Convergence of Min-Sum Message Passing for Quadratic Optimization

Ciamac C. Moallemi, *Member, IEEE,* and Benjamin Van Roy, *Member, IEEE*

***Abstract*—We establish the convergence of the min-sum message passing algorithm for minimization of a quadratic objective function given a convex decomposition. Our results also apply to the equivalent problem of the convergence of Gaussian belief propagation.**

***Index Terms*—message-passing algorithms, decentralized optimization**

I. INTRODUCTION

CONSIDER an optimization problem that is characterized by a set $\mathcal{X}$ and a hypergraph $(V, \mathcal{C})$. There are $|V|$ decision variables; each is associated with a vertex $i \in V$ and takes values in a set $\mathcal{X}$. The set $\mathcal{C}$ is a collection of subsets (or, "hyperedges") of the vertex set V; each hyperedge $C \in \mathcal{C}$ is associated with a real-valued "component function" (or, "factor") $f_C: \mathcal{X}^C \to \mathbb{R}$. The optimization problem takes the form

$$\min_{x \in \mathcal{X}^{|V|}} f(x),$$

where

$$f(x) = \sum_{C \in \mathcal{C}} f_C(x_C).$$

Here, $x_C \in \mathcal{X}^{|C|}$ is the vector of variables associated with vertices in the subset C. We refer to an optimization program of this form as a *graphical model*. While this formulation may seem overly broad—indeed, almost any optimization problem can be cast in this framework—we are implicitly assuming that the graph is sparse and that the hyperedges are small.

Over the past few years, there has been significant interest in a heuristic optimization algorithm for graphical models. We will call this algorithm the min-sum message passing algorithm, or the min-sum algorithm, for short. This is equivalent to the so-called max-product algorithm, also known as belief revision, and is closely related to the sum-product algorithm, also known as belief propagation. Interest in such algorithms has to a large extent been triggered by the success of message passing algorithms for decoding low-density parity-check codes and turbo codes [1], [2], [3]. Message passing algorithms are now used routinely to solve NP-hard decoding problems in communication systems. It was a surprise that this simple and efficient approach offers sufficing solutions.

C.C. Moallemi is with the Graduate School of Business, Columbia University, New York, NY 10025 USA e-mail: ciamac@gsb.columbia.edu.

B. Van Roy is with the Departments of Management Science & Engineering and Electrical Engineering, Stanford University, Stanford, CA 94305 USA e-mail: bvr@stanford.edu.

Manuscript received March 14, 2006; revised December 11, 2008.

The majority of literature has been focused on the case where the set $\mathcal{X}$ is discrete and the resulting optimization problem is combinatorial in nature. We, however, are interested in the the case where $\mathcal{X} = \mathbb{R}$ and the optimization problem is continuous. In particular, many continuous optimization problems that are traditionally approached using methods of linear programming, convex programming, etc. also possess graphical structure, with objectives defined by sums of component functions. We believe the min-sum algorithm leverages this graphical structure in a way that can complement traditional optimization algorithms, and that combining strengths will lead to algorithms that are able to scale to larger instances of linear and convex programs.

One continuous case that has been considered in the literature is that of pairwise quadratic graphical models. Here, the objective function is a positive definite quadratic function

$$f(x) = \frac{1}{2} x^\top \Gamma x - h^\top x, \quad \Gamma \succ 0. \tag{1}$$

This function is decomposed in a pairwise fashion according to an undirected graph (V, E), so that

$$f(x) = \sum_{i \in V} f_i(x_i) + \sum_{(i,j) \in E} f_{ij}(x_i, x_j),$$

where the functions $\{f_i(\cdot), f_{ij}(\cdot, \cdot)\}$ are quadratic. It has been shown that, if the min-sum algorithm converges, it computes the global minimum of the quadratic [4], [5], [6]. The question of convergence, however, has proved difficult. Sufficient conditions for convergence have been established [4], [5], but these conditions are abstract and difficult to verify. Convergence has also been established for classes of quadratic programs arising in certain applications [7], [8].

In recent work, Johnson, et al. [9], [10] have introduced the notion of *walk-summability* for pairwise quadratic graphical models. They establish convergence of the min-sum algorithm for walk-summable pairwise quadratic graphical models when the particular set of component functions

$$f_{ij}(x_i, x_j) = \Gamma_{ij} x_i x_j, \quad \forall\ (i,j) \in E, \tag{2}$$

is employed by the algorithm and the algorithm is initialized with zero-valued messages. Further, they give examples outside this class for which the min-sum algorithm does not converge.

Note that there may be many ways to decompose a given objective function into component functions. The min-sum algorithm takes the specification of component functions as an input and exhibits different behavior for different decompositions of the same objective function. Alternatively, the choice of a decomposition can be seen to be equivalent to the choice of initial conditions for the min-sum algorithm [6], [11]. A limitation of the convergence result of Johnson, et al. [9], [10] is that it requires use of a particular decomposition of the objective function of the form (2) and with zero-valued initial messages. The analysis presented does not hold in other situation. For example, the result does not establish convergence of the min-sum algorithm in the applied context considered in [7].

We will study the convergence of the min-sum algorithm given a convex decomposition:

Definition 1. (Convex Decomposition)
A convex decomposition of a quadratic function $f(\cdot)$ *is a set of quadratic functions* $\{f_i(\cdot), f_{ij}(\cdot)\}$ *such that*

$$f(x) = \sum_{i \in V} f_i(x_i) + \sum_{(i,j) \in E} f_{ij}(x_i, x_j),$$

each function $f_i(\cdot)$ *is strictly convex, and each function* $f_{ij}(\cdot, \cdot)$ *is convex (although not necessarily strictly so).*

We will say that a quadratic objective function is *convex decomposable* if there exists a convex decomposition. This condition implies strict convexity of the quadratic objective function, however, not all strictly convex, quadratic functions are convex decomposable.

The primary contribution of this paper is in establishing that the min-sum algorithm converges given *any* convex decomposition or even decompositions that are in some sense "dominated" by convex decompositions. This result can be equivalently restated as a sufficient condition on the initial messages used in the min-sum algorithm. Convergence is established under both synchronous and asynchronous models of computation. We believe that this is the most general convergence result available for the min-sum algorithm with a quadratic objective function.

The walk-summability condition of Johnson, et al. is equivalent to the existence of a convex decomposition [10]. In this way, our work can be viewed as a generalization of their convergence results to a broad class of decompositions or initial conditions. This generalization is of more than purely theoretical interest. The decentralized and asynchronous settings in which such optimization algorithms are deployed are typically dynamic. Consider, for example, a sensor network which seeks to estimate some environmental phenomena by solving an optimization problem of the form (1). As sensors are added or removed from the network, the objective function in (1) will change slightly. Reinitializing the optimization algorithm after each such change would require synchronization across the entire network and a large delay to allow the algorithm to converge. If the change in the objective function is small, it is likely that the change in the optimum of the optimization problem is small also. Hence, using the current state of the algorithm (the set of messages) as an initial condition may result in much quicker convergence. In this way, understanding the robustness of the min-sum algorithm over different initial conditions is important to assessing it's practical value.

Beyond this, however, our work suggests path towards understanding the convergence of the min-sum algorithm in the context of general convex (i.e., not necessarily quadratic) objective functions. The notion of a convex decomposition is easily generalized, while it is not clear how to interpret the walk-summability condition or a decomposition of the form (2) in the general convex case. In follow-on work [12], we have been able to establish such a generalization and develop conditions for the convergence of the min-sum algorithm in a broad range of general convex optimization problems. When specialized to the quadratic case, however, those results are not as general as the results presented herein.

The optimization of quadratic graphical models can be stated as a problem of inference in Gaussian graphical models. In this case, the min-sum algorithm is mathematically equivalent to sum-product algorithm (belief propagation), or the max-product algorithm. Our results therefore also apply to Gaussian belief propagation. However, since Gaussian belief propagation, in general, computes marginal distributions that have correct means but incorrect variances, we believe that the optimization perspective is more appropriate than the inference perspective. As such, we state our results in the language of optimization.

Finally, note that solution of quadratic programs of the form (1) is equivalent to the solution of the sparse, symmetric, positive definite linear system $\Gamma x = h$. This is a well-studied problem with an extensive literature. The important feature of the min-sum algorithm in this context is that it is decentralized and totally asynchronous. The comparable algorithms from the literature fall into the class of classical iterative methods, such as the Jacobi method or the Gauss-Seidel method [13]. In an optimization context, these methods can be interpreted as local search algorithms, such as gradient descent or coordinate descent. While these methods are quite robust, they suffer from a notoriously slow rate of convergence. Our hope is that message-passing algorithms will provide faster decentralized solutions to such problems than methods based on local search. In application contexts where a comparison can be made [7], preliminary results show that this may indeed be the case.

II. The Min-Sum Algorithm

Consider a connected undirected graph with vertices $V = \{1, \ldots, n\}$ and edges E. Let $N(i)$ denote the set of neighbors of a vertex i. Consider an objective function $f : \mathbb{R}^n \to \mathbb{R}$ that decomposes according to pairwise cliques of (V, E); that is

$$f(x) = \sum_{i \in V} f_i(x_i) + \sum_{(i,j) \in E} f_{ij}(x_i, x_j). \tag{3}$$

The min-sum algorithm attempts to minimize $f(\cdot)$ by an iterative, message passing procedure. In particular, at time t, each vertex i keeps track of a "message" from each neighbor $u \in N(i)$. This message takes the form of a function $J^{(t)}_{u\to i} : \mathbb{R} \to \mathbb{R}$. These incoming messages are combined to compute new outgoing messages for each neighbor. In particular, the message $J^{(t+1)}_{i\to j}(\cdot)$ from vertex i to vertex $j \in N(i)$ evolves according to

$$J_{i\to j}^{(t+1)}(x_j) = \kappa + \min_{y_i}\left(f_i(y_i) + f_{ij}(y_i,x_j) + \sum_{u\in N(i)\setminus j} J_{u\to i}^{(t)}(y_i)\right). \quad (4)$$

Here, κ represents an arbitrary offset term that varies from message to message. Only the relative values of the function $J_{i\to j}^{(t+1)}(\cdot)$ matter, so κ does not influence relevant information. Its purpose is to keep messages finite. One approach is to select κ so that $J_{i\to j}^{(t+1)}(0) = 0$. The functions $\{J_{i\to j}^{(0)}(\cdot)\}$ are initialized arbitrarily; a common choice is to set $J_{i\to j}^{(0)}(\cdot) = 0$ for all messages.

At time t, each vertex j forms a local objective function $f_j^{(t)}(\cdot)$ by combining incoming messages according to

$$f_j^{(t)}(x_j) = \kappa + f_j(x_j) + \sum_{i\in N(j)} J_{i\to j}^{(t)}(x_j).$$

The vertex then generates a running estimate of the jth component of an optimal solution to the original problem according to

$$x_j^{(t)} = \operatorname*{argmin}_{y_j} f_j^{(t)}(y_j).$$

By dynamic programming arguments, it is easy to see that this procedure converges and is exact given a convex decomposition when the graph (V,E) is a tree. We are interested in the case where the graph has arbitrary topology.

A. Reparameterizations

An alternative way to view iterates of the min-sum algorithm is as a series of "reparameterizations" of the objective function $f(\cdot)$ [6], [11]. Each reparameterization corresponds to a different decomposition of the objective function. In particular, at each time t, we define a function $f_j^{(t)} : \mathbb{R}\to\mathbb{R}$, for each vertex $j\in V$, and a function $f_{ij}^{(t)} : \mathbb{R}^2\to\mathbb{R}$, for each edge $(i,j)\in E$, so that

$$f(x) = \sum_{i\in V} f_i^{(t)}(x_i) + \sum_{(i,j)\in E} f_{ij}^{(t)}(x_i,x_j).$$

The functions evolve jointly according to

$$\begin{aligned} f_i^{(t+1)}(x_i) &= \kappa + f_i^{(t)}(x_i) + \sum_{j\in N(i)} \min_{x_j}\left(f_j^{(t)}(x_j) + f_{ij}^{(t)}(x_i,x_j)\right),\\ f_{ij}^{(t+1)}(x_i,x_j) &= \kappa + f_{ij}^{(t)}(x_i,x_j) \\ &\quad - \min_{y_i}\left(f_i^{(t)}(y_i) + f_{ij}^{(t)}(y_i,x_j)\right) \\ &\quad - \min_{y_j}\left(f_j^{(t)}(y_j) + f_{ij}^{(t)}(x_i,y_j)\right). \end{aligned} \quad (5)$$

They are initialized at time $t=0$ according to

$$f_i^{(0)}(x_i) = \kappa + f_i(x_i) + \sum_{j\in N(i)} J_{j\to i}^{(0)}(x_i),$$

$$f_{ij}^{(0)}(x_i,x_j) = \kappa + f_{ij}(x_i,x_j) - J_{j\to i}^{(0)}(x_i) - J_{i\to j}^{(0)}(x_j).$$

In the common case, where the functions $\{J_{i\to j}^{(0)}(\cdot)\}$ are all set to zero, the initial component functions $\{f_i^{(0)}(\cdot), f_{ij}^{(0)}(\cdot,\cdot)\}$ are identical to $\{f_i(\cdot), f_{ij}(\cdot,\cdot)\}$, modulo constant offsets. A running estimate of the jth component of an optimal solution to the original problem is generated according to

$$x_j^{(t)} = \operatorname*{argmin}_{y_j} f_j^{(t)}(y_j). \quad (6)$$

The message passing interpretation and the reparameterization interpretation can be related by

$$f_j^{(t)}(x_j) = \kappa + f_j(x_j) + \sum_{i\in N(j)} J_{i\to j}^{(t)}(x_j),$$

$$f_{ij}^{(t)}(x_i,x_j) = \kappa + f_{ij}(x_i,x_j) - J_{j\to i}^{(t)}(x_i) - J_{i\to j}^{(t)}(x_j),$$

$$J_{i\to j}^{(t+1)}(x_j) = \kappa + J_{i\to j}^{(0)}(x_j) + \sum_{s=0}^{t} \min_{y_i}\left(f_i^{(s)}(y_i) + f_{ij}^{(s)}(y_i,x_j)\right).$$

These relations are easily established by induction on t. As they indicate, the message passing interpretation and the reparameterization interpretation are completely equivalent in the sense that convergence of one implies convergence of the other, and that they compute the same estimates of an optimal solution to the original optimization problem.

Reparameterizations are more convenient for our purposes for the following reason: Note that the decomposition (3) of the objective $f(\cdot)$ is not unique. Indeed, many alternate factorizations can be obtained by moving mass between the single vertex functions $\{f_i(\cdot)\}$ and the pairwise functions $\{f_{ij}(\cdot,\cdot)\}$. Since the message passing update (4) depends on the factorization, this would seem to suggest that the each choice of factorization results in a different algorithm. However, in the reparameterization interpretation, the choice of factorization only enters via the initial conditions. Moreover, it is clear that the choice of factorization is equivalent to the initial choice of messages $\{J_{i\to j}^{(0)}(\cdot)\}$. Our results will identify sufficient conditions on these choices so that the min-sum algorithm converges.

III. The Quadratic Case

We are concerned with the case where the objective function f is quadratic, i.e.

$$f(x) = \frac{1}{2}x^\top \Gamma x - h^\top x.$$

Here, $\Gamma\in\mathbb{R}^{n\times n}$ is a symmetric, positive definite matrix and $h\in\mathbb{R}^n$ is a vector. Since f must decompose relative to the graph (V,E) according to (3), we must have the non-diagonal entries satisfy $\Gamma_{ij}=0$ if $(i,j)\notin E$. Without loss of generality, we will assume that $\Gamma_{ij}\neq 0$ for all $(i,j)\in E$ (otherwise, each such edge (i,j) can be deleted from the graph) and that $\Gamma_{ii}=1$ for all $i\in V$ (otherwise, the variables can be rescaled so that this is true).

Let $\vec{E}\subset V\times V$ be the set of directed edges. That is, $(i,j)\in E$ iff $\{i,j\}\in\vec{E}$ and $(i,j)\in E$ iff $\{j,i\}\in\vec{E}$. (We use braces and parentheses to distinguish directed and undirected edges, respectively.) Quadratic component functions

$\{f_i(\cdot), f_{ij}(\cdot)\}$ that sum to $f(\cdot)$ can be parameterized by two vectors of parameters, $\gamma = (\gamma_{ij}) \in \mathbb{R}^{|\vec{E}|}$ and $z = (z_{ij}) \in \mathbb{R}^{|\vec{E}|}$, according to

$$f_{ij}(x_i, x_j) = \frac{1}{2}\left(\gamma_{ji}\Gamma_{ij}^2 x_i^2 + 2\Gamma_{ij}x_i x_j + \gamma_{ij}\Gamma_{ij}^2 x_j^2\right) - z_{ji}x_i - z_{ij}x_j,$$

$$f_j(x_j) = \frac{1}{2}\left(1 - \sum_{i\in N(j)} \Gamma_{ij}^2 \gamma_{ij}\right) x_j^2 - \left(h_j - \sum_{i\in N(j)} z_{ij}\right) x_j.$$

Given such a representation, we will refer to the components of γ as the quadratic parameters and the components of z as the linear parameters.

Iterates $\{f_i^{(t)}(\cdot), f_{ij}^{(t)}(\cdot,\cdot)\}$ of the min-sum algorithm can be represented by quadratic parameters $\gamma^{(t)}$ and linear parameters $z^{(t)}$. By explicit computation of the minimizations involved in the reparameterization update (5), we can rewrite the update equations in terms of the parameters $\gamma^{(t)}$ and $z^{(t)}$. In particular, if $\sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}^{(t)} < 1$, then

$$\gamma_{ij}^{(t+1)} = \frac{1}{1 - \sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}^{(t)}}, \tag{7}$$

$$z_{ij}^{(t+1)} = \frac{\Gamma_{ij}}{1 - \sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}^{(t)}} \left(h_i - \sum_{u\in N(i)\setminus j} z_{ui}^{(t)}\right). \tag{8}$$

If, on the other hand, $\sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}^{(t)} \geq 1$, then the minimization

$$\min_{y_i} f_i^{(t)}(y_i) + f_{ij}^{(t)}(y_i, x_j)$$

is unbounded and the update equation is ill-posed. Further, the estimate of the jth component of the optimal solution, defined by (6), becomes

$$x_j^{(t)} = \frac{1}{1 - \sum_{i\in N(j)} \Gamma_{ij}^2 \gamma_{ij}^{(t)}} \left(h_j - \sum_{i\in N(j)} z_{ij}^{(t)}\right), \tag{9}$$

when $\sum_{i\in N(j)} \Gamma_{ij}^2 \gamma_{ij}^{(t)} < 1$, and is ill-posed otherwise.

We define a generalization to the notion of a convex decomposition.

Definition 2. (Convex-Dominated Decomposition) *A convex-dominated decomposition of a quadratic function $f(\cdot)$ is a set of quadratic functions $\{f_i(\cdot), f_{ij}(\cdot,\cdot)\}$ that form a decomposition of $f(\cdot)$, such that for some convex decomposition $\{g_i(\cdot), g_{ij}(\cdot,\cdot)\}$,*

$$g_{ij}(x_i, x_j) - f_{ij}(x_i, x_j)$$

is convex, for all edges $(i,j) \in E$.

Note that any convex decomposition is also convex-dominated.

The following theorem is the main result of this paper.

Theorem 1. (Quadratic Min-Sum Convergence) *If $f(\cdot)$ is convex decomposable and $\{f_i^{(0)}(\cdot), f_{ij}^{(0)}(\cdot,\cdot)\}$ is a convex-dominated decomposition, then the quadratic parameters $\gamma^{(t)}$, the linear parameters $z^{(t)}$, and the running estimates $x^{(t)}$ converge. Moreover,*

$$\lim_{t\to\infty} f(x^{(t)}) = \min_x f(x).$$

This result is more general than required to capture the "typical" situation. In particular, consider a situation where a problem formulation gives rise to component functions $\{f_i(\cdot), f_{ij}(\cdot)\}$ that form a convex decomposition of an objective function f. Then, initialize the min-sum algorithm with $\{f_i^{(0)}(\cdot), f_{ij}^{(0)}(\cdot,\cdot)\} = \{f_i(\cdot), f_{ij}(\cdot,\cdot)\}$. Since the initial iterate is a convex decomposition, it certifies that $f(\cdot)$ is convex decomposable, and it is also a convex-dominated decomposition.

We will prove Theorem 1 in Section VI. Before doing so, we will study the parameter sequences $\gamma^{(t)}$ and $z^{(t)}$ independently.

IV. Convergence of Quadratic Parameters

The update (7) for the the quadratic parameters $\gamma^{(t)}$ does not depend on the linear parameters $z^{(t)}$. Hence, it is natural to study their evolution independently, as in [5], [7]. In this section, we establish existence and uniqueness of a fixed point of the update (7). Further, we characterize initial conditions under which $\gamma^{(t)}$ converges to this fixed point.

Whether or not a decomposition is convex depends on quadratic parameters but not the linear ones. Let $\mathcal{V}$ be the set of quadratic parameters $\gamma \in \mathbb{R}^{|\vec{E}|}$ that correspond to convex decompositions.

We have the following theorem establishing convergence for the quadratic parameters. The proof relies on certain monotonicity properties of the update (7), and extends the method developed in [5], [7].

Theorem 2. (Quadratic Parameter Convergence) *Assume that $f(\cdot)$ is convex decomposable. The system of equations*

$$\gamma_{ij} = \frac{1}{1 - \sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}}, \quad \forall\ \{i,j\} \in \vec{E},$$

has a solution γ^* such that

$$\mathbf{0} < \gamma^* < v, \quad \forall\ v \in \mathcal{V}.$$

Moreover, γ^* is the unique such solution.

If we initialize the min-sum algorithm so that $\gamma^{(0)} \leq v$, for some $v \in \mathcal{V}$, then $0 < \gamma^{(t)} < v$, for all $t > 0$, and

$$\lim_{t\to\infty} \gamma^{(t)} = \gamma^*.$$

Proof: See Appendix A. ■

The key condition for the convergence is that the initial quadratic parameters $\gamma^{(0)}$ must be dominated by those of a convex decomposition. Such initial conditions are easy to find, for example $\gamma^{(0)} = \mathbf{0}$ or $\gamma^{(0)} \in \mathcal{V}$ satisfy this requirement.

Note that we should not expect the algorithm to converge for arbitrary $\gamma^{(0)}$. For the update (7) to even be well-defined at time t, we require that

$$\sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}^{(t)} < 1, \quad \forall\ \{i,j\} \in \vec{E}.$$

The condition on $\gamma^{(0)}$ in Theorem 2 guarantees this at time $t = 0$, and the theorem guarantees that it continue to hold for all $t > 0$. Similarly, the computation (9) of the estimate $x^{(t)}$ requires that

$$\sum_{i \in N(j)} \Gamma_{ij}^2 \gamma_{ij}^{(t)} < 1, \quad \forall\, j \in V.$$

The theorem guarantees that this is true for all $t \geq 0$, given suitable choice of $\gamma^{(0)}$.

V. Convergence of Linear Parameters

In this section, we will assume that the quadratic parameters $\gamma^{(t)}$ are set to the fixed point γ^*, and study the evolution of the linear parameters $z^{(t)}$. In this case, the update (8) for the linear parameters takes the particularly simple form

$$z_{ij}^{(t+1)} = \gamma_{ij}^* \Gamma_{ij} \left(h_i - \sum_{u \in N(i) \setminus j} z_{ui}^{(t)} \right).$$

This linear equation can be written in vector form as

$$z^{(t+1)} = -Dy + Az^{(t)},$$

where $y \in \mathbb{R}^{|\vec{E}|}$ is a vector with

$$y_{ij} = h_i, \tag{10}$$

$D \in \mathbb{R}^{|\vec{E} \times \vec{E}|}$ is a diagonal matrix with

$$D_{ij,ij} = -\gamma_{ij}^* \Gamma_{ij}, \tag{11}$$

and $A \in \mathbb{R}^{|\vec{E} \times \vec{E}|}$ is a matrix such that

$$A_{ij,uk} = \begin{cases} -\gamma_{ij}^* \Gamma_{ij} & \text{if } (u,i),(i,j) \in E,\ k = i,\ j \neq u, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

If the spectral radius of A is less than 1, then we have convergence of $z^{(t)}$ independent of the initial condition $z^{(0)}$ by

$$\lim_{t \to \infty} z^{(t)} = -\sum_{t=0}^{\infty} A^t D y.$$

We will show that existence of a convex decomposition of $f(\cdot)$ is a sufficient condition for this to be true. In order to proceed, we first introduce the notion of walk-summability.

A. Walk-Summability

Note that the optimization problem we are considering,

$$\min_x \frac{1}{2} x^\top \Gamma x - h^\top x,$$

has the unique solution

$$x^* = \Gamma^{-1} h.$$

Define $R = I - \Gamma$, so $R_{ii} = 0$ and $R_{ij} = -\Gamma_{ij}$, if $i \neq j$. If we assume that the matrix R has spectral radius less than 1, we can express the solution x^* by the infinite series

$$x^* = \sum_{t=0}^{\infty} R^t h. \tag{13}$$

The idea of walk-sums, introduced by Johnson, et al. [9], allows us to interpret this solution as a sum of weights of walks on the graph.

To be precise, define a *walk* of length k to be a sequence of vertices

$$w = \{w_0, \ldots, w_k\},$$

such that $(w_i, w_{i+1}) \in E$, for all $0 \leq i < k$. Given a walk w, we can define a weight by the product

$$\rho(w) = R_{w_0 w_1} \cdots R_{w_{|w|-1} w_{|w|}}.$$

(We adopt the convention that $\rho(w) = 1$ for walks of length 0, which consist of a single vertex.) Given a set of walks $\mathcal{W}$, we define the weight of the set to be the sum of the weights of the walks in the set, that is

$$\rho(\mathcal{W}) = \sum_{w \in \mathcal{W}} \rho(w).$$

Define $\mathcal{W}_{i \to j}$ to be the (infinite) set of all walks from vertex i to vertex j. If the quantity $\rho(\mathcal{W}_{i \to j})$ was well-defined, examining the structure of R and (13), we would have

$$x_j^* = \sum_{i \in V} \rho(\mathcal{W}_{i \to j}) h_i. \tag{14}$$

Definition 3. (Walk-Summability) *Given a matrix $\Gamma \succ 0$ with $\Gamma_{ii} = 1$, define $|R|$ by $|R|_{ij} = |[I - \Gamma]_{ij}|$. We say Γ is walk-summable if the spectral radius of $|R|$ is less than 1.*

Walk-summability of Γ guarantees the the function $\rho(\cdot)$ is well-defined even for infinite sets of walks, since in this case, the series $\sum_{t=0}^{\infty} R^t$ is absolutely convergent. It is not difficult to see that existence of a convex decomposition of $f(\cdot)$ implies walk-summability [9]. More recent work [10] shows that these two conditions are in fact equivalent.

We introduce a different weight function $\nu(\cdot)$ defined by

$$\nu(w) = \gamma_{w_0 w_1}^* R_{w_0 w_1} \cdots \gamma_{w_{|w|-1} w_{|w|}}^* R_{w_{|w|-1} w_{|w|}}.$$

$\nu(\cdot)$ can be extends to sets of walks as before. However, we interpret this function only over *non-backtracking* walks, where a walk w is non-backtracking if $w_{i-1} \neq w_{i+1}$, for $1 \leq i < |w|$. Denote by $\mathcal{W}^{nb}$ the set of non-backtracking walks. The following combinatorial lemma establishes a correspondence between $\nu(\cdot)$ on non-backtracking walks and $\rho(\cdot)$.

Lemma 1. *Assume that $f(\cdot)$ is convex decomposable. For each $w \in \mathcal{W}^{nb}$, there exists a set of walks $\mathcal{W}_w$, all terminating at the same vertex as w, such that*

$$\nu(w) = \rho(\mathcal{W}_w).$$

Further, if $w' \in \mathcal{W}^{nb}$ and $w' \neq w$, then $\mathcal{W}_w$ and $\mathcal{W}_{w'}$ are disjoint.

Proof: See Appendix B. ∎

The above lemma reveals that $\nu(\cdot)$ is well-defined on infinite sets of non-backtracking walks. Indeed, if $\mathcal{W} \subset \mathcal{W}^{nb}$,

$$\sum_{w \in \mathcal{W}} |\nu(w)| = \sum_{w \in \mathcal{W}} |\rho(\mathcal{W}_w)| \leq \sum_{w \in \mathcal{W}} \sum_{u \in \mathcal{W}_w} |\rho(u)|, \tag{15}$$

and the latter sum is finite since Γ is walk-summable.

We can make the correspondence between $\nu(\cdot)$ and $\rho(\cdot)$ stronger with the following lemma.

Lemma 2. *Assume that $f(\cdot)$ is convex decomposable. If we define $\mathcal{W}^{nb}_{i\to r}$ to be the set of all non-backtracking walks from vertex i to vertex r, we have*

$$\rho(\mathcal{W}_{i\to r}) = \frac{\nu(\mathcal{W}^{nb}_{i\to r})}{1-\sum_{u\in N(r)} R^2_{ur}\gamma^*_{ur}}.$$

Proof: See Appendix B. ■

B. Spectral Radius of A

Examining the structure of the matrix A from (12), it is clear that if $\mathcal{W}^{nb,t}_{uk\to ij}$ is defined to be the set of all length t non-backtracking walks w with $\{w_0, w_1\} = \{u,k\}$ and $\{w_{|w|-1}, w_{|w|}\} = \{i,j\}$, then

$$[A^t D]_{ij,uk} = \nu(\mathcal{W}^{nb,t+1}_{uk\to ij}).$$

Thus, if $\mathcal{W}^{nb,1+}_{uk\to ij}$ is the set of all non-backtracking walks w of length at least 1 satisfying $\{w_0, w_1\} = \{u,k\}$ and $\{w_{|w|-1}, w_{|w|}\} = \{i,j\}$,

$$\begin{aligned}\sum_{t=0}^{\infty}[A^t D]_{ij,uk} &= \sum_{t=0}^{\infty}\nu(\mathcal{W}^{nb,t+1}_{uk\to ij}) = \nu(\mathcal{W}^{nb,1+}_{uk\to ij}) \\ &= \sum_{w\in\mathcal{W}^{nb,1+}_{uk\to ij}} \nu(w).\end{aligned}$$

Lemma 1 and (15) assure us that the later sum must be absolutely convergent. Then, we have established the following lemma.

Lemma 3. *Assume that $f(\cdot)$ is convex decomposable. The spectral radius of $|A|$ is less than 1.*

C. Exactness

From Lemma 3, we have

$$z^{(\infty)} = \lim_{t\to\infty} z^{(t)} = -\sum_{t=0}^{\infty} A^t D y.$$

For each vertex j, define the quantity

$$\bar{\Gamma}_j = \frac{1}{1-\sum_{i\in N(j)}\Gamma^2_{ij}\gamma^*_{ij}}.$$

In this case, the estimate $x^{(t)}_j$ for each vertex j, defined by (6), converges to

$$\begin{aligned}x^{(\infty)}_j &= \bar{\Gamma}_j\left(h_j - z^{(\infty)}\right) \\ &= \bar{\Gamma}_j\left(h_j + \sum_{i\in N(j)}\sum_{t=0}^{\infty}[A^t D y]_{ij}\right) \\ &= \bar{\Gamma}_j\left(h_j + \sum_{i\in N(j)}\sum_{\{u,k\}\in\vec{E}}\nu(\mathcal{W}^{nb,1+}_{uk\to ij})h_u\right) \\ &= \bar{\Gamma}_j\left(h_j + \sum_{u\in V}\nu(\mathcal{W}^{nb,1+}_{u\to j})h_u\right).\end{aligned}$$

Here, we define $\mathcal{W}^{nb,1+}_{u\to j}$ is the set of non-backtracking walks of length at least 1 starting at u and ending at j. Note that if $u\neq j$, then a non-backtracking walk from u to j must have length at least 1. Thus,

$$\nu(\mathcal{W}^{nb}_{u\to j}) = \nu(\mathcal{W}^{nb,1+}_{u\to j}).$$

If $u=j$, there is a single non-backtracking walk of length 0 from j to j, namely $w=\{j\}$, and $\nu(w)=1$. Thus,

$$\nu(\mathcal{W}^{nb}_{u\to j}) = 1 + \nu(\mathcal{W}^{nb,1+}_{u\to j}).$$

Hence,

$$x^{(\infty)}_j = \frac{1}{1-\sum_{i\in N(j)}\Gamma^2_{ij}\gamma^*_{ij}}\sum_{u\in V}\nu(\mathcal{W}^{nb}_{u\to j})h_u.$$

Comparing with Lemma 2, and (14), we have

$$x^{(\infty)}_j = \sum_{u\in V}\rho(\mathcal{W}_{u\to j})h_u = x^*_j.$$

Thus, $x^{(\infty)} = x^*$.

Putting together the results in this section, we have the following theorem.

Theorem 3. (Linear Parameter Convergence) *Assume that $f(\cdot)$ is convex decomposable and that $\gamma^{(0)} = \gamma^*$. Then, for arbitrary initial conditions $z^{(0)}$, the linear parameters $z^{(t)}$ converge. Further, the corresponding estimates $x^{(t)}$ converge to the global optimum x^*.*

VI. Overall Convergence

In Section IV, we established the convergence of the quadratic parameters $\gamma^{(t)}$. In Section V, we established the convergence of the linear parameters $z^{(t)}$ assuming the quadratic parameters were set to their fixed point. Here, we will combine these results in order to prove Theorem 1, which establishes convergence of the full min-sum algorithm, where the linear parameters evolve jointly with the quadratic parameters.

It suffices to establish convergence of the linear parameters $z^{(t)}$. Define the matrix $A^{(t)}\in\mathbb{R}^{|\vec{E}\times\vec{E}|}$ by

$$A^{(t)}_{ij,uk} = \begin{cases}-\gamma^{(t+1)}_{ij}\Gamma_{ij} & \text{if } (u,i),(i,j)\in E,\ k=i,\ j\neq u,\\ 0 & \text{otherwise.}\end{cases}$$

Define the diagonal matrix $D^{(t)}\in\mathbb{R}^{|\vec{E}\times\vec{E}|}$ by $D^{(t)}_{ij,ij} = -\gamma^{(t+1)}_{ij}\Gamma_{ij}$. Then, the min-sum update (8) becomes

$$z^{(t+1)} = -D^{(t)}y + A^{(t)}z^{(t)},$$

where y is defined by (10). From Theorem 2, it is clear that $A^{(t)}\to A$ and $D^{(t)}\to D$ (where A and D are defined by (12) and (11), respectively).

From Lemma 3, the spectral radius of $|A|$ is less than 1. Hence, there is a vector norm $\|\cdot\|$ on $\mathbb{R}^{|\vec{E}|}$ and a corresponding induced operator norm such that $\|A\|<\alpha$, for some $\alpha<1$ [14]. Pick K_1 sufficiently large so that $\|A^{(t)}\|<\alpha$ for all $t\geq K_1$. Then, the series

$$\sum_{s=0}^{\infty}\left(A^{(t)}\right)^s$$

converges for $t \geq K_1$. Set

$$w^{(t)} = -\sum_{s=0}^{\infty} \left(A^{(t)}\right)^s D^{(t)} y = -(I - A^{(t)})^{-1} D^{(t)} y,$$

$$z^{(\infty)} = -\sum_{s=0}^{\infty} A^s D y = -(I - A)^{-1} D y.$$

Then, for $t \geq K_1$,

$$\begin{aligned}\|z^{(t+1)} - z^{(\infty)}\| &\leq \|A^{(t)}(z^{(t)} - w^{(t)})\| + \|z^{(\infty)} - w^{(t)}\| \\ &\leq \alpha \|z^{(t)} - w^{(t)}\| + \|z^{(\infty)} - w^{(t)}\| \\ &\leq \alpha \|z^{(t)} - z^{(\infty)}\| + (1+\alpha)\|z^{(\infty)} - w^{(t)}\|.\end{aligned}$$

Since $w^{(t)} \to z^{(\infty)}$, for any $\epsilon > 0$ we can pick $K_2 \geq K_1$ so that if $t > K_2$, $\|w^{(t)} - z^{(\infty)}\| < \epsilon$. Then, for $t > K_2$,

$$\|z^{(t+1)} - z^{(\infty)}\| < \alpha \|z^{(t)} - z^{(\infty)}\| + (1+\alpha)\epsilon.$$

Repeating over t,

$$\|z^{(t)} - z^{(\infty)}\| < \alpha^{t-K_2} \|z^{(K_2)} - z^{(\infty)}\| + \frac{1+\alpha}{1-\alpha}\epsilon.$$

Thus,

$$\limsup_{t\to\infty} \|z^{(t)} - z^{(\infty)}\| \leq \frac{1+\alpha}{1-\alpha}\epsilon.$$

Since ϵ is arbitrary, it is clear that $z^{(t)}$ converges to $z^{(\infty)}$.

The fact that $x^{(t)}$ converges to x^* follows from the same argument as in Theorem 3.

A. Asynchronous Convergence

The work we have presented thus far considers the convergence of a synchronous variation of the min-sum algorithm. In that case, every component of each of the parameter vectors $\gamma^{(t)}$ and $z^{(t)}$ is update at every time step. However, the min-sum algorithm has a naturally parallel nature and can be applied in distributed contexts. In such implementations, different processors may be responsible for updating different components of the parameter vector. Further, these processors may not be able to communicate at every time step, and thus may have insufficient information to update the corresponding components of the parameter vectors. There may not even be a notion of a shared clock. As such, it is useful to consider the convergence properties of the min-sum algorithm under an *asynchronous* model of computation.

In such a model, we assume that a processor associated with vertex i is responsible for updating the parameters $\gamma_{ij}^{(t)}$ and $z_{ij}^{(t)}$ for each neighbor $j \in N(i)$. We define the T^i to be the set of times at which these parameters are updated. We define $0 \leq \tau_{ji}(t) \leq t$ to be the last time the processor at vertex j communicated to the processor at vertex i. Then, the parameters evolve according to

$$\gamma_{ij}^{(t+1)} = \begin{cases} \frac{1}{1-\sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}^{(\tau_{ui}(t))}} & \text{if } t \in T^i, \\ \gamma_{ij}^{(t)} & \text{otherwise,} \end{cases}$$

$$z_{ij}^{(t+1)} = \begin{cases} \frac{\Gamma_{ij}\left(h_i - \sum_{u\in N(i)\setminus j} z_{ui}^{(\tau_{ui}(t))}\right)}{1-\sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}^{(\tau_{ui}(t))}} & \text{if } t \in T^i, \\ z_{ij}^{(t)} & \text{otherwise,} \end{cases}$$

Note that the processor at vertex i is not computing its updates with the most recent values of the other components of the parameter vector. It uses the values of components from the last time it communicated with a particular processor.

We will make the assumption of *total asynchronism* [13]: we assume that each set T^i is infinite, and that if $\{t_k\}$ is a sequence in T^i tending to infinity, then $\lim_{k\to\infty} \tau_{ij}(t_k) = \infty$, for each neighbor $j \in N(i)$. This mild assumption guarantees that each component is updated infinitely often, and that processors eventually communicate with neighboring processors. It allows for arbitrary delays in communication, and even the out-of-order arrival of messages between processors.

We can extend the convergence result of Theorem 1 to this setting. The proof is straightforward given the results we have already established and standard results on asynchronous algorithms (see [13], for example). We will provide an outline here. For the convergence of the quadratic parameters, note that the synchronous iteration (7) is a monotone mapping (see Lemma 4 in Appendix A). For such monotone mappings, synchronous convergence implies totally asynchronous convergence by Proposition 6.2.1 in [13]. The linear parameter update equation for the synchronous algorithm has the form

$$z^{(t+1)} = -D^{(t)} y + A^{(t)} z^{(t)}.$$

For t sufficiently large, by the convergence of the quadratic parameters, the matrix $A^{(t)}$ becomes arbitrarily close to A. From Lemma 3, the matrix $|A|$ has spectral radius less than one. In this case, by Corollary 2.6.2 in [13], it must correspond to a weighted maximum norm contraction. Then, one can establish asynchronous convergence of the linear parameters by appealing again to Proposition 6.2.1 in [13].

VII. Discussion

The following corollary is a restatement of Theorem 1 in terms of message passing updates of the form (4).

Corollary 1. (Convergence of Message Passing Updates) *Let $\{g_i(\cdot), g_{ij}(\cdot,\cdot)\}$ be a convex decomposition of $f(\cdot)$, and let $\{f_i(\cdot), f_{ij}(\cdot)\}$ be a decomposition of $f(\cdot)$ into quadratic functions such that*

$$g_{ij}(x_i, x_j) + J_{i\to j}^{(0)}(x_j) + J_{j\to i}^{(0)}(x_i) - f_{ij}(x_i, x_j) \qquad (16)$$

is a convex function of (x_i, x_j), for all $(i,j) \in E$. Then, using the decomposition $\{f_i(\cdot), f_{ij}(\cdot,\cdot)\}$ and quadratic initial messages $\{J_{i\to j}^{(0)}(\cdot)\}$, the running estimates $x^{(t)}$ generated by the min-sum algorithm converge. Further,

$$\lim_{t\to\infty} f(x^{(t)}) = \min_x f(x).$$

The work of Johnson, et al. [9] identifies existence of convex decomposition of the objective as a important condition for such convergence results and also introduces the notion of walk-summability. However, the convergence analysis presented there only establishes a special case of the above corollary, where

$$f_{ij}(x_i, x_j) = \Gamma_{ij} x_i x_j, \ \forall\ (i,j) \in E,$$

$$J_{i\to j}^{(0)}(x_j) = 0, \ \forall\ \{i,j\} \in \vec{E}.$$

In addition, they present a quadratic program that is not convex decomposable, and where the min-sum algorithm fails to converge.

The prior work of the current authors in [7] considers a case that arises in distributed averaging applications. There, convergence is established when

$$f_{ij}(x_i, x_j) = \frac{1}{2}\Gamma_{ij}(x_i - x_j)^2,\ \Gamma_{ij} > 0,\ \forall\ (i,j) \in E,$$

$$J^{(0)}_{i\to j}(\cdot) \text{ is convex},\ \forall\ \{i,j\} \in \vec{E},$$

This is also a special case of Corollary 1. The work in [7] further develops complexity bounds on the rate of convergence in certain special cases. Study of the rate of convergence of the min-sum algorithm in more general cases remains an open issue.

Note that the main convexity condition (16) of Corollary 1 can also be interpreted in the context of general convex objectives. While our analysis is very specific to the quadratic case, the result may be illuminating in the broader context of convex programs.

Finally, although every quadratic program can be decomposed over pairwise cliques, as we assume in this paper, there may also be decompositions involving higher order cliques. Our analysis does not apply to that case, and this is an interesting question for future consideration.

Appendix A
Proof of Theorem 2

Define the domain

$$\mathcal{D} = \left\{ \gamma \in \mathbb{R}^{|\vec{E}|} \ \middle|\ \sum_{u \in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui} < 1,\ \forall\ \{i,j\} \in \vec{E} \right\},$$

and the operator $F : \mathcal{D} \to \mathbb{R}^{|\vec{E}|}$ by

$$F_{ij}(\gamma) = \frac{1}{1 - \sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}}, \quad \forall\ \{i,j\} \in \vec{E}.$$

This operator corresponds to a single min-sum update (7) of the quadratic parameters. We will first establish some properties of this operator.

Lemma 4. *The following hold:*

(i) *The operator $F(\cdot)$ is continuous.*

(ii) *The operator $F(\cdot)$ is monotonic. That is, if $\gamma, \gamma' \in \mathcal{D}$ and $\gamma \le \gamma'$, $F(\gamma) \le F(\gamma')$.*

(iii) *The operator $F(\cdot)$ is positive. That is, if $\gamma \in \mathcal{D}$, $F(\gamma) > \mathbf{0}$.*

(iv) *If $v \in \mathcal{V}$ and $\gamma \le v$,*

$$\alpha F(\gamma) < (\alpha - 1)v + F(v - \alpha(v - \gamma)), \quad \forall\ \alpha > 1.$$

(v) *If $v \in \mathcal{V}$, $F(v) < v$.*

Proof: Parts (i)-(iii) follow from the corresponding properties of the function

$$x \mapsto \frac{1}{1-x},$$

for $x \in (-\infty, 1)$. Part (v) follows from setting $\gamma = v$ in Part (iv).

Part (iv) remains. For notational convenience, define

$$R_{ij}(\gamma) = 1 - \sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui},$$

$$z = v - \gamma \ge 0.$$

We have

$$\begin{aligned}
&(\alpha-1)v_{ij} + F_{ij}(v - \alpha(v-\gamma)) - \alpha F_{ij}(\gamma) \\
&= (\alpha-1)v_{ij} + \frac{1}{R_{ij}(v-\alpha z)} - \frac{\alpha}{R_{ij}(v-z)} \\
&= \frac{1}{R_{ij}(v-\alpha z) R_{ij}(v-z)} \\
&\quad \times \big\{(\alpha-1)v_{ij} R_{ij}(v-\alpha z) R_{ij}(v-z) \\
&\quad\quad + R_{ij}(v-z) - \alpha R_{ij}(v-\alpha z)\big\}.
\end{aligned}$$

Denote the numerator of the last expression by Δ. Since the denominator is positive, it suffices to show that $\Delta > 0$. Define

$$V_j = 1 - \sum_{i\in N(j)} \Gamma_{ij}^2 v_{ij} > 0,$$

$$S_{ij} = \sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 z_{ui} \ge 0.$$

Note that

$$R_{ij}(v - \alpha z) = V_i + \Gamma_{ij}^2 v_{ji} + \alpha S_{ij},$$

$$R_{ij}(v - z) = V_i + \Gamma_{ij}^2 v_{ji} + S_{ij}.$$

Since $v \in \mathcal{V}$, we have $\Gamma_{ij}^2 v_{ij} v_{ji} \ge 1$, for each $\{i,j\} \in \vec{E}$. Then, we can derive the chain of inequalities

$$\begin{aligned}
\Delta &= (\alpha-1)v_{ij}(V_i + \Gamma_{ij}^2 v_{ji} + \alpha S_{ij})(V_i + \Gamma_{ij}^2 v_{ji} + S_{ij}) \\
&\quad + V_i + \Gamma_{ij}^2 v_{ji} + S_{ij} - \alpha(V_i + \Gamma_{ij}^2 v_{ji} + \alpha S_{ij}) \\
&\ge (\alpha-1)v_{ij}(V_i + \alpha S_{ij})(V_i + \Gamma_{ij}^2 v_{ji} + S_{ij}) \\
&\quad + (\alpha-1)(V_i + \Gamma_{ij}^2 v_{ji} + S_{ij}) + V_i + \Gamma_{ij}^2 v_{ji} + S_{ij} \\
&\quad - \alpha(V_i + \Gamma_{ij}^2 v_{ji} + \alpha S_{ij}) \\
&= (\alpha-1)v_{ij}(V_i + \alpha S_{ij})(V_i + \Gamma_{ij}^2 v_{ji} + S_{ij}) \\
&\quad - \alpha(\alpha-1) S_{ij} \\
&\ge (\alpha-1)v_{ij}(V_i + \alpha S_{ij})(V_i + S_{ij}) \\
&\quad + (\alpha-1)(V_i + \alpha S_{ij}) - \alpha(\alpha-1)S_{ij} \\
&= (\alpha-1)v_{ij}(V_i + \alpha S_{ij})(V_i + S_{ij}) + (\alpha-1)V_i \\
&> 0.
\end{aligned}$$

■

We are now ready to prove Theorem 2.

Theorem 2. *Assume that $f(\cdot)$ is convex decomposable. The set of system of equations*

$$\gamma_{ij} = \frac{1}{1 - \sum_{u\in N(i)\setminus j} \Gamma_{ui}^2 \gamma_{ui}}, \quad \forall\ \{i,j\} \in \vec{E},$$

has a solution γ^* such that

$$\mathbf{0} < \gamma^* < v, \quad \forall\ v \in \mathcal{V}.$$

Moreover, γ^* is the unique such solution.

If we initialize the min-sum algorithm so that $\gamma^{(0)} \leq v$, for some $v \in \mathcal{V}$, then $0 < \gamma^{(t)} < v$, for all $t > 0$, and

$$\lim_{t\to\infty} \gamma^{(t)} = \gamma^*.$$

Proof: Pick some $v \in \mathcal{V}$. Then, $F(v) < v$ from Part (v) of Lemma 4. Thus, we have $F^t(v) \leq F^{t-1}(v)$, for all $t > 0$, by monotonicity. (Here, $F^t(\cdot)$ denotes t applications of the operator $F(\cdot)$.) Then, the sequence $\{F^t(v)\}$ is a monotonically decreasing sequence, which by the positivity of $F(\cdot)$, is bounded below by zero. Hence, the limit $F^\infty(v)$ exists. By continuity, it must be a fixed point of $F(\cdot)$.

Now, note that, by positivity, $\mathbf{0} \leq F^\infty(v)$. Thus, by monotonicity, $F^t(\mathbf{0}) \leq F^\infty(v)$, for all $t > 0$. Since $\mathbf{0} < F(\mathbf{0}) = \mathbf{1}$, we have $F^{t-1}(\mathbf{0}) \leq F^t(\mathbf{0})$, for all $t > 0$, and this sequence converges to a fixed point $F^\infty(\mathbf{0}) \leq F^\infty(v)$.

We wish to show that $F^\infty(\mathbf{0}) = F^\infty(v)$. Assume otherwise. Define

$$\beta = \inf\{\alpha \geq 1 \mid v - \alpha(v - F^\infty(v)) \leq F^\infty(\mathbf{0})\}.$$

Since $F^\infty(v) < v$, the set in the above infimum is not empty. Since $F^\infty(\mathbf{0}) \leq F^\infty(v)$ and $F^\infty(\mathbf{0}) \neq F^\infty(v)$, we must have $\beta > 1$. Then, we have

$$F^\infty(\mathbf{0}) \geq v - \beta(v - F^\infty(v)).$$

Applying $F(\cdot)$ and using Part (iv) of Lemma 4,

$$\begin{aligned} F^\infty(\mathbf{0}) &\geq F(v - \beta(v - F^\infty(v))) \\ &> \beta F^\infty(v) - (\beta - 1)v \\ &= v - \beta(v - F^\infty(v)). \end{aligned}$$

This contradicts the definition of β. Thus, we must have $F^\infty(\mathbf{0}) = F^\infty(v)$.

Set $\gamma^* = F^\infty(\mathbf{0})$. From the above argument, we have $\mathbf{0} < \gamma^* = F^\infty(v) < v$, for all $v \in \mathcal{V}$. Thus, γ^* satisfies the conditions of the lemma.

Assume there is some other fixed point γ' satisfying the conditions of the lemma. Positivity implies $\gamma' > 0$. Then, since $\mathbf{0} < \gamma' < v$ for some $v \in \mathcal{V}$, by repeatedly applying $F(\cdot)$, we have

$$F^t(\mathbf{0}) \leq \gamma' \leq F^t(v),$$

for all $t > 0$. Taking a limit as $t \to \infty$, it is clear that $\gamma' = \gamma*$.

It remains to prove the final statement of the lemma. Consider $\gamma^{(0)}$, with $\gamma^{(0)} \leq v$, for some $v \in \mathcal{V}$. Note that $\mathbf{0} < F(\gamma) \leq F(v) < v$. Then,

$$\mathbf{0} < F^t(\mathbf{0}) \leq \gamma^{(t+1)} = F^{t+1}(\gamma^{(0)}) \leq F^{t+1}(v) < v.$$

for all $t > 0$. Taking limits,

$$\lim_{t\to\infty} \gamma^{(t)} = \gamma^*.$$

■

APPENDIX B
PROOF OF LEMMAS 1 AND 2

For the balance of this section, we assume that $f(\cdot)$ admits a convex decomposition.

In order to prove Lemma 1, we first fix an arbitrary vertex r, and consider an infinite computation tree rooted at a vertex $\tilde{r}$ corresponding to r. Such a tree is constructed in an iterative process, first starting with a single vertex $\tilde{r}$. As each step, vertices are added to leaves on the tree corresponding to the neighbors of the leaf in the original graph other than its parent. Hence, the tree's vertices consist of replicas of vertices in the original graph, and the local structure around each vertex is the same as that in the original graph. We can extend both functions $\rho(\cdot)$ and $\nu(\cdot)$ to walks on the computation tree by defining weights on edges in the computation tree according to the weights of the corresponding edges in the original graph. We will use the tilde symbol to distinguish vertices and subsets of the computation tree from those in the underlying graph.

We begin with a lemma.

Lemma 5. *Given connected vertices $\tilde{i}, \tilde{j}$ in the computation tree, with labels i,j, respectively, let $\tilde{\mathcal{W}}_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}$ be the set of walks starting at $\tilde{i}$ and returning to $\tilde{i}$ but never crossing the edge $(\tilde{i}, \tilde{j})$. Then,*

$$\rho(\tilde{\mathcal{W}}_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}) = \gamma^*_{ij}.$$

Proof: First, note that walks in $\tilde{\mathcal{W}}_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}$ can be mapped to disjoint walks on the original graph. Hence, by walk-summability, the infinite sum

$$\sum_{\tilde{w}\in\tilde{\mathcal{W}}_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}} \rho(\tilde{w})$$

converges absolutely.

Now, define the set $\tilde{\mathcal{W}}^d_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}$ to be the set of walks in $\tilde{\mathcal{W}}_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}$ that travel at most a distance d away from $\tilde{i}$ in the computation tree. A walk $\tilde{w} \in \tilde{\mathcal{W}}^d_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}$ can be decomposed into a series of traversals to neighbors $\tilde{u} \in N(\tilde{i}) \setminus \tilde{j}$, self-returning walks from $\tilde{u}$ to $\tilde{u}$ that do not cross $(\tilde{u}, \tilde{i})$ and travel at most distance $d-1$ from $\tilde{u}$, and then returns to $\tilde{i}$. Letting t index the total number of such traversals, we have the expression

$$\rho(\tilde{\mathcal{W}}^d_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}) = \sum_{t=0}^{\infty}\left(\sum_{\tilde{u}\in N(\tilde{i})\setminus\tilde{j}} R^2_{\tilde{u}\tilde{i}}\rho(\tilde{\mathcal{W}}^{d-1}_{\tilde{u}\to\tilde{u}\setminus\tilde{i}})\right)^t.$$

By walk-summability, this infinite sum must converge. Thus,

$$\rho(\tilde{\mathcal{W}}^d_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}) = \frac{1}{1 - \sum_{\tilde{u}\in N(\tilde{i})\setminus\tilde{j}} R^2_{\tilde{u}\tilde{i}}\rho(\tilde{\mathcal{W}}^{d-1}_{\tilde{u}\to\tilde{u}\setminus\tilde{i}})}.$$

By the symmetry of the computation tree, the quantity $\rho(\tilde{\mathcal{W}}^d_{\tilde{i}\to\tilde{i}\setminus\tilde{j}})$ depends only on the labels of $\tilde{i}$ and $\tilde{j}$ in the original graph. Set $\gamma^{(0)}_{ij} = \rho(\tilde{\mathcal{W}}^0_{\tilde{i}\to\tilde{i}\setminus\tilde{j}}) = 1$ and $\gamma^{(d)}_{ij} = \rho(\tilde{\mathcal{W}}^d_{\tilde{i}\to\tilde{i}\setminus\tilde{j}})$, for each $\{i,j\} \in \vec{E}$ and integer $d > 0$. Then, we have

$$\gamma^{(d)}_{ij} = \frac{1}{1 - \sum_{u\in N(i)\setminus j} R^2_{ui}\gamma^{(d-1)}_{ui}}.$$

By Theorem 2, we have

$$\lim_{d\to\infty} \gamma_{ij}^{(d)} = \gamma_{ij}^*.$$

Then, since $\tilde{\mathcal{W}}^d_{\tilde{i}\to\tilde{i}\backslash\tilde{j}} \subset \tilde{\mathcal{W}}^{d+1}_{\tilde{i}\to\tilde{i}\backslash\tilde{j}}$, and

$$\tilde{\mathcal{W}}_{\tilde{i}\to\tilde{i}\backslash\tilde{j}} = \bigcup_{d=0}^{\infty} \tilde{\mathcal{W}}^d_{\tilde{i}\to\tilde{i}\backslash\tilde{j}},$$

we have

$$\rho(\tilde{\mathcal{W}}_{\tilde{i}\to\tilde{i}\backslash\tilde{j}}) = \lim_{d\to\infty} \rho(\tilde{\mathcal{W}}^d_{\tilde{i}\to\tilde{i}\backslash\tilde{j}}) = \gamma_{ij}^*.$$

■

We call a walk on the computation tree a *shortest-path* walk if it is the unique shortest path between its endpoints. Given a shortest-path walk $\tilde{p}$ define $\tilde{\mathcal{W}}_{\tilde{p}}$ to be the set of all walks of the form

$$\{\tilde{p_0}, \tilde{w}^0, \tilde{p_1}, \tilde{w}^1, \ldots, \tilde{w}^{|\tilde{p}|-1}, \tilde{p}_{|\tilde{p}|}\},$$

where $\tilde{w}^i \in \tilde{\mathcal{W}}_{\tilde{p}_i\to\tilde{p}_i\backslash\tilde{p}_{i+1}}$, for $0 \le i < |\tilde{p}|$. Intuitively, these walks proceed along the path p, but at each point $\tilde{p}_i$, they may also take a self-returning walk from vertex $\tilde{p}_i$ to vertex $\tilde{p}_i$ that does not cross the edge $(\tilde{p}_i, \tilde{p}_{i+1})$.

Lemma 6. *Given a shortest-path walk $\tilde{p}$,*

$$\rho(\tilde{\mathcal{W}}_{\tilde{p}}) = \nu(\tilde{p}).$$

Proof:

$$\begin{aligned}\rho(\tilde{\mathcal{W}}_{\tilde{p}}) &= \sum_{\tilde{w}^0 \in \tilde{\mathcal{W}}_{\tilde{p}_0\to\tilde{p}_0\backslash\tilde{p}_1}} \cdots \sum_{\tilde{w}^{|\tilde{p}|-1} \in \tilde{\mathcal{W}}_{\tilde{p}_0\to\tilde{p}_{|\tilde{p}|-1}\backslash\tilde{p}_{|\tilde{p}|}}} \rho(\{\tilde{p_0}, \tilde{w}^0, \tilde{p_1}, \tilde{w}^1, \ldots, \tilde{w}^{|\tilde{p}|-1}, \tilde{p}_{|\tilde{p}|}\}) \\ &= \rho(\tilde{p}) \prod_{i=0}^{|\tilde{p}|-1} \rho(\tilde{\mathcal{W}}_{\tilde{p}_i\to\tilde{p}_i\backslash\tilde{p}_{i+1}}) \\ &= \nu(\tilde{p}).\end{aligned}$$

■

We are now ready to prove Lemma 1.

Lemma 1. *Assume that $f(\cdot)$ is convex decomposable. For each $w \in \mathcal{W}^{nb}$, there exists a set of walks $\mathcal{W}_w$, all terminating at the same vertex as w, such that*

$$\nu(w) = \rho(\mathcal{W}_w).$$

Further, if $w' \in \mathcal{W}^{nb}$ and $w' \neq w$, then $\mathcal{W}_w$ and $\mathcal{W}_{w'}$ are disjoint.

Proof: Take a vertex i in the original graph. Given a walk from i to r in the original graph, there is a unique corresponding walk from a replica of i to $\tilde{r}$ in the computation tree. Also notice that non-backtracking walks in the original graph that terminate at r correspond uniquely to shortest-path walks in the computation tree that terminate at $\tilde{r}$.

Now, assume that $w \in \mathcal{W}^{nb}$ terminates at r. Let $\tilde{p}$ be the corresponding shortest-path walk in the computation tree, and consider the set $\tilde{W}_{\tilde{p}}$. We will define $\mathcal{W}_w$ to be the set of walks in the original graph corresponding to $\tilde{W}_{\tilde{p}}$. From Lemma 6,

$$\nu(w) = \nu(\tilde{p}) = \rho(\tilde{W}_{\tilde{p}}) = \rho(\mathcal{W}_w).$$

Now, consider another walk $w' \in \mathcal{W}^{nb}$, $w' \neq w$, that also terminates at r. We would like to show that $\mathcal{W}_w$ and $\mathcal{W}_{w'}$ are disjoint. Let $\tilde{p}'$ be the shortest-path walk corresponding to w'. Equivalently, we can show $\tilde{W}_{\tilde{p}}$ and $\tilde{W}_{\tilde{p}'}$ are

disjoint. Assume there is some walk $\tilde{u} \in \tilde{W}_{\tilde{p}} \cap \tilde{W}_{\tilde{p}'}$. Then, both $\tilde{p}$ and $\tilde{p}'$ must be the shortest-path from the origin of $\tilde{u}$ to $\tilde{r}$. Since shortest-paths between a pair of vertices on the computation tree are unique, we must have $\tilde{p} = \tilde{p}'$ and this $w = w'$, which is a contradiction.

Note that we only considered non-backtracking walks terminating at a fixed vertex r. However, our choice or r was arbitrary hence we can repeat the construction for each $r \in V$. Moreover, if w and w' terminate at different vertices r and r', respectively, the sets $\mathcal{W}_w$ and $\mathcal{W}_{w'}$ will contain only walks that terminate at r and r', respectively, thus they will be disjoint.

■

Using similar arguments as above, we can prove Lemma 2.

Lemma 2. *Assume that $f(\cdot)$ is convex decomposable. If we define $\mathcal{W}^{nb}_{i\to r}$ to be the set of all non-backtracking walks from vertex i to vertex r, we have*

$$\rho(\mathcal{W}_{i\to r}) = \frac{\nu(\mathcal{W}^{nb}_{i\to r})}{1 - \sum_{u\in N(r)} R^2_{ur}\gamma^*_{ur}}.$$

Proof: Consider a walk $w \in \mathcal{W}_{i\to r}$, and let $\tilde{w}$ be the unique corresponding walk in the computation tree terminating at $\tilde{r}$. Let $\tilde{p}$ be the unique shortest-path walk corresponding to $\tilde{w}$. Note that $\tilde{p}$ will originate at a replica of i, and end at $\tilde{r}$. Thus, $\tilde{p}$ uniquely corresponds to a non-backtracking walk $w' \in \mathcal{W}^{nb}_{i\to r}$.

Now, $\tilde{w}$ can be uniquely decomposed according to

$$\{\tilde{p_0}, \tilde{w}^0, \tilde{p_1}, \tilde{w}^1, \ldots, \tilde{w}^{|\tilde{p}|-1}, \tilde{p}_{|\tilde{p}|}, \tilde{v}\},$$

where $\tilde{w}^i \in \tilde{\mathcal{W}}_{\tilde{p}_i\to\tilde{p}_i\backslash\tilde{p}_{i+1}}$, for $0 \le i < |\tilde{p}|$, and $\tilde{v}$ is a self-returning walk from $\tilde{r}$ to $\tilde{r}$. Applying Lemma 6, we have

$$\rho(\mathcal{W}_{i\to r}) = \nu(\mathcal{W}^{nb}_{i\to r})\rho(\tilde{\mathcal{W}}_{\tilde{r}\to\tilde{r}}),$$

where $\tilde{\mathcal{W}}_{\tilde{r}\to\tilde{r}}$ is the set of self-returning walks from $\tilde{r}$ to $\tilde{r}$.

However, a walk $\tilde{v} \in \tilde{\mathcal{W}}_{\tilde{r}\to\tilde{r}}$ can be uniquely decomposed into a series of traversals to neighbors $\tilde{u} \in N(\tilde{r})$, self-returning walks from $\tilde{u}$ to $\tilde{u}$ that do not cross $(\tilde{u}, \tilde{r})$, and then returns to $\tilde{i}$. Letting t index the total number of such traversals, we have the expression

$$\rho(\tilde{\mathcal{W}}_{\tilde{r}\to\tilde{r}}) = \sum_{t=0}^{\infty} \left(\sum_{\tilde{u}\in N(\tilde{r})} R^2_{\tilde{u}\tilde{r}} \rho(\tilde{\mathcal{W}}_{\tilde{u}\to\tilde{u}\backslash\tilde{r}}) \right)^t.$$

From Lemma 5,

$$\rho(\tilde{\mathcal{W}}_{\tilde{u}\to\tilde{u}\backslash\tilde{r}}) = \gamma^*_{ur}.$$

Thus,

$$\begin{aligned}\rho(\mathcal{W}_{i\to r}) &= \nu(\mathcal{W}^{nb}_{i\to r}) \sum_{t=0}^{\infty} \left(\sum_{u\in N(r)} R^2_{ur}\gamma^*_{ur} \right)^t \\ &= \frac{\nu(\mathcal{W}^{nb}_{i\to r})}{1 - \sum_{u\in N(r)} R^2_{ur}\gamma^*_{ur}}.\end{aligned}$$

■

ACKNOWLEDGMENT

The first author wishes to thank Jason Johnson for a helpful discussion. The first author was supported by a Benchmark Stanford Graduate Fellowship. This material is based upon work supported by the National Science Foundation under Grant No. IIS-0428868.

REFERENCES

[1] R. G. Gallager, *Low-Density Parity Check Codes*. Cambridge, MA: M.I.T. Press, 1963.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding,” in *Proc. Int. Communications Conf.*, Geneva, Switzerlang, May 1993, pp. 1064–1070.

[3] T. Richardson and R. Urbanke, “The capacity of low-density parity check codes under message-passing decoding,” *IEEE Transactions on Information Theory*, vol. 47, pp. 599–618, 2001.

[4] Y. Weiss and W. T. Freeman, “Correctness of belief propagation in Gaussian graphical models of arbitrary topology,” *Neural Computation*, vol. 13, pp. 2173–2200, 2001.

[5] P. Rusmevichientong and B. Van Roy, “An analysis of belief propagation on the turbo decoding graph with Gaussian densities,” *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 745–765, 2001.

[6] M. J. Wainwright, T. Jaakkola, and A. S. Willsky, “Tree-based reparameterization framework for analysis of sum-product and related algorithms,” *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1120–1146, 2003.

[7] C. C. Moallemi and B. Van Roy, “Consensus propagation,” *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 4753–4766, 2006.

[8] A. Montanari, B. Prabhakar, and D. Tse, “Belief propagation based multi-user detection,” in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, 2005.

[9] J. K. Johnson, D. M. Malioutov, and A. S. Willsky, “Walk-sum interpretation and analysis of Gaussian belief propagation,” in *Advances in Neural Information Processing Systems 18*, 2006.

[10] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, “Walk-sums and belief propagation in Gaussian graphical models,” *Journal of Machine Learning Research*, vol. 7, pp. 2031–2064, October 2006.

[11] M. J. Wainwright, T. Jaakkola, and A. S. Willsky, “Tree consistency and bounds on the performance of the max-product algorithm and its generalizations,” *Statistics and Computing*, vol. 14, pp. 143–166, 2004.

[12] C. C. Moallemi and B. Van Roy, “Convergence of the min-sum algorithm for convex optimization,” Management Science & Engineering Department, Stanford University, Tech. Rep., 2007, uRL: http://moallemi.com/ciamac/papers/cc-2007.pdf.

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

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

PLACE PHOTO HERE

Ciamac C. Moallemi Ciamac C. Moallemi is an Assistant Professor at the Graduate School of Business of Columbia University, where he has been since 2007. He received SB degrees in Electrical Engineering & Computer Science and in Mathematics from the Massachusetts Institute of Technology (1996). He studied at the University of Cambridge, where he earned a Certificate of Advanced Study in Mathematics, with distinction (1997). He received a PhD in Electrical Engineering from Stanford University (2007). He is a member of the IEEE and INFORMS. He is the receipient of a British Marshall Scholarship (1996) and a Benchmark Stanford Graduate Fellowship (2003).

PLACE PHOTO HERE

Benjamin Van Roy Benjamin Van Roy is an Associate Professor of Management Science and Engineering, Electrical Engineering, and, by courtesy, Computer Science, at Stanford University. He has held visiting positions as the Wolfgang and Helga Gaul Visiting Professor at the University of Karlsruhe and as the Chin Sophonpanich Foundation Professor of Banking and Finance at Chulalongkorn University. He received the SB (1993) in Computer Science and Engineering and the SM (1995) and PhD (1998) in Electrical Engineering and Computer Science, all from MIT. He is a member of INFORMS and IEEE. He has served on the editorial boards of Discrete Event Dynamic Systems, Machine Learning, Mathematics of Operations Research, and Operations Research. He has been a recipient of the MIT George C. Newton Undergraduate Laboratory Project Award (1993), the MIT Morris J. Levin Memorial Master’s Thesis Award (1995), the MIT George M. Sprowls Doctoral Dissertation Award (1998), the NSF CAREER Award (2000), and the Stanford Tau Beta Pi Award for Excellence in Undergraduate Teaching (2003). He has been a Frederick E. Terman Fellow and a David Morgenthaler II Faculty Scholar.