Walk-Summable Gauss-Markov Random Fields*

Jason K. Johnson

Febuary, 2002 (Corrected)†

Abstract

This note introduces an interesting class of Gauss-Markov Random Fields designated as *walk-summable*. Several equivalent characterizations of this class of GMRFs are established. Also, several important subclasses of GMRFs are identified as being walk-summable. These include (i) diagonally dominant, (ii) pairwise normalizable, (iii) regular singly-connected, and (iv) regular bipartite with only positive (or only negative) interactions. The utility of this walk-summable property for inference of GMRFs is indicated. In particular, it is shown that several iterative estimation algorithms for GMRFs submit to a walk-sum interpretation and hence the convergence of these iterative methods is assured for walk-summable systems. Also, several interesting conjectures and ideas concerning walk-summable systems are briefly summarized in the conclusion.

1 Introduction

Gauss-Markov Random Fields (GMRFs)[1] are jointly Gaussian processes which admit compact description as graphical models. Consider a Gaussian process $\mathrm{x} \sim \mathcal{N}(\mu, P)$ with *moment parameterization* given by the *mean vector* $\mu \equiv E\{\mathrm{x}\}$ and the *covariance matrix* $P \equiv E\{(\mathrm{x}-\mu)(\mathrm{x}-\mu)'\}$. The probability density function is given in the usual form below.

$$p(x) = \frac{1}{\sqrt{|2\pi P|}} \exp\{-\frac{1}{2}(x-\mu)'P^{-1}(x-\mu)\} \tag{1}$$

An alternative parameterization is provided by the so-call *information filter* form specified in terms of a *field vector* $h \equiv P^{-1}\mu$ and an *information matrix* $J \equiv P^{-1}$. The pdf of the

*This technical note was originally written as a term paper for a graduate seminar on graphical models organized by Sanjoy Mitter, MIT, EECS Dept., Fall 2001. I would also like to thank Mike Schneider, Erik Sudderth and Alan Willsky for useful discussions during this time this paper was written.

†The following corrections were made November, 2005: Added "unique representation" condition to discussion, p.9; Deleted incorrect part of conjecture regarding LBP, p.22.

[1]For an introduction to the theory and application of Markov random fields (MRFs) see Mitter [Mit92] and Guyon [Guy95].

process is given within the information filter parameterization as

$$p(x) = \exp\{-\frac{1}{2}x'Jx + h'x + \phi(h, J)\} \tag{2}$$

with normalization constant

$$\phi(h, J) = \frac{1}{2}\{\ln|J/2\pi| - \mu' J\mu\}. \tag{3}$$

This may be regarded as an exponential family model[2] with natural parameters $\theta = (h, -J/2)$ and sufficient statistics $t(x) = (x, xx')$. The moment parameters $\eta \equiv E\{t(\mathrm{x})\}$ of this exponential family are then trivially related to those of the Gaussian density as $\eta = (\mu, P + \mu\mu')$.

This information filter form proves to be a convenient parameterization for modeling GMRFs in several regards. First, this information filter form of the density provides a compact graphical model[3] of GMRFs in the following sense. Consider a Gaussian process which satisfies the so-called *Markov* property with respect to an undirected graph $\mathcal{G} = (\mathcal{S}, \mathcal{I})$ having vertices $\mathcal{S}$ corresponding to the sites of the field and edges $\mathcal{I} \subset \{\langle st\rangle | s, t \in \mathcal{S}\}$[4] representing direct interactions among sites.

This Markov property may be stated in a number of equivalent forms (see, for instance, Lauritzen [Lau96]). One useful form is stated with respect to the neighborhood systems defined by the interaction graph. Let x_s^b denote the joint-state of all the neighboors of site s in the interaction graph $\mathcal{G} = (\mathcal{S}, \mathcal{I})$ (the "b" superscript is meant to suggest the "Markov blanket" of site s consisting of all sites t such that $\langle st\rangle \in \mathcal{I}$). The Markov property then asserts that the state x_s is conditionally independent of the state outside the blanket $\mathrm{x}_{\mathcal{S}\setminus\{s,s^b\}}$ given the state of the blanket x_s^b. Mathematically, this is expressed as $p(x_s|x_{\mathcal{S}\setminus s}) = p(x_s|x_s^b)$.

By the *Hammersley-Clifford theorem* [Gri73, Mit92] (specialized somewhat for Gaussian processes)[5] the pdf of any Gaussian process which is Markov with respect to $\mathcal{G}$ may be factored in terms of singleton and pairwise positive potential functions defined on the vertices and edges of $\mathcal{G}$. This means the pdf may be represented as

$$p(x) = \frac{1}{Z}\prod_{s\in\mathcal{S}}\psi_s(x_s)\prod_{\langle st\rangle\in\mathcal{I}}\psi_{\langle st\rangle}(x_s, x_t) \tag{4}$$

where $\psi_s(x_s)$ are the singleton potentials specifying "external" influences upon each site s, $\psi_{\langle st\rangle}(x_s, x_t)$ are the pairwise potentials specifying "interactions" between pairs of sites and Z is just a normalization constant. Moreover, the converse is also asserted such that all Gaussian processes with pdfs of the above form are Markov with respect to $\mathcal{G}$.

Observe that the information filter form of the density may be put into this form by defining the single-site influence functions as

$$\psi_s(x_s) \equiv \exp\{-\frac{1}{2}J_{s,s}x_s^2 + h_s x_s\} \tag{5}$$

[2]For a comprehensive discussion of the exponential family see Barndorff-Nielsen [BN78].

[3]For an introduction to the field of graphical models see, for instance, Lauritzen [Lau96] or Jordan [Jor99].

[4]The notation $\langle st\rangle$ denotes unordered pairs of distinct sites such that $\langle st\rangle$ is not distinguished from $\langle ts\rangle$.

[5]For Gaussian processes, only pairwise "cliques" (a subset of the vertices of the interaction graph which are fully connected) need be considered whereas maximal cliques must be considered more generally.

and by defining pairwise influences as

$$\psi_{\langle st\rangle}(x_s, x_t) \equiv \exp\{-J_{s,t} x_s x_t\}. \tag{6}$$

where these pairwise influences are only needed for those "interactions" $\langle st\rangle$ where $J_{s,t}$ is nonzero. Hence, the sparsity structure of J dictates the Markov structure of the field. Conversely, the Markov structure of the field given by $\mathcal{G}$ imposes a corresponding sparsity structure upon J. Hence, throughout the remainder of this note we shall consider the interaction graph $\mathcal{G}$ as being defined by the sparsity structure of J such that $\langle st\rangle \in \mathcal{I} \Leftrightarrow J_{s,t} \neq 0$ this being the minimal graph for which the Gaussian process is Markov.

The above equivalence asserted by Hammersley-Clifford is not formally proved here. However, this connection is briefly indicated by the following analysis of the partial correlation coefficients which summarize the Markov structure of Gaussian processes. These coefficients play a central role in the walk-summable analysis to follow.

Definition 1 *The* partial correlation coefficient *denoted ρ_{st} between sites $s, t \in \mathcal{S}$ of a GMRF* x *is the correlation coefficient measuring the residual correlation remaining between states x_s and x_t after conditioning upon the state of the rest of the system x_{st}^c.*

$$\rho_{st} = \rho(\mathrm{x}_s, \mathrm{x}_t | \mathrm{x}_{st}^c) = \frac{\mathrm{var}(\mathrm{x}_s, \mathrm{x}_t|\mathrm{x}_{st}^c)}{\sqrt{\mathrm{var}(\mathrm{x}_t|\mathrm{x}_{st}^c)\mathrm{var}(\mathrm{x}_t|\mathrm{x}_{st}^c)}} \tag{7}$$

Also, let R denote the (zero-diagonal) partial correlation matrix *having zeros along the diagonal and partial correlation coefficients off the diagonal.*

$$R_{st} = \begin{cases} \rho_{st}, & s \neq t \\ 0, & s = t \end{cases} \tag{8}$$

Since $\rho_{ss} = 1$ for all s, we may consider R defined as $R_{st} = \rho_{st} - \delta_{st}$ (the full partial correlation matrix minus the identity).

The connection between the information matrix and the partial correlation coefficients is made explicit by the following proposition.

Proposition 1 *Let $\tilde{J} \equiv D^{-1/2} J D^{-1/2}$ denote the* normalized information matrix *constructed by applying the diagonal congruence transform $D^{-1/2}$ where $D = \mathrm{diag}(J)$ is the positive diagonal matrix with diagonal entries as in the information matrix J. Then, the normalized information matrix $\tilde{J}$ has unit-diagonal and minus the partial correlation coefficients elsewhere.*

$$(\tilde{J})_{st} \equiv \frac{J_{st}}{\sqrt{J_{ss}J_{tt}}} \tag{9}$$

$$= \begin{cases} 1, & s = t \\ -\rho(\mathrm{x}_s, \mathrm{x}_t|\mathrm{x}_{st}^c), & s \neq t \end{cases} \tag{10}$$

In terms of the zero-diagonal partial correlations matrix R we have $\tilde{J} = I - R$.

Proof. The conditional covariance of the bivariate state $\mathrm{x}_{st} \equiv (\mathrm{x}_s, \mathrm{x}_t)$ given the state of the complement process x_{st}^c is just the inverse of the associated principle minor of the information matrix $J_{\{s,t\}}$.

$$P_{\mathrm{x}_{st}|\mathrm{x}_{st}^c} = (J_{\{s,t\}})^{-1} \tag{11}$$

$$\propto \begin{pmatrix} J_{tt} & -J_{st} \\ -J_{st} & J_{ss} \end{pmatrix} \tag{12}$$

The associated correlation coefficient is calculated below.

$$\rho(\mathrm{x}_s, \mathrm{x}_t|\mathrm{x}_{st}^c) = \frac{(P_{\mathrm{x}_{st}|\mathrm{x}_{st}^c})_{st}}{\sqrt{(P_{\mathrm{x}_{st}|\mathrm{x}_{st}^c})_{ss}(P_{\mathrm{x}_{st}|\mathrm{x}_{st}^c})_{tt}}} \tag{13}$$

$$= \frac{-J_{st}}{\sqrt{J_{tt}J_{ss}}} \tag{14}$$

□

A second advantage of the information filter form of the model is that evidence in the form of local observations are easily absorbed into this representation by locally adjusting the information filter parameters. This corresponds to factoring the local likelihood functions into those potential functions described earlier. For instance, given a local observation $y = Cx_\Lambda + v$ with $v \sim \mathcal{N}(0, Q)$ we may locally update the information filter parameters (h_Λ, J_Λ) to construct a graphical model for the full conditional density $p(x|y)$ (for fixed y) as shown below.

$$h_\Lambda \rightarrow h_\Lambda + CQ^{-1}y \tag{15}$$

$$J_\Lambda \rightarrow J_\Lambda + CQ^{-1}C \tag{16}$$

Note that the parameters are unchanged elsewhere. Moreover, so long as the subfield Λ is a clique (fully connected) then this update does not modify the graphical structure of the model. Otherwise, these updates may introduce additional interactions to the field. Independent observations may then be sequentially absorbed into the model in this manner. In this regard, the information filter form of the model provides a natural representation for accruing evidence in the form of local observations.

Hence, for the purposes of this note, we shall regard the *inference problem* as follows. Given a graphical model of the GMRF in the information filter form (h, J), evaluate the marginal densities $\mathrm{x}_s \sim \mathcal{N}(\mu_s, P_{s,s})$ for all sites s within the field. Note that this corresponds to recovering the mean vector μ and the diagonal of the covariance matrix P. If evidence has been absorbed into the model, then this corresponds to inferring the conditional marginals $\mathrm{x}_s|y \sim \mathcal{N}(E\{\mathrm{x}_s|y\}, \mathrm{cov}(\mathrm{x}_s|y))$ given all available evidence y.

2 Characterization of Walk-Summable GMRFs

This section defines the property of walk-summable for GMRFs and also provides several equivalent (necessary and sufficient) characterizations of this subclass of GMRFs.

2.1 Walk Sums

In order to define the notion of walk-summable we will need the following definition from graph theory.

Definition 2 *A* walk *of length l in a graph $\mathcal{G} = (\mathcal{S}, \mathcal{I})$ is any ordered sequence of sites $w = (w_0, w_1, \ldots, w_l)$ in $\mathcal{S}$ having the property that consecutive sites are adjacent in $\mathcal{G}$ such that $\langle w_i w_{i+1} \rangle \in \mathcal{I}$ for $i = 0, \ldots, l-1$.*

Let $\mathcal{W}_{s \xrightarrow{l} t}$ denote the set of all walks of length $l > 0$ from $w_0 = s$ to $w_l = t$.

$$\mathcal{W}_{s \xrightarrow{l} t} \equiv \{w = (w_0, \ldots, w_l) | w_0 = s, \, w_l = t\} \tag{17}$$

It proves convenient to extend this definition for $l = 0$ as below.

$$\mathcal{W}_{s \xrightarrow{0} t} \equiv \begin{cases} \{(s)\}, & s = t \\ \emptyset, & s \neq t \end{cases} \tag{18}$$

Then let $\mathcal{W}_{s \to t}$ denote the set of all walks (of arbitrary lengths) from s to t.

$$\mathcal{W}_{s \to t} \equiv \cup_{l=0}^{\infty} \mathcal{W}_{s \xrightarrow{l} t} \tag{19}$$

Given a walk w, let the *walk-product* be defined as the product of the partial correlation coefficients associated with the edges traversed by that walk.

$$\rho(w) = \prod_{i=0}^{l-1} \rho_{w_i, w_{i+1}} \tag{20}$$

For those zero-length walks $w = (s)$ let $\rho(w) = 1$. Given a set of walks $\mathcal{W}$, define the *walk-sum* as the sum of the walk-products taken over that set.

$$\rho(\mathcal{W}) = \sum_{w \in \mathcal{W}} \rho(w) \tag{21}$$

Similarly, define the *absolute walk-sum* as the sum of the absolute walk-products.

$$\bar{\rho}(\mathcal{W}) = \sum_{w \in \mathcal{W}} |\rho(w)| \tag{22}$$

Note that it is not yet clear if either of these sums are well-defined when $\mathcal{W}$ is a countably infinite set. Hence, the following definition.

Definition 3 *Let $\mathcal{W}$ denote a countably infinite set of walks. Let us then say that the walk-sum $\rho(\mathcal{W})$* exists *when for all possible orderings $\mathcal{W} = \{w^{(k)}\}_{k=0}^{\infty}$ the series $\sum_{k=0}^{\infty} \rho(w^{(k)})$ converges to a given finite value independent of the choice of ordering. This value is then unambiguously denoted by $\rho(\mathcal{W})$. Likewise, the absolute walk-sum $\bar{\rho}(\mathcal{W})$* exists *when the series converges absolutely such that $\sum_{k=0}^{\infty} |\rho(w^{(k)})|$ converges to a fixed finite value independent of the ordering.*

Lemma 1 *Let $\mathcal{W}$ denote a countably infinite set of walks. The following two result from analysis will prove useful.*

(i) The series $\sum_{k=0}^{\infty} |\rho(w^{(k)})|$ converges for some ordering $\mathcal{W} = \{w^{(k)}\}_{k=0}^{\infty}$ if and only if the absolute walk-sum $\bar{\rho}(\mathcal{W})$ exists.

(ii) If the absolute walk-sum $\bar{\rho}(\mathcal{W})$ exists then the walk-sum $\rho(\mathcal{W})$ exists.

Hence, absolute convergence for one ordering is sufficient to demonstrate existence of the walk-sum.

Proof. Standard conditional and absolute convergence results (see Rudin [Rud76]). □

These considerations then motivate the following definition of a walk-summable system.

Definition 4 *A GMRF with partial correlation matrix R is said to be* walk-summable *if the absolute walk-sums $\bar{\rho}(\mathcal{W}_{s\to t})$ exist for all $s, t \in \mathcal{S}$.*

$$\text{Walk Summable} \equiv \forall s, t : \bar{\rho}(\mathcal{W}_{s\to t}) < \infty \tag{23}$$

The remainder of this note is devoted to the study and characterization of this class of GMRFs. The main result of this note is the following equivalent characterization of the walk-summable property.

Definition 5 *Let A by a symmetric matrix. Then define the* spectral radius *of A denoted by $\varrho(A)$ to be the maximum magnitude of the eigenvalues $\{\lambda_s\}$ of A.*

$$\varrho(A) \equiv \max_s |\lambda_s| \tag{24}$$

Proposition 2 *A GMRF with partial correlations matrix R is walk-summable if and only if the element-wise absolute value matrix $\bar{R}$ has spectral radius less than one.*

$$\text{Walk Summable} \Longleftrightarrow \varrho(\bar{R}) < 1 \tag{25}$$

This is equivalent to both of the following positive-definite conditions being simultaneously satisfied.

(i) $I + \bar{R} > 0$

(ii) $I - \bar{R} > 0$

Proof. By Lemma 1, to determine if R corresponds to a walk-summable system it is necessary and sufficient to test the convergence of the absolute walk-sums $\{\bar{\rho}(\mathcal{W}_{s\to t})\}_{s,t\in\mathcal{S}}$ for any convenient ordering of the walks. Such an ordering is constructed by the matrix power series

$$\bar{P} \equiv \sum_{l=0}^{\infty} \bar{R}^l \tag{26}$$

since $(\bar{R}^l)_{st} = \bar{\rho}(\mathcal{W}_{s \xrightarrow{l} t})$ such that

$$\begin{aligned}
\bar{P}_{st} &= \sum_{l=0}^{\infty} (\bar{R}^l)_{st} && (27)\\
&= \sum_{l=0}^{\infty} \bar{\rho}(\mathcal{W}_{st}^l) && (28)\\
&= \bar{\rho}(\mathcal{W}_{s \to t}) && (29)
\end{aligned}$$

The following lemma arises in the analysis of the Gauss-Jacobi iteration.

Lemma 2 *Let A be a symmetric matrix. Then the matrix power series*

$$\sum_{n=0}^{\infty} A^n \qquad (30)$$

converges if and only if $\varrho(A) < 1$ which is equivalent to both of the following conditions being simultaneously satisfied.

(i) $I + A > 0$,

(ii) $I - A > 0$.

Moreover, when the series converges it converges to $(I - A)^{-1}$.

Proof. This is just the geometric series for symmetric matrices. Let $A = S\Lambda S'$ be the eigendecomposition of A with the columns of S being the orthogonal eigenvectors of A and Λ the diagonal matrix of real eigenvalues $\{\lambda_i\}$. The matrix $I + A$ is similar to the matrix $I + \Lambda$ such that (i) implies $1 + \lambda_i > 0$ for all i. Likewise, (ii) implies $1 - \lambda_i > 0$ for all i. Hence $|\lambda_i| < 1$ for all i such that the series $\sum_n \Lambda^n$ converges to $(I - \Lambda)^{-1}$ by the geometric series $\sum_n \lambda_i^n = 1/(1 - \lambda_i)$ for real numbers. Conjugation by S then yields that the series of the lemma converges to $(I - A)^{-1}$ as claimed. Conversely, the geometric series $\sum \lambda_i^n$ converges absolutely if and only if $|\lambda_i| < 1$ for all i, or $-I < \Lambda < I$. Conjugation by S then recovers the conditions (i) and (ii). This proves the lemma. □

Then if we take $A = \bar{R}$, the two necessary and sufficient conditions of the lemma are equivalent to those of the proposition. The matrix series converges absolutely if and only if all absolute walk-sums converge which is the definition of walk-summable. This proves the proposition. □

2.2 Path Sums

Further insight into the class of walk-summable GMRFs may be obtained by defining a consistent notion of path-sums and to relate these path-sums to the walk-sum computations considered earlier. First, we need another definition from graph theory.

Definition 6 *A (directed)* path γ *in* $\mathcal{G} = (\mathcal{S}, \mathcal{I})$ *from site* s *to site* t *of length* l *is a sequence of distinct sites* $(\gamma_0 = s, \gamma_1, \ldots, \gamma_l = t)$ *such that* $\gamma_i \neq \gamma_j$ *for* $i \neq j$.

A key notion we will exploit is that a path may be thought of as an excursion-deleted walk.

Definition 7 *Given a walk* $w = (w_0, \ldots, w_l)$ *let* $\phi(w)$ *denote the so-called* excursion deleted walk *constructed by the following recursive excursion-deletion procedure. Let* $w_{i:j} \equiv (w_i, w_{i_1}, ..., w_j)$ *for* $0 \leq i \leq j \leq l(w)$ *(where* $l(w)$ *indicates the length of walk* w*) denote subwalks of walk* w*. An excursion is a subwalk* $w_{i:j}$ *which begins and ends at the same site* $w_i = w_j$*. The excursion-deleted walk* $\phi(w)$ *is constructed by repeatedly deleting the earliest available excursion until there are no more such excursions. This is formally specified by the following recursive definition.*

$$\phi(w) \equiv \begin{cases} w, & l(w) = 0 \\ \phi(w_{k:l(w)}), & \exists k > 0 : w_k = w_0, w_k \neq w_j \text{ for } 0 < j < k \\ (w_0, \phi(w_{1:l(w)})), & \text{otherwise} \end{cases} \tag{31}$$

Note that every path γ is the excursion deleted path of a countably infinite set of walks $\phi^{-1}(\gamma)$ which reduce to this path under the above excursion-deletion procedure. The inverse images $\phi^{-1}(\gamma)$ defines a natural partitioning of the set all walks. In particular, let Γ_{st} denote the set of all paths from s to t. Then the family of inverse images $\{\phi^{-1}(\gamma)\}_{\gamma \in \Gamma_{st}}$ partitions the set $\mathcal{W}_{s \to t}$, the set all walks from s to t. This then leads to a natural decomposition of those walk-sums considered earlier. First, let us define a corresponding path sum notion consistent with the walk-sums introduced earlier.

Definition 8 *Given a path* γ *define the* path value $\rho(\gamma)$ *to be the walk-sum taken over all walks having excursion deleted path* γ.

$$\rho(\gamma) = \rho(\phi^{-1}(\gamma)) \tag{32}$$

Definition 9 *Given a set of paths* Γ*, define the* path-sum $\rho(\Gamma)$ *to be the sum of the path-values taken over that set.*

$$\rho(\Gamma) = \sum_{\gamma \in \Gamma} \rho(\gamma) \tag{33}$$

These definitions are such that the measure $\rho(\cdot)$ is consistent with the excursion deleted path operator $\phi : \mathcal{W}_{\mathcal{G}} \to \Gamma_{\mathcal{G}}$ in the sense that if $\Gamma = \phi(\mathcal{W})$ then $\rho(\Gamma) = \rho(\mathcal{W})$. This then leads to the natural reduction of (countably infinite) site-to-site walks-sums to (finite) path-sums since we have that $\rho(\mathcal{W}_{s \to t}) = \rho(\Gamma_{st})$. This leads to a significant simplification of the walk-summable property for finite GMRFs. To see this, we develop a factorization of the walk-sum $\rho(\gamma)$ in terms of closed-walk sums. But first, we introduce the notion of concatenation of walks and of walk-sets.

Definition 10 *Given two walks $u = (u_0, ..., u_l)$ and $v = (v_0, \dots, v_m)$ beginning and ending at the same site $u_l = v_0$, let uv denote the* concatenate *of u and v formed by chaining together the walks but omitting the redundant endpoint $u_l = v_0$.*

$$uv = (u_0, \dots, u_l, v_1, \dots, v_m) \tag{34}$$

Likewise, given two walks sets $\mathcal{U}$ and $\mathcal{V}$ with all walks in $\mathcal{U}$ ending at some common site u_f and all walks in $\mathcal{V}$ beginning at this same site $v_0 = u_f$, let $\mathcal{U} \cdot \mathcal{V}$ denote the set of all walks formed by concatenating a walk from $\mathcal{U}$ with a walk from $\mathcal{V}$.

$$\mathcal{U} \cdot \mathcal{V} = \{uv | u \in \mathcal{U}, v \in \mathcal{V}\} \tag{35}$$

It is immediately apparent that the walk-product operator is consistent with concatenation of walks such that $\rho(uv) = \rho(u)\rho(v)$. Suppose that the set $\mathcal{U} \cdot \mathcal{V}$ is one-to-one with the cartesian product $\mathcal{U} \otimes \mathcal{V} = \{(u, v) | u \in \mathcal{U}, v \in \mathcal{V}\}$ so that every walk $w \in \mathcal{U} \cdot \mathcal{W}$ has a unique representation as $w = uv$ with $u \in \mathcal{U}$ and $v \in \mathcal{V}$. Then, the walk-sum operator is likewise consistent with concatenation of sets such that $\rho(\mathcal{U} \cdot \mathcal{V}) = \rho(\mathcal{U})\rho(\mathcal{V})$.

$$\begin{aligned} \rho(\mathcal{U} \cdot \mathcal{V}) &= \sum_{uv \in \mathcal{U} \cdot \mathcal{V}} \rho(uv) && (36) \\ &= \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \rho(uv) && (37) \\ &= \sum_u \sum_v \rho(u)\rho(v) && (38) \\ &= \sum_u \rho(u) \sum_v \rho(v) && (39) \\ &= \rho(\mathcal{U})\rho(\mathcal{V}) && (40) \end{aligned}$$

If the one-to-one correspondence does not hold, we still have the property $\bar{\rho}(\mathcal{U} \cdot \mathcal{V}) \leq \bar{\rho}(\mathcal{U})\bar{\rho}(\mathcal{V})$ since some walks are added multiple times in the latter expression and the absolute walk-products are positive.

Now consider how we may form the walk set $\phi^{-1}(\gamma)$, the set of all walks which reduce to path γ under the excursion-deletion procedure. For notational brevity, label the sites visited by path γ sequentially as $(0, 1, 2, ..., l)$. Also, let $\mathcal{W}_{s \to s}^{\backslash \mathcal{A}}$ with $s \in \mathcal{S}$ and $\mathcal{A} \subset \mathcal{S}$ denote the set of walks which begin and end at site s without ever visiting any site in $\mathcal{A}$. The set of walks $\phi^{-1}(\gamma)$ which reduce to γ may then be reconstructed from γ by concatenation as follows.

$$\phi^{-1}(\gamma) = \mathcal{W}_{0,0} \cdot (0, 1) \cdot \mathcal{W}_{1,1}^{\backslash \{0\}} \cdot (1, 2) \cdot \mathcal{W}_{2,2}^{\backslash \{0,1\}} \cdot \dots \cdot \mathcal{W}_{l,l}^{\backslash \{0,1,\dots,l-1\}} \tag{41}$$

This just "inserts" all of those walk-segments which would have been deleted under the excursion-deletion procedure. The corresponding walk-sum then factors accordingly.

$$\rho(\gamma) = \rho(\mathcal{W}_{0,0}) \prod_{s=1}^{l} \rho_{s-1,s} \rho(\mathcal{W}_{s \to s}^{\backslash \{0,\dots,s-1\}}) \tag{42}$$

Observe also that $\mathcal{W}_{s\to s}^{\setminus\mathcal{A}} \subset \mathcal{W}_{s\to s}$ such that $\bar{\rho}(\mathcal{W}_{s\to s}^{\setminus\mathcal{A}}) \leq \bar{\rho}(\mathcal{W}_{s\to s})$ and consequently we have the following upper-bound of the absolute walk-sum associated with a given path.

$$\bar{\rho}(\gamma) \leq \bar{\rho}(\mathcal{W}_{0,0}) \prod_{s=1}^{l} |\rho_{s-1,s}| \bar{\rho}(\mathcal{W}_{s\to s}) \tag{43}$$

This leads to the following proposition.

Proposition 3 *A finite GMRF (having a finite number of sites) is walk-summable if and only if the closed absolute walk-sums $\bar{\rho}(\mathcal{W}_{s\to s})$ exists for all sites $s \in \mathcal{S}$.*

$$\text{Walk Summable} \Longleftrightarrow \forall s \in \mathcal{S} : \ \bar{\rho}(\mathcal{W}_{s\to s}) < \infty \tag{44}$$

Proof. First note that existence of the absolute walk-sums $\bar{\rho}(\mathcal{W}_{s\to s})$ is required by the definition of walk-summable. We only need to show that existence of these absolute closed walk-sums

$$\forall s \ : \ \bar{\rho}(\mathcal{W}_{s\to s}) < \infty \tag{45}$$

is sufficient to demonstrate existence of all absolute site-to-site walk-sums.

$$\begin{aligned} \bar{\rho}(\mathcal{W}_{s\to t}) &= \sum_{\gamma \in \Gamma_{s,t}} \bar{\rho}(\gamma) & (46)\\ &\leq \sum_{\gamma \in \Gamma_{s,t}} \prod_{u\in\gamma} \bar{\rho}(\mathcal{W}_{u,u}) \prod_{\langle vw \rangle \in \gamma} |\rho_{v,w}| & (47)\\ &\leq \infty & (48) \end{aligned}$$

The second inequality follows from (43) while the final existence bound follows from (45) and that there are only a finite number of paths each having finite length. □

2.3 Single-Revisit Walk Sums

We may simplify this analysis even further by more carefully considering the set of all closed walks $\mathcal{W}_{s\to s}$ beginning and ending at a given site s. Note that this set has redundant structure in that it includes walks which revisit site s an infinite number of times. As shown below, we may ascertain if the system is walk-summable by considering only the single-revisit walk-sums. Towards this end, let $\mathcal{W}_s^r$ denote the set of all closed walks (which begin and end at site s) which revisit site s exactly r times (such that there are a total of $r+1$ occurrences of s in the walk). By convention, take $\mathcal{W}_s^0 = \{(s)\}$ such that $\rho(\mathcal{W}_s^0) = 1$. The single-revisit walk-set $\mathcal{W}_s^1$ then consists of all walks which depart from s, visit other sites in the field and then return to s terminating the walk.

Consider how we might then construct the double-revisit walk set $\mathcal{W}_s^2$ from the single-revisit walk set $\mathcal{W}_s^1$. The double-revisit walk set is formed by concatenating pairs of single-revisit walks.

$$\mathcal{W}_s^2 = \mathcal{W}_s^1 \cdot \mathcal{W}_s^1 \tag{49}$$

The double-revisit walk-sum then factors accordingly as $\rho(\mathcal{W}_s^2) = \rho(\mathcal{W}_s^1)^2$. Higher revisit walk sets may be formed recursively according to

$$\mathcal{W}_s^r = \mathcal{W}_s^{r-1} \cdot \mathcal{W}_s^1 \tag{50}$$

such that $\rho(\mathcal{W}_s^r) = \rho(\mathcal{W}_s^{r-1})\rho(\mathcal{W}_s^1)$ and, by induction, the r-revisit walk sums as just the powers of the single-revisit walk-sum.

$$\rho(\mathcal{W}_s^r) = \rho(\mathcal{W}_s^1)^r \tag{51}$$

By partitioning the set of all closed walks accordingly as

$$\mathcal{W}_{s\to s} = \cup_{r=0}^{\infty} \mathcal{W}_s^r \tag{52}$$

the corresponding walk-sum then reduces to a geometric series in the single-revisit walk-sum.

$$\rho(\mathcal{W}_{s\to s}) = \sum_{r=0}^{\infty} \rho(\mathcal{W}_s^r) \tag{53}$$

$$= \sum_{r=0}^{\infty} \rho(\mathcal{W}_s^1)^r \tag{54}$$

$$= (1-\rho(\mathcal{W}_s^1))^{-1},\ |\rho(\mathcal{W}_s^1)| < 1 \tag{55}$$

Of coarse, the walk-sum $\rho(\mathcal{W}_{s\to s})$ is well-defined (independent of the ordering of the walks) if and only if the system is walk-summable. Repetition of the above argument with absolute walk-sums then leads to another characterization of the walk-summable property.

Proposition 4 *A finite GMRF is walk-summable if and only if the single-revisit walks-sums $\bar{\rho}(\mathcal{W}_s^1)$ exist and are less than one for all sites $s \in \mathcal{S}$.*

$$\text{Walk Summable} \iff \forall s \in \mathcal{S} :\ \bar{\rho}(\mathcal{W}_s^1) < 1 \tag{56}$$

Note the similarity between the above characterization of the walk-summable property and the earlier characterization requiring that the absolute values of all eigenvalues of the matrix $\bar{R}$ be less than one. This suggests that their may be some relationship between the eigenvalues of the matrix $\bar{R}$ and the single-revisit absolute walk-sums $\bar{\rho}(\mathcal{W}_s^1)$. In particular, all of those eigenvalues are less than one if and only if all single-revisit absolute walk-sums are less than one.

3 Subclasses of Walk-Summable GMRFs

In this section several significant classes of GMRFs are defined and shown to be walk-summable. The following lemma will prove useful for several of these proofs.

Lemma 3 *If R is the zero-diagonal partial correlation matrix of a regular GMRF then the off-diagonal negated partial correlation matrix $I - R$ is positive definite.*

$$J > 0 \implies I - R > 0 \tag{57}$$

Proof. The regularity hypothesis assures $J > 0$. By Proposition 1, J is congruent to $I - R$ such that, by Sylvester's law [HJ85], the matrices J and $I - R$ respectively have the same number of positive, zero, and negative eigenvalues. Hence, $J > 0$ implies $I - R > 0$. □

Definition 11 *A GMRF with information matrix J and corresponding interaction graph $\mathcal{G} = (\mathcal{S}, \mathcal{I})$ is said to be* pairwise-normalizable *if the information matrix may be decomposed as a sum of 1×1 positive singleton influences $\{J_s\}_{s \in \mathcal{S}}$ and 2×2 positive-definite pairwise interactions $\{J_{\langle st \rangle}\}_{\langle st \rangle \in \mathcal{I}}$ as shown below.*

$$J = \sum_{s \in \mathcal{S}} (J_s)_{\mathcal{S}} + \sum_{\langle st \rangle \in \mathcal{I}} (J_{\langle st \rangle})_{\mathcal{S}} \tag{58}$$

Above, $(\cdot)_{\mathcal{S}}$ denotes zero-padded expansion with respect to the full index set $\mathcal{S}$.

The "pairwise-normalizable" terminology here indicates that the pdf of the GMRF $p(\mathrm{x})$ may be factored into a product of positive potential functions involving interactions of order not higher than pairwise (all Gaussian processes have this property) and where the individual potential functions are each normalizable in the following sense.

$$\int \psi_s(x_s) dx_s \quad < \quad \infty \tag{59}$$

$$\int \psi_{\langle st \rangle}(x_{st}) dx_{st} \quad < \quad \infty \tag{60}$$

To see the equivalence, consider the node and edge potentials defined by

$$\psi_s(x_s) \quad = \quad \exp\{-\frac{1}{2} J_s x_s^2\} \tag{61}$$

$$\psi_{\langle st \rangle}(x_{st}) \quad = \quad \exp\{-\frac{1}{2} x_{st}' J_{\langle st \rangle} x_{st}\} \tag{62}$$

which are respectively normalizable if and only if $J_s > 0$ and $J_{\langle st \rangle} > 0$. Note also that any information matrix J constructed in this manner is then positive definite and thus realizable by some GMRF. The converse does not hold and thus the importance of the distinction (i.e., there exist regular GMRFs which are not pairwise-normalizable).

Proposition 5 *All pairwise-normalizable GMRFs are walk-summable.*

Proof. Given a pairwise-normalizable GMRF x with partial correlations R we may select an arbitrary interaction $\langle st \rangle \in \mathcal{I}$ and construct a second pairwise-normalizable GMRF which has exactly the same partial correlations as the original but with just ρ_{st} negated. This is accomplished by negating the off-diagonal entry of the corresponding 2×2 information matrix $J_{\langle st \rangle}$ of the pairwise potential $\psi_{\langle st \rangle}$. Note that this preserves the positive-definiteness of both the constituent pairwise information $J_{\langle st \rangle}$ and hence the composite information J such that this actually does correspond to a realizable pairwise-normalizable GMRF. By performing a sequence of these sign-flips we may construct a new GMRF x_+ which has all positive partial correlations $R_+ = \bar{R}$. Likewise, we may construct a GMRF x_- which has all negative partial correlations $R_- = -\bar{R}$. Since both of these fields are pairwise-normalizable and hence regular, then by Lemma 3 we have that $I - R_+ = I - \bar{R} > 0$ and $I - R_- = I + \bar{R} > 0$ thus satisfying conditions (i) and (ii) of Proposition 2. Hence the original GMRF x is shown to be walk-summable. □

Definition 12 *A GMRF is* diagonally-dominant *if the information matrix (or equivalently, the covariance matrix) has the property that the diagonal elements are greater than the diagonal-deleted absolute row sums.*

$$J_{ss} > \sum_{t \neq s} |J_{st}| \tag{63}$$

Proposition 6 *Diagonally-dominant GMRFs are pairwise normalizable.*

Proof. Given diagonal dominance we may explicitly construct a pairwise normalizable potential description. Details omitted for now (discussed in an earlier note concerning simulation of pairwise normalizable fields). □

Corollary 6.1 *Diagonally dominant GMRFs are walk-summable.*

Definition 13 *An undirected simple graph $\mathcal{G} = (\mathcal{S}, \mathcal{I})$ is said to be* singly-connected *if for every pair of vertices $s, t \in \mathcal{S}$ there is at most one path from s to t.*

Proposition 7 *Any regular GMRF which has a singly-connected interaction graph is walk-summable.*

Proof. Given a regular GMRF x with a singly-connected interaction graph $\mathcal{G} = (\mathcal{S}, \mathcal{I})$ and partial correlations R we may again select an arbitrary interaction $\langle uv \rangle \in \mathcal{I}$ constructing a new regular GMRF with precisely the same partial correlations but where the sign of ρ_{uv} is negated. This is accomplished as follows. Since the interaction graph is singly-connected, there exists exactly one path connecting any two sites (assume for simplicity that the graph is connected, otherwise perform the following procedure just for the connected subgraph containing the indicated edge). Hence, deleting the edge $\langle uv \rangle$ from the graph disconnects the graph into two components. Let $\mathcal{U}$ and $\mathcal{V}$ denote site sets associated with these two components where $u \in \mathcal{U}$ and $v \in \mathcal{V}$. Now define the diagonal matrix D with nonzero entries as below.

$$D_{ss} = \begin{cases} +1, & s \in \mathcal{U} \\ -1, & s \in \mathcal{V} \end{cases} \tag{64}$$

Now define the second GMRF as $y = Dx$ having covariance $P_{\mathrm{y}} = DP_{\mathrm{x}}D$. By the regularity hypothesis, $P_{\mathrm{x}} > 0$ implies $P_{\mathrm{y}} > 0$ such that y is also a regular GMRF. Furthermore, the partial correlation coefficients of y are identical to those of x except that the sign of ρ_{uv} is flipped. By applying a sequence of these sign-flip transformations we may construct either a new process x_+ which has all positive coefficients $R_+ = \bar{R}$ or a new process x_- which has all negative coefficients $R_- = -\bar{R}$. Since both of these are regular processes, Lemma 3 applies so that $I - R_+ = I - \bar{R} > 0$ and $I - R_- = I + \bar{R} > 0$ satisfying conditions (i) and (ii) of Proposition 2. Hence the original GMRF x is shown to be walk-summable. □

Definition 14 *An undirected simple graph $\mathcal{G} = (\mathcal{S}, \mathcal{I})$ is said to be* bipartite *if the sites may be partitioned into two subsets R and $\mathcal{B}$ (such that $R \cup \mathcal{B} = \mathcal{S}$ and $R \cap \mathcal{B} = \emptyset$) where no two sites in R are neighboors and likewise for $\mathcal{B}$.*

Pictorially, this means we could paint all of the nodes in the graph using only two colors (say red or blue) without any two adjacent vertices being the same color. Intuitively, this is equivalent to requiring that any cycles of the graph must have even length.

Proposition 8 *Any regular GMRF which has a bipartite interaction graph and has either all positive interactions $R = \bar{R}$ or all negative interactions $R = -\bar{R}$ is walk-summable.*

Proof. In the case that the GMRF has all positive (or all negative) interactions then we have that $(I - \bar{R}) > 0$ (or $I + \bar{R} > 0$) thus establishing one of the two walk-summable conditions of Proposition 2. We may then exploit the bipartite nature of the interaction graph to demonstrate the other. Given such a partition $(R, \mathcal{B})$ of the sites $\mathcal{S}$ we may use another sign-flip trick to demonstrate that both $I + R$ and $I - R$ are both positive definite. Let x denote the first regular GMRF and define a second GMRF $\mathrm{y} = D\mathrm{x}$ where D is a diagonal matrix as defined below.

$$D_{ss} = \begin{cases} +1, & s \in R \\ -1, & s \in \mathcal{B} \end{cases} \tag{65}$$

Then y is also a regular GMRF with partial correlations $R_{\mathrm{y}} = -R_{\mathrm{x}}$. By Lemma 3 we have that $I - R_{\mathrm{y}} = I + R_{\mathrm{x}} > 0$. So, in either case, both $I + \bar{R} > 0$ and $I - \bar{R} > 0$ which establishes that the GMRF is walk-summable by Proposition 2. □

4 Estimation of Walk-Summable GMRFs

In this section we demonstrate the utility of the previous development for designing and analyzing iterative algorithms for the estimation of GMRFs. First, the relevance of walk-sums to estimation and inference of GMRFs is shown. Next, several existing iterative methods are shown to submit to a walk-sum interpretation thus guaranteeing the convergence of these algorithms for walk-summable systems.

Suppose we are given an arbitrary GMRF x with model specified in the information filter parameterization as (h, J) and that we then wish to evaluate (possibly a subset of) the associated moment parameters (μ, P). For the purpose of this discussion, it is convenient to first normalize the process by a diagonal scaling operation to have unit-diagonal information. This normalized process is defined by

$$\tilde{\mathrm{x}} \equiv D^{1/2}\, \mathrm{x} \tag{66}$$

where $D = \mathrm{diag}(J)$ is the positive diagonal matrix having non-zero entries $D_{ss} = J_{ss}$. The information filter model of this normalized process $(\tilde{h}, \tilde{J})$ is then easily obtained from the original as below.

$$\begin{aligned}\tilde{h} &= D^{-1/2}h &(67)\\ \tilde{J} &= D^{-1/2}JD^{-1/2} &(68)\\ & &(69)\end{aligned}$$

Note that the information matrix $\tilde{J}$ of the new process is precisely the normalized (unit-diagonal) information J defined previously. The inference problem is now recast as evaluating the moment parameters of this normalized process $(\tilde{\mu}, \tilde{P})$ from the normalized information model $(\tilde{h}, \tilde{J})$. As before, this is straight-forward in principle.

$$\begin{aligned}\tilde{\mu} &= \tilde{J}^{-1}\,\tilde{h} &(70)\\ \tilde{P} &= \tilde{J}^{-1} &(71)\end{aligned}$$

But this is computationally burdensome requiring cubic computation and is not feasible for very large fields. Below we will consider approximate techniques for estimating these moments by partially evaluating certain related walk-sums. Once these moment parameters of the normalized process are known (atleast approximately), then those of the original process are easily recovered (estimated) by applying the diagonal inverse scaling operation.

$$\begin{aligned}\mu &= D^{-1/2}\,\tilde{\mu} &(72)\\ P &= D^{-1/2}\,\tilde{P}\,D^{-1/2} &(73)\end{aligned}$$

The fundamental relevance of the walk-sum formalism to estimation is given by the following proposition.

Proposition 9 *Given a walk-summable GMRF with partial correlation matrix R, let K denote the matrix of associated walks sums defined by $K_{st} = \rho(\mathcal{W}_{s\to t})$. The walk-sum matrix K is then identical to the covariance $\tilde{P}$ of the normalized process $\tilde{\mathrm{x}}$ as defined above: $K = \tilde{P}$.*

Proof. By similar analysis as in the proof of Proposition 2, we may construct the walk-sum matrix K as a power series in R.

$$K = \sum_{l=0}^{\infty} R^l \tag{74}$$

By Lemma 2, this series converges to $(I-R)^{-1}$ when it converges. The walk-summable hypothesis is equivalent to the absolute convergence of this series. So $K = (I-R)^{-1}$. As shown previously, $\tilde{J} = (I-R)$ so $K = \tilde{J}^{-1} = \tilde{P}$ as claimed. □

Hence, selected elements of the covariance matrix may be evaluated (approximately) by performing the associated (partial) walk-sums. Similarly, the state estimates $\tilde{\mu} = \tilde{P}\tilde{h}$ may be constructed by accruing input-weighted walk-sums. This may be defined as an input-weighted walk-product operator defined for a walk $w = (s_0, s_1, ..., s_l)$ and input-vector $\tilde{h}$ as

$$\rho(w; \tilde{h}) \equiv \rho(w) \times \tilde{h}_{s_l} \tag{75}$$

which may be considered as the effect site s_l exerts upon site s_0 by virtue of the sequence of interactions toured by walk w. We may then define a corresponding input-weighted walk-sum operator as below.

$$\rho(\mathcal{W};\tilde{h}) = \sum_{w\in\mathcal{W}} \rho(w;\tilde{h}) \tag{76}$$

Let $\mathcal{W}_{s\to *}$ denote the set of all walks beginning at a specified site s. The state-estimates $\tilde{\mu} = \tilde{P}\tilde{h}$ are then given by applying this input-weighted walk-sum operator respectively to these walk sets.

$$\begin{aligned}
\tilde{\mu}_s &= \sum_{t\in\mathcal{S}} \tilde{P}_{st}\tilde{h}_t && (77)\\
&= \sum_{t\in\mathcal{S}} \rho(\mathcal{W}_{s\to t})\tilde{h}_t && (78)\\
&= \sum_{t\in\mathcal{S}} \rho(\mathcal{W}_{s\to t};\tilde{h}) && (79)\\
&= \rho(\cup_{t\in\mathcal{S}}\mathcal{W}_{s\to t};\tilde{h}) && (80)\\
&= \rho(\mathcal{W}_{s\to *};\tilde{h}) && (81)
\end{aligned}$$

Intuitively, this is just summing all effects on site s. Truncating the walks-sums then provides a basis for sub-optimal estimation.

Much of the following analysis benefited greatly from discussions with Erik Sudderth especially in regards to the connection to the embedded-trees algorithm.

For the remainder of this section, we assume that $J = \tilde{J}$ and hence drop the tilde notation.

4.1 Gauss-Jacobi Algorithm

The Gauss-Jacobi Algorithm[6] is an iterative estimation algorithm for computing the mean-vector. This algorithm is fundamentally connected to the expansion (74) of the covariance P as a power-series in the zero-diagonal partial correlations matrix R since the algorithm simply evaluates the product of the truncated power series

$$P^{(l)} \equiv \sum_{k=0}^{l} R^k \tag{82}$$

applied to the input vector h thus providing an estimate

$$\mu^l \equiv P^{(l)}h \tag{83}$$

of the mean vector μ. These estimates may be constructed by summing the partial effects

$$\delta\mu^l \equiv R^l h \tag{84}$$

[6]Here "Gauss-Jacobi Algorithm" refers to performing a Richardson's iteration for solving a sparse linear system employing Gauss-Jacobi preconditioner [GVL96].

which are iteratively constructed by the recursion

$$\delta\mu^{l+1} = R\delta\mu^l \tag{85}$$

initialized by $\delta\mu^0 = h$. The L-th order estimate is then given by

$$\mu^L = \sum_{l=0}^{L} \delta\mu^l \tag{86}$$

This specifies the Gauss-Jacobi algorithm. The efficiency of this algorithm for estimating the state of GMRFs is provided by the the sparsity of the matrix R. For instance, if d is the maximum degree (number of adjacent edges to a given site) of the interaction graph, then the computational complexity Gauss-Jacobi is $\mathcal{O}(dlN)$ and hence linear in the size of the field $N = |\mathcal{S}|$ for fixed l.

The walk-sum interpretation of this iteration is now clarified. Let $\mathcal{W}_{s \xrightarrow{l} *} \equiv \cup_{t \in \mathcal{S}} \mathcal{W}_{s \xrightarrow{l} t}$ denote the set of all walks of length l beginning at site s. We then have the partitioning of $\mathcal{W}_{s \to *}$ as

$$\mathcal{W}_{s \to *} = \cup_{l=0}^{\infty} \mathcal{W}_{s \xrightarrow{l} *} \tag{87}$$

where $\mathcal{W}_{s \xrightarrow{k} *} \cap \mathcal{W}_{s \xrightarrow{l} *} = \emptyset$ for $k \neq l$. The Gauss-Jacobi algorithm is then interpreted as evaluating and accruing the partial input-weighted walk-sums

$$\delta\mu_s^l = \rho(\mathcal{W}_{s \xrightarrow{l} *}; h) \tag{88}$$

which are recursively computed as

$$\rho(\mathcal{W}_{s \xrightarrow{l} *}; h) = \sum_{\langle st \rangle \in \mathcal{I}} \rho_{st} \times \rho(\mathcal{W}_{t \xrightarrow{l-1} *}; h) \tag{89}$$

where the sum may be taken over just those sites t which are adjacent to s such that $\langle st \rangle \in \mathcal{I}$ (otherwise $\rho_{st} = 0$). Due to this walk-sum interpretation of the Gauss-Jacobi algorithm, we are thus assured the convergence of the algorithm for all walk-summable systems. Of coarse, the hope is that lower-order estimates $l << N$ will provide near-optimal estimation.

4.2 Embedded-Trees Algorithm

This section develops a walk-sum interpretation for the embedded-trees [WSW00, Sud02] and related algorithms. Throughout, we consider a partitioning of the interactions $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$ into two disjoint subsets $\mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset$. These subsets of the interactions of the process constitute embedded graphs of the original interaction graph. For walk-summable systems, we may also consider these as defining "embedded" systems having partial correlations matrices R_k for $k = 1, 2$ defined as

$$(R_k)_{st} = \begin{cases} \rho_{st}, & \langle st \rangle \in \mathcal{I}_k \\ 0, & \text{otherwise.} \end{cases} \tag{90}$$

such that $R = R_1 + R_2$. This follows as the walk-summable property of either matrix $\bar{R}_k$ is assured by the walk-summable property of the matrix $\bar{R}$ since those walks restricted to $\mathcal{I}_k$ are a subset of those walks in $\mathcal{I}$ such that the corresponding absolute walk-sums of the former are bounded by those of the latter and hence finite. But walk-summability implies convergence of $\sum_l R_k^l$ such that $\varrho(R_k) < 1$ which is equivalent to both $I - R_k > 0$ and $I + R_k > 0$. The former condition $I - R_k > 0$ assures us that the corresponding information matrices $J_k \equiv I - R_k$ are positive definite and hence realizable. Hence, if R is the partial correlations matrix of a walk-summable GMRF then so are R_1 and R_2. Hence, we may consider performing estimation with respect to these embedded subsystems. This corresponds to implementing the operation of multiplying an arbitrary input vector b by the covariance matrix $P_k \equiv J_k^{-1}$ of the embedded system (equivalently, solving the sparse linear system $J_k \mu = b$ for μ given b).

The algorithms considered below are appropriate when one or both of these embedded systems are tractable such that an efficient algorithm exists for applying the matrix P_k to an arbitrary input vector given the sparse model J_k. Some examples of such tractable subsystems include: (i) many small disconnected components which may be inferred independently using brute-force inversion of each component, (ii) singly-connected systems which may be inferred using two-sweep message passing algorithms such as in Pearl's belief propagation algorithm [Pea88], and (iii) low "tree-width" systems which may be "clustered" into a singly-connected process model having nodes with higher-dimensional state spaces but where the maximal dimension can be kept small such that the system may be efficiently solved by the junction tree algorithm [LS88, Daw92, Lau96, Jor99].

Single-Tree Version. Let $\mathcal{U}_{s\to*}^k \subset \mathcal{W}_{s\to*}$ denote the subset of $\mathcal{W}_{s\to*}$ (the set of all walks beginning at site s) which traverse exactly k of those "cut" edges in $\mathcal{I}_2$. We then have the partitioning $\mathcal{W}_{s\to*} = \cup_{k=0}^{\infty} \mathcal{U}_{s\to*}^k$ where $\mathcal{U}_{s\to*}^n \cap \mathcal{U}_{s\to*}^m = \emptyset$ for $n \neq m$. The associated partial walk sums $\delta\mu_s^k \equiv \rho(\mathcal{U}_{s\to*}^k; h)$ may then be constructed recursively as

$$\delta\mu_s^{k+1} = P_1 R_2 \delta\mu_s^k \tag{91}$$

for $k > 0$ which is initialized for $k = 0$ by

$$\delta\mu_s^0 = P_1 h \tag{92}$$

The K-th order estimate is then constructed by summing all of those effects associated with the set of all walks traversing K or fewer such "cut" edges

$$\mathcal{U}_{s\to*}^K \equiv \cup_{k=0}^K \mathcal{U}_{s\to*}^k \tag{93}$$

yielding the estimates

$$\mu_s^K \equiv \rho(\mathcal{U}_{s\to*}^K; h) \tag{94}$$

$$= \sum_{k=0}^K \delta\mu_s^k \tag{95}$$

This amounts to multiplication of the input h by a truncated version of the series expansion

$$P = P_1 + P_1 R_2 P_1 + P_1 R_2 P_1 R_2 P_1 + \cdots \tag{96}$$

which holds for all walk-summable systems. This implies convergence of the estimates.

The above algorithm may be shown to be equivalent to the single-tree version of the embedded trees algorithm defined previously. The analysis provided here shows that this version of the ET algorithm converges for all walk-summable systems. Stronger (both necessary and sufficient) conditions for convergence of the ET algorithm have been provided elsewhere, but the above result is useful in that the walk-summable property might be more easily identified in some cases.

Alternating-Tree Version. The above algorithm only assumes that one of the embedded systems R_1 is tractable (such as a tree). If those "cut" edges $\mathcal{I}_2$ also constitutes a tractable system then we may apparently accelerate the convergence of our estimation algorithms by accruing more walks per iteration as shown below.

Consider the subset of walks $w \in \mathcal{W}_{s\to *}$ which "switch trees" exactly k times. These walks consist of a sequence of $k+1$ subwalks each of which reside entirely in either $\mathcal{I}_1$ or in $\mathcal{I}_2$ and where adjacent subwalks alternate between these two sets. Partition this set into two disjoint subsets $\mathcal{U}^k_{s\to *}$ and $\mathcal{V}^k_{s\to *}$ which respectively begin in either $\mathcal{I}_1$ or $\mathcal{I}_2$. Let $\delta\mu^k_s \equiv \rho(\mathcal{U}^k_{s\to *}; h)$ and $\delta\nu^k_s \equiv \rho(\mathcal{V}^k_{s\to *}; h)$ denote the associated walk-sums. Define R^*_k for $k = 1, 2$ to be the sum of all positive powers of R_k.

$$R^*_k \equiv \sum_{l=1}^{\infty} R^l_k \tag{97}$$

Multiplication by R^*_k then corresponds to accruing walk-sums for all walks (of non-zero length) in $\mathcal{I}_k$. The walk-sums defined above may then be iteratively constructed according to the recursions

$$\begin{aligned} \delta\mu^{k+1} &= R^*_{k+1}\delta\mu^k & (98) \\ \delta\nu^{k+1} &= R^*_k\delta\nu^k & (99) \end{aligned}$$

where R^*_k alternates between R^*_1 and R^*_2 for $k > 2$ (for instance, $R^*_3 = R^*_1$ and so forth). These recursion are initialized for $k = 0$ as below.

$$\begin{aligned} \delta\mu^0 &= R^*_1 h & (100) \\ \delta\nu^0 &= R^*_2 h & (101) \end{aligned}$$

Note that this assumes that the zero-length "walk" (s) is omitted from the sets $\mathcal{U}^0_s$ and $\mathcal{V}^0_s$. The K-th order estimates are then generated by summing all effects which switch trees K or fewer times but also picking up the "self-effects" h_s due to the zero-length walks (s).

$$\begin{aligned} \mu^K_s &\equiv \rho(\{(s)\} \cup \mathcal{U}^K_{s\to *} \cup \mathcal{V}^K_{s\to *}; h) & (102) \\ &= h_s + \sum_{k=0}^{K} (\delta\mu^k_s + \delta\nu^k_s) & (103) \end{aligned}$$

This amounts to multiplication of h by a truncated version of the following series with the terms grouped as indicated.

$$P = I + (R^*_1 + R^*_2) + (R^*_1 R^*_2 + R^*_2 R^*_1) + (R^*_1 R^*_2 R^*_1 + R^*_2 R^*_1 R^*_2) + \cdots \tag{104}$$

This series is constructed from the earlier power series by grouping terms and factoring. The validity of these types of manipulations follows from the walk-summable property of the field.

Observe that the operators $R_k^* = P_k - I$ applied to a given input vector may be implemented by taking the difference between the output and input of the inference operator P_k. An algorithm implementing the above iteration must store three vectors between each iteration and performs two inference computations per iteration. Alternatively, we may reduce these requirements to just two vectors of storage and one inference computation if we combine the two parallel iterations by favoring one tree in the initialization as shown below. Define

$$\sigma_s^k \equiv \delta\mu_s^k + \delta\nu_s^{k-1} \tag{105}$$

for $k > 1$ which may be iteratively constructed according to the recursion

$$\sigma^{k+1} = R_{k+1}^* \sigma^k \tag{106}$$

where the iteration is initialized at $k = 1$ by defining $\sigma^1 \equiv h + \delta\mu_s^1 = P_1 h$. This then lends itself to the computation to the K-th order estimates defined below.

$$\begin{aligned} \delta\mu_s^K &\equiv \rho(\{(s)\} \cup \mathcal{U}_{s\to*}^K \cup \mathcal{V}_{s\to*}^{K-1}; h) && (107) \\ &= \sum_{k=1}^{K} \sigma_s^k && (108) \end{aligned}$$

This corresponds to regrouping and factoring the series (104) as shown below.

$$\begin{aligned} P &= (I + R_1^*) + (R_2^* + R_2^* R_1^*) + (R_1^* R_2^* + R_1^* R_2^* R_1^*) + \cdots && (109) \\ &= (I + R_1^*) + R_2^*(I + R_1^*) + R_1^* R_2^*(I + R_1^*) + \cdots && (110) \\ &= (I + R_2^* + R_1^* R_2^* + R_2^* R_1^* R_2^* + \cdots)(I + R_1^*) && (111) \\ &= (I + R_2^* + R_1^* R_2^* + R_2^* R_1^* R_2^* + \cdots) P_1 && (112) \end{aligned}$$

This latter version of the alternating-trees walk-sum approach is equivalent to a special-case of the embedded-trees algorithm. Specifically, the above analysis applies to the case where the embedded-trees algorithm is set up to alternate between two embedded trees which exactly partition the edges of the graph between them (such that every edge of the interaction graph is included in exactly one of the two trees). This walk-sum interpretation then guarantees the convergence of that version of the embedded trees algorithm for all walk-summable systems. Also, I see no reason the above two-tree analysis couldn't be extended to any finite number of embedded trees so long as these trees have the property of partitioning the edges of the interaction graph among them.

4.3 Marginalization of Walk-Summable GMRFs

Another approach to inference is based upon marginalization. This involves eliminating a subfield $\Lambda \subset \mathcal{S}$ of the GMRF producing a graphical model for the marginal pdf $p(x_\Lambda^c)$. Below, this marginalization is shown to submit to a straight-forward walk-sum interpretation where the requisite walk-sums are restricted to the eliminated subgraph $\mathcal{G}_\Lambda = (\Lambda, \mathcal{I}_\Lambda)$.

First, this elimination operation is expressed in the information filter parameterization. Suppose we are given the model (h, J) and we wish to evaluate the marginal model $(\hat{h}, \hat{J})$ formed by eliminating subfield Λ so as to form the marginal model for the rest of the field Λ^c. This marginal model $(\hat{h}, \hat{J})$ is related to the associated marginal moments $(\mu_{\Lambda^c}, P_{\Lambda^c})$ as $(\hat{h} = P_{\Lambda^c}^{-1}\mu_{\Lambda^c}, \hat{J} = P_{\Lambda^c}^{-1})$ such that, by the Schur complement formula for $P_{\Lambda^c}^{-1}$, the marginal model $(\hat{h}, \hat{J})$ is constructed from the full model (h, J) as shown below.

$$\begin{aligned} \hat{h} &= h_{\Lambda^c} - J_{\Lambda^c,\Lambda} J_\Lambda^{-1} h_\Lambda && (113) \\ \hat{J} &= J_{\Lambda^c} - J_{\Lambda^c,\Lambda} J_\Lambda^{-1} J_{\Lambda,\Lambda^c} && (114) \end{aligned}$$

Due to the Markov structure of the field, this only modifies the Markov blanket $\Lambda^b = \{\langle st \rangle \in \mathcal{I} | s \in \Lambda, t \notin \Lambda\}$ as shown below.

$$\begin{aligned} \hat{h}_{\Lambda^b} &= h_{\Lambda^b} - J_{\Lambda^b,\Lambda} J_\Lambda^{-1} h_\Lambda && (115) \\ \hat{J}_{\Lambda^b} &= J_{\Lambda^b} - J_{\Lambda^b,\Lambda} J_\Lambda^{-1} J_{\Lambda,\Lambda^b} && (116) \end{aligned}$$

This follows as only the submatrix $J_{\Lambda^b,\Lambda}$ of $J_{\Lambda^c,\Lambda}$ contains non-zero entries. The walk-sum interpretation of these operations follows as the submatrix J_Λ is itself walk-summable such that

$$(J_\Lambda^{-1})_{st} = \frac{\rho(\mathcal{W}_{s\to t}^\Lambda)}{\sqrt{J_{ss} J_{tt}}} \qquad (117)$$

where the walk-sums are restricted to those walks $\mathcal{W}_{s\to t}^\Lambda$ which only tour the subgraph $\mathcal{G}_\Lambda$. Recall also that the multiplication by the matrix $J_{\Lambda^b,\Lambda}$ constructs single-step extensions of these walks touring the subfield into the blanket since $(J_{\Lambda^b,\Lambda})_{s,t} = -\rho_{s,t}\sqrt{J_{ss}J_{tt}}$. For sites $s, t \in \Lambda^b$, let $\mathcal{W}_{s\to t}^\Lambda$ denote the set of all walks which depart form s, tour Λ and then terminate at t. Also, let $\mathcal{W}_{s,*}^\Lambda$ denote the set of all walks which depart from s, tour the subfield Λ but then terminate within Λ. The marginalization operation may then be expressed as below

$$\begin{aligned} \hat{h}_s &= h_s + \sqrt{J_{ss}} \times \rho(\mathcal{W}_{s\to *}^\Lambda; h_\Lambda) && (118) \\ \hat{J}_{st} &= J_{st} - \sqrt{J_{ss} J_{tt}} \times \rho(\mathcal{W}_{s\to t}^\Lambda) && (119) \end{aligned}$$

Hence, the singleton influences at sites in the blanket are updated to absorb all effects due to influences originating from within the eliminated subfield not mitigated by sites outside that subfield. Likewise, the connections between sites in the blanket are updated (or created when they did not exist previously) to reflect the coupling between those sites due to all indirect interactions through that subfield.

This walk-sum interpretation seems relevant to approximate inference in that it would seemingly point the way to approximation schemes based upon approximate marginalization by partial evaluation of the walk-sums $\rho(\mathcal{W}_{s\to t}^\Lambda)$ and $\rho(\mathcal{W}_{s\to *}^\Lambda; h)$. But we will not pursue this idea further here. These marginalization techniques are relevant to the stated inference problem in the sense that the technique of marginalizing subfields may be implemented with respect to a tree-structured partitioning of the field to infer all site marginals. This is essentially the idea behind the junction tree algorithm and provides a model for approximate inference such as in the recursive cavity modeling approach [Joh03].

5 Conclusion

This note has introduced the notion of a walk-summable GMRF and provided several equivalent characterizations of this property. Also, several interesting subsclasses of this class of walk-summable GMRFs were identified. Finally, the utility of this notion for estimation was demonstrated by analyzing several existing estimation algorithms from this perspective. In closing, I'd like to suggest some further interesting possibilities suggested by this walk-sum formalism.

First, I offer the conjecture that the Loopy Belief Propagation algorithm (LBP) for GMRFs (see Weiss and Freeman [WF01]) ought to also submit to a walk-sum interpretation. This is because the estimates generated by LBP for GMRFs have been shown to be equivalent to those estimates generated by the associated "Bethe trees" formed by "unrolling" the interaction graph relative to each site. Estimates generated by each such Bethe tree then submit to a walk-sum interpretation where all effects should be captured as the Bethe tree is grown indefinitely. If the details of this reasoning pan out, then we should be able to demonstrate that LBP converges to the correct means for all walk-summable systems. Futher analysis is required to determine if LBP can actually be interpreted in this manner and why the LBP variances are incorrect.

Furthermore, I would like to suggest that the walk-sum picture is an insightful and intuitive basis for thinking about approximate inference of GMRFs and provides fertile grounds for developing new methods in addition to providing novel interpretations of existing methods. As an example of this, consider estimating the site variances by partial evaluation of the single-revisit walk sums $\rho(\mathcal{W}_s^1)$ and formula (54). Also, exact path-based factorizations such as in (41) are highly evocative and may suggest approximate factorizations as providing a basis for estimating the correlation between distant sites within the field based upon known statistics along the "most significant" path connecting those sites. Finally, I would also like to suggest that the statistics of a walk-summable GMRFs are closely related to the equilibrium distribution of a finite-state Markov chain having states corresponding to the sites of the interaction graph and transition probabilities related to the partial correlation coefficients. Formalizing this connection may provide additional insight into the properties of such fields as well as a possible basis for inference of the walk-summable GMRF by Monte-Carlo simulation of the associated finite-state Markov chain. I will not be able explore these ideas further in this note. Yet, I do hope that these evocative ideas further motivate consideration of this interesting class of GMRFs.

References

[BN78] O. Barndorff-Nielsen. *Information and Exponential Families.* Wiley series in probability and mathematical statistics. John Wiley, 1978.

[Daw92] A.P. Dawid. Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing*, 2:25–36, 1992.

[Gri73] G.R. Grimmett. A theorem about random fields. *Bulletin of the London Mathematical Society*, 5:81–84, 1973.

[Guy95] X. Guyon. *Random Fields on a Network: modeling, statistics, and applications.* Springer-Verlag, 1995.

[GVL96] G.H. Golub and C.F. Van Loan. *Matrix Computations, 3rd ed.* Johns Hopkins University Press, 1996.

[HJ85] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[Joh03] J. Johnson. Estimation of GMRFs by recursive cavity modeling. Master's thesis, MIT, March 2003.

[Jor99] M.I. Jordan, editor. *Learning in Graphical Models.* Adaptive Computation and Machine Learning Series. MIT Press, 1999.

[Lau96] S.L. Lauritzen. *Graphical Models.* Oxford Statistical Science Series. Oxford University Press, 1996.

[LS88] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, 50:157–224, 1988.

[Mit92] S.K. Mitter. Modelling and estimation for random fields. Technical Report LIDS-P-2167, Laboratory for Information and Decision Systems, MIT, November 1992.

[Pea88] J. Pearl. *Probabilistic inference in intelligent systems.* Morgan Kaufmann, 1988.

[Rud76] W. Rudin. *Principles of Mathematical Analysis, 3rd ed.* McGraw Hill, 1976.

[Sud02] E. Sudderth. Multiscale modeling and estimation using graphs with loops. Master's thesis, Laboratory for Information and Decision Systems, MIT, February 2002.

[WF01] Y. Weiss and W.T. Freeman. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation*, 13:2173–2200, 2001.

[WSW00] M.J. Wainwright, E. Sudderth, and A.S. Willsky. Tree-based modeling and estimation of Gaussian processes on graphs with cycles. *Advances in Neural Information Processing Systems*, 11, 2000.