2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

A Control Perspective for Centralized and Distributed Convex Optimization

Jing Wang and Nicola Elia

Abstract— **In this paper, we want to study how natural and engineered systems could perform complex optimizations with limited computational and communication capabilities. We adopt a continuous-time dynamical system view rooted in early work on optimization and more recently in network protocol design, and merge it with the dynamic view of distributed averaging systems. We obtain a general approach, based on the control system viewpoint, that allows to analyze and design (distributed) optimization systems converging to the solution of given convex optimization problems. The control system viewpoint provides many insights and new directions of research. We apply the framework to a distributed optimal location problem and demonstrate the natural tracking and adaptation capabilities of the system to changing constraints.**

I. Introduction

In recent years, there has been a renewed research interest towards finding efficient algorithms for solving convex optimization problems in parallel or distributed fashion [1]. This trend has been mainly motivated by the explosion in size and complexity of data-sets used in statistical machine learning [2] and applications in modern communication networks [3] and other applications in networked systems.

However, new applications are emerging where the discrete-time algorithmic paradigm may no longer be appropriate. These are situations where the computational capability is elementary, severely limited, and distributed among many nodes physically separated and connected over (possibly noisy) networks, or where we want to analyze how natural, social, and biological systems could collectively optimize. In this paper, we depart from the classical algorithmic view and focus on the control system view. We turn back to early work on optimization done by economists [4] and unfortunately neglected by most current optimization textbooks, which shows that there is a very natural continuous-time optimization system associated with the Lagrangian of the convex optimization problem at hand. Such dynamical system is guaranteed to converge to an optimal solution under mild conditions and can be studied as a feedback control system. The control system viewpoint provides many insights and new directions of research. In particular, we can now study disturbance rejection properties of optimization systems, the robustness to parameter and model uncertainty, and discuss tracking and adaptation capabilities. We can also envision designing controllers for optimization systems.

We demonstrate that the proposed approach can unveil the natural distributed structure in the optimization problem by applying it to a basic Network Utility Maximization problem, [5]–[8]. We then turn our attention to problems where the distributed structure needs to be imposed as these problems need to be solved over networks, [9]-[14]. We show that the Laplacian of the available network interconnection, assumed to be undirected and strongly connected, can be used to guarantee convergence. Expanding on our previous work [10], we show that, differently from most available approaches, the network needs to be used twice. This guarantees also some good (communication) noise rejection properties of the optimization system. The result can be seen as the natural extension of [15] from dynamic distributed computation of averages to dynamic distributed computation of optimal solutions to convex optimization problems. This approach contrasts with existing algorithmic methods which often use dominated convex mixing with diminishing step-size, or primal-dual methods where the primal step requires each agent to solving an optimization problem often constrained. Finally, we extend the framework to more general distributed optimization problems and illustrate it by an application to distributed location.

II. Centralized Problems with Equality Constraints

Consider the following constrained optimization problem.

$$\begin{aligned} p^* \quad &= \quad \min_{x\in\mathbb{R}^n} f(x) \\ s.t. \quad &Ax = b \end{aligned} \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. To provide the main idea and to simplify the derivations, in this section, we assume that A has full row rank, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is strictly convex and differentiable, and that Problem (1) is feasible and has a finite optimal cost $-\infty < p^* < \infty$. We use x^* to denote the optimal solution to problem (1).

From the classical Lagrange multiplier theory, we construct the Lagrangian function $F(x,\nu) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ as

$$F(x,\nu) = f(x) + \nu^T(Ax - b)$$

We consider the following dual problem associated with the primal Problem (1).

$$d^* = \max_{\nu\in\mathbb{R}^m} \min_{x\in\mathbb{R}^n} F(x,\nu) \tag{2}$$

Under current assumptions, $F(x,\nu)$ is convex in x and concave in ν, and $p^* = d^*$, i.e., the duality gap is zero.

To solve the problem, we consider the following dynamical system which has been investigated in [4].

$$\begin{aligned} \dot{x} \quad &= \quad -\nabla_x F(x,\nu) \\ \dot{\nu} \quad &= \quad \nabla_\nu F(x,\nu) \end{aligned} \tag{3}$$

This research has been supported by NSF under grant number ECS-0901846. The authors are with the Dept. of Electrical and Computer engineering, Iowa State University, Ames, IA 50010. Emails : {criswang,nelia}@iastate.edu

where ∇_x and ∇_ν represent the gradients with respect to x and ν respectively. Substituting $F(x,\nu)$ in the above equations, we have

$$\begin{array}{rcl}\dot{x} &=& -\nabla_x f(x) - A^T\nu \\ \dot{\nu} &=& Ax - b\end{array} \tag{4}$$

It is important to note that the above dynamical system always converges to the optimal solution x^* under the current assumptions.

Theorem 2.1: Consider system (4), for any initial values of x and ν, we have $\lim_{t\to\infty} x(t) = x^*$.

Here, we present a modern version of the proof in [4], which is not a readily available reference.

Proof. The equilibrium points of the system are the solutions to the equations

$$\begin{array}{rcl}0 &=& -\nabla_x f(\bar{x}) - A^T\bar{\nu} \\ 0 &=& A\bar{x} - b\end{array} \tag{5}$$

In other words, the equilibrium points satisfy the KKT conditions(see, e.g., [16], Chap. 5). Under the current assumptions we know that there is only one set of vectors (x^*, ν^*) which satisfies the KKT conditions, namely the optimal primal and dual solutions. Thus let $(\bar{x}, \bar{\nu}) = (x^*, \nu^*)$.

Next, we study the stability property of the equilibrium point. Let $\tilde{x} = x - x^*$ and $\tilde{\nu} = \nu - \nu^*$. In the new variables, the differential equations (4) can be written as follows

$$\begin{array}{rcl}\dot{\tilde{x}} &=& -\nabla_x f(x) + \nabla_x f(x^*) - A^T\tilde{\nu} \\ \dot{\tilde{\nu}} &=& A\tilde{x}\end{array} \tag{6}$$

Consider the quadratic candidate Lyapunov Function $V(\tilde{x},\tilde{\nu}) = \frac{1}{2}\tilde{x}^T\tilde{x} + \frac{1}{2}\tilde{\nu}^T\tilde{\nu}$. Then

$$\dot{V}(\tilde{x},\tilde{\nu}) = \tilde{x}^T\dot{\tilde{x}} + \tilde{\nu}^T\dot{\tilde{\nu}} = -\tilde{x}^T\nabla_x f(x) + \tilde{x}^T\nabla_x f(x^*)$$

From the global under-estimator property of the gradient [16], we know that

$$f(y) \geq f(x) + \nabla_x^T f(x)(y - x), \quad \forall x, y \in dom\, f$$

In particular, the inequality is strict when $y \neq x$ if f is strictly convex.

Therefore, $\forall x \neq x^*$,

$$f(x) - f(x^*) > \nabla_x^T f(x^*)\tilde{x}, \quad f(x^*) - f(x) > -\nabla_x^T f(x)\tilde{x}.$$

Adding both side of the equations, we have that

$$\dot{V}(\tilde{x},\tilde{\nu}) = -\nabla_x^T f(x)\tilde{x} + \nabla_x^T f(x^*)\tilde{x} < 0, \quad \forall x \neq x^*, (\tilde{x} \neq 0)$$

However, $\dot{V}(\tilde{x},\tilde{\nu}) = 0$ on the set $\mathcal{E} = \{\tilde{x},\tilde{\nu} \,|\, \tilde{x} = 0\}$. To shows that the equilibrium point (x^*, ν^*) is globally asymptotically stable, we need to invoke LaSalle invariant principle and show that there are no trajectories of (6) in $\mathcal{E}$ besides the equilibrium point $\tilde{x} = 0, \tilde{\nu} = 0$.

When $\tilde{x} = 0$, we have that (6) reduces to

$$\begin{array}{rcl}\dot{\tilde{x}} &=& -A^T\tilde{\nu} \\ \dot{\tilde{\nu}} &=& 0.\end{array}$$

It follows that under the standard assumption that A has full row rank, the only solution $\mathcal{E}$ is indeed $(0, 0)$, which completes the proof. ■

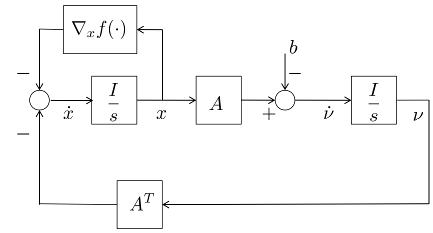


Fig. 1. The block diagram of system (4).

For a proof requiring twice differentiable cost and using a different Lyapunov function see [7]. The result is quite interesting as it links the global under-estimator property of the gradient of a convex function with a dynamical property of convex functions as dissipators. When we will apply this result to distributed optimization problems, it will allow to simple interconnected agents to solve complex optimization problems with only local gradient sensing capabilities.

A. Control Perspective

It is instructive to consider the block diagram of the dynamical system (4). Figure 1 shows the block diagram describing the interconnection among the various components. This has allowed us to focus on the feedback system and its properties. It shows fairly apparent that the dynamical systems for convex optimization is subject to the fundamental limitation of feedback. Note that how the vector b part of the constraint can be interpreted as an input command or disturbance to the dynamical system, as shown later in Section VI-B. This naturally bears the question of dynamic performance of the system like tracking and rejection. Note that b can change over time and correspondingly the system can adapt the convergence point leading to the possibility of real-time adaptive optimization.

For the large class of quadratic programming problems the cost function $f(x) = x^T Px + q^T x + r$ where $P > 0$ is positive definite. Then, $\nabla_x f(x) = Px + q$. In this case, the system is Linear Time Invariant, and both b and q represent exogenous inputs. The scheme can then be analyzed and improved using the vast array of advanced analysis and design methodologies from control theory. This may require us to move away from the general purpose optimization algorithms, which work for most problems, and will allow to design special purpose algorithms for specific applications, especially in the cases of network distributed optimization problems.

III. Problems with Inequality Constraints

In this section, we consider problem (1) with additional inequality constraints defined by convex functions. We use logarithmic barrier to handle the inequality constraints although other approaches are possible. Consider the problem

$$\begin{array}{rcl}p^* &=& \min\limits_{x\in\mathbb{R}^n} f(x) \\ & s.t. & Ax = b \\ && f_i(x) \leq 0, \quad i = 1, \ldots, p\end{array} \tag{7}$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is a convex function. We assume that there is an x which is strictly feasible and the slater's condition is satisfied [16], i.e., an x that satisfies the equality and all strict inequality constraints, so the duality gap is zero.

Problem (7) can be approximated by the following one.

$$\begin{aligned} p^*(\gamma) &= \min_{x\in\mathbb{R}^n} f(x) - \frac{1}{\gamma}\sum_{i=1}^{p}\log(-f_i(x)) \\ s.t. &\quad Ax = b \end{aligned} \tag{8}$$

which reduces to the previous case. It is well known that $p^*(\gamma) \to p^*$ as $\gamma \to \infty$ as $p^*(\gamma) - p^* < p/\gamma$. Since the inequality constraints are replaced by the barrier cost function, the block diagram of Problem (8) is similar to that of Figure 1. The inequality constraints affect the gradient feedback loop in this case.

When $f_i(x)$ is linear in x, i.e., $f_i(x) = C_i^T x$, we can use slack variables to obtain an equivalent form of Problem (8). Namely:

$$\begin{aligned} p^* &= \min_{x\in\mathbb{R}^n} f(x) \\ s.t. &\quad Ax = b \\ &\quad C_i^T x + z_i = 0, \quad i = 1,\dots,p \\ &\quad z_i \geq 0 \end{aligned} \tag{9}$$

Then, the barrier function approximation for this problem is the following one.

$$\begin{aligned} \gamma p^*(\gamma) &= \min_{x\in\mathbb{R}^n} \gamma f(x) - \sum_{i=1}^{p}\log(z_i) \\ s.t. &\quad Ax = b \\ &\quad C_i^T x + z_i = 0, \quad i = 1,\dots,p \end{aligned} \tag{10}$$

which has the same optimal solution, x_γ^*, for $\gamma > 0$.

Let $f_0(x,z) = \gamma f(x) - \sum_{i=1}^{p}\log(z_i)$. Then, $\nabla_x f_0(x) = \gamma\nabla_x f(x)$ and $\nabla_{z_i} f_0(x,z) = -\frac{1}{z_i}$. The new dynamical system is then given, following (3), by

$$\begin{aligned} \dot{x} &= -\nabla_x F(x,z,\nu,\lambda) &&= -\gamma\nabla_x f(x) - A^T\nu - C^T\lambda \\ \dot{z} &= -\nabla_z F(x,z,\nu,\lambda) &&= \mathbf{vec}\left[\frac{1}{z_i}\right] - \lambda \\ \dot{\nu} &= \nabla_\nu F(x,z,\nu,\lambda) &&= Ax - b \\ \dot{\lambda} &= \nabla_\lambda F(x,z,\nu,\lambda) &&= Cx + z \end{aligned} \tag{11}$$

We have the following result.

Theorem 3.1: Consider system (11) with $\gamma > 0$, for all initial conditions $z(0) > 0$, $x(0)$, $\lim_{t\to\infty} x(t) = x_\gamma^*$.

Note that λ_i and z_i are coupled together for each i. Moreover, the bank of integrators and the other diagonal blocks that indicate that the variable updates are mostly decoupled. The coupling can only come from the cost and the constraints. This observation directly affects the derivation of distributed solutions to certain optimization problems, as we will see next.

IV. Naturally Distributed Problems

The optimization system approach can unveil the fundamentally distributed nature of certain problems.

A. Network Utility Maximization

As an example, we consider a basic network utility maximization problem, which has attracted a lot interest from networking community, see, e.g., [5]–[8]. The problem is to find the optimal allocation of transmission rate which could be deliverable by the network. This translates into affine inequality constraints, subject to positive variables. Namely:

$$\begin{aligned} p^* &= \max_{x\in\mathbb{R}^n} \sum_{j=1}^{N} f_j(x_j) \\ s.t. &\quad Cx \leq d \\ &\quad x_j \geq 0 \end{aligned} \tag{12}$$

where $f_j(x_j)$s are strictly concave functions.

Because, $f(x) = \sum_{j=1}^{N} f_j(x_j)$, $\nabla_x f(x)$ is block diagonal, where the size of each block is equal to the size of each variable x_j. C is a matrix of zeros and ones, where the ones represent feasible relay nodes on a viable path. The nodes is assumed to have information about the congestion on the path, or in other words, the i^{th} node has access to the i^{th} row of $Cx - d$.

Employing the log barrier function with minor rearrangement, we set to solve the approximation

$$\begin{aligned} \gamma p^*(\gamma) = -\min_{x\in\mathbb{R}^n} \sum_{j=1}^{N}(-\gamma f_j(x_j) - \log(x_j)) - \sum_{i=1}^{M}\log(z_i) \\ s.t.\ Cx + z - d = 0 \end{aligned}$$

and applying (3), we have the dynamical systems solver for problem (12) as

$$\begin{aligned} \dot{x}_j &= \gamma\nabla_x f_j(x_j) + \tfrac{1}{x_j} - [C^T\lambda]_j, \ j = 1,\dots,N \\ \dot{z}_i &= \tfrac{1}{z_i} - \lambda_i, \qquad\qquad j = 1,\dots,M \\ \dot{\lambda}_i &= [Cx]_i + z_i - d_i \end{aligned} \tag{13}$$

The reader should notice how the updates of x_j and z_i, and λ_i depends only on local information. However, feedback from the nodes on the path is needed to compute $[C^T\lambda]_j$. Finally, we see the emergent natural separation between access control and congestion control protocols. Further note that we do not require each node to solve an optimization problem at each step, as done in primal-dual algorithms e.g., [8], but only to move along a favorable direction. This last point may have important implications, as we see that no much intelligence or computational capabilities are required at each node.

V. Network Distributed Convex Optimization

The key feature of the problem in the previous section is in the agent's utility function, which only depends on the agent's local variables, i.e., $f_j(x_j)$. In this section we consider problems where the agent's utility function depends on a global variable, i.e. x. Of course, this is a generalization of the previous case as the utility function may depend on a smaller subset of the variables x. Thus, we consider the following problem:

$$\begin{array}{rcl} p^* & = & \min_{x\in\mathbb{R}^n} \sum_{j=1}^{N} f_j(x) \\ & s.t. & Ax = b \\ & & Cx \leq d \end{array} \tag{14}$$

The above problem without or with more general constraints has been considered in [9]–[14] where most discrete time algorithms adopt a local convex mixing and vanishing step size on local gradient searching with exception of [10] that uses constant step size in certain situations. Here, we extend the computation model of [10] to solve constrained problem (14) using the framework developed before.

Assumption 5.1: We make the following assumptions.

1) Each agent knows (or is subject to) a subset of the constraints $A_j x = b_j$ and $C_j x \leq d_j$;
2) The union of the agents' constraints is given by $Ax = b$ and $Cx \leq d$;
3) The sets of equality constraints for each agent may overlap;
4) The agents/nodes are connected over a strongly connected communication network, with graph Laplacian L, assumed to be symmetric.

Related to the above optimization problem are the following local problems

$$\begin{array}{rcl} p_j^* & = & \min_{x\in\mathbb{R}^n} f_j(x^j) \\ & s.t. & A_j x^j = b_j \\ & & C_j x^j \leq d_j \end{array} \tag{15}$$

where each agent solves a local optimization problem based on the local constraints, and x^j denotes the agent's variables. A key distinctive feature of our approach is that we use the available network graph Laplacian to implicitly impose the agreement among the various local x^js. It is not difficult to verify the following formulation is equivalent to the original Problem (14) under the current assumptions.

$$\begin{array}{rcl} p_d^* & = & \min_{x\in\mathbb{R}^n} \sum_{j=1}^{N} f_j(x^j) \\ & s.t. & A_j x^j = b_j, \quad j = 1, \ldots, N \\ & & C_j x^j + z^j = d_j, \quad j = 1, \ldots, N \\ & & \sum_{j=1}^{N} L_{:,j} x^j = 0 \\ & & z^j \geq 0 \quad j = 1, \ldots, N \end{array} \tag{16}$$

where $z^j \geq 0$ is the slack variable, $L_{:,j}$ is the j column of L. (with appropriate Kronecker product when x^j has dimension larger than 1). Note that the last constraint requires $x^j = x^i$ for all $i, j \in \{1 \ldots N\}$. In the literature various other approaches, based on the Method of Multipliers (MoM) [1], have been proposed to impose these last constraints. The approach in here, although related to MoM, is more natural when a network interconnection is already in place and is completely distributed as it does not require a global network collector connected to all nodes as in [2].

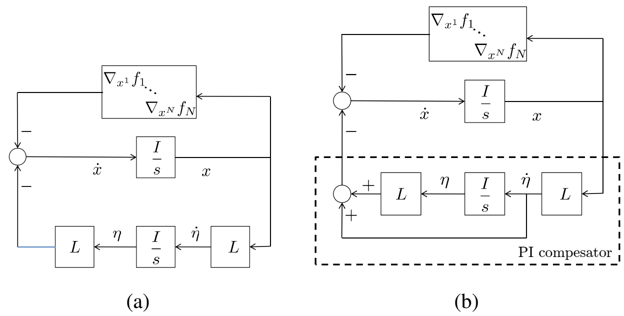


Fig. 2. Augmented Lagrangian corresponds to a PI (instead of an Integrator) feedback controller action

The dynamic equations of the approximate distributed solver are the following ones.

$$\begin{array}{rcl} \dot{x}^j & = & -\gamma \nabla_{x^j} f_j(x^j) - A_j^T \nu^j - C_j^T \lambda^j - \sum_{i=1}^{N} L_{ij}^T \eta^i \\ \dot{z}^j & = & \mathbf{vec}\left[\frac{1}{z_i^j}\right] - \lambda^j \\ \dot{\nu}^j & = & A_j x^j - b_j \\ \dot{\lambda}^j & = & C_j x^j + z^j - d_j \\ \dot{\eta}^j & = & \sum_{i=1}^{N} L_{ij} x^i \end{array} \tag{17}$$

Note that the only coupling in the dynamical equations comes from the existing equality constraints, through A, and the network L. System (17) may be better suited to describe optimization behaviors in biological and social systems, and it can be seen as the natural generalization of the dynamic consensus system of [15].

Theorem 5.2: Consider system (17) with $\gamma > 0$, for all initial conditions $z^j(0) = z^j > 0$, $x^j(0) = x^j$, $j = 1, \ldots, N$, $\lim_{t\to\infty} x^j(t) = x^*_\gamma$.

A. Augmented Lagrangian and Control Interpretation

In the special case of unconstrained optimization, Problem (16) reduces to

$$\begin{array}{rcl} p_d^* & = & \min_{x\in\mathbb{R}^n} \sum_{j=1}^{N} f_j(x^j) \\ & & \sum_{j=1}^{N} L_{:,j} x^j = 0 \end{array} \tag{18}$$

Therefore the associated dynamical system is

$$\begin{array}{rcl} \dot{x} & = & \mathbf{vec}\left[\nabla_{x^j} f_j(x^j)\right] - L\eta \\ \dot{\eta} & = & Lx \end{array} \tag{19}$$

where we have used $L = L^T$. Figure 2(a) shows the resulting block diagram. In [10], the following dynamical system was proposed to solve Problem (18) instead.

$$\begin{array}{rcl} \dot{x} & = & \mathbf{vec}\left[\nabla_{x^j} f_j(x^j)\right] - Lx - L\eta \\ \dot{\eta} & = & Lx \end{array} \tag{20}$$

These equations were derived from a different approach than the one proposed in this paper. It is interesting to note that the extra term, namely Lx in the first equation is equal to $\dot{\eta}$ from the second equation. Therefore, the dynamical system (20) includes a proportional action. Figure 2(b) shows the resulting scheme where the Proportional Integral compensation is highlighted.

We next show that system (20) can be the result of the approach proposed in this paper. Consider the following optimization problem

$$p^*_{au} \;=\; \min_{x\in\mathbb{R}^n} \sum_{j=1}^{N} f_j(x^j) + \frac{1}{2}x^T L x \qquad (21)$$
$$\sum_{j=1}^{N} L_{:,j} x^j = 0$$

Clearly $p^*_{au} = p^*_d$ since the constraint force $Lx = 0$, thus the added extra cost is zero if the problem is feasible. This technique is known as augmented lagrangian method, see, e.g., [17]. It is easy to verify that the dynamical system for solving the above problem is (20). From a control system prospective, having a PI compensator instead of just an integrator in the loop, leads to improved stability/convergence properties. The augmented lagrangian method is known for its better convergence properties. We now see that these properties are justified from the feedback control system prospective.

Remark 5.3: 1) The augmented scheme, by introducing extra damping, allows us to relax the assumption that the f_i be strictly convex to just convex. This property extends the approach to solving distributed linear programs.
2) The PI compensation associated with the augmented scheme generalizes to constrained problems (15). This leads to the question of finding more general and powerful controllers and points to a new research direction of controller design for (distributed) optimization systems.
3) In contrast with the classical algorithmic methods, it may now be possible to uncover these newly derived networked structures with local gradient sensing (or others like them) in biological networks and realize that some systems are indeed cooperatively optimizing.

VI. More General Distributed Problems

In this section we consider problems of the following form. For simplicity we restrict our attention to unconstrained optimization problems. Let $\mathcal{J} = \{1, \ldots, M\}$, and $x_j \in \mathbb{R}^k$, $j \in \mathcal{J}$. Also, let $\mathcal{J}_i \subset \mathcal{J}$, for $i = 1, \ldots N$ and $y_i = \mathbf{vec}[x_j]_{j\in\mathcal{J}_i}$. Consider

$$p^* \;=\; \min_{x_j\in\mathbb{R}^k, j\in\mathcal{J}} \sum_{i=1}^{N} f_i(y_i). \qquad (22)$$

In this case, each f_i may be a function of a subset of the x_js and for $i \neq k$, f_i and f_n may share a subset of variables. When $\mathcal{J}_i = \mathcal{J}$ of all $i = 1, \ldots N$, we have recover problems of the form (14), while when $\mathcal{J}_i = \{i\}$, we recover formulations like (12) (albeit the constraints).

Since x_j may appear in several f_i's, Let $\mathcal{I}_j = \{i \,|\, f_i \text{ is function of } x_j\}$, and let p_j be the cardinality of $\mathcal{I}_j$.

We then let each node/agent to be associated to one f_i. The agent is in charge of updating its variables y_i. If f_i is function of x_j for some j, then we denote the estimate of node i of x_j as x^i_j. Then all the agents in $\mathcal{I}_j$ will have local estimates of x_j, which they need to reconcile through communication.

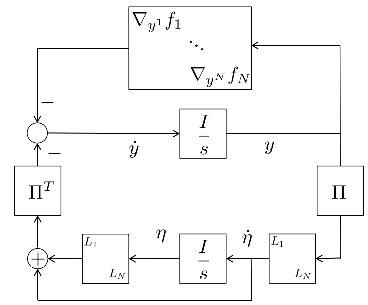


Fig. 3. Block diagram of the optimization system (24). The LTI network distributed controller is in feedback with the gradient system.

We describe such network by $\mathcal{L}_j$, the Laplacian of a strongly connected and undirected graph. If $p_j \geq 2$, $\mathcal{L}_j$ has size p_j, is symmetric and has only one eigenvalue at zero and associate eigenvector $\mathbf{1}_{p_j}$. If $p_j = 1$, indicating the variable is local to only one of the f_i's, then $\mathcal{L}_j = []$. To keep the exposition simple and without loss of generality, we assume that no variable is local to only one function.

Let $L_j = \mathcal{L}_j \otimes \mathbf{1}_k$. Based on L_j we define $L = Diag\,[L_j]_{j\in\mathcal{J}}$. Let $z_j = \mathbf{vec}[x^i_j]_{i\in\mathcal{I}_j}$ and $z = \mathbf{vec}[z_j]_{j\in\mathcal{J}}$. Finally, we introduce a permutation operator mapping $y = \mathbf{vec}\,[y_i]_{i\in\mathcal{I}}$ to z. $z = \Pi y$. Consider the following optimization problem equivalent to (22)

$$p^* \;=\; \min_{x_j\in\mathbb{R}^k, j\in\mathcal{J}} \sum_{i=1}^{N} f_i(y_i) + \frac{1}{2} y^T \Pi^T L \Pi y \qquad (23)$$
$$s.t. \quad L\Pi y = 0$$

Then we have the following dynamical system described in Figure 3.

$$\begin{aligned} \dot{y} &= -\Pi^T L \Pi y - \Pi^T L \eta - \mathbf{vec}\,[\nabla_{y_i} f_i(y_i)]_{i=1}^N \\ \dot{\eta} &= L\Pi y. \end{aligned} \qquad (24)$$

which converges to the optimal solution of Problem (22). Note how the resulting block diagram is a generalization of that of Figure 2(b), and reduces to it when $L_j = L$ for all $j = 1, \ldots N$.

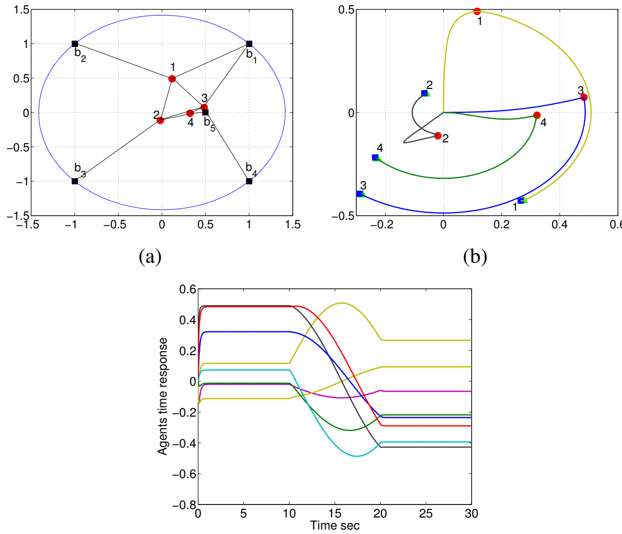
A. Network Distributed Placement and Location

The above derivation directly applies to the optimal location and placement problems [16]. Consider

$$p^* \;=\; \min_{x_j\in\mathbb{R}^k, j\in\mathcal{J}} \sum_{(i,j)\in\mathcal{A}} f_{ij}(x_i, x_j) \qquad (25)$$
$$s.t. \quad x_\ell = b_\ell,\; \ell \in \mathcal{F} \subset \mathcal{J}$$

where b_ℓs are fixed locations (anchors), $\mathcal{A}$ is the set of all links in a graph (see Figure 4(a)) and $f_{ij} : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ is associated with the length of the arc (i, j). When $k = 2$ or 3, x_i represents the location of node i. $\mathcal{F}$ is a subset of the nodes representing those nodes that are at fixed locations. The anchors could represent targets to be monitored. We will later allow the constraints (target locations) to change.

Assume we have 4 mobile nodes in the plane $x_1, \ldots, x_4$, and 5 anchors at $b_1 \ldots, b_5$. Let the length of each arc in the



(a) (b) (c)

Fig. 4. An example of distributed placement/location problem.

graph be the squared Euclidian norm. Define

$$\begin{aligned} g_1(x_1,x_2) &= \|x_1-b_1\|_2^2+\|x_1-b_2\|_2^2+\|x_1-x_2\|_2^2 \\ g_2(x_2,x_3,x_4) &= \|x_2-b_3\|_2^2+\|x_2-x_3\|_2^2+\|x_2-x_4\|_2^2 \\ g_3(x_3,x_1) &= \|x_3-b_1\|_2^2+\|x_3-b_4\|_2^2+\|x_3-x_1\|_2^2 \\ g_4(x_4,x_3) &= \|x_4-b_5\|_2^2+\|x_4-x_3\|_2^2. \end{aligned}$$

This association, consistent with the graph, respects the information available to each mobile node. We then want to minimize $g_1+g_2+g_3+g_4$. In this case, we have $N=M=4$, $y_1=[x_1^1,x_2^1]^T$, $y_2=[x_2^2,x_3^2,x_4^2]^T$, $y_3=[x_3^3,x_1^3]^T$, $y_4=[x_4^4,x_3^4]^T$, and $\mathcal{I}_1=\{1,3\}$, $\mathcal{I}_2=\{1,2\}$, $\mathcal{I}_3=\{2,3,4\}$, $\mathcal{I}_4=\{2,4\}$. We have chosen the Laplacian for each j to be $\mathcal{L}_1=\mathcal{L}_2=\mathcal{L}_4=circul[1,-1]$ and $\mathcal{L}_3=circul[2,-1,-1]$ where $circul$ denotes a circulant matrix. Figure 4(a) shows the location of the anchors (squares) and the optimal placement of the four nodes circles.

B. Tracking changing constraints

Since the optimization is a least square problem, the optimization system is LTI with the b_is as inputs coming from constant term of the gradient. Therefore, the inputs can be changed over time and the system is expected to track, within its bandwidth, the set of optimal trajectories, as the anchors now move.

Figure 4(b) shows the traces of the four agents. The circles represent the optimal locations corresponding to those of Figure 4(a) obtained when the anchors are kept fixed.

After the first 10 seconds, the anchors are all rotated by 135 degrees at constant speed for another 10 seconds. The triangles represent the locations of the agents after 20 seconds. Finally, the squares represent the final location of the agents after the movement of the anchors is stopped for other 10 seconds. Note how the triangles and the squares are close to each others denoting the good tracking properties in this case. The time evolution of the locations is shown in Figure 4(c).

VII. Conclusions

The ideas and results presented in this paper provide a theory for agents with simple dynamics and local gradient sensing abilities to collaboratively solve complicated convex optimization problems through simple nearest neighbor interactions. We hope they could help the discovery of naturally emerging optimization systems in biological, social, and economical systems and the engineering of new cooperative networked systems.

References

[1] D.P. Bertsekas and J. N. Tsitsiklis "Parallel and Distributed Computation: Numerical Methods" Prentice-Hall, 1989.
[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, 3(1), pp. 1-122, 2011.
[3] M. Chiang, S. H. Low, A. R. Calderbank and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proc. of the IEEE*, vol. 95, pp. 255-312, Jan. 2007.
[4] K. J. Arrow, L. Hurwicz and H. Uzawa, *Studies in Linear and Nonlinear Programming,* Stanford University Press, Stanford, 1958.
[5] F. P. Kelly, Am Maulloo, and D. Tan, "Rate control in communication networks: Shadow prices, proportional fairness and stability," Journal of the Operational Research Society, vol. 49, pp. 237-252, 1998.
[6] S. Kunniyur and R. Srikant, "End-to-end congestion control schemes: utility functions, random losses and ECN marks," *IEEE/ACM transaction on networking*, vol. 11, no. 5, pp. 689-702, 2003.
[7] D. Feijer and F. Paganini,"Stability of primal-dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46. no. 12, pp.1974-1981, 2010.
[8] S. H. Low and D. E. Lapsley,"Optimal flow control, I:Basic algorithm and convergence," Vol. 7, no. 6, pp. 861-874, 1999.
[9] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Automat. Contr.*, vol. 54, Jan. 2009.
[10] J. Wang and N. Elia, "Control approach to distributed optimization," in 48-*th Annual Allerton Conference*, pp. 557-561, 2010.
[11] I. Lobel, A. Ozdaglar, and D. Feijer, "Distributed Multi-agent Optimization with State-Dependent Communication," LIDS report 2834, submitted for publication, 2010.
[12] J. C. Duchi, A. Agarwal and M. J. Weinwright, "Dual Averaging for Distributed Optimization: Convergence Analysis and Network Scaling," available on line: `http://arxiv.org/abs/1005.2012`.
[13] S.S. Ram, A. Nedic, and V. V. Veeravalli, "Distributed Stochastic Subgradient Projection Algorithms for Convex Optimization," to appear in *Journal of Optimization Theory and Applications*, 2010.
[14] M. Zhu and S. Martinez, "On distributed optimization under inequality and equality constraints via penalty primal-dual methods," in *Proc. of 2010 American Contr. Conf.*, pp. 2434-2439, June 30-July 2, 2010.
[15] R. Olfati-Saber and R. M. Murray, "Consensus problems in network of agents with switching topology and time-delays," *IEEE Trans. Automat. Contr.*, vol. 49, no. 9, pp. 1520-1533, Sep., 2004.
[16] S. Boyd, and L. Vandenberghe, *Convex Optimization*. New York: Cambrige Univ. Press, 2004.
[17] M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization and Applications,* vol. 4, pp. 302-320, 1969.