16:2 2025

Algebraic Statistics

A TREK RULE FOR THE LYAPUNOV EQUATION

NIELS RICHARD HANSEN

msp

Algebraic Statistics
Vol. 16, No. 2, 2025
https://doi.org/10.2140/astat.2025.16.95

A TREK RULE FOR THE LYAPUNOV EQUATION

NIELS RICHARD HANSEN

The Lyapunov equation is a linear matrix equation characterizing the cross-sectional steady-state covariance matrix of a Gaussian Markov process. We show a new version of the trek rule for this equation, which links the graphical structure of the drift of the process to the entries of the steady-state covariance matrix. In general, the trek rule is a power series expansion of the covariance matrix in the entries of the drift and volatility matrices. For acyclic models it simplifies to a polynomial in the off-diagonal entries of the drift matrix. Using the trek rule we can give relatively explicit formulas for the entries of the covariance matrix for some special cases of the drift matrix. These results illustrate notable differences between covariance models resulting from the Lyapunov equation and those resulting from linear additive noise models. To further explore differences and similarities between these two model classes, we use the trek rule to derive a new lower bound on the marginal variances in the acyclic case. This sheds light on the phenomenon, well known for the linear additive noise model, that the variances in the acyclic case tend to increase along a topological ordering of the variables.

1. Introduction

With M any $d \times d$ matrix and C a $d \times d$ positive semidefinite matrix, we can define a Gaussian Markov process $(X_t)_{t \geq 0}$ on $\mathbb{R}^d$ as a solution to the stochastic differential equation

$$\mathrm{d}X_t = MX_t \,\mathrm{d}t + C^{1/2}\,\mathrm{d}W_t, \tag{1}$$

where W_t is a d-dimensional Brownian motion; see, e.g., [7] or Section 3.7 in [11]. We call M the *drift matrix* and C is called the *diffusion matrix* or the *volatility matrix*.

The Markov process has a steady-state distribution if and only if the matrix M is stable, that is, if and only if all eigenvalues of M have strictly negative real parts. In this case, the steady-state distribution is Gaussian with mean 0 and covariance matrix Σ, which is the unique solution to the (continuous) Lyapunov equation

$$M\Sigma + \Sigma M^T + C = 0; \tag{2}$$

see, e.g., [7, Theorem 2.12].

When the process is stationary, that is, when it is started in its (unique) Gaussian steady-state distribution $\mathcal{N}(0, \Sigma)$, with Σ solving (2), the cross-sectional distribution of X_t is $\mathcal{N}(0, \Sigma)$ at any time t. Questions regarding estimation and identification of M and/or C from cross-sectional observations of the process were treated in [2; 16].

MSC2020: 15A24, 62A09, 62R01.
Keywords: Gaussian Markov processes, graphical models, Lyapunov equation, treks.

The main contribution of this paper is Theorem 2.5, which gives a novel representation of Σ in terms of the drift and volatility matrices M and C. The nonzero entries of M and C define a mixed graph, see Section 1.2, which is used in Section 2 to define a collection of treks between any two nodes. These are special walks in the graph, and we give two versions of the trek rule in Proposition 2.3 and Theorem 2.5, respectively, that express the entries of the solution Σ to (2) as a sum of trek weights over all treks.

Trek rules are well known for linear additive noise models, also known as linear structural equation models; see [15] and Section 4 in [3]. A first version of the trek rule for the Lyapunov equation was presented by [16], but our new version in Theorem 2.5 is more explicit and easier to use and interpret. A special case of our general trek rule, valid for certain acyclic models, was recently presented as Proposition 4.3 in [1], where it was used to characterize the conditional independencies that hold in the steady-state distribution. We treat the acyclic case in detail in Section 3, where we use the trek rule to derive a novel lower bound on the marginal variances.

1.1. ***The Lyapunov equation.*** The Lyapunov equation (2) is a linear matrix equation. Using the Kronecker product, we can rewrite it as

$$(M \otimes I + I \otimes M)\operatorname{vec}(\Sigma) = -\operatorname{vec}(C), \tag{3}$$

where I is the $d \times d$ identity matrix and $\operatorname{vec}(A)$ denotes the vectorization of the matrix A. When M is stable, the $d^2 \times d^2$ matrix $(M \otimes I + I \otimes M)$ is invertible, and the unique solution to (3) is given by

$$\operatorname{vec}(\Sigma) = -(M \otimes I + I \otimes M)^{-1}\operatorname{vec}(C). \tag{4}$$

The representation (4) shows that Σ_{ij} is a rational function of the entries of M and C, but (4) is not very explicit, nor is it efficient for numerical computation of the solution. There is a vast literature on numerical methods for solving the Lyapunov equation efficiently; see, e.g., [13] for a review. See also [6] for a comprehensive treatment of the Lyapunov equation and its applications.

We will not be particularly concerned with numerical methods but rather with giving a new representation of the solution that is interpretable in terms of the graphical structure of the drift matrix M. To this end, we will use the well-known integral representation of the solution given by

$$\Sigma = \int_0^\infty e^{tM} C e^{tM^T}\,\mathrm{d}t. \tag{5}$$

Here, e^{tM} denotes the matrix exponential of tM, and the integral is understood as a matrix integral. It is convergent when M is stable. See [7; 16] for further details on the integral representation of the solution.

1.2. ***Graphs.*** We will represent the nonzero entries of the drift matrix M and the volatility matrix C by a mixed graph $\mathcal{G}$ with nodes $[d] = \{1, \dots, d\}$, directed edges, $\to$, and blunt edges, $\mapsto$. The blunt edges, introduced in [9; 16], will be used to represent the covariance structure of the noise process. In the literature on linear additive noise models, bidirected edges, rather than blunt edges, are conventionally used to represent the covariance structure of the noise. The primary reason for using a different notation is that a bidirected edge is also used to represent a common latent factor, and in the context of stochastic

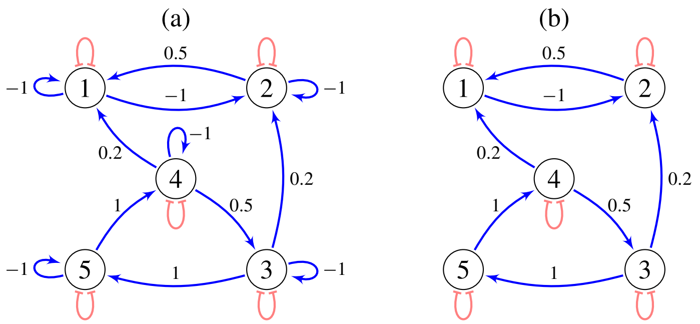

Figure 1. Mixed graph (a) representing a model with $d = 5$ variables and with a diagonal C-matrix. The edge labels of the directed (blue) edges are the values of the nonzero entries in the M-matrix, while all the blunt (light red) edges correspond to the diagonal entries of C all being 1. The mixed graph (b) is the same as (a) but with all directed self-loops removed. We call (b) the base graph of (a).

processes, a latent factor cannot in general be captured by a correlated noise process. See [9] for further details.

Definition 1.1. A pair of matrices (M, C) is compatible with a mixed graph $\mathcal{G}$ if $m_{ji} \neq 0$ implies $i \to j$ and $c_{ij} \neq 0$ implies $i \mapsto j$.

Example 1.2. We consider the specific model with $d = 5$ and, using $\cdot$ to denote zero entries,

$$M = \begin{pmatrix} -1 & 0.5 & \cdot & 0.2 & \cdot \\ -1 & -1 & 0.2 & \cdot & \cdot \\ \cdot & \cdot & -1 & 0.5 & \cdot \\ \cdot & \cdot & \cdot & -1 & 1 \\ \cdot & \cdot & 1 & \cdot & -1 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}.$$

This pair (M, C) is compatible with the graph $\mathcal{G} = ([5], E)$ as given in Figure 1(a). The eigenvalues of M are

$$-0.206, \quad -1 \pm 0.707i, \quad \text{and} \quad -1.397 \pm 0.687i,$$

with all real parts strictly negative, and M is thus stable. Solving the Lyapunov equation gives the steady-state covariance matrix (rounded to three decimals)

$$\Sigma = \begin{pmatrix} 0.496 & -0.091 & 0.123 & 0.207 & 0.151 \\ -0.091 & 0.594 & 0.013 & -0.038 & -0.005 \\ 0.123 & 0.013 & 0.838 & 0.676 & 0.647 \\ 0.207 & -0.038 & 0.676 & 1.412 & 0.912 \\ 0.151 & -0.005 & 0.647 & 0.912 & 1.147 \end{pmatrix}.$$

We use standard graph terminology; see, e.g., [8]. Specifically we let a walk be a sequence of (not necessarily unique) nodes where each node is connected to the next by an edge. A walk from i to j is directed if all edges are directed toward j. For example, in Figure 1(a), the walk $1 \leftarrow 4 \mapsto 4 \to 3 \to 5 \to 4$ is a walk from 1 to 4, and $3 \to 5 \to 5 \to 4 \to 1$ is a directed walk from 3 to 1.

We will by convention assume that a mixed graph includes all directed self-loops unless otherwise stated. It will, however, also be convenient to have a version of the graph with all directed self-loops removed, which we call the *base graph*.

Definition 1.3. The base graph $\mathcal{G}_0$ of a mixed graph $\mathcal{G}$ is obtained from $\mathcal{G}$ by removing all directed self-loops.

In Figure 1(b) we have the base graph of the mixed graph (a). A walk in the base graph is called a base walk. The walk $3 \to 5 \to 5 \to 4 \to 1$ in Figure 1(a) is not a base walk due to the self-loop $5 \to 5$, while $1 \leftarrow 4 \mapsto 4 \to 3 \to 5 \to 4$ is a base walk.

1.3. ***Linear additive noise models.*** It is instructive to compare and contrast our results for the Lyapunov equation with the well-known results for linear additive noise models. In the linear additive noise model, the random variable $X \in \mathbb{R}^d$ is given as a solution to the equation

$$X = BX + \varepsilon, \tag{6}$$

where B is a $d \times d$ matrix with $B_{ii} = 0$ and ε has mean zero and covariance matrix Ω. Consequently, the covariance matrix, Σ, of X solves the equation

$$(I - B)\Sigma(I - B)^T = \Omega. \tag{7}$$

When $I - B$ is invertible, the solution is unique and given by $\Sigma = (I - B)^{-1}\Omega(I - B)^{-T}$. Similarly to the Lyapunov equation, the matrix equation (7) is linear, and its solution yields a parametrization of the covariance matrix in terms of the entries of B and Ω. Zero-constraints on the entries of B and Ω can, moreover, be represented by a mixed graph according to Definition 1.1, just as for the Lyapunov equation. The covariance models resulting from the two equations are, however, quite different, though they also share some structural similarities.

2. Treks and trek rules

The trek rule for the solution of (7) is a well-known power series representation of the entries Σ_{ij} in the covariance matrix in terms of the entries of B and Ω. We derive in this section a corresponding trek rule for the solution of the Lyapunov equation (2). To this end, we need to define treks in a mixed graph.

Definition 2.1. A trek τ in a mixed graph $\mathcal{G}$ is a walk of the form

$$\tau : \underbrace{i \leftarrow \cdots \leftarrow i_1}_{n(\tau)} \leftarrow i_0 \mapsto j_0 \to \underbrace{j_1 \to \cdots \to j}_{m(\tau)}. \tag{8}$$

Here $n(\tau)$ and $m(\tau)$ denote the number of nodes to the left of i_0 and to the right of j_0, respectively. Let $l(\tau) = n(\tau) + m(\tau)$.

The trek given by (8) is said to be a trek from i to j. The pair (i_0, j_0) is the *top* of the trek, and if $j_0 = i_0$ we refer to i_0 as the top. The nodes $i_0, \ldots, i$ and $j_0, \ldots, j$ are the *left-hand* and *right-hand* sides of the trek, respectively. The top nodes are connected by a blunt edge, possibly a blunt self-loop, while all other edges are directed. The left-hand side of the trek forms a directed walk from i_0 to i with $n(\tau)+1$ nodes, and the right-hand side forms a directed walk from j_0 to j with $m(\tau)+1$ nodes. It is possible for a trek to have $n(\tau)=m(\tau)=0$, in which case the trek is just $i_0 \mapsto j_0$, which is possibly a blunt self-loop $i_0 \mapsto i_0$.

Definition 2.2. Let (M, C) be compatible with a mixed graph $\mathcal{G}$ and let τ be a trek in $\mathcal{G}$ with top (i_0, j_0). The weight of the trek τ is the product of the edge weights along the trek, that is,

$$\omega(M, C, \tau) = c_{i_0, j_0} \prod_{k \to l \in \tau} m_{lk}.$$

The walk $1 \leftarrow 4 \mapsto 4 \to 3 \to 5 \to 4$ in Figure 1(a) is an example of a trek τ from 1 to 4 with top 4 and with weight $\omega(M, C, \tau) = 0.2 \cdot 1 \cdot 0.5 \cdot 1 \cdot 1 = 0.1$.

With the definitions above, suppose that (B, Ω) is compatible with a mixed graph $\mathcal{G}$ and that B has spectral radius less than 1. Then the trek rule for the linear additive noise model, with Σ the solution of (7), is

$$\Sigma_{ij} = \sum_{\tau \in \mathcal{T}(i,j)} \omega(B, \Omega, \tau),$$

where $\mathcal{T}(i, j)$ denotes the set of all treks from i to j in $\mathcal{G}$. See, for example, Theorem 4.2 in [3] or Proposition 14.2.13 in [14]. The proof is a simple application of the Neumann series expansion of $(I-B)^{-1}$.

For the Lyapunov equation, the following variant of a trek rule was obtained as Proposition 2.2 in [16]. Suppose that (M, C) is compatible with the mixed graph $\mathcal{G}$, and M is stable and C is positive semidefinite. For any trek τ in $\mathcal{G}$ we introduce the monomial

$$\kappa(s, \tau) = \frac{s^{l(\tau)+1}}{((l(\tau)+1)n(\tau)!\, m(\tau)!)}$$

in the auxiliary variable $s \in \mathbb{R}$, and the solution of the Lyapunov equation (2) is then given by

$$\Sigma_{ij} = \lim_{s \to \infty} \sum_{\tau \in \mathcal{T}(i,j)} \kappa(s, \tau) \omega(M, C, \tau). \tag{9}$$

The series appearing in (9) is an infinite sum over all treks from i to j of the trek weights multiplied by the factors $\kappa(s, \tau)$ (not depending on M and C). Besides these factors, the series is similar to the classical trek rule. However, to obtain Σ_{ij}, we have to take the limit $s \to \infty$ of the resulting sum. This makes the representation (9) somewhat clumsy and difficult to use and interpret.

The new trek rule that we give below essentially interchanges the summation and limit operation by a translation of M by the identity matrix I (and by possibly rescaling M). Throughout we will let

$$\Lambda = M + I. \tag{10}$$

Since the Lyapunov equation is invariant to rescaling, we can without loss of generality assume that Λ has spectral radius strictly smaller than 1 when M is stable. If not, we can always rescale M and C to ensure this without changing Σ.

Proposition 2.3. *Let (M, C) be compatible with a mixed graph $\mathcal{G}$, and let M be stable and C be positive semidefinite. When $\Lambda = M + I$ has spectral radius strictly smaller than* 1, *the solution of the Lyapunov equation* (2) *is given by*

$$\Sigma_{ij} = \sum_{\tau \in \mathcal{T}(i,j)} 2^{-l(\tau)-1} \binom{l(\tau)}{n(\tau)} \omega(\Lambda, C, \tau), \tag{11}$$

where $\mathcal{T}(i, j)$ denotes the set of all treks from i to j in $\mathcal{G}$. The convergence of the series (11) *is absolute.*

Proof. Using the series expansion of the exponential function we have

$$\Sigma = \int_0^\infty e^{tM} C e^{tM^T} \, \mathrm{d}t = \int_0^\infty e^{-2t} e^{t\Lambda} C e^{t\Lambda^T} \, \mathrm{d}t = \int_0^\infty \sum_{n=0}^\infty \sum_{m=0}^\infty \frac{e^{-2t} t^n t^m}{n!\, m!} \Lambda^n C (\Lambda^m)^T \, \mathrm{d}t. \tag{12}$$

With $\|\cdot\|$ denoting any matrix norm, the assumption that the spectral radius of Λ is strictly smaller than 1, combined with the spectral radius formula, implies the existence of constants $K > 0$ and $r \in [0, 1)$ such that

$$\|\Lambda^n\| \leq K r^n.$$

This bound implies absolute convergence of (12), which justifies interchanging the summation and integration. Thus,

$$\begin{aligned}
\Sigma &= \sum_{n=0}^\infty \sum_{m=0}^\infty \Lambda^n C (\Lambda^m)^T \frac{1}{n!\, m!} \int_0^\infty t^{n+m} e^{-2t} \, \mathrm{d}t \\
&= \sum_{n=0}^\infty \sum_{m=0}^\infty \Lambda^n C (\Lambda^m)^T \, \frac{2^{-n-m-1} \Gamma(n+m+1)}{n!\, m!} \\
&= \sum_{n=0}^\infty \sum_{m=0}^\infty \Lambda^n C (\Lambda^m)^T 2^{-n-m-1} \binom{n+m}{n}.
\end{aligned}$$

Since $\mathcal{G}$, by convention, includes all directed self-loops, (Λ, C) is also compatible with $\mathcal{G}$. Now note that

$$(\Lambda^n C (\Lambda^m)^T)_{ij}$$

is precisely the sum over all trek weights, $\omega(\Lambda, C, \tau)$, for treks τ from i to j with $n + 1$ nodes on the left-hand side and $m + 1$ nodes on the right-hand side. By the arguments above, the series representation of Σ is absolutely convergent and the summation order does not matter. Therefore, the ij-th entry of Σ can be written as the trek representation (11). □

The trek weights are monomials in the M and C entries, and (11) provides a (generally infinite) power series representation of the covariances in terms of these entries. Some factors in the monomials are the diagonal entries $\Lambda_{ii} = m_{ii} + 1$, while the remaining factors are off-diagonal entries of Λ (that coincide with off-diagonal entries of M) and entries of C. It is possible, as we will show, to derive a trek rule where the contributions from the diagonal entries are separated from the off-diagonal entries.

Recall that the mixed graph $\mathcal{G}$ is assumed to include all directed self-loops, while the base graph $\mathcal{G}_0$ is obtained by removing the directed self-loops. A base trek is then a trek in the base graph $\mathcal{G}_0$, and it is a trek without any self-loops. Any trek in $\mathcal{G}$ can be regarded as a base trek combined with a total of $\alpha_1, \ldots, \alpha_d \geq 0$

self-loops on the left-hand side and $\beta_1, \dots, \beta_d \geq 0$ self-loops on the right-hand side. Here α_i and β_i denote the total number of self-loops for node i present on the trek on the left- and right-hand sides, respectively.

To elaborate on the definition of the self-loop counts α_i and β_i, consider the example trek

$$1 \leftarrow 1 \leftarrow 2 \leftarrow 2 \leftarrow 2 \leftarrow 1 \mapsto 1 \to 1 \to 2 \to 1 \to 1$$

with nodes $\{1, 2\}$. The corresponding base trek is $1 \leftarrow 2 \leftarrow 1 \mapsto 1 \to 2 \to 1$, and the trek includes one self-loop at node 1 and two self-loops at node 2 on the left-hand side and two self-loops at node 1 on the right-hand side, whence $\alpha_1 = 1,\ \alpha_2 = 2,\ \beta_1 = 2$, and $\beta_2 = 0$.

The base trek and self-loop counts do not uniquely determine a trek, but there are some constraints. If node i is not present on the left-hand side of the base trek, then $\alpha_i = 0$. If node i is present once on the left-hand side of the base trek, then α_i is exactly the number of self-loops at that position. In the example above, node 2 is present once on the left-hand side of the base trek, and $\alpha_2 = 2$ gives the number of self-loops of node 2 at that position. If node i is present more than once on the left-hand side of the base trek, then the total number of self-loops α_i can be partitioned in several ways among the different positions of node i, and similiarly for β_i on the right-hand side. This situation can only occur when the graph contains directed cycles beyond self-loops. In the example above, node 1 is present twice on the right-hand side of the base trek and $\beta_1 = 2$ can be partitioned in three different ways among the two occurrences of node 1.

For a base trek $\tau \in \mathcal{G}_0$ and multiindices $\alpha, \beta \in \mathbb{N}_0^d$ we define $\rho(\tau, \alpha, \beta)$ as the number of ways the α and β self-loops can be partitioned among the nodes of the base trek. If the base trek τ does not contain repeated nodes in its left- or right-hand sides, $\rho(\tau, \alpha, \beta) \in \{0, 1\}$, but when τ contains repeated nodes, it is possible that $\rho(\tau, \alpha, \beta) > 1$. Define also

$$\alpha_\bullet = \sum_{i=1}^{d} \alpha_i$$

and similarly for β.

Definition 2.4. For $\lambda_1, \dots, \lambda_d \in (-1, 1)$ and a base trek τ define

$$D(\lambda_1, \dots, \lambda_d, \tau) = \sum_{\alpha, \beta \in \mathbb{N}_0^d} \rho(\tau, \alpha, \beta) \binom{l(\tau) + \alpha_\bullet + \beta_\bullet}{n(\tau) + \alpha_\bullet} \prod_{i=1}^{d} \left(\frac{\lambda_i}{2}\right)^{\alpha_i + \beta_i}. \tag{13}$$

It may not be obvious that the series in (13) converges. Its absolute convergence does, however, follow from Theorem 2.5 below by taking $C = I$ and letting $m_{ii} \in [-1, 0)$ and $m_{ij} = \varepsilon$ for $i \neq j$ for a small $\varepsilon > 0$. Then $M + I$ has spectral radius strictly smaller than 1 by Perron–Frobenius theory, and $D(m_{11} + 1, \dots, m_{dd} + 1, \tau) < \infty$ for all base treks τ in the complete graph since (15) below is finite and all terms in that sum are positive.

Theorem 2.5. *Let (M, C) be compatible with a mixed graph $\mathcal{G}$, and let M be stable and C be positive semidefinite. If $\Lambda = M + I$ has spectral radius strictly smaller than 1 and $\Lambda_{ii} = m_{ii} + 1 \in (-1, 1)$, then*

$$\Sigma_{ij} = \sum_{\tau \in \mathcal{T}_0(i,j)} 2^{-l(\tau)-1} D(\Lambda_{11}, \dots, \Lambda_{dd}, \tau) \omega(\Lambda, C, \tau) \tag{14}$$

$$= \sum_{\tau \in \mathcal{T}_0(i,j)} 2^{-l(\tau)-1} D(m_{11} + 1, \dots, m_{dd} + 1, \tau) \omega(M, C, \tau), \tag{15}$$

where $\mathcal{T}_0(i, j)$ denotes the set of all treks from i to j in the base graph $\mathcal{G}_0$ of $\mathcal{G}$. In the special case where $m_{ii} = -1$ for all i, then

$$\Sigma_{ij} = \sum_{\tau \in \mathcal{T}_0(i,j)} 2^{-l(\tau)-1} \binom{l(\tau)}{n(\tau)} \omega(M, C, \tau). \tag{16}$$

Proof. Since the convergence of the series (11) is absolute, we can reorder the terms. The formula (14) follows by splitting the sum (11) into an outer sum over the base treks and an inner sum over self-loops. Specifically,

$$\begin{aligned}
\Sigma_{ij} &= \sum_{\tau \in \mathcal{T}(i,j)} 2^{-l(\tau)-1} \binom{l(\tau)}{n(\tau)} \omega(\Lambda, C, \tau) \\
&= \sum_{\tau \in \mathcal{T}_0(i,j)} \sum_{\alpha,\beta \in \mathbb{N}_0^d} \rho(\tau, \alpha, \beta) 2^{-l(\tau)-\alpha_\bullet-\beta_\bullet-1} \binom{l(\tau)+\alpha_\bullet+\beta_\bullet}{n(\tau)+\alpha_\bullet} \prod_{i=1}^{d} \Lambda_{ii}^{\alpha_i+\beta_i} \omega(\Lambda, C, \tau) \\
&= \sum_{\tau \in \mathcal{T}_0(i,j)} 2^{-l(\tau)-1} \left(\sum_{\alpha,\beta \in \mathbb{N}_0^d} \rho(\tau, \alpha, \beta) \binom{l(\tau)+\alpha_\bullet+\beta_\bullet}{n(\tau)+\alpha_\bullet} \prod_{i=1}^{d} \left(\frac{\Lambda_{ii}}{2} \right)^{\alpha_i+\beta_i} \right) \omega(\Lambda, C, \tau) \\
&= \sum_{\tau \in \mathcal{T}_0(i,j)} 2^{-l(\tau)-1} D(\Lambda_{11}, \dots, \Lambda_{dd}, \tau) \omega(\Lambda, C, \tau).
\end{aligned}$$

Moreover, (15) follows by the definition of $\Lambda_{ii} = m_{ii} + 1$ and by noting that $\omega(\Lambda, C, \tau) = \omega(M, C, \tau)$ for any base trek τ. Finally, if $m_{ii} = -1$ then $\Lambda_{ii} = m_{ii} + 1 = 0$ for $i = 1, \dots, d$, and since

$$D(0, \dots, 0, \tau) = \binom{l(\tau)}{n(\tau)}$$

for all base treks τ, (16) follows from (15). □

Note that for a base trek, the monomial $\omega(M, C, \tau)$ does not depend on the diagonal elements of M, and (15) provides a certain disentanglement of how the total covariance depends on the self-loop entries and the other edge entries.

Example 2.6. To illustrate the trek rule, we revisit Example 1.2, specifically the entry $\Sigma_{13} = 0.123$. Note that $m_{ii} = -1$ for all i. Table 1 lists all 12 base treks from 1 to 3 in the base graph in Figure 1(b) with $l(\tau) \leq 5$ together with their corresponding trek weights and term values contributing to Σ_{13} in the sum (16).

We see from Table 1 that

$$\sum_{\tau \in \mathcal{T}_0(1,3): l(\tau) \leq 5} 2^{-l(\tau)-1} \binom{l(\tau)}{n(\tau)} \omega(M, C, \tau) = 0.07890625,$$

which is still a bit from the limit value $\Sigma_{13} = 0.123$. Proceeding up to $l(\tau) = 10$ gives 74 base treks, and the sum of the corresponding terms is 0.10992737, while the sum of the 515 terms with $l(\tau) \leq 20$ is 0.12127330, which is accurate up to the second decimal. The purpose of this example *is not* to claim that the trek rule is useful for computing the solution of the Lyapunov equation in the general case. On the contrary, there is quite a lot of bookkeeping involved in computing the treks and the corresponding terms in the sum, and you may need a fairly large number of terms to get an accurate finite sum approximation.

trek	ω	ω factorization	l	n	$2^{-l-1}\binom{l}{n}\cdot\omega$	term value
$1 \leftarrow 4 \mapsto 4 \rightarrow 3$	0.1	$0.2\cdot 0.5$	2	1	$2^{-3}\cdot\binom{2}{1}\cdot 0.1$	0.02500000
$1 \leftarrow 2 \leftarrow 3 \mapsto 3$	0.1	$0.5\cdot 0.2$	2	2	$2^{-3}\cdot\binom{2}{2}\cdot 0.1$	0.01250000
$1 \leftarrow 4 \leftarrow 5 \leftarrow 3 \mapsto 3$	0.2	$0.2\cdot 1\cdot 1$	3	3	$2^{-4}\cdot\binom{3}{3}\cdot 0.2$	0.01250000
$1 \leftarrow 2 \leftarrow 1 \leftarrow 2 \leftarrow 3 \mapsto 3$	−0.05	$0.5\cdot(-1)\cdot 0.5\cdot 0.2$	4	4	$2^{-5}\cdot\binom{4}{4}\cdot(-0.05)$	−0.00156250
$1 \leftarrow 2 \leftarrow 1 \leftarrow 4 \mapsto 4 \rightarrow 3$	−0.05	$0.5\cdot(-1)\cdot 0.2\cdot 0.5$	4	3	$2^{-5}\cdot\binom{4}{3}\cdot(-0.05)$	−0.00625000
$1 \leftarrow 2 \leftarrow 3 \leftarrow 4 \mapsto 4 \rightarrow 3$	0.025	$0.5\cdot 0.2\cdot 0.5\cdot 0.5$	4	3	$2^{-5}\cdot\binom{4}{3}\cdot 0.025$	0.00312500
$1 \leftarrow 4 \leftarrow 5 \mapsto 5 \rightarrow 4 \rightarrow 3$	0.1	$0.2\cdot 1\cdot 1\cdot 0.5$	4	2	$2^{-5}\cdot\binom{4}{2}\cdot 0.1$	0.01875000
$1 \leftarrow 2 \leftarrow 3 \leftarrow 4 \leftarrow 5 \leftarrow 3 \mapsto 3$	0.05	$0.5\cdot 0.2\cdot 1\cdot 1\cdot 0.5$	5	5	$2^{-6}\cdot\binom{5}{5}\cdot 0.05$	0.00078125
$1 \leftarrow 2 \leftarrow 1 \leftarrow 4 \leftarrow 5 \leftarrow 3 \mapsto 3$	−0.1	$0.5\cdot(-1)\cdot 0.2\cdot 1\cdot 1$	5	5	$2^{-6}\cdot\binom{5}{5}\cdot(-0.1)$	−0.00156250
$1 \leftarrow 4 \leftarrow 5 \leftarrow 3 \leftarrow 4 \mapsto 4 \rightarrow 3$	0.05	$0.2\cdot 1\cdot 1\cdot 0.5\cdot 0.5$	5	4	$2^{-6}\cdot\binom{5}{4}\cdot 0.05$	0.00390625
$1 \leftarrow 2 \leftarrow 3 \mapsto 3 \rightarrow 5 \rightarrow 4 \rightarrow 3$	0.05	$0.5\cdot 0.2\cdot 1\cdot 1\cdot 0.5$	5	2	$2^{-6}\cdot\binom{5}{2}\cdot 0.05$	0.00781250
$1 \leftarrow 4 \mapsto 4 \rightarrow 3 \rightarrow 5 \rightarrow 4 \rightarrow 3$	0.05	$0.2\cdot 0.5\cdot 1\cdot 1\cdot 0.5$	5	1	$2^{-6}\cdot\binom{5}{1}\cdot 0.05$	0.00390625
Total						0.07890625

Table 1. Base treks τ with $l = l(\tau) \leq 5$ in Figure 1(b) and their corresponding weights $\omega = \omega(M, C, \tau)$ and term values $2^{-l-1}\binom{l}{n}\cdot\omega$ in the sum (16).

The purpose of the example is rather to illustrate how the trek rule breaks down the total covariance into contributions from the individual treks, with each term being a monomial in the edge entries.

3. Acyclic models

If the directed part of the base graph $\mathcal{G}_0$ is acyclic, and thus a DAG, we say that the model given by M and C is acyclic. Choosing any topological order of the nodes will make the M-matrix lower triangular. We assume throughout this section that the model is acyclic and that the nodes $1, \dots, d$ are ordered in a topological order such that M is lower triangular, that is,

$$M = \begin{pmatrix} m_{11} & \cdot & \cdot & \cdots & \cdot \\ m_{21} & m_{22} & \cdot & \cdots & \cdot \\ m_{31} & m_{32} & m_{33} & \cdots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{d1} & m_{d2} & m_{d3} & \cdots & m_{dd} \end{pmatrix}.$$

The diagonal entries of M are then the eigenvalues, and M is stable if and only $m_{ii} < 0$ for all i. By rescaling, we can assume that $m_{ii} \in [-1, 0)$ for all i so that $\Lambda_{ii} = m_{ii} + 1 \in [0, 1)$, in which case Λ also has spectral radius strictly less than 1.

3.1. ***Simplifying the trek rule for acyclic models.*** For an acyclic model any base trek has no nodes repeated on either side (though the same node can be present once on the left- and once on the right-hand

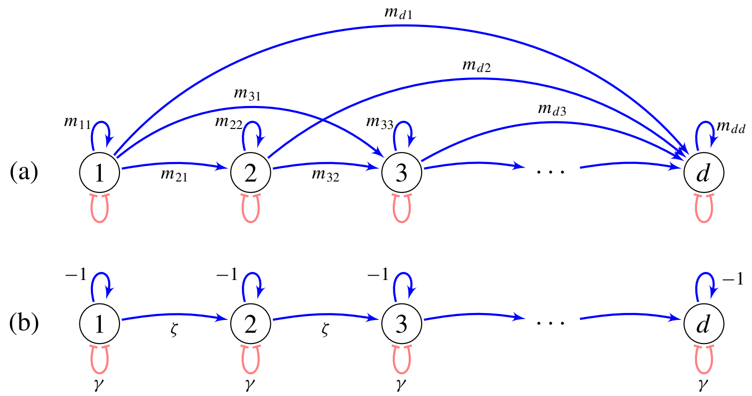

Figure 2. The mixed graph (a) for a general acyclic model with the nodes in a topological order, and the mixed graph (b) for the specific model in Example 3.1.

side). This means that $\rho(\tau, \alpha, \beta) \in \{0, 1\}$, and if we define

$$\mathcal{N}(\tau) = \{(\alpha, \beta) \in \mathbb{N}_0^d \times \mathbb{N}_0^d \mid \rho(\tau, \alpha, \beta) = 1\}$$

then

$$D(m_{11}+1, \dots, m_{dd}+1, \tau) = \sum_{(\alpha,\beta)\in\mathcal{N}(\tau)} \binom{l(\tau)+\alpha_\bullet+\beta_\bullet}{n(\tau)+\alpha_\bullet} \prod_{i=1}^{d} \left(\frac{m_{ii}+1}{2}\right)^{\alpha_i+\beta_i}. \tag{17}$$

The expression (17) does not appear to simplify further in general. However, if all the diagonal entries of M are equal, in which case we can assume them all equal to -1 by rescaling, the trek rule simplifies to (16). This special case is effectively the same as Proposition 4.3 in [1], which was obtained directly by an induction argument.

For any acyclic model, the directed part of $\mathcal{G}_0$ forms a DAG, and there are no loops but self-loops in $\mathcal{G}$. The number of base treks from i to j is thus finite, and (15) gives a *finite* sum representation of Σ_{ij}. That is, the representation (15) is a polynomial in the off-diagonal entries of M and the entries of C.

Example 3.1 (path model). Suppose $m_{ii} = -1$, $m_{(i+1)i} = \zeta \in \mathbb{R}$ and $c_{ii} = \gamma \geq 0$. All other entries are 0. That is,

$$M = \begin{pmatrix} -1 & \cdot & \cdot & \cdots & \cdot & \cdot \\ \zeta & -1 & \cdot & \cdots & \cdot & \cdot \\ \cdot & \zeta & -1 & \cdots & \cdot & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \cdots & -1 & \cdot \\ \cdot & \cdot & \cdot & \cdots & \zeta & -1 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \gamma & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \gamma & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \gamma & \cdots & \cdot & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \cdots & \gamma & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \gamma \end{pmatrix}.$$

The mixed graph for this model is shown in Figure 2(b). We have $\Lambda_{(i+1)i} = \zeta$ and all other entries of Λ are 0. The only base treks from i to j have top $i_0 \leq \min\{i, j\}$ and are of the form

$$\tau : i \leftarrow i-1 \leftarrow \cdots i_0+1 \leftarrow i_0 \mapsto i_0 \rightarrow i_0+1 \rightarrow \cdots \rightarrow j-1 \rightarrow j,$$

for which $n(\tau) = i - i_0,\ m(\tau) = j - i_0,\ l(\tau) = i + j - 2i_0$. This shows that

$$\Sigma_{ij} = \sum_{i_0=1}^{\min\{i,j\}} 2^{2i_0-i-j-1}\binom{i+j-2i_0}{i-i_0}\zeta^{i+j-2i_0}\gamma = \frac{\gamma}{2}\left(\frac{\zeta}{2}\right)^{i+j}\sum_{i_0=1}^{\min\{i,j\}}\left(\frac{4}{\zeta^2}\right)^{i_0}\binom{i+j-2i_0}{i-i_0}. \tag{18}$$

We see that $\Sigma_{i1} = \Sigma_{1i} = \gamma/2(\zeta/2)^{i-1}$, and by induction the following recursion holds for $i,\ j > 1$:

$$\Sigma_{ij} = \frac{\zeta}{2}\Sigma_{i-1,j} + \frac{\zeta}{2}\Sigma_{i,j-1} + \frac{\gamma}{2}\delta_{ij}, \tag{19}$$

where δ_{ij} is the Kronecker delta.

If we take $\zeta = \gamma = 2$, we get the simple expression

$$\Sigma_{ij} = \sum_{i_0=1}^{\min\{i,j\}}\binom{i+j-2i_0}{i-i_0}, \tag{20}$$

involving only a sum of binomial coefficients. Then $\Sigma_{i1} = \Sigma_{1i} = 1$, the recursion is

$$\Sigma_{ij} = \Sigma_{i-1,j} + \Sigma_{i,j-1} + \delta_{ij},$$

and the corresponding covariance matrix becomes

$$\Sigma = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & \cdots \\ 1 & 3 & 4 & 5 & 6 & \cdots \\ 1 & 4 & 9 & 14 & 20 & \cdots \\ 1 & 5 & 14 & 29 & 49 & \cdots \\ 1 & 6 & 20 & 49 & 99 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{21}$$

with the antidiagonal elements being similar to Pascal's triangle except that an extra 1 is added to the diagonal. This is OEIS A013580.

Example 3.2 (factor model). Another simple acyclic model is the factor model with $m_{11} = -1,\ m_{ii} \in [-1, 0)$ for $i = 2, \ldots, d$, C diagonal, and the only possible off-diagonal nonzero entries of M being m_{i1} for $i = 2, \ldots, d$. That is,

$$M = \begin{pmatrix} -1 & \cdot & \cdot & \cdots & \cdot \\ m_{21} & m_{22} & \cdot & \cdots & \cdot \\ m_{31} & \cdot & m_{33} & \cdots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{d1} & \cdot & \cdot & \cdots & m_{dd} \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} c_{11} & \cdot & \cdot & \cdots & \cdot \\ \cdot & c_{22} & \cdot & \cdots & \cdot \\ \cdot & \cdot & c_{33} & \cdots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \cdots & c_{dd} \end{pmatrix}.$$

The mixed graph for this model is shown in Figure 3. The only base treks are of the form $i \mapsto i$ and $i \leftarrow 1 \mapsto 1 \rightarrow j$ for $i,\ j > 1$. Using (15) we get for $i,\ j > 1$ that

$$\Sigma_{ij} = \begin{cases} d_i^0 c_{ii} + d_{ii} m_{i1}^2 c_{11} & \text{if } i = j, \\ d_{ij} m_{i1} m_{j1} c_{11} & \text{if } i \neq j, \end{cases} \tag{22}$$

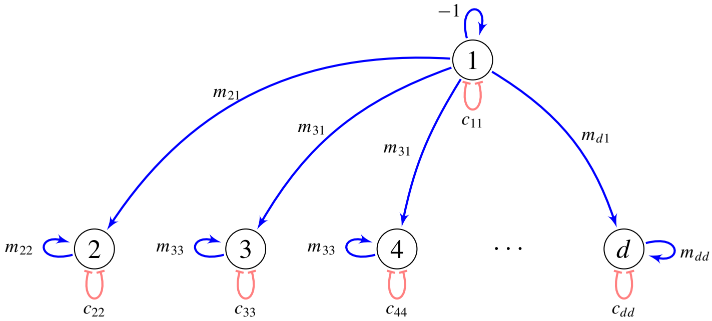

Figure 3. The mixed graph for the factor model in Example 3.2.

where

$$d_i^0 = 2^{-1} D(0, m_{22}+1, \dots, m_{dd}+1, i \mapsto i),$$
$$d_{ij} = 2^{-3} D(0, m_{22}+1, \dots, m_{dd}+1, i \leftarrow 1 \mapsto 1 \rightarrow j).$$

A somewhat tedious computation, similar to the one given in the proof of Proposition 3.3, shows that $d_i^0 = -1/(2m_{ii})$ and

$$d_{ij} = \frac{1}{2} \frac{m_{ii}+m_{jj}-2}{(1-m_{ii})(1-m_{jj})(m_{ii}+m_{jj})} > 0. \tag{23}$$

Another way to write the off-diagonal entries is

$$\Sigma_{ij} = \left(\frac{1}{2} - \frac{1}{m_{ii}+m_{jj}}\right) \frac{m_{i1}}{1-m_{ii}} \frac{m_{j1}}{1-m_{jj}} c_{11}, \quad i, j > 1,\ i \neq j, \tag{24}$$

which reveals that the model does not generally have the same low-rank structure as the classical one-factor model based on linear additive noise. Indeed, the tetrad $\Sigma_{ij}\Sigma_{kl} - \Sigma_{il}\Sigma_{kj}$ is nonzero unless $(m_{ii} + m_{jj})(m_{kk}+m_{ll}) = (m_{ii}+m_{ll})(m_{jj}+m_{kk})$; see [4; 5] for further details on algebraic invariants for factor models. It is an open problem to characterize invariants for factor models based on the Lyapunov equation.

3.2. ***A lower bound on the marginal variance.*** Consider an *acyclic* linear additive noise model with the nodes in a topological order such that

$$B = \begin{pmatrix} 0 & \cdot & \cdot & \cdots & \cdot \\ \beta_{21} & 0 & \cdot & \cdots & \cdot \\ \beta_{31} & \beta_{32} & 0 & \cdots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{d1} & \beta_{d2} & \beta_{d3} & \cdots & 0 \end{pmatrix} = \begin{pmatrix} B_{11} & 0 \\ B_{d1} & 0 \end{pmatrix},$$

where B_{11} is the $(d-1)\times(d-1)$ upper left block of B and $B_{d1} = (\beta_{d1}, \dots, \beta_{d(d-1)})$ consists of the $d-1$ first entries of the last row of B. It then follows directly from (6) that for this linear additive noise model,

$$X_d = \sum_{i=1}^{d-1} B_{di} X_i + \varepsilon_d = B_{d1}(X_1, \dots, X_{d-1})^T + \varepsilon_d.$$

Hence, since ε_d is independent of $(X_1, \dots, X_{d-1})^T$ and $\Omega_{dd} = \mathrm{Var}(\varepsilon_d)$, the marginal variance of X_d is

$$\Sigma_{dd} = \Omega_{dd} + B_{d1}\boldsymbol{\Sigma}_{11}(B_{d1})^T, \tag{25}$$

where $\boldsymbol{\Sigma}_{11}$ is the covariance matrix of $(X_1, \dots, X_{d-1})^T$.

The marginal variance of the last node d is by (25) decomposed into the variance of the residual error ε_d and the variance propagated forward from all parents of d. As these parents have variances that again decompose into residual error variances and the variances propagated from their parents, the marginal variances Σ_{ii} tend to increase along the topological order of the nodes. The paper [12] introduced *varsortability* as a measure that captures this phenomenon, and the authors demonstrated that commonly used simulation models typically have marginal variances that increase along the topological order. In this case, the topological order can be identified from the order of the marginal variances. See also [10] for related ideas.

In this section we explore if a similar phenomenon holds for the solution of the Lyapunov equation in the acyclic case, and we will, in particular, investigate if the variance Σ_{dd} has a decomposition similar to (25). Specifically, we derive a lower bound on Σ_{dd} in terms of the other (co)variances, which is similar to the identity (25). To state the result, we write the matrices M and Σ in block form as

$$M = \begin{pmatrix} M_{11} & 0 \\ M_{d1} & m_{dd} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{1d} \\ \boldsymbol{\Sigma}_{d1} & \Sigma_{dd} \end{pmatrix},$$

where M_{11} is a $(d-1) \times (d-1)$ matrix, and M_{d1} is a $(d-1)$-dimensional row vector, and similarly for Σ.

Proposition 3.3. *Let M be a lower triangular matrix with $m_{ii} \in [-1, 0)$ and $m_{ij} \geq 0$ for $j < i$, and let C be a diagonal positive semidefinite matrix. With Σ the solution of the Lyapunov equation, the following lower bound holds for the marginal variance of node d*:

$$\Sigma_{dd} \geq -\frac{c_{dd}}{2m_{dd}} + \frac{1}{2} M_{d1}\boldsymbol{\Sigma}_{11}(M_{d1})^T. \tag{26}$$

The proof is given in Section 3.3. Note that (26) does not hold generally if the off-diagonal entries m_{ij} are allowed to be negative.

Example 3.4. To illuminate what the lower bound in Proposition 3.3 says, and to illustrate the phenomenon that marginal variances tend to increase along the topological order, we reconsider the simple acyclic model from Example 3.1 with $\zeta = \gamma = 1$. Then

$$\Sigma_{ij} = 2^{-i-j-1} \sum_{i_0=1}^{\min\{i,j\}} 4^{i_0} \binom{i+j-2i_0}{i-i_0}. \tag{27}$$

From (27) it follows directly that, for $d \geq 2$,

$$\begin{aligned}\Sigma_{dd} &= 2^{-2d-1} \sum_{i_0=1}^{d} 4^{i_0} \binom{2(d-i_0)}{d-i_0} \\ &= 2^{-2(d-1)-1} \binom{2(d-1)}{d-1} + \underbrace{2^{-2d-1} \sum_{i_0=2}^{d} 4^{i_0} \binom{2(d-i_0)}{d-i_0}}_{=\Sigma_{(d-1)(d-1)}} > \Sigma_{(d-1)(d-1)},\end{aligned} \tag{28}$$

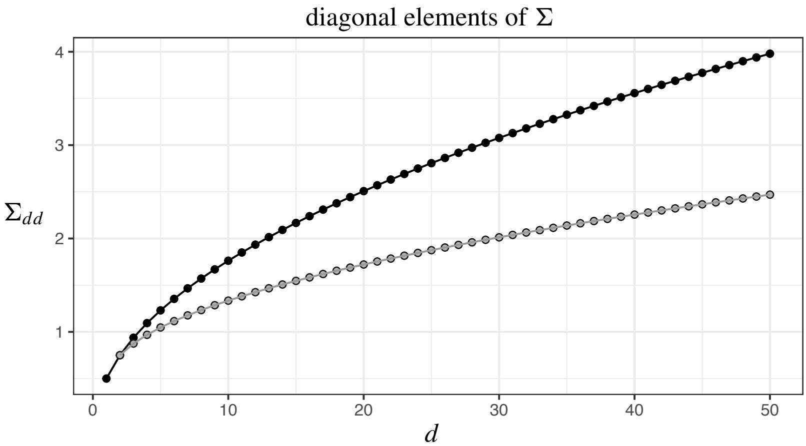

Figure 4. The variance Σ_{dd} in Example 3.4 given by (28) as a function of d (black) and the lower bound given by (29) (gray).

which shows that the marginal variance Σ_{dd} is strictly increasing as a function of d; see Figure 4. This is not surprising, since the sink node d accumulates the variances from all other nodes.

To compare Σ_{dd} to the lower bound in (26) we first derive a slightly different representation of Σ_{dd}. Using that

$$\binom{n}{k} = \frac{n(n-1)}{k(n-k)}\binom{n-2}{k-1}$$

for the second identity below, we get from (28) that, for $d \geq 2$,

$$\begin{aligned}
\Sigma_{dd} &= \frac{1}{2} + 2^{-2d-1}\sum_{i_0=1}^{d-1} 4^{i_0}\binom{2(d-i_0)}{d-i_0} \\
&= \frac{1}{2} + 2^{-2(d-1)-1}2^{-2}\sum_{i_0=1}^{d-1} 4^{i_0}\frac{2(d-i_0)(2(d-i_0)-1)}{(d-i_0)^2}\binom{2(d-1-i_0)}{d-1-i_0} \\
&= \frac{1}{2} + 2^{-2(d-1)-1}\sum_{i_0=1}^{d-1} 4^{i_0}\left(1 - \frac{1}{2(d-i_0)}\right)\binom{2(d-1-i_0)}{d-1-i_0}.
\end{aligned}$$

The second term above is $\Sigma_{(d-1)(d-1)}$ except from the factors $(1 - 1/(2(d-i_0)))$ within the sum. Observe that

$$\frac{1}{2} \leq \left(1 - \frac{1}{2(d-i_0)}\right) < 1,$$

with the factor being equal to $\frac{1}{2}$ for $i_0 = d-1$. This shows that

$$\frac{1}{2} + \frac{1}{2}\Sigma_{(d-1)(d-1)} \leq \Sigma_{dd} < \frac{1}{2} + \Sigma_{(d-1)(d-1)}.$$

To compute the lower bound in (26), we first note that $m_{dd} = -1$ and $c_{dd} = 1$; thus

$$-\frac{c_{dd}}{2m_{dd}} = \frac{1}{2}.$$

Moreover, $M_{d1} = (0, 0, \ldots, 1)$; thus $M_{d1}\mathbf{\Sigma}_{11}(M_{d1})^T = \Sigma_{(d-1)(d-1)}$. This shows that the lower bound in (26) is indeed

$$\Sigma_{dd} \geq \tfrac{1}{2} + \tfrac{1}{2}\Sigma_{(d-1)(d-1)}, \tag{29}$$

as we also found above. The computations in this example show that we cannot in general replace the factor $\frac{1}{2}$ in the bound by a larger constant, and certainly not by 1. The lower bound is also shown in Figure 4.

3.3. *Proof of Proposition 3.3.*

Before giving the proof, we will need a few auxiliary results.

Definition 3.5. For $a, b \in \mathbb{N}_0$ and $|z| < \frac{1}{2}$ define

$$H(a, b, z) = \sum_{n,m\in\mathbb{N}_0} \frac{(a+b+1)_{n+m+2}}{(a+1)_{n+1}(b+1)_{m+1}} z^{n+m}. \tag{30}$$

Here $(a)_n = a(a+1)\cdots(a+n-1)$ denotes the rising Pochhammer symbol.

Proposition 3.6. *Let $\tau = d \leftarrow \tilde{\tau} \rightarrow d$ be a base trek from d to d. Then*

$$D(\lambda_1, \ldots, \lambda_d, \tau) = \sum_{\alpha,\beta\in\mathcal{N}(\tilde{\tau})} H\left(n(\tilde{\tau}) + \alpha_\bullet, m(\tilde{\tau}) + \beta_\bullet, \frac{\lambda_d}{2}\right)\binom{l(\tilde{\tau})+\alpha_\bullet+\beta_\bullet}{n(\tilde{\tau})+\alpha_\bullet}\prod_{i=1}^{d-1}\left(\frac{\lambda_i}{2}\right)^{\alpha_i+\beta_i}.$$

Proof. By Definition 2.4 we have

$$\begin{aligned} D(\lambda_1, \ldots, \lambda_d, \tau) &= \sum_{\alpha,\beta\in\mathcal{N}(\tau)} \binom{l(\tau)+\alpha_\bullet+\beta_\bullet}{n(\tau)+\alpha_\bullet}\prod_{i=1}^{d}\left(\frac{\lambda_i}{2}\right)^{\alpha_i+\beta_i} \\ &= \sum_{\alpha,\beta\in\mathcal{N}(\tilde{\tau})}\sum_{\alpha_d,\beta_d\in\mathbb{N}_0} \binom{l(\tilde{\tau})+\alpha_\bullet+\beta_\bullet+\alpha_d+\beta_d+2}{n(\tilde{\tau})+\alpha_\bullet+\alpha_d+1}\prod_{i=1}^{d}\left(\frac{\lambda_i}{2}\right)^{\alpha_i+\beta_i}. \end{aligned}$$

Recall that $l(\tilde{\tau}) = n(\tilde{\tau}) + m(\tilde{\tau})$, so with $a = n(\tilde{\tau}) + \alpha_\bullet$ and $b = m(\tilde{\tau}) + \beta_\bullet$ we have the following identity for the binomial coefficient:

$$\begin{aligned} \binom{l(\tilde{\tau})+\alpha_\bullet+\beta_\bullet+\alpha_d+\beta_d+2}{n(\tilde{\tau})+\alpha_\bullet+\alpha_d+1} &= \binom{a+b+\alpha_d+\beta_d+2}{a+\alpha_d+1} = \frac{(a+b+\alpha_d+\beta_d+2)!}{(a+\alpha_d+1)!\,(b+\beta_d+1)!} \\ &= \frac{(a+b+1)_{\alpha_d+\beta_d+2}}{(a+1)_{\alpha_d+1}(b+1)_{\beta_d+1}}\frac{(a+b)!}{a!\,b!} = \frac{(a+b+1)_{\alpha_d+\beta_d+2}}{(a+1)_{\alpha_d+1}(b+1)_{\beta_d+1}}\binom{l(\tilde{\tau})+\alpha_\bullet+\beta_\bullet}{n(\tilde{\tau})+\alpha_\bullet}. \end{aligned}$$

Plugging this into the sum above and using the definition of H in (30) gives

$$\begin{aligned} D(\lambda_1, \ldots, \lambda_d, \tau) &= \sum_{\alpha,\beta\in\mathcal{N}(\tilde{\tau})}\sum_{\alpha_d,\beta_d\in\mathbb{N}_0} \frac{(a+b+1)_{\alpha_d+\beta_d+2}}{(a+1)_{\alpha_d+1}(b+1)_{\beta_d+1}}\left(\frac{\lambda_d}{2}\right)^{\alpha_d+\beta_d}\binom{l(\tilde{\tau})+\alpha_\bullet+\beta_\bullet}{n(\tilde{\tau})+\alpha_\bullet}\prod_{i=1}^{d-1}\left(\frac{\lambda_i}{2}\right)^{\alpha_i+\beta_i} \\ &= \sum_{\alpha,\beta\in\mathcal{N}(\tilde{\tau})} H\left(n(\tilde{\tau})+\alpha_\bullet, m(\tilde{\tau})+\beta_\bullet, \frac{\lambda_d}{2}\right)\binom{l(\tilde{\tau})+\alpha_\bullet+\beta_\bullet}{n(\tilde{\tau})+\alpha_\bullet}\prod_{i=1}^{d-1}\left(\frac{\lambda_i}{2}\right)^{\alpha_i+\beta_i}. \end{aligned}$$

□

Lemma 3.7. *For $a, b \in \mathbb{N}_0$ and $z \in \left[0, \frac{1}{2}\right)$ we have*

$$H(a, b, z) \geq 2. \tag{31}$$

Proof. Note that $H(a,b,z) = H(b,a,z)$, so we may assume that $b \geq a$. Since $z \geq 0$, all terms in the sum defining $H(a,b,z)$ are nonnegative and we find that

$$H(a,b,z) = \sum_{n,m\in\mathbb{N}_0} \frac{(a+b+1)_{n+m+2}}{(a+1)_{n+1}(b+1)_{m+1}} z^{n+m} \geq \frac{(a+b+1)_2}{(a+1)_1(b+1)_1}$$
$$= \frac{(a+b+2)(a+b+1)}{(a+1)(b+1)} = \left(1+\frac{b+1}{a+1}\right)\left(1+\frac{a}{b+1}\right) \geq 2. \qquad \square$$

Proof of Proposition 3.3. The base treks from d to d are either of the form $d \mapsto d$ or $d \leftarrow \tilde{\tau} \rightarrow d$ for a base trek $\tilde{\tau}$ from i to j with $i, j < d$. Since $\omega(M, C, d \mapsto d) = c_{dd}$, we have the following representation of Σ_{dd}:

$$\Sigma_{dd} = \frac{c_{dd}}{2} \sum_{\alpha_d,\beta_d\in\mathbb{N}_0} \binom{\alpha_d+\beta_d}{\alpha_d}\left(\frac{m_{dd}+1}{2}\right)^{\alpha_d+\beta_d}$$
$$+ \underbrace{\sum_{i=1}^{d-1}\sum_{j=1}^{d-1}\sum_{\tilde{\tau}\in\mathcal{T}_0(i,j)} 2^{-l(\tilde{\tau})-2-1} D(m_{11}+1,\ldots,m_{dd}+1, d\leftarrow\tilde{\tau}\rightarrow d)\,\omega(M,C,d\leftarrow\tilde{\tau}\rightarrow d)}_{=R}.$$

Using the negative binomial and geometric series, the sum in the first term equals

$$\sum_{\alpha_d,\beta_d\in\mathbb{N}_0} \binom{\alpha_d+\beta_d}{\alpha_d}\left(\frac{m_{dd}+1}{2}\right)^{\alpha_d+\beta_d} = \sum_{\beta_d\in\mathbb{N}_0}\left(\frac{m_{dd}+1}{2}\right)^{\beta_d}\sum_{\alpha_d\in\mathbb{N}_0}\binom{\alpha_d+\beta_d}{\alpha_d}\left(\frac{m_{dd}+1}{2}\right)^{\alpha_d}$$
$$= \sum_{\beta_d\in\mathbb{N}_0}\left(\frac{m_{dd}+1}{2}\right)^{\beta_d}\left(1-\frac{m_{dd}+1}{2}\right)^{-\beta_d-1}$$
$$= \frac{2}{1-m_{dd}}\sum_{\beta_d\in\mathbb{N}_0}\left(\frac{1+m_{dd}}{1-m_{dd}}\right)^{\beta_d} = \frac{2}{(1-m_{dd})}\frac{1}{\left(1-\frac{1+m_{dd}}{1-m_{dd}}\right)} = -\frac{1}{m_{dd}}.$$

For the second term R, we first note that for $\tilde{\tau} \in \mathcal{T}_0(i,j)$,

$$\omega(M,C,d\leftarrow\tilde{\tau}\rightarrow d) = m_{di}m_{dj}\,\omega(M,C,\tilde{\tau}).$$

By Proposition 3.6 and Lemma 3.7 we have

$$D(m_{11}+1,\ldots,m_{dd}+1,d\leftarrow\tilde{\tau}\rightarrow d) \geq 2\sum_{\alpha,\beta\in\mathcal{N}(\tilde{\tau})}\binom{l(\tilde{\tau})+\alpha_\bullet+\beta_\bullet}{n(\tilde{\tau})+\alpha_\bullet}\prod_{k=1}^{d-1}\left(\frac{m_{kk}+1}{2}\right)^{\alpha_k+\beta_k}$$
$$= 2D(m_{11}+1,\ldots,m_{dd}+1,\tilde{\tau}).$$

As $m_{ij} \geq 0$ for $j < i$, $\omega(M,C,\tilde{\tau}) \geq 0$ for a base trek $\tilde{\tau}$, and using the above inequality within the sum in the second term R gives

$$R \geq \frac{1}{2}\sum_{i=1}^{d-1}\sum_{j=1}^{d-1} m_{di}m_{dj}\sum_{\tilde{\tau}\in\mathcal{T}_0(i,j)} 2^{-l(\tilde{\tau})-1} D(m_{11}+1,\ldots,m_{dd}+1,\tilde{\tau})\,\omega(M,C,\tilde{\tau})$$
$$= \frac{1}{2}\sum_{i=1}^{d-1}\sum_{j=1}^{d-1} m_{di}m_{dj}\Sigma_{ij} = \frac{1}{2}M_{d1}\mathbf{\Sigma}_{11}(M_{d1})^T.$$

Combining these results gives

$$\Sigma_{dd} = -\frac{c_{dd}}{2m_{dd}} + R \geq -\frac{c_{dd}}{2m_{dd}} + \frac{1}{2} M_{d1} \mathbf{\Sigma}_{11} (M_{d1})^T. \qquad \square$$

Remark 3.8. The proof of Proposition 3.3 shows that we can always write

$$\Sigma_{dd} = -\frac{c_{dd}}{2m_{dd}} + R,$$

which is directly comparable to (25) for the linear additive noise model. The general expression for R is somewhat complicated, but the proof above shows that it can be lower bounded by a simpler expression whenever $m_{ij} \geq 0$. It is an open question if other assumptions lead to either bounds or simplifications of R.

4. Concluding remarks

Trek rules are useful for linking the entries of the solution to the Lyapunov equation to graphical properties of the underlying mixed graph. A straightforward observation is that $\Sigma_{ij} = 0$ if there is no trek from i to j. The general trek rule as stated in Proposition 2.3 and Theorem 2.5 is, however, more complicated than the well-known trek rule for the linear additive noise model, and even in the acyclic case it does not simplify completely to a polynomial representation in general. This is due to the self-loops. For acyclic models the trek rule in Theorem 2.5 is, nevertheless, a polynomial in the off-diagonal entries of M and the entries of C.

The trek rule is not generally useful for the numerical computation of the solution to the Lyapunov equation, but it can be used in special cases to derive either explicit formulas for solutions or to derive bounds on entries of the solution. We have illustrated this for the acyclic models in Examples 3.2 and 3.4, where much simpler polynomial trek rules are possible. As an example of a nontrivial application of the trek rule, we derived the lower bound in Proposition 3.3 on the marginal variance Σ_{dd} for a stable acyclic model with a diagonal C matrix and off-diagonal drift entries being nonnegative.

Acknowledgements

The trek rules in Proposition 2.3 and Theorem 2.5 were discovered while I was visiting Mathias Drton at the Technical University of Munich in November and December 2021. I am grateful to Mathias for hosting me. More recently, Alexander Reisach mentioned to me that lower bounds on the variances for acyclic models would be of interest, and the trek rule seemed well suited for this purpose. I am grateful to Alexander for this suggestion. I also thank Mathias Drton, Sarah Lumpp and the three anonymous reviewers for a number of suggestions that improved the manuscript. I was supported by a research grant (NNF20OC0062897) from Novo Nordisk Fonden.

References

[1] T. Boege, M. Drton, B. Hollering, S. Lumpp, P. Misra, and D. Schkoda, "Conditional independence in stationary distributions of diffusions", *Stochastic Process. Appl.* **184** (2025), art. id. 104604.

[2] P. Dettling, R. Homs, C. Améndola, M. Drton, and N. R. Hansen, "Identifiability in continuous Lyapunov models", *SIAM J. Matrix Anal. Appl.* **44**:4 (2023), 1799–1821.

[3] M. Drton, "Algebraic problems in structural equation modeling", pp. 35–86 in *The* 50*th anniversary of Gröbner bases*, edited by T. Hibi, Adv. Stud. Pure Math. **77**, Math. Soc. Japan, Tokyo, 2018.

[4] M. Drton, B. Sturmfels, and S. Sullivant, "Algebraic factor analysis: tetrads, pentads and beyond", *Probab. Theory Related Fields* **138**:3-4 (2007), 463–493.

[5] M. Drton, A. Grosdos, I. Portakal, and N. Sturma, "Algebraic sparse factor analysis", *SIAM J. Appl. Algebra Geom.* **9**:2 (2025), 279–309.

[6] Z. Gajic and M. T. J. Qureshi, *Lyapunov matrix equation in system stability and control*, Mathematics in Science and Engineering **195**, Academic Press, San Diego, CA, 1995.

[7] M. Jacobsen, "A brief account of the theory of homogeneous Gaussian diffusions in finite dimensions", pp. 86–94 in *Frontiers in Pure and Applied Probability* (Turku, Finland), edited by H. Niemi et al., VSP/TVP, Utrecht, The Netherlands/Moscow, Russia, 1993.

[8] M. Maathuis, M. Drton, S. Lauritzen, and M. Wainwright (editors), *Handbook of graphical models*, CRC Press, Boca Raton, FL, 2019.

[9] S. W. Mogensen and N. R. Hansen, "Graphical modeling of stochastic processes driven by correlated noise", *Bernoulli* **28**:4 (2022), 3023–3050.

[10] G. Park, "Identifiability of additive noise models using conditional variances", *J. Mach. Learn. Res.* **21** (2020), art. id. 75.

[11] G. A. Pavliotis, *Stochastic processes and applications*: *diffusion processes, the Fokker–Planck and Langevin equations*, Texts in Applied Mathematics **60**, Springer, 2014.

[12] A. G. Reisach, C. Seiler, and S. Weichwald, "Beware of the simulated DAG! Causal discovery benchmarks may be easy to game", in *Advances in neural information processing systems* 34, edited by A. Beygelzimer et al., 2021.

[13] V. Simoncini, "Computational methods for linear matrix equations", *SIAM Rev.* **58**:3 (2016), 377–441.

[14] S. Sullivant, *Algebraic statistics*, Graduate Studies in Mathematics **194**, Amer. Math. Soc., Providence, RI, 2018.

[15] S. Sullivant, K. Talaska, and J. Draisma, "Trek separation for Gaussian graphical models", *Ann. Statist.* **38**:3 (2010), 1665–1685.

[16] G. Varando and N. Hansen, "Graphical continuous Lyapunov models", pp. 989–998 in *Proceedings of the* 36*th Conference on Uncertainty in Artificial Intelligence*, edited by J. Peters and D. Sontag, Proceedings of Machine Learning Research **124**, 2020.

Received 2024-12-02. Revised 2025-07-02. Accepted 2025-07-17.

NIELS RICHARD HANSEN: niels.r.hansen@math.ku.dk
Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

See inside back cover or msp.org/astat for submission instructions.

The subscription price for 2025 is US $250/year for the electronic version, and $310/year (+$15, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebraic Statistics (ISSN 2693-3004 electronic, 2693-2997 printed) at Mathematical Sciences Publishers, 2000 Allston Way # 59, Berkeley, CA 94701-4004, is published continuously online.

AStat peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

Algebraic Statistics

2025 16 : 2