otimes x_{\min})$. In particular, choosing $\boldsymbol{z}^* \in \mathbb{R}^d$ such that $\mathbf{M}\boldsymbol{z}^* = \mathbf{M}\boldsymbol{z}_0$ shows that the convergence of

the trajectory starting from $(\boldsymbol{x}_0, \boldsymbol{z}_0)$ *to the point* $(\mathbb{1} \otimes x_{\min}, \boldsymbol{z}^*)$ *is exponential.*

Proof. Since $\Sigma^1 = \Sigma^2 = 0$, the system of SDEs (4.2) becomes a system of ordinary differential equations. Let $\boldsymbol{z}_{\text{agree}}(t) := \mathbf{M}\boldsymbol{z}(t)$. By left-multiplying the dynamics of $\boldsymbol{z}(t)$ in (4.2) by $\mathbf{M}$, we obtain that $\dot{\boldsymbol{z}}_{\text{agree}} = 0$ and therefore $\boldsymbol{z}_{\text{agree}}(t) = \boldsymbol{z}_{\text{agree}}(t_0)$ for all $t \geq t_0$. Using that $\mathbf{M}$ is symmetric and $\mathbf{M} = \mathbf{M}^2$, if we define $\boldsymbol{z}^*_{\text{agree}} := \mathbf{M}\,\boldsymbol{z}^*$, then

$$\begin{aligned}(\boldsymbol{z}(t) - \boldsymbol{z}^*)^\top \mathbf{M} (\boldsymbol{z}(t) - \boldsymbol{z}^*) &= (\boldsymbol{z}_{\text{agree}}(t) - \boldsymbol{z}^*_{\text{agree}})^\top \mathbf{M} (\boldsymbol{z}_{\text{agree}}(t) - \boldsymbol{z}^*_{\text{agree}}) \\ &= (\boldsymbol{z}_{\text{agree}}(t_0) - \boldsymbol{z}^*_{\text{agree}})^\top \mathbf{M} \, (\boldsymbol{z}_{\text{agree}}(t_0) - \boldsymbol{z}^*_{\text{agree}}) \\ &= (\boldsymbol{z}_0 - \boldsymbol{z}^*)^\top \mathbf{M} \, (\boldsymbol{z}_0 - \boldsymbol{z}^*) = \|\boldsymbol{z}_0 - \boldsymbol{z}^*\|^2_{\mathbf{M}}.\end{aligned}$$

On the other hand, using that $\mathrm{I}_{Nd} = \mathbf{L}_{\mathcal{K}} + \mathbf{M}$ and $\mathbf{L}^2_{\mathcal{K}} = \mathbf{L}_{\mathcal{K}}$, we obtain

$$\|\boldsymbol{z}(t) - \boldsymbol{z}^*\|_2^2 = (\boldsymbol{z}(t) - \boldsymbol{z}^*)^\top \big(\mathbf{L}_{\mathcal{K}} + \mathbf{M}\big)(\boldsymbol{z}(t) - \boldsymbol{z}^*) = \|\boldsymbol{z}(t) - \boldsymbol{z}^*\|^2_{\mathbf{L}_{\mathcal{K}}} + \|\boldsymbol{z}_0 - \boldsymbol{z}^*\|^2_{\mathbf{M}}.$$

Equation (4.6) follows from this fact together with (4.5). Finally, noting that $\mathbf{M}\boldsymbol{z} = \mathbb{1} \otimes (\frac{1}{N}\sum_{i=1}^N z^i)$ and $\mathbf{L}(\mathbb{1} \otimes a) = 0$ for any $a \in \mathbb{R}^d$, it is clear that, given an initial condition $\boldsymbol{z}_0 \in (\mathbb{R}^d)^N$, one can choose $\boldsymbol{z}^*$ that satisfies at the same time $\mathbf{L}\boldsymbol{z}^* = -\nabla \tilde{f}(\mathbb{1} \otimes x_{\min})$ and $\mathbf{M}\boldsymbol{z}^* = \mathbf{M}\boldsymbol{z}_0$. If this is the case, $\|\boldsymbol{z}_0 - \boldsymbol{z}^*\|_{\mathbf{M}} = 0$, and (4.6) shows exponential convergence of the trajectory starting from $(\boldsymbol{x}_0, \boldsymbol{z}_0)$ to $(\mathbb{1} \otimes x_{\min}, \boldsymbol{z}^*)$. ∎

5. Algorithm properties and stability analysis. In this section, we establish a series of properties of the distributed coordination algorithm (4.2) leading up to the characterization of its asymptotic correctness stated in Theorem 4.2. We begin by expressing the dynamics in compact form. Let $v := (\boldsymbol{x}^\top, \boldsymbol{z}^\top)^\top \in \mathbb{R}^{2Nd}$ and consider

$$\begin{aligned}&\mathrm{d}v = (\mathbf{A}v + \mathbf{N}(\boldsymbol{x}))\mathrm{d}t + G(v,t)\Sigma(t)\mathrm{dB} \\ &:= \left(\begin{bmatrix} -\gamma\mathbf{L} & -\mathbf{L} \\ \mathbf{L} & 0 \end{bmatrix} v + \begin{bmatrix} -\nabla \tilde{f}(\boldsymbol{x}) - 2\epsilon \mathbf{L} x \\ 0 \end{bmatrix}\right)\mathrm{d}t + \begin{bmatrix} G^1(\boldsymbol{x}, \boldsymbol{z}, t) & 0 \\ 0 & G^2(\boldsymbol{x}, \boldsymbol{z}, t) \end{bmatrix} \Sigma(t)\mathrm{dB},\end{aligned} \tag{5.1}$$

where, for convenience, we have split the parameter $\tilde{\gamma}$ as $\tilde{\gamma} = \gamma + 2\epsilon$. This dynamics fits the model (2.1) with $g(v) := \mathbf{A}v + \mathbf{N}(\boldsymbol{x})$. As mentioned earlier, $\Sigma : [t_0, \infty) \to \mathbb{R}^{q \times m}$ is measurable and essentially locally bounded, and $G : \mathbb{R}^{2Nd} \times [t_0, \infty) \to \mathbb{R}^{2Nd \times q}$ is measurable in time, uniformly globally Lipschitz in the first argument, say with Lipschitz constant $\kappa_1 \in \mathbb{R}_{>0}$, and bounded in its domain (essentially in time) by $\kappa_2 \in \mathbb{R}_{>0}$. Formally,

$$|G(v,t) - G(v',t)|_{\mathcal{F}} \leq \kappa_1 \|v - v'\|_2, \qquad \sup_{v \in \mathbb{R}^{2Nd}} \operatorname*{ess\,sup}_{t \geq t_0} |G(v,t)|_{\mathcal{F}} \leq \kappa_2, \tag{5.2}$$

for all $v, v' \in \mathbb{R}^{2Nd}$.

5.1. Equilibrium points. In this section we show, for completeness, the correspondence between the equilibrium points of (4.2) in the absence of noise and the solutions of the optimization problem stated in Section 3.

LEMMA 5.1. (Equilibrium points and Karush-Kuhn-Tucker conditions). *Let* $\mathcal{G}$ *be weight-balanced and strongly connected. Then, there exists* $\boldsymbol{x}^*$ *such that* $[\boldsymbol{x}^{*\top}, \boldsymbol{z}^{*\top}]^\top$ *satisfies the equilibrium conditions for the dynamics* (4.2) *without noise,*

$$\nabla \tilde{f}(\boldsymbol{x}^*) + \mathbf{L}\boldsymbol{z}^* = 0_{Nd} \quad \textit{and} \quad \mathbf{L}\boldsymbol{x}^* = 0_{Nd}, \tag{5.3}$$

for some $\boldsymbol{z}^* \in (\mathbb{R}^d)^N$*, if and only if there exists* $\boldsymbol{x}_{KKT}$ *such that* $[\boldsymbol{x}_{KKT}^\top, \boldsymbol{z}_{KKT}^\top]^\top$ *satisfies the Karush-Kuhn-Tucker conditions for the minimization of* $\tilde{f}$ *in* (4.1) *subject to* $\mathbf{L}\boldsymbol{x} = 0$*,*

$$\nabla \tilde{f}(\boldsymbol{x}_{KKT}) + \mathbf{L}^\top \boldsymbol{z}_{KKT} = 0_{Nd} \quad \textit{and} \quad \mathbf{L}\boldsymbol{x}_{KKT} = 0_{Nd}, \tag{5.4}$$

for some $\boldsymbol{z}_{KKT} \in (\mathbb{R}^d)^N$*. Moreover, both* (5.3) *and* (5.4) *are equivalent to*

$$(\mathbb{1}^\top \otimes \mathrm{I}_d)\nabla \tilde{f}(\boldsymbol{x}) = 0_{Nd} \quad \textit{and} \quad \mathbf{L}\boldsymbol{x} = 0_{Nd}, \tag{5.5}$$

and, if either $\boldsymbol{x}^*$ *or* $\boldsymbol{x}_{KKT}$ *exists and is unique, then so is the other one and* $\boldsymbol{x}^* = \boldsymbol{x}_{KKT}$*.*

Proof. Since $\mathcal{G}$ is weight-balanced and strongly connected, then $\mathcal{N}(\mathsf{L}+\mathsf{L}^\top) = \mathrm{span}\{\mathbb{1}\}$. The first equation in (5.5) follows by left-multiplying the first equation in (5.3) and (5.4) by $(\mathbb{1}^\top \otimes \mathrm{I}_d)$ and using that $\mathbb{1}^\top \mathsf{L} = 0$ because $\mathcal{G}$ is weight-balanced. The reason why (5.5) is equivalent to both (5.3) and (5.4) is the following: if there exists any $\boldsymbol{x}$ such that $(\mathbb{1}^\top \otimes \mathrm{I}_d)\nabla \tilde{f}(\boldsymbol{x}) = 0_d$, then $\nabla \tilde{f}(\boldsymbol{x})$ is in the column space of both $\mathbf{L}$ and $\mathbf{L}^\top$, which means that there exist $\boldsymbol{z}^*$ and $\boldsymbol{z}_{\mathrm{KKT}}$, respectively, that satisfy (5.3) and (5.4). This is because $\mathbf{L}(\mathbb{1} \otimes \mathrm{I}_d) = \mathbf{L}^\top(\mathbb{1} \otimes \mathrm{I}_d) = 0_{Nd\times d}$, and $\mathrm{rank}(\mathbf{L}) + \mathrm{rank}(\mathbb{1} \otimes \mathrm{I}_d) = \mathrm{rank}(\mathbf{L}^\top) + \mathrm{rank}(\mathbb{1} \otimes \mathrm{I}_d) = (N-1)d + d = Nd$. The result now follows by observing that $\boldsymbol{x}^*$ and $\boldsymbol{x}_{\mathrm{KKT}}$ are both defined by (5.5). ∎

As a consequence of this result and since there exists a unique minimizer $x_{\min}$ of (3.3), we deduce that the equilibrium points of the dynamics (4.2) in the absence of noise are $\boldsymbol{x}^* = \mathbb{1} \otimes x_{\min} \in (\mathbb{R}^d)^N$ and any $\boldsymbol{z}^* \in (\mathbb{R}^d)^n$ with $\mathbf{L}\boldsymbol{z}^* = -\nabla \tilde{f}(\mathbb{1} \otimes x_{\min})$.

5.2. Co-coercivity properties of the dynamics. In this section, we study the co-coercivity properties of the vector field $\mathbf{N}$ in the dynamics (5.1). Our results here play a key role later in establishing the global existence and uniqueness of the solutions and the noise-to-state stability properties of the dynamics. We first provide a general discussion on co-coercivity and then focus our attention on the properties of the dynamics (5.1). Given $S \in \mathbb{R}^{m\times m}$ and $\delta > 0$, we refer to a vector field $F : \mathbb{R}^m \to \mathbb{R}^m$ as $(S, \delta) - $ *co-coercive* with respect to $\bar{\boldsymbol{x}} \in \mathbb{R}^m$ if,

$$(\boldsymbol{x} - \bar{\boldsymbol{x}})^\top S(F(\boldsymbol{x}) - F(\bar{\boldsymbol{x}})) \geq \delta \, \|F(\boldsymbol{x}) - F(\bar{\boldsymbol{x}})\|_2^2, \tag{5.6}$$

for all $\boldsymbol{x} \in \mathbb{R}^m$. This corresponds to the notion of co-coercivity of $S^\top F$ as defined in [31] but here we define it for a vector field that is not necessarily the gradient of a scalar function. The following result provides sufficient conditions for a family of vector fields to be co-coercive under transformations that are small perturbations of the identity.

THEOREM 5.2. (Sufficient conditions for $(\mathrm{I}+\beta^2\tilde{S}, \delta)-$co-coercivity). *Let* $\mathbf{G} : (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N$ *be a continuously differentiable vector field such that* $\mathbf{DG}(\boldsymbol{x}) \in \mathbb{R}^{Nd\times Nd}$ *is symmetric positive semidefinite for all* $\boldsymbol{x} \in (\mathbb{R}^d)^N$*. Also, let* $\mathbf{T} : (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N$ *be the linear vector field* $\mathbf{T}(\boldsymbol{x}) = 2(\mathsf{L}\otimes \mathrm{I}_d)\boldsymbol{x}$*, where* L *is the Laplacian matrix of a strongly connected and weight-balanced digraph. Assume that there exist* $i_0 \in \{1, \dots, N\}$ *and* $r, R > 0$ *such that* $r\,(e_{i_0}e_{i_0}^\top) \otimes \mathrm{I}_d \preccurlyeq \mathbf{DG}(\boldsymbol{x}) \preccurlyeq R\,\mathrm{I}_{Nd}$ *for all* $\boldsymbol{x} \in (\mathbb{R}^d)^N$*. Given* $\epsilon > 0$*, let* $K_1 := \lambda_{\min}\left(r\, e_{i_0}e_{i_0}^\top + \epsilon\,(\mathsf{L}+\mathsf{L}^\top)\right)$*,* $K_2 := R + 2\epsilon\,\sigma_{\max}(\mathsf{L})$*, and* $\mathbf{F} := \mathbf{G} + \epsilon\mathbf{T}$*. Then,*

(i) $K_1 > 0$ *and* $2K_1\,\mathrm{I}_{Nd} \preccurlyeq \mathbf{DF}(\boldsymbol{x}) + (\mathbf{DF}(\boldsymbol{x}))^\top$ *for any* $\boldsymbol{x} \in (\mathbb{R}^d)^N$*.*
(ii) $K_1\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2 \leq \|\mathbf{F}(\boldsymbol{x}) - \mathbf{F}(\bar{\boldsymbol{x}})\|_2 \leq K_2\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2$ *for any* $\boldsymbol{x}, \bar{\boldsymbol{x}} \in (\mathbb{R}^d)^N$*.*

(iii) $\mathbf{F}$ *is* $(\mathrm{I}+\beta^2\tilde{S},\delta)-$*co-coercive with respect to every* $\bar{\boldsymbol{x}}\in(\mathbb{R}^d)^N$ *for any nonzero matrix* $\tilde{S}\in\mathbb{R}^{Nd\times Nd}$ *if* $\delta\in[0,K_1K_2^{-2})$ *and*

$$\beta\in\big[\,0,\ \sqrt{\big(K_1K_2^{-2}-\delta\big)/(\|\tilde{S}\|_2\,K_1^{-1})}\ \big].$$

Proof. Regarding *(i)*, we first show that $\lambda_{\min}\big(r\,e_{i_0}e_{i_0}^\top+\epsilon\,(\mathsf{L}+\mathsf{L}^\top)\big)>0$. For this, note that the matrices $r\,e_{i_0}e_{i_0}^\top$ and $\epsilon(\mathsf{L}+\mathsf{L}^\top)$ are positive semidefinite. In addition, their sum has rank N as we show next. Arguing by contradiction, assume that $y\in\mathbb{R}^N\setminus\{0\}$ is in its nullspace, i.e., $\big(r\,e_{i_0}e_{i_0}^\top+\epsilon(\mathsf{L}+\mathsf{L}^\top)\big)y=0$. Pre-multiplying by $y^\top$, it follows then that $0\le\epsilon y^\top(\mathsf{L}+\mathsf{L}^\top)y^\top=-r\,(y_{i_0})^2\le 0$, which implies that $y_{i_0}=0$ and $y^\top(\mathsf{L}+\mathsf{L}^\top)y^\top=0$. As $\mathsf{L}+\mathsf{L}^\top$ is symmetric positive semidefinite (because the graph is weight-balanced), we have $y\in\mathcal{N}(\mathsf{L}+\mathsf{L}^\top)$. Since $\mathcal{N}(\mathsf{L}+\mathsf{L}^\top)=\mathrm{span}\{\mathbb{1}_N\}$, because the graph is strongly connected, and $y_{i_0}=0$, we obtain that $y=0_N$, which is a contradiction. Therefore, $r\,(e_{i_0}e_{i_0}^\top)\otimes\mathrm{I}_d+\epsilon(\mathsf{L}+\mathsf{L}^\top)\otimes\mathrm{I}_d$ is positive definite, and hence $K_1>0$. On the other hand,

$$\begin{aligned}2K_1\,\mathrm{I}_{Nd}&\preccurlyeq 2\big(r\,e_{i_0}e_{i_0}^\top+\epsilon(\mathsf{L}+\mathsf{L}^\top)\big)\otimes\mathrm{I}_d\\&\preccurlyeq 2\mathbf{DG}(\boldsymbol{x})+\mathbf{DT}(\boldsymbol{x})+(\mathbf{DT}(\boldsymbol{x}))^\top\preccurlyeq\mathbf{DF}(\boldsymbol{x})+(\mathbf{DF}(\boldsymbol{x}))^\top,\end{aligned}$$

for any $\boldsymbol{x}\in(\mathbb{R}^d)^N$, as required. Before proving *(ii)* and *(iii)*, we derive some useful expressions. We start by defining $j:[0,1]\to(\mathbb{R}^d)^N$ as $j(t):=\mathbf{F}\big(\bar{\boldsymbol{x}}+t(\boldsymbol{x}-\bar{\boldsymbol{x}})\big)-\mathbf{F}(\bar{\boldsymbol{x}})$. By the Fundamental Theorem of Calculus, we have that

$$j(1)=j(1)-j(0)=\int_0^1 j'(t)\mathrm{d}t=\mathrm{E}(\boldsymbol{x})(\boldsymbol{x}-\bar{\boldsymbol{x}}),\tag{5.7}$$

where the integral is taken component-wise and the matrix-valued function $\mathrm{E}:(\mathbb{R}^d)^N\to\mathbb{R}^{Nd\times Nd}$ is defined by

$$\begin{aligned}\mathrm{E}(\boldsymbol{x})\ &:=\int_0^1\mathbf{DF}\big(\bar{\boldsymbol{x}}+t(\boldsymbol{x}-\bar{\boldsymbol{x}})\big)\mathrm{d}t\ =\ \int_0^1\mathbf{DG}\big(\bar{\boldsymbol{x}}+t(\boldsymbol{x}-\bar{\boldsymbol{x}})\big)\mathrm{d}t+2\epsilon(\mathsf{L}\otimes\mathrm{I}_d)\\&:=\mathrm{D}(\boldsymbol{x})+2\epsilon(\mathsf{L}\otimes\mathrm{I}_d),\end{aligned}$$

for $\boldsymbol{x}\in(\mathbb{R}^d)^N$. We derive next some useful facts about E.

(a) Since $\mathrm{D}(\boldsymbol{x})$ is symmetric positive semidefinite and $\mathrm{D}(\boldsymbol{x})\preccurlyeq R\,\mathrm{I}$ for all $\boldsymbol{x}\in(\mathbb{R}^d)^N$, using [1, Fact 5.11.2], we deduce

$$\begin{aligned}\sigma_{\max}(\mathrm{E}(\boldsymbol{x}))&\le\sigma_{\max}(\mathrm{D}(\boldsymbol{x}))+\sigma_{\max}(2\epsilon(\mathsf{L}\otimes\mathrm{I}_d))\\&=\lambda_{\max}(\mathrm{D}(\boldsymbol{x}))+2\epsilon\,\sigma_{\max}(\mathsf{L}\otimes\mathrm{I}_d)\le R+2\epsilon\,\sigma_{\max}(\mathsf{L})=K_2,\end{aligned}\tag{5.8}$$

where in the last inequality we have used $\sigma_{\max}(\mathsf{L}\otimes\mathrm{I}_d)=\sqrt{\lambda_{\max}((\mathsf{L}^\top\mathsf{L})\otimes\mathrm{I}_d)}=\sqrt{\lambda_{\max}(\mathsf{L}^\top\mathsf{L})}=\sigma_{\max}(\mathsf{L})$.

(b) Using *(i)*, we deduce

$$2K_1\,\mathrm{I}\preccurlyeq\mathrm{E}(\boldsymbol{x})+\mathrm{E}(\boldsymbol{x})^\top.\tag{5.9}$$

(c) Using [1, Fact 8.14.4] and (5.9), we get

$$\sigma_{\min}(\mathrm{E}(\boldsymbol{x}))\ge\tfrac{1}{2}\,\lambda_{\min}(\mathrm{E}(\boldsymbol{x})+\mathrm{E}(\boldsymbol{x})^\top)\ge K_1>0.\tag{5.10}$$

(d) Since $\mathrm{E}(\boldsymbol{x})$ is a square matrix, we have $\lambda_i(\mathrm{E}(\boldsymbol{x})\mathrm{E}(\boldsymbol{x})^\top) = \lambda_i(\mathrm{E}(\boldsymbol{x})^\top\mathrm{E}(\boldsymbol{x})) = \big(\sigma_i(\mathrm{E}(\boldsymbol{x}))\big)^2$ for $i = 1, \ldots, Nd$, and, therefore, both $\mathrm{E}(\boldsymbol{x})\mathrm{E}(\boldsymbol{x})^\top$ and $\mathrm{E}(\boldsymbol{x})^\top\mathrm{E}(\boldsymbol{x})$ are lower and upper bounded by $(\sigma_{\min}(\mathrm{E}(\boldsymbol{x})))^2\,\mathrm{I}$ and $(\sigma_{\max}(\mathrm{E}(\boldsymbol{x})))^2\,\mathrm{I}$, respectively.

(e) Taking the invertible congruence given by the matrix $\mathrm{E}(\boldsymbol{x})^{-1} \in \mathbb{R}^{Nd\times Nd}$ (which is invertible by (c)) on both sides of (5.9), that is, multiplying on the left by $(\mathrm{E}(\boldsymbol{x})^\top)^{-1} = (\mathrm{E}(\boldsymbol{x})^{-1})^\top := \mathrm{E}(\boldsymbol{x})^{-\top}$ and on the right by $\mathrm{E}(\boldsymbol{x})^{-1}$, we get

$$2K_1\,\mathrm{E}(\boldsymbol{x})^{-\top}\mathrm{E}(\boldsymbol{x})^{-1} \preccurlyeq \mathrm{E}(\boldsymbol{x})^{-\top} + \mathrm{E}(\boldsymbol{x})^{-1}. \tag{5.11}$$

Now, since $\mathrm{E}(\boldsymbol{x})^{-\top}\mathrm{E}(\boldsymbol{x})^{-1} = \big(\mathrm{E}(\boldsymbol{x})\mathrm{E}(\boldsymbol{x})^\top\big)^{-1}$ we obtain from (5.11) that

$$\mathrm{E}(\boldsymbol{x})^{-\top} + \mathrm{E}(\boldsymbol{x})^{-1} \succcurlyeq \frac{2K_1}{\lambda_{\max}\big(\mathrm{E}(\boldsymbol{x})\mathrm{E}(\boldsymbol{x})^\top\big)}\,\mathrm{I} = 2K_1\big(\sigma_{\max}(\mathrm{E}(\boldsymbol{x}))\big)^{-2}\,\mathrm{I} \succcurlyeq 2K_1K_2^{-2}\,\mathrm{I}, \tag{5.12}$$

for all $\boldsymbol{x} \in (\mathbb{R}^d)^N$, where we used (d) in the identity and (a) in the last inequality. Equipped with these facts, we are ready to establish items *(ii)* and *(iii)*.

Regarding *(ii)*, notice that $\|\mathbf{F}(\boldsymbol{x}) - \mathbf{F}(\bar{\boldsymbol{x}})\|_2^2 = \|j(1)\|_2^2 = (\boldsymbol{x}-\bar{\boldsymbol{x}})^\top\mathrm{E}(\boldsymbol{x})^\top\mathrm{E}(\boldsymbol{x})(\boldsymbol{x}-\bar{\boldsymbol{x}})$, and therefore the result follows from (d) using the bound for $\sigma_{\min}(\mathrm{E}(\boldsymbol{x}))$ in (5.10) and for $\sigma_{\max}(\mathrm{E}(\boldsymbol{x}))$ in (5.8).

Regarding *(iii)*, we rewrite the inequality (5.6), which we need to establish for the vector field $\mathbf{F}$ and the matrix transformation $S := \mathrm{I} + \beta^2\tilde{S}$, as $(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top S j(1) \geq \delta\, j(1)^\top j(1)$, for all $\boldsymbol{x} \in (\mathbb{R}^d)^N$. Using (5.7), this becomes

$$(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top S\mathrm{E}(\boldsymbol{x})(\boldsymbol{x}-\bar{\boldsymbol{x}}) \geq \delta\,(\boldsymbol{x}-\bar{\boldsymbol{x}})^\top\mathrm{E}(\boldsymbol{x})^\top\mathrm{E}(\boldsymbol{x})(\boldsymbol{x}-\bar{\boldsymbol{x}}), \quad \forall \boldsymbol{x} \in (\mathbb{R}^d)^N,$$

which follows from the stronger condition given by

$$\tfrac{1}{2}\big(\mathrm{E}(\boldsymbol{x})^\top S^\top + S\mathrm{E}(\boldsymbol{x})\big) \succcurlyeq \delta\,\mathrm{E}(\boldsymbol{x})^\top\mathrm{E}(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in (\mathbb{R}^d)^N. \tag{5.13}$$

We now proceed verifying an equivalent linear matrix inequality. Taking now on both sides of (5.13) the same congruence as in (e) and substituting $S = \mathrm{I} + \beta^2\tilde{S}$, we get

$$(\mathrm{I} + \beta^2\tilde{S}^\top)\mathrm{E}(\boldsymbol{x})^{-1} + \mathrm{E}(\boldsymbol{x})^{-\top}(\mathrm{I} + \beta^2\tilde{S}) \succcurlyeq 2\delta\mathrm{I}, \quad \forall \boldsymbol{x} \in (\mathbb{R}^d)^N,$$

which, after reordering terms and defining $\tilde{\mathrm{E}}(\boldsymbol{x}) := \mathrm{E}(\boldsymbol{x})^{-1} - \delta\mathrm{I}$, becomes

$$\tilde{\mathrm{E}}(\boldsymbol{x}) + \tilde{\mathrm{E}}(\boldsymbol{x})^\top \succcurlyeq -\beta^2\big(\tilde{S}^\top\mathrm{E}(\boldsymbol{x})^{-1} + \mathrm{E}(\boldsymbol{x})^{-\top}\tilde{S}\big), \quad \forall \boldsymbol{x} \in (\mathbb{R}^d)^N. \tag{5.14}$$

To guarantee that (5.14) holds, we seek bounds on both sides that are uniform. Regarding the left-hand side of (5.14), we get from (5.12) that

$$\tilde{\mathrm{E}}(\boldsymbol{x}) + \tilde{\mathrm{E}}(\boldsymbol{x})^\top = \mathrm{E}(\boldsymbol{x})^{-1} + \mathrm{E}(\boldsymbol{x})^{-\top} - 2\delta\mathrm{I} \succcurlyeq 2\big(K_1K_2^{-2} - \delta\big)\,\mathrm{I}, \quad \forall \boldsymbol{x} \in (\mathbb{R}^d)^N. \tag{5.15}$$

Regarding the right-hand side of (5.14), using (5.10) we first observe that

$$\|\mathrm{E}(\boldsymbol{x})^{-1}\|_2 = \sigma_{\max}(\mathrm{E}(\boldsymbol{x})^{-1}) = \big(\sigma_{\min}(\mathrm{E}(\boldsymbol{x}))\big)^{-1} \leq K_1^{-1}, \quad \forall \boldsymbol{x} \in (\mathbb{R}^d)^N.$$

Thus, using the triangular inequality, the fact that $\|A\|_2 = \|A^\top\|_2$, and the submultiplicativity of the norm, we get that for all $\boldsymbol{x} \in (\mathbb{R}^d)^N$,

$$\|\tilde{S}^\top\mathrm{E}(\boldsymbol{x})^{-1} + \mathrm{E}(\boldsymbol{x})^{-\top}\tilde{S}\|_2 \leq 2\|\tilde{S}^\top\mathrm{E}(\boldsymbol{x})^{-1}\|_2 \leq 2\|\tilde{S}\|_2\,\|\mathrm{E}(\boldsymbol{x})^{-1}\|_2 \leq 2\|\tilde{S}\|_2\,K_1^{-1}.$$

Since $\pm A \preccurlyeq \|A\|_2\,\mathrm{I}$, we deduce

$$-\big(\tilde{S}^\top \mathrm{E}(\boldsymbol{x})^{-1} + \mathrm{E}(\boldsymbol{x})^{-1}\tilde{S}\big) \preccurlyeq 2\,\|\tilde{S}\|_2\,K_1^{-1}\,\mathrm{I}, \quad \forall \boldsymbol{x} \in (\mathbb{R}^d)^N. \tag{5.16}$$

Therefore, relating the uniform bounds (5.15) and (5.16), we conclude that if $\beta \leq \beta_1^*$, then (5.14) holds for every $\boldsymbol{x} \in (\mathbb{R}^d)^N$ because

$$\tilde{\mathrm{E}}(\boldsymbol{x}) + \tilde{\mathrm{E}}(\boldsymbol{x})^\top \succcurlyeq 2\big(K_1K_2^{-2} - \delta\big)\,\mathrm{I} \succcurlyeq 2\,\|\tilde{S}\|_2\,K_1^{-1}\,\beta^2\,\mathrm{I} \succcurlyeq -\beta^2\big(\tilde{S}^\top \mathrm{E}(\boldsymbol{x})^{-1} + \mathrm{E}(\boldsymbol{x})^{-1}\tilde{S}\big),$$

which concludes the proof. □

Note that, under the hypotheses of Theorem 4.2, the above result is applicable to $\mathbf{G} = \nabla\tilde{f}$ (so that $\mathbf{DG} = \nabla^2\tilde{f}$ is symmetric and conveniently lower and upper bounded by the hypotheses on the local functions), $\mathbf{F}(\boldsymbol{x}) = \mathbf{F}_\epsilon(\boldsymbol{x}) := \nabla\tilde{f}(\boldsymbol{x}) + 2\epsilon\mathbf{L}\boldsymbol{x}$, and $\tilde{S} = \mathbf{L}_\mathcal{K}$ (which has $\|\tilde{S}\|_2 = \|\mathbf{L}_\mathcal{K}\|_2 = 1$). In particular, we have

$$K_1\|\boldsymbol{x}' - \boldsymbol{x}\|_2 \leq \|\mathbf{F}_\epsilon(\boldsymbol{x}') - \mathbf{F}_\epsilon(\boldsymbol{x})\|_2 \leq K_2\|\boldsymbol{x}' - \boldsymbol{x}\|_2, \tag{5.17}$$

for all $\boldsymbol{x}', \boldsymbol{x} \in (\mathbb{R}^d)^N$, and

$$(\boldsymbol{x} - \boldsymbol{x}^*)^\top(\mathrm{I} + \beta^2\mathbf{L}_\mathcal{K})(\mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}^*)) \geq \delta\,\|\mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}^*)\|_2^2, \tag{5.18}$$

for all $\boldsymbol{x} \in (\mathbb{R}^d)^N$, $\delta \in (0,\, K_1K_2^{-2})$ and $\beta \in \big[\,0,\, \sqrt{K_1^2K_2^{-2} - K_1\delta}\,\big]$.

5.3. Global existence and uniqueness of solutions. Here we establish the global existence and uniqueness of solutions of the dynamics (4.2) by verifying the hypotheses in Lemma 2.1. We obtain the following bound for almost every $t \geq t_0$:

$$\begin{aligned}
&\max\big\{\,\|g(v) - g(v')\|_2\,,\,|G(v,t) - G(v',t)|_\mathcal{F}\,\big\}\\
&\leq \|\mathbf{A}(v - v')\|_2 + \|\mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}')\|_2 + |G(v,t) - G(v',t)|_\mathcal{F}\\
&\leq \|\mathbf{A}\|_2\|v - v'\|_2 + K_2\|\boldsymbol{x} - \boldsymbol{x}'\|_2 + \kappa_1\|v - v'\|_2 \leq \big(\|\mathbf{A}\|_2 + K_2 + \kappa_1\big)\|v - v'\|_2,
\end{aligned}$$

where in the second inequality we have used (5.17) and the Lipschitz condition for G in (5.2). In addition, for almost every $t \geq t_0$,

$$\begin{aligned}
&v^\top g(v) + \tfrac{1}{2}|G(v,t)|_\mathcal{F}^2 = v^\top\mathbf{A}v + v^\top\mathbf{N}(\boldsymbol{x}) + \tfrac{1}{2}|G(v,t)|_\mathcal{F}^2\\
&\leq \|\mathbf{A}\|_2\|v\|_2^2 + \|v\|_2\|\mathbf{F}_\epsilon(\boldsymbol{x})\|_2 + \tfrac{1}{2}\kappa_2^2\\
&\leq \|\mathbf{A}\|_2\|v\|_2^2 + \|v\|_2\big(K_2\|\boldsymbol{x} - \boldsymbol{x}^*\|_2 + \|\mathbf{F}_\epsilon(\boldsymbol{x}^*)\|_2\big) + \tfrac{1}{2}\kappa_2^2\\
&\leq \|\mathbf{A}\|_2\|v\|_2^2 + \|v\|_2\big(K_2\|\boldsymbol{x}\|_2 + K_2\|\boldsymbol{x}^*\|_2 + \|\mathbf{F}_\epsilon(\boldsymbol{x}^*)\|_2\big) + \tfrac{1}{2}\kappa_2^2\\
&\leq \big(\|\mathbf{A}\|_2 + K_2\big)\|v\|_2^2 + \|v\|_2\big(K_2\|\boldsymbol{x}^*\|_2 + \|\mathbf{F}_\epsilon(\boldsymbol{x}^*)\|_2\big) + \tfrac{1}{2}\kappa_2^2\\
&\leq \big(1 + \|v\|_2^2\big)\big(\|\mathbf{A}\|_2 + K_2 + K_2\|\boldsymbol{x}^*\|_2 + \|\mathbf{F}_\epsilon(\boldsymbol{x}^*)\|_2 + \tfrac{1}{2}\kappa_2^2\big).
\end{aligned}$$

The global existence and uniqueness of the solutions of the dynamics (4.2) now follows from Lemma 2.1 as a consequence of these two facts.

5.4. NSS Lyapunov function. Our strategy to establish the noise-to-state stability properties of the distributed coordination algorithm (4.2) is based on identifying a suitable NSS Lyapunov function for the dynamics. Our first result of this section identifies a candidate Lyapunov function whose derivative in the sense of Itô can be conveniently upper bounded. To obtain this bound, we build on the co-coercivity

properties stated in Theorem 5.2 of the vector fields that combine local gradient descent and local consensus.

PROPOSITION 5.3. (Candidate second moment NSS-Lyapunov function). *Under the hypotheses of Theorem 4.2, let*

$$\mathrm{P}_\beta := \begin{bmatrix} \mathrm{I} + \beta^2 \mathbf{L}_\mathcal{K} & \beta \mathbf{L}_\mathcal{K} \\ \beta \mathbf{L}_\mathcal{K} & \mathbf{L}_\mathcal{K} \end{bmatrix} \in \mathbb{R}^{2Nd \times 2Nd},$$
$$\mathrm{Q}_\beta := \begin{bmatrix} \begin{bmatrix} \beta^3 + 2\beta + \frac{2}{\beta} & (1+\beta^2) \\ (1+\beta^2) & \beta \end{bmatrix} \otimes (\mathbf{L} + \mathbf{L}^\top) & 0 \\ 0 & \beta \mathbf{L}_\mathcal{K} \end{bmatrix} \begin{bmatrix} 0 \\ 2\delta \mathrm{I} \end{bmatrix} \in \mathbb{R}^{3Nd \times 3Nd},$$

and define the functions $\mathrm{V}, \mathrm{W} : \mathbb{R}^{2Nd} \to \mathbb{R}$ *by*

$$\mathrm{V}(v) := \tfrac{1}{2}[(\boldsymbol{x} - \boldsymbol{x}^*)^\top, (\boldsymbol{z} - \boldsymbol{z}^*)^\top] \, \mathrm{P}_\beta \begin{bmatrix} \boldsymbol{x} - \boldsymbol{x}^* \\ \boldsymbol{z} - \boldsymbol{z}^* \end{bmatrix},$$
$$\mathrm{W}(v) := \tfrac{1}{2}[(\boldsymbol{x} - \boldsymbol{x}^*)^\top, (\boldsymbol{z} - \boldsymbol{z}^*)^\top, (\mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}^*))^\top] \, \mathrm{Q}_\beta \begin{bmatrix} \boldsymbol{x} - \boldsymbol{x}^* \\ \boldsymbol{z} - \boldsymbol{z}^* \\ \mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}^*) \end{bmatrix},$$

where $\boldsymbol{x}^* = \mathbb{1} \otimes x_{\min} \in (\mathbb{R}^d)^N$ *and* $\boldsymbol{z}^* \in (\mathbb{R}^d)^n$ *is such that* $\mathbf{L}\boldsymbol{z}^* = -\nabla \tilde{f}(\mathbb{1} \otimes x_{\min})$. *Then the following holds:*

(i) The matrix P_β *is positive semidefinite for any* $\beta \in \mathbb{R}$ *with nullspace*

$$\mathcal{N}(\mathrm{P}_\beta) = \operatorname{span}\big\{ \begin{bmatrix} 0 & (\mathbb{1} \otimes b)^\top \end{bmatrix}^\top : b \in \mathbb{R}^d \big\}.$$

(ii) The matrix Q_β *is positive semidefinite for the range of values of* β *specified in Theorem 4.2, and has nullspace*

$$\mathcal{N}(\mathrm{Q}_\beta) = \operatorname{span}\big\{ \begin{bmatrix} (\mathbb{1} \otimes b_1)^\top & 0 & 0 \end{bmatrix}^\top, \begin{bmatrix} 0 & (\mathbb{1} \otimes b_2)^\top & 0 \end{bmatrix}^\top : b_1, b_2 \in \mathbb{R}^d \big\}.$$

(iii) The function V *is twice continuously differentiable and bounded by*

$$\alpha_1\big(\|v - v^*\|_{\hat{\mathrm{I}}}^2\big) \le \mathrm{V}(v) \le \alpha_2\big(\|v - v^*\|_{\hat{\mathrm{I}}}^2\big), \tag{5.19}$$

where $v^* := (\boldsymbol{x}^*, \boldsymbol{z}^*)$, $\alpha_1(r) := \lambda_{(2N-1)d}(\mathrm{P}_\beta) r$, $\alpha_2(r) := \lambda_{\max}(\mathrm{P}_\beta) r$, *and the matrix* $\hat{\mathrm{I}} \in \mathbb{R}^{2Nd}$ *is defined as*

$$\hat{\mathrm{I}} := \operatorname{diag}\big(\mathrm{I}_{Nd}, \mathbf{L}_\mathcal{K}\big).$$

(iv) The function W *is continuous, and the following dissipation inequality holds,*

$$\mathcal{L}[\mathrm{V}](v, t) \le -\mathrm{W}(v) + \sigma\big(|\Sigma(t)|_\mathcal{F}\big), \tag{5.20}$$

for all $(v, t) \in \mathbb{R}^{2Nd} \times [t_0, \infty)$, *where* $\sigma(r) := \operatorname{trace}(\mathrm{P}_\beta) \, \kappa_2^2 \, r^2$.

Proof. To show *(i)*, we note that P_β is a congruence by an invertible matrix of the positive semidefinite matrix $\hat{\mathrm{I}}$,

$$\mathrm{P}_\beta = \begin{bmatrix} \mathrm{I} & 0 \\ \beta \mathrm{I} & \mathrm{I} \end{bmatrix}^\top \begin{bmatrix} \mathrm{I} & 0 \\ 0 & \mathbf{L}_\mathcal{K} \end{bmatrix} \begin{bmatrix} \mathrm{I} & 0 \\ \beta \mathrm{I} & \mathrm{I} \end{bmatrix}.$$

Therefore, $\operatorname{rank}(\mathrm{P}_\beta) = \operatorname{rank}(\hat{\mathrm{I}}) = \operatorname{rank}(\mathrm{I}) + \operatorname{rank}(\mathbf{L}_\mathcal{K}) = Nd + (N-1)d = (2N-1)d$. The statement follows now by noting that the subspace $\operatorname{span}\left\{ \begin{bmatrix} 0 & (\mathbb{1}\otimes b)^\top \end{bmatrix}^\top : b \in \mathbb{R}^d \right\}$ has dimension d and lies in the nullspace of P_β.

To establish *(ii)*, we show that $-\mathrm{Q}_\beta$ is negative semidefinite for the range of values of β in the statement. For convenience, define the matrices

$$\begin{aligned} \mathrm{B} &:= \begin{bmatrix} \beta^3 + 2\beta + \frac{2}{\beta} & (1+\beta^2) \\ (1+\beta^2) & \beta \end{bmatrix}, \\ \mathrm{Q}_1 &:= -\mathrm{B} \otimes (\mathbf{L} + \mathbf{L}^\top), \end{aligned}$$

and note that Q_1 corresponds to the first block of $-\mathrm{Q}_\beta$. Since B is symmetric, $\det(\mathrm{B}) = 1$, and $\operatorname{trace}(\mathrm{B}) = \beta^3 + 3\beta + \frac{2}{\beta} > 0$ for $\beta > 0$, we deduce that $-\mathrm{B} \prec 0$ for any $\beta > 0$. Therefore, Q_1 is symmetric negative semidefinite with nullspace

$$\mathcal{N}(\mathrm{Q}_1) = \operatorname{span}\left\{ \begin{bmatrix} (\mathbb{1}\otimes b_1)^\top & 0 \end{bmatrix}^\top, \begin{bmatrix} 0 & (\mathbb{1}\otimes b_2)^\top \end{bmatrix}^\top \,:\, b_1, b_2 \in \mathbb{R}^d \right\}.$$

Next, defining

$$\mathrm{Q}_2 := \begin{pmatrix} 0 & 0 \\ 0 & \frac{\beta^2}{2\delta} \end{pmatrix} \otimes \mathbf{L}_\mathcal{K}$$

and using $\mathbf{L}_\mathcal{K}^2 = \mathbf{L}_\mathcal{K}$, we simplify the following invertible congruence,

$$\begin{aligned} &- \begin{bmatrix} \mathrm{I} & 0 & 0 \\ 0 & \mathrm{I} & 0 \\ 0 & -\frac{\beta}{2\delta}\mathbf{L}_\mathcal{K} & \mathrm{I} \end{bmatrix}^\top \mathrm{Q}_\beta \begin{bmatrix} \mathrm{I} & 0 & 0 \\ 0 & \mathrm{I} & 0 \\ 0 & -\frac{\beta}{2\delta}\mathbf{L}_\mathcal{K} & \mathrm{I} \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{I} & 0 & 0 \\ 0 & \mathrm{I} & -\frac{\beta}{2\delta}\mathbf{L}_\mathcal{K} \\ 0 & 0 & \mathrm{I} \end{bmatrix} \left(\begin{bmatrix} \mathrm{Q}_1 & & 0 \\ & 0 & \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\beta\mathbf{L}_\mathcal{K} \\ 0 & -\beta\mathbf{L}_\mathcal{K} & -2\delta\mathrm{I} \end{bmatrix} \right) \begin{bmatrix} \mathrm{I} & 0 & 0 \\ 0 & \mathrm{I} & 0 \\ 0 & -\frac{\beta}{2\delta}\mathbf{L}_\mathcal{K} & \mathrm{I} \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Q}_1 + \begin{pmatrix} 0 & 0 \\ 0 & \frac{\beta^2}{2\delta} \end{pmatrix} \otimes \mathbf{L}_\mathcal{K} & 0 \\ 0 & -2\delta\mathrm{I} \end{bmatrix} = \begin{bmatrix} \mathrm{Q}_1 + \mathrm{Q}_2 & 0 \\ 0 & -2\delta\mathrm{I} \end{bmatrix}. \end{aligned}$$

Since this is a block-diagonal matrix whose lower block, $-2\delta\mathrm{I}$, is negative definite, to establish the result is sufficient to show that for the specified values of β, the sum $\mathrm{Q}_1 + \mathrm{Q}_2$ is negative semidefinite. Note the maximum nonzero eigenvalue of Q_1, denoted $\lambda_{\max}^{\varnothing}(\mathrm{Q}_1)$, is

$$\left(-\tfrac{\beta^4+3\beta^2+2}{2\beta} + \sqrt{\left(\tfrac{\beta^4+3\beta^2+2}{2\beta}\right)^2 - 1} \right) \lambda_2(\mathsf{L} + \mathsf{L}^\top).$$

On the other hand, Q_2 is symmetric positive semidefinite with $\operatorname{rank}(\mathrm{Q}_2) = \operatorname{rank}(\mathbf{L}_\mathcal{K}) = (N-1)d$ and $\operatorname{spec}(\mathrm{Q}_2) = \{0, \frac{\beta^2}{2\delta}\}$, so the maximum eigenvalue of Q_2 is $\lambda_{\max}(R) = \frac{\beta^2}{2\delta}$. Now, since $\mathcal{N}(\mathrm{Q}_1) \subseteq \mathcal{N}(\mathrm{Q}_2)$, it follows that $\mathcal{N}(\mathrm{Q}_1) \subseteq \mathcal{N}(\mathrm{Q}_1 + \mathrm{Q}_2)$. Thus, in order to check the semidefiniteness of $\mathrm{Q}_1 + \mathrm{Q}_2$, we can restrict our attention to the subspace $\mathcal{U}^\perp := \mathcal{N}(\mathrm{Q}_1)^\perp$. By Weyl's Theorem [8, Theorem 4.3.7],

$$\begin{aligned} \lambda_{\max}^{\varnothing}(\mathrm{Q}_1 + \mathrm{Q}_2) &= \lambda_{\max}^{\mathcal{U}^\perp}(\mathrm{Q}_1 + \mathrm{Q}_2) \le \lambda_{\max}^{\mathcal{U}^\perp}(\mathrm{Q}_1) + \lambda_{\max}^{\mathcal{U}^\perp}(\mathrm{Q}_2) \\ &= \lambda_{\max}^{\varnothing}(\mathrm{Q}_1) + \lambda_{\max}(\mathrm{Q}_2) = h(\beta, \delta). \end{aligned}$$

Since, by Lemma A.1, $h(\beta,\delta)<0$ for $\delta\in(0,\,K_1K_2^{-2})$ and $\beta\in(0,\min\{\beta_1^*(\delta,\epsilon),\beta_2^*(\delta)\})$, we deduce that $\mathrm{Q}_1+\mathrm{Q}_2$ is negative definite in the subspace $\mathcal{N}(\mathrm{Q}_1)^\perp$. Therefore, $\mathcal{N}(\mathrm{Q}_1{+}\mathrm{Q}_2)=\mathcal{N}(\mathrm{Q}_1)$, which in turn implies that $\mathcal{N}(\mathrm{Q}_\beta)=\operatorname{span}\big\{[u^\top,0]^\top\,:u\in\mathcal{N}(\mathrm{Q}_1)\big\}$.

Regarding *(iii)*, it is clear from its definition that V is twice (in fact, infinitely) continuously differentiable. Furthermore, notice that $\hat{\mathrm{I}}$ and P_β are symmetric positive semidefinite with the same nullspace, so that

$$\tfrac{\lambda_{(2N-1)d}(\mathrm{P}_\beta)}{\lambda_{\max}(\hat{\mathrm{I}})}\,y^\top\hat{\mathrm{I}}\,y\le y^\top\mathrm{P}_\beta\,y\le\tfrac{\lambda_{\max}(\mathrm{P}_\beta)}{\lambda_{(2N-1)d}(\hat{\mathrm{I}})}\,y^\top\hat{\mathrm{I}}\,y,$$

for all $y\in\mathbb{R}^{2Nd}$. Since $\hat{\mathrm{I}}$ is idempotent, $\hat{\mathrm{I}}=\hat{\mathrm{I}}^2$, we have $y^\top\hat{\mathrm{I}}\,y=\|y\|_{\hat{\mathrm{I}}}^2$. The result now follows by observing that all nonzero eigenvalues of $\hat{\mathrm{I}}$ are 1.

Finally, we turn our attention to *(iv)*. We first compute the elements of $\mathcal{L}[\mathrm{V}]$ in (2.2). With the notation of (5.1), using that $\mathrm{P}_\beta=\mathrm{P}_\beta^\top$ and the sub-multiplicativity of the Frobenius norm, the diffusion term yields

$$\begin{aligned}&\tfrac12\operatorname{trace}\big(\Sigma(t)^\top G(v,t)^\top\nabla_v^2\mathrm{V}(v)G(v,t)\Sigma(t)\big)\\&\quad=\tfrac12\operatorname{trace}\big(\Sigma(t)^\top G(v,t)^\top\mathrm{P}_\beta G(v,t)\Sigma(t)\big)\\&\quad=|\mathrm{P}_\beta^{1/2}G(v,t)\Sigma(t)|_{\mathcal{F}}^2\;\le\;|\mathrm{P}_\beta^{1/2}|_{\mathcal{F}}^2\,|G(v,t)|_{\mathcal{F}}^2\,|\Sigma(t)|_{\mathcal{F}}^2\\&\quad\le\operatorname{trace}(\mathrm{P}_\beta)\,\kappa_2^2\,|\Sigma(t)|_{\mathcal{F}}^2\;=\;\sigma(|\Sigma(t)|_{\mathcal{F}}).\end{aligned}$$

On the other hand, defining $\tilde{\mathrm{Q}}_1:=2\operatorname{sym}\big(\mathrm{P}_\beta\mathbf{A}\big):=\mathrm{P}_\beta\mathbf{A}+\mathbf{A}^\top\mathrm{P}_\beta$ and $\tilde v:=v-v^*$, and subtracting the quantity $\mathbf{A}v^*+\mathbf{N}(\boldsymbol{x}^*)=0$, the drift term yields

$$\begin{aligned}\nabla_v\mathrm{V}(v)^\top\big(\mathbf{A}v+\mathbf{N}(\boldsymbol{x})\big)&=\nabla_v\mathrm{V}(v)^\top\big(\mathbf{A}\tilde v-\mathbf{N}(\boldsymbol{x}^*)+\mathbf{N}(\boldsymbol{x})\big)\\&=\tfrac12\tilde v^\top\tilde{\mathrm{Q}}_1\tilde v+\tilde v^\top\mathrm{P}_\beta(-\mathbf{N}(\boldsymbol{x}^*)+\mathbf{N}(\boldsymbol{x})).\end{aligned}$$

Summarizing, we have

$$\mathcal{L}[\mathrm{V}](v,t)\le\tfrac12\tilde v^\top\tilde{\mathrm{Q}}_1\tilde v+\tilde v^\top\mathrm{P}_\beta(-\mathbf{N}(\boldsymbol{x}^*)+\mathbf{N}(\boldsymbol{x}))+\sigma\big(|\Sigma(t)|_{\mathcal{F}}\big)\tag{5.21}$$

for all $(v,t)\in\mathbb{R}^{2Nd}\times[t_0,\infty)$. We look first at the quadratic term in (5.21) arising from the linear part of the dynamics. Since $\mathbf{L}_{\mathcal{K}}\mathbf{L}=\mathrm{I}\,\mathbf{L}=\mathbf{L}$, splitting the matrix P_β, we obtain the factorization

$$\begin{aligned}\tilde{\mathrm{Q}}_1&=2\operatorname{sym}\Big(\Big(\begin{bmatrix}1&0\\0&0\end{bmatrix}\otimes\mathrm{I}_{Nd}+\begin{bmatrix}\beta^2&\beta\\\beta&1\end{bmatrix}\otimes\mathbf{L}_{\mathcal{K}}\Big)\Big(\begin{bmatrix}-\gamma&-1\\1&0\end{bmatrix}\otimes\mathbf{L}\Big)\Big)\\&=2\operatorname{sym}\Big(\Big(\begin{bmatrix}1+\beta^2&\beta\\\beta&1\end{bmatrix}\begin{bmatrix}-\gamma&-1\\1&0\end{bmatrix}\Big)\otimes\big(\mathbf{L}_{\mathcal{K}}\mathbf{L}\big)\Big)\\&=2\operatorname{sym}\Big(\begin{bmatrix}-\gamma(1+\beta^2)+\beta&-(1+\beta^2)\\-\gamma\beta+1&-\beta\end{bmatrix}\otimes\mathbf{L}\Big).\end{aligned}$$

Now, recalling that $(2+\beta^2)/\beta+2\epsilon=\tilde\gamma=\gamma+2\epsilon$, we have $\gamma=(2+\beta^2)/\beta$, so the first matrix is indeed symmetric and we can factor out $2\operatorname{sym}(\mathbf{L}):=\mathbf{L}+\mathbf{L}^\top$ using the Kronecker product. In fact, $-\gamma(1+\beta^2)+\beta=-\beta^3-2\beta-\frac{2}{\beta}$, and we deduce

$$\tilde{\mathrm{Q}}_1=-\begin{bmatrix}\beta^3+2\beta+\frac{2}{\beta}&(1+\beta^2)\\(1+\beta^2)&\beta\end{bmatrix}\otimes(\mathbf{L}+\mathbf{L}^\top)=\mathrm{Q}_1.\tag{5.22}$$

Next, we turn our attention to the nonlinear term in (5.21). Note that

$$\begin{aligned}
&\tilde{v}^\top \mathrm{P}_\beta\big(-\mathbf{N}(\boldsymbol{x}^*)+\mathbf{N}(\boldsymbol{x})\big)\\
&\quad= \tilde{v}^\top \begin{bmatrix} \mathrm{I}+\beta^2\mathbf{L}_\mathcal{K} & \beta\mathbf{L}_\mathcal{K} \\ \beta\mathbf{L}_\mathcal{K} & \mathbf{L}_\mathcal{K}\end{bmatrix}\begin{bmatrix}\mathrm{I}\\0\end{bmatrix}\big(-\mathbf{F}_\epsilon(\boldsymbol{x}^*)+\mathbf{F}_\epsilon(\boldsymbol{x})\big)\\
&\quad= -(\boldsymbol{x}-\boldsymbol{x}^*)^\top(\mathrm{I}+\beta^2\mathbf{L}_\mathcal{K})\big(\mathbf{F}_\epsilon(\boldsymbol{x})-\mathbf{F}_\epsilon(\boldsymbol{x}^*)\big)-(\boldsymbol{z}-\boldsymbol{z}^*)^\top\beta\mathbf{L}_\mathcal{K}\big(\mathbf{F}_\epsilon(\boldsymbol{x})-\mathbf{F}_\epsilon(\boldsymbol{x}^*)\big)\\
&\quad\le -\delta\|\mathbf{F}_\epsilon(\boldsymbol{x})-\mathbf{F}_\epsilon(\boldsymbol{x}^*)\|_2^2-(\boldsymbol{z}-\boldsymbol{z}^*)^\top\beta\mathbf{L}_\mathcal{K}\big(\mathbf{F}_\epsilon(\boldsymbol{x})-\mathbf{F}_\epsilon(\boldsymbol{x}^*)\big)\,.
\end{aligned}$$

Here, the last inequality follows from (5.18). Therefore, the nonlinear term can be expressed as

$$\begin{aligned}
(5.23)\quad &\tilde{v}^\top \mathrm{P}_\beta\big(-\mathbf{N}(\boldsymbol{x}^*)+\mathbf{N}(\boldsymbol{x})\big)\\
&\qquad= \frac{1}{2}\big[\tilde{v}^\top,\ \big(\mathbf{F}_\epsilon(\boldsymbol{x})-\mathbf{F}_\epsilon(\boldsymbol{x}^*)\big)^\top\big]\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & -\beta\mathbf{L}_\mathcal{K}\\ 0 & -\beta\mathbf{L}_\mathcal{K} & -2\delta\mathrm{I}\end{bmatrix}\begin{bmatrix}\tilde{v}\\ \mathbf{F}_\epsilon(\boldsymbol{x})-\mathbf{F}_\epsilon(\boldsymbol{x}^*)\end{bmatrix}.
\end{aligned}$$

The result now follows from substituting (5.22) and (5.23) into (5.21). ∎

Given the result in Proposition 5.3, the missing piece to establish that V is a second moment NSS-Lyapunov function with respect to span$\big\{\begin{bmatrix}0 & (\mathbb{1}\otimes b)^\top\end{bmatrix}^\top : b\in\mathbb{R}^d\big\}$ is to relate its value to that of W. To this end, we define the constraint set

$$\mathcal{D}_{\boldsymbol{x}^*} := \big\{y\in\mathbb{R}^{3Nd} : y^3 = \mathbf{F}_\epsilon(y^1+\boldsymbol{x}^*)-\mathbf{F}_\epsilon(\boldsymbol{x}^*)\big\}\subset\mathbb{R}^{3Nd},$$

and the quadratic functions $\mathrm{V}_{\tilde{\mathrm{P}}_\beta}, \mathrm{W}_{\mathrm{Q}_\beta} : \mathcal{D}_{\boldsymbol{x}^*}\to\mathbb{R}_{\ge 0}$,

$$\begin{aligned}
\mathrm{V}_{\tilde{\mathrm{P}}_\beta}(y) &:= \tfrac{1}{2}y^\top\tilde{\mathrm{P}}_\beta y, \qquad \tilde{\mathrm{P}}_\beta := \tfrac{1}{2}y^\top\begin{bmatrix}\mathrm{P}_\beta & 0\\ 0 & 0\end{bmatrix}y,\\
\mathrm{W}_{\mathrm{Q}_\beta}(y) &:= \tfrac{1}{2}y^\top\mathrm{Q}_\beta y.
\end{aligned}$$

Note that $\mathrm{V}(v)=\mathrm{V}_{\tilde{\mathrm{P}}_\beta}(v-v^*,\mathbf{F}_\epsilon(\boldsymbol{x})-\mathbf{F}_\epsilon(\boldsymbol{x}^*))$ and $\mathrm{W}(v)=\mathrm{W}_{\mathrm{Q}_\beta}(v-v^*,\mathbf{F}_\epsilon(\boldsymbol{x})-\mathbf{F}_\epsilon(\boldsymbol{x}^*))$ for all $v\in\mathbb{R}^{2Nd}$. The following result relates the value of these quadratic functions.

Proposition 5.4. (Bound on candidate second moment NSS-Lyapunov function). *Under the hypotheses of Theorem 4.2, the next bound holds,*

$$\mathrm{V}_{\tilde{\mathrm{P}}_\beta}(y)\le\eta(\mathrm{W}_{\mathrm{Q}_\beta}(y)),\quad \forall y\in\mathcal{D}_{\boldsymbol{x}^*}, \tag{5.24}$$

with linear gain $\eta(r):=C_\eta\, r$, for $r\ge 0$, where

$$C_\eta := \frac{\lambda_{\max}(\hat{\mathrm{Q}})\,\lambda_{\max}(\tilde{\mathrm{P}}_\beta)}{\min\{1,\frac{K_1}{(1+K_2^2)}\}\,\lambda_{(3N-2)d}(\mathrm{Q}_\beta)\,\lambda_{(2N-1)d}(\hat{\mathrm{P}})},$$

$\hat{\mathrm{Q}} := \mathrm{diag}\big(\mathbf{L}+\mathbf{L}^\top,\mathbf{L}+\mathbf{L}^\top,\mathrm{I}\big)$ *and* $\hat{\mathrm{P}} := \mathrm{diag}\big(\mathrm{I},\mathbf{L}+\mathbf{L}^\top,0\big)$.

Proof. For $A := \mathrm{diag}\big(\mathrm{I},\sqrt{\mathbf{L}+\mathbf{L}^\top},\mathrm{I}\big)\in\mathbb{R}^{3Nd\times 3Nd}$, whose nullspace is $\mathcal{N}(A)=\mathrm{span}\big\{\begin{bmatrix}0 & (\mathbb{1}\otimes b)^\top & 0\end{bmatrix}^\top : b\in\mathbb{R}^d\big\}$, we define the functions $\phi_{2,A}, \psi_{2,A}:\mathbb{R}_{\ge 0}\to\mathbb{R}_{\ge 0}$,

$$\begin{aligned}
\phi_{2,A}(s) &:= \sup_{\{y\in\mathcal{D}_{\boldsymbol{x}^*}:\|y\|_A^2\le s\}}\mathrm{V}_{\tilde{\mathrm{P}}_\beta}(y),\\
\psi_{2,A}(s) &:= \inf_{\{y\in\mathcal{D}_{\boldsymbol{x}^*}:\|y\|_A^2\ge s\}}\mathrm{W}_{\mathrm{Q}_\beta}(y).
\end{aligned}$$

Before preceding with our proof strategy we show that the infimum and supremum are taken over nonempty sets. Consider the bijective map $\ell : \mathbb{R}^{2Nd} \to \mathbb{R}^{3Nd}$ given by

$$\ell(\boldsymbol{x}, \boldsymbol{z}) := (\boldsymbol{x} - \boldsymbol{x}^*, \boldsymbol{z} - \boldsymbol{z}^*, \mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}^*)),$$

which is continuous with image $\ell(\mathbb{R}^{2Nd}) = \mathcal{D}_{\boldsymbol{x}^*}$. We deduce that, as $(\boldsymbol{x}, \boldsymbol{z})$ ranges over $\mathbb{R}^{2Nd}$, the norm $\|\ell(\boldsymbol{x}, \boldsymbol{z})\|_A^2 = \|\boldsymbol{x} - \boldsymbol{x}^*\|_2^2 + q(\boldsymbol{z} - \boldsymbol{z}^*) + \|\mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}^*)\|_2^2$ takes all the values in $\mathbb{R}_{\geq 0}$ (because the composition is continuous), where

$$q(y) := y^\top (\mathbf{L} + \mathbf{L}^\top) y$$

for $y \in (\mathbb{R}^d)^N$. Therefore, the sets $\{y \in \mathcal{D}_{\boldsymbol{x}^*} : \|y\|_A \geq s\}$ and $\{y \in \mathcal{D}_{\boldsymbol{x}^*} : \|y\|_A \leq s\}$ are nonempty for each $s \geq 0$.

Our proof strategy consists of showing that for all $y \in \mathcal{D}_{\boldsymbol{x}^*}$ it holds that

$$\mathrm{V}_{\tilde{\mathrm{P}}_\beta}(y) \leq \phi_{2,A}\big(\|y\|_A^2\big) \leq \bar{\alpha}_2\big(\|y\|_A^2\big), \tag{5.25a}$$

$$\bar{\alpha}_3\big(\|y\|_A^2\big) \leq \psi_{2,A}\big(\|y\|_A^2\big) \leq \mathrm{W}_{\mathrm{Q}_\beta}(y)\,. \tag{5.25b}$$

If this were the case, then the result would follow by defining $\eta(r) = \bar{\alpha}_2(\bar{\alpha}_3^{-1}(r))$. For convenience, we use the shorthand notation $\tilde{\boldsymbol{x}} := \boldsymbol{x} - \boldsymbol{x}^*$, $\tilde{\boldsymbol{z}} := \boldsymbol{z} - \boldsymbol{z}^*$, and $\Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}}) := \mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}^*)$. Regarding (5.25b), note that $\hat{\mathrm{Q}}$ and Q_β are positive semidefinite with $\mathcal{N}(\hat{\mathrm{Q}}) = \mathcal{N}(\mathrm{Q}_\beta)$ by Proposition 5.3*(ii)*, and hence $c_1 \tilde{w}^\top \hat{\mathrm{Q}}\, \tilde{w} \leq \tilde{w}^\top \mathrm{Q}_\beta\, \tilde{w}$ for all $\tilde{w} \in \mathcal{D}_{\boldsymbol{x}^*}$, with $c_1 := \lambda_{(3N-2)d}(\mathrm{Q}_\beta)/\lambda_{\max}(\hat{\mathrm{Q}})$. For each $s > 0$, we then have

$$\begin{aligned}
\psi_{2,A}(s) &= \inf_{\{\|\tilde{\boldsymbol{x}}\|_2^2 + q(\tilde{\boldsymbol{z}}) + \|\Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})\|_2^2 \geq s\}} \mathrm{W}_{\mathrm{Q}_\beta}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})) \\
&\geq \inf_{\{\|\tilde{\boldsymbol{x}}\|_2^2 + q(\tilde{\boldsymbol{z}}) + \|\Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})\|_2^2 \geq s\}} c_1\big(q(\tilde{\boldsymbol{x}}) + q(\tilde{\boldsymbol{z}}) + \|\Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})\|_2^2\big) \\
&\geq \inf_{\{q(\tilde{\boldsymbol{z}}) + (1+K_2^2)\|\tilde{\boldsymbol{x}}\|_2^2 \geq s\}} c_1\big(q(\tilde{\boldsymbol{x}}) + q(\tilde{\boldsymbol{z}}) + K_1\|\tilde{\boldsymbol{x}}\|_2^2\big) \\
&\geq \inf_{\{q(\tilde{\boldsymbol{z}}) + (1+K_2^2)\|\tilde{\boldsymbol{x}}\|_2^2 \geq s\}} \min\{c_1\hat{c}s\,,\, c_1 s\} = c_1 \min\{\hat{c}, 1\}\, s,
\end{aligned}$$

where $\hat{c} := K_1/(1 + K_2^2)$, in the second inequality we have used that for each $s > 0$,

$$\big\{\|\tilde{\boldsymbol{x}}\|_2^2 + q(\tilde{\boldsymbol{z}}) + \|\Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})\|_2^2 \geq s\big\} \subseteq \big\{q(\tilde{\boldsymbol{z}}) + (1 + K_2^2)\|\tilde{\boldsymbol{x}}\|_2^2 \geq s\big\},$$

(which follows from (5.17)), and in the last inequality we have used

$$q(\tilde{\boldsymbol{x}}) + q(\tilde{\boldsymbol{z}}) + K_1\|\tilde{\boldsymbol{x}}\|_2^2 \geq \min\{\hat{c}, 1\}\big(q(\tilde{\boldsymbol{z}}) + (1 + K_2^2)\|\tilde{\boldsymbol{x}}\|_2^2\big).$$

Thus, the linear gain in (5.25b) is $\bar{\alpha}_3(r) := c_1 \min\{\hat{c}, 1\} r$. Regarding (5.25a), we proceed similarly: $\hat{\mathrm{P}}$ and $\tilde{\mathrm{P}}_\beta$ are positive semidefinite with $\mathcal{N}(\hat{\mathrm{P}}) = \mathcal{N}(\tilde{\mathrm{P}}_\beta)$ by Proposition 5.3*(i)*, and hence $\tilde{w}^\top \tilde{\mathrm{P}}_\beta \tilde{w} \leq \bar{c}_2 \tilde{w}^\top \hat{\mathrm{P}} \tilde{w}$ for all $\tilde{w} \in \mathcal{D}_{\boldsymbol{x}^*}$, with $\bar{c}_2 := \lambda_{\max}(\tilde{\mathrm{P}}_\beta)/\lambda_{(2N-1)d}(\hat{\mathrm{P}})$. We then have

$$\begin{aligned}
\phi_{2,A}(s) &= \sup_{\{\|\tilde{\boldsymbol{x}}\|_2^2 + q(\tilde{\boldsymbol{z}}) + \|\Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})\|_2^2 \leq s\}} \mathrm{V}_{\tilde{\mathrm{P}}_\beta}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{z}}, \Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})) \\
&\leq \sup_{\{\|\tilde{\boldsymbol{x}}\|_2^2 + q(\tilde{\boldsymbol{z}}) + \|\Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})\|_2^2 \leq s\}} \bar{c}_2\big(\|\tilde{\boldsymbol{x}}\|_2^2 + q(\tilde{\boldsymbol{z}})\big) \\
&\leq \sup_{\{\|\tilde{\boldsymbol{x}}\|_2^2 + q(\tilde{\boldsymbol{z}}) + \|\Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})\|_2^2 \leq s\}} \bar{c}_2\big(\|\tilde{\boldsymbol{x}}\|_2^2 + q(\tilde{\boldsymbol{z}}) + \|\Delta\mathbf{F}_\epsilon(\tilde{\boldsymbol{x}})\|_2^2\big) = \bar{c}_2 s.
\end{aligned}$$

Thus, the linear gain in (5.25a) is $\bar{\alpha}_2(r) := \bar{c}_2 r$. □

5.5. Proof of Theorem 4.2. The combination of the above developments leads us here to the proof of Theorem 4.2.

Proof. [Proof of Theorem 4.2] By Proposition 5.3, the function V also satisfies (5.19) and (5.20). Additionally, from Proposition 5.4, for all $v \in \mathbb{R}^{2Nd}$ we have

$$\mathrm{V}(v) = \mathrm{V}_{\bar{\mathrm{P}}_\beta}(v - v^*, \mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}^*)) \leq \eta(\mathrm{W}_{\mathrm{Q}_\beta}(v - v^*, \mathbf{F}_\epsilon(\boldsymbol{x}) - \mathbf{F}_\epsilon(\boldsymbol{x}^*))) = \eta(\mathrm{W}(v)).$$

Therefore, V is a second moment NSS-Lyapunov function (as defined in Theorem 2.2) for the dynamics (4.2) with respect to the affine subspace

$$[\mathbb{1}^\top \otimes x_{\min}^\top,\, \boldsymbol{z}^{*\top}]^\top + \mathcal{N}(\hat{\mathrm{I}}) \,=\, [\mathbb{1}^\top \otimes x_{\min}^\top,\, \boldsymbol{z}^{*\top}]^\top + \operatorname{span}\big\{ \begin{bmatrix} 0 & (\mathbb{1} \otimes b)^\top \end{bmatrix}^\top : b \in \mathbb{R}^d \big\}.$$

Applying Theorem 2.2, we conclude that the dynamics (4.2) is second moment NSS stable with respect to the same affine subspace with

$$\mu(r,s) := \alpha_1^{-1}\big(2\tilde{\mu}(\alpha_2(r^p), s)\big) = \frac{2\lambda_{\max}(\mathrm{P}_\beta) r^2}{\lambda_{(2N-1)d}(\mathrm{P}_\beta)} \exp\big(-\tfrac{1}{2C_\eta}\, s\big),$$

$$\theta(r) := \alpha_1^{-1}\big(2\eta(2\sigma(r))\big) = \frac{4C_\eta \operatorname{trace}(\mathrm{P}_\beta)\, \kappa_2^2}{\lambda_{(2N-1)d}(\mathrm{P}_\beta)}\, r^2,$$

where κ_2 is such that (5.2) holds and C_η is defined in Proposition 5.4. ∎

6. Conclusions. We have considered a multi-agent network communicating over a weight-balanced, strongly connected digraph that seeks to collectively solve a convex optimization problem defined by a sum of local functions, one per agent, in the presence of noise both in the communication channels and in the agent computations. We have studied the robustness properties against additive persistent noise of a family of distributed continuous-time algorithms that have each agent update its estimate by following the gradient of its local cost function while, at the same time, seeking to agree with its neighbors' estimates via proportional-integral feedback on their disagreement. Specifically, we have established that the proposed class of algorithms is noise-to-state exponentially stable in second moment. Our strategy to establish this result has relied on constructing a function whose nullset is the solution to the optimization problem plus a direction of variance accumulation in some auxiliary variables, and then showing that in fact this is a NSS-Lyapunov function relying on the co-coercivity properties of the vector fields that define the dynamics. Future work will include the design of distributed procedures to determine the values of the design parameter for convergence and disturbance attenuation, relaxing the weight-balanced property of the directed communication topology, and extensions to scenarios with discrete-time communication, delays, and bandwidth limitations.

REFERENCES

[1] D. S. Bernstein, *Matrix Mathematics*, Princeton University Press, Princeton, NJ, 2005.

[2] D. P. Bertsekas, *Incremental gradient, subgradient, and proximal methods for convex optimization: a survey*, in Optimization for Machine Learning, S. Sra, S. Nowozin, and S. J. Wright, eds., MIT Press, Cambridge, MA, 2011, pp. 85–120.

[3] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, 1997.

[4] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks*, Applied Mathematics Series, Princeton University Press, 2009. Electronically available at http://coordinationbook.info.

[5] H. Deng and M. Krstić, *Output-feedback stabilization of stochastic nonlinear systems driven by noise of unknown covariance*, Systems & Control Letters, 39 (2000), pp. 173–182.
[6] J. C. Duchi, A. Agarwal, and M. J. Wainwright, *Dual averaging for distributed optimization: convergence analysis and network scaling*, IEEE Transactions on Automatic Control, 57 (2012), pp. 592–606.
[7] B. Gharesifard and J. Cortés, *Distributed continuous-time convex optimization on weight-balanced digraphs*, IEEE Transactions on Automatic Control, 59 (2014), pp. 781–786.
[8] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
[9] B. Johansson, M. Rabi, and M. Johansson, *A randomized incremental subgradient method for distributed optimization in networked systems*, SIAM Journal on Control and Optimization, 20 (2009), pp. 1157–1170.
[10] R. Khasminskii, *Stochastic Stability of Differential Equations*, vol. 66 of Stochastic Modelling and Applied Probability, Springer, 2012.
[11] S. S. Kia, J. Cortés, and S. Martínez, *Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication*, Automatica, (2014). Submitted.
[12] J. Lu and C. Y. Tang, *Zero-gradient-sum algorithms for distributed convex optimization: the continuous-time case*, IEEE Transactions on Automatic Control, 57 (2012), pp. 2348–2354.
[13] J. Lu, C. Y. Tang, P. R. Regier, and T. D. Bow, *Gossip algorithms for convex consensus optimization over networks*, IEEE Transactions on Automatic Control, 56 (2011), pp. 2917–2923.
[14] X. Mao, *Stochastic versions of the LaSalle theorem*, Journal of Differential Equations, 153 (1999), pp. 175–195.
[15] ———, *Stochastic Differential Equations and Applications*, Woodhead Publishing, 2nd ed., 2011.
[16] D. Mateos-Núñez and J. Cortés, *Noise-to-state stable distributed convex optimization on weight-balanced digraphs*, in IEEE Conf. on Decision and Control, Florence, Italy, 2013, pp. 2781–2786.
[17] ———, *pth moment noise-to-state stability of stochastic differential equations with persistent noise*, SIAM Journal on Control and Optimization, (2014). To appear.
[18] A. Nedic and A. Ozdaglar, *Distributed subgradient methods for multi-agent optimization*, IEEE Transactions on Automatic Control, 54 (2009), pp. 48–61.
[19] S. Nikookhoy, J. Lu, and C. Y. Tang, *Distributed convex optimization with identical constraints*, in IEEE Conf. on Decision and Control, Orlando, FL, Dec. 2011, pp. 2926–2931.
[20] B. Öksendal, *Stochastic Differential Equations - An Introduction with Applications*, Universitext, Springer-Verlag, 2010.
[21] M. Rabbat and R. Nowak, *Distributed optimization in sensor networks*, in Symposium on Information Processing of Sensor Networks, Berkeley, CA, Apr. 2004, pp. 20–27.
[22] S. S. Ram, A. Nedić, and V. V. Veeravalli, *Distributed stochastic subgradient projection algorithms for convex optimization*, Journal of Optimization Theory & Applications, 147 (2010), pp. 516–545.
[23] K. Srivastava and A. Nedić, *Distributed asynchronous constrained stochastic optimization*, IEEE Journal of Selected Topics in Signal Processing, 5 (2011), pp. 772–790.
[24] U. H. Thygesen, *A survey of Lyapunov techniques for stochastic differential equations*, Tech. Rep. 18-1997, Department of Mathematical Modeling, Technical University of Denmark, 1997.
[25] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, *Push-sum distributed dual averaging for convex optimization*, in IEEE Conf. on Decision and Control, Maui, HI, 2012, pp. 5453–5458.
[26] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Transactions on Automatic Control, 31 (1986), pp. 803–812.
[27] P. Wan and M. D. Lemmon, *Event-triggered distributed optimization in sensor networks*, in Symposium on Information Processing of Sensor Networks, San Francisco, CA, 2009, pp. 49–60.
[28] J. Wang and N. Elia, *Control approach to distributed optimization*, in Allerton Conf. on Communications, Control and Computing, Monticello, IL, Oct. 2010, pp. 557–561.
[29] E. Wei and A. Ozdaglar, *Distributed alternating direction method of multipliers*, in IEEE Conf. on Decision and Control, Maui, HI, 2012, pp. 5445–5450.
[30] F. Zanella, D. Varagnolo, A. Cenedese, G. Pillonetto, and L. Schenato, *Newton-Raphson consensus for distributed convex optimization*, in IEEE Conf. on Decision and Control, Orlando, Florida, Dec. 2011, pp. 5917–5922.
[31] D. L. Zhu and P. Marcotte, *Co-coercivity and its role in the convergence of iterative schemes*

for solving variational inequalities, SIAM Journal on Optimization, 6 (1996), pp. 714–726.
[32] M. Zhu and S. Martínez, *On distributed convex optimization under inequality and equality constraints*, IEEE Transactions on Automatic Control, 57 (2012), pp. 151–164.

Appendix. The following result concerning the function h defined by (4.4) is employed in the proof of Proposition 5.3.

Lemma A.1. *For $\delta > 0$, let $h(.,\delta) : (0,\infty) \to \mathbb{R}$ be defined by* (4.4) *and L be the Laplacian matrix of a strongly connected and weight-balanced digraph. Then, there exists $\hat{\beta} \equiv \hat{\beta}(\delta) > 0$ such that $h(\beta,\delta) < 0$ for all $\beta \in (0,\hat{\beta})$.*

Proof. Since the function $h(.,\delta)$ is continuous in the first argument, it is enough to show that the next two limits hold,

$$\lim_{\beta\to+\infty} h(\beta,\delta) = \infty, \quad \text{and} \quad \lim_{\beta\to 0^+} h(\beta,\delta) = 0^-,$$

to deduce the result from the by Bolzano Intermediate Value Theorem. Note that

$$-r + \sqrt{r^2-1} = \frac{(-r+\sqrt{r^2-1})(-r-\sqrt{r^2-1})}{-r-\sqrt{r^2-1}} = \frac{|r^2-1|-r^2}{r+\sqrt{r^2-1}},$$

which behaves asymptotically as $-\frac{1}{2r}$ when $r \to \infty$. Since $r := \frac{\beta^4+3\beta^2+2}{2\beta}$ goes to ∞ for both cases in which $\beta \to \infty$ or $\beta \to 0$, it follows that $-\frac{\beta^4+3\beta^2+2}{2\beta} + \sqrt{\big(\frac{\beta^4+3\beta^2+2}{2\beta}\big)^2 - 1}$ behaves as $-\frac{\beta}{\beta^4+3\beta^2+2}$ in both cases. Therefore,

$$\begin{aligned}\lim_{\beta\to+\infty} h(\beta,\delta) &= \lim_{\beta\to+\infty} \Big(-\tfrac{\beta}{\beta^4+3\beta^2+2}\lambda_2(\mathsf{L}+\mathsf{L}^\top) + \tfrac{\beta^2}{2\delta}\Big) \\ &= \lim_{\beta\to+\infty} \Big(-\tfrac{1}{\beta^3}\lambda_2(\mathsf{L}+\mathsf{L}^\top) + \tfrac{\beta^2}{2\delta}\Big) = \infty,\end{aligned}$$

and

$$\begin{aligned}&\lim_{\beta\to 0^+} h(\beta,\delta) = \lim_{\beta\to 0^+} \Big(-\tfrac{\beta}{\beta^4+3\beta^2+2}\lambda_2(\mathsf{L}+\mathsf{L}^\top) + \tfrac{\beta^2}{2\delta}\Big) \\ &= \lim_{\beta\to 0^+} \Big(-\tfrac{\beta}{2}\lambda_2(\mathsf{L}+\mathsf{L}^\top) + \tfrac{\beta^2}{2\delta}\Big) = \lim_{\beta\to 0^+} \beta\big(-\tfrac{\lambda_2(\mathsf{L}+\mathsf{L}^\top)}{2} + \beta\big) = 0^-,\end{aligned}$$

and the result follows. □