SIAM J. Appl. Algebra Geometry
Vol. 10, No. 2, pp. 261–302


# Complete Collineations for Maximum Likelihood Estimation*

Gergely Bérczi†, Eloise Hamilton‡, Philipp Reichenbach§, and Anna Seigal¶

**Abstract.** We import the algebro-geometric notion of a complete collineation into the study of maximum likelihood estimation in directed Gaussian graphical models. A complete collineation produces a perturbation of sample data, which we call a stabilization of the sample. While a maximum likelihood estimate (MLE) may not exist or be unique given sample data, it is always unique given a stabilization. We relate the MLE given a stabilization to the MLE given original sample data, when one exists, providing necessary and sufficient conditions for the MLE given a stabilization to be one given the original sample. We show that the MLE has a well-defined limit as the stabilization of a sample tends to the original sample, and that the limit is an MLE given the original sample, when one exists. Finally, we address the question of which MLEs given a sample can arise as such limits.

**Key words.** complete collineation, maximum likelihood estimation, directed Gaussian graphical model, sample stabilization, Tikhonov regularization

**MSC codes.** 62F10, 62H22, 62R01, 14L30, 14E15

**DOI.** 10.1137/24M1694793

## 1. Introduction.

We study maximum likelihood estimation in directed Gaussian graphical models. The existence or uniqueness of a maximum likelihood estimate (MLE) given observed data is known to depend on the number of samples and their genericity [6, 11, 15]. Several approaches have been proposed to compute MLEs given data that is insufficient or nongeneric, including regularization [7], dividing the model into subnetworks [32], and reducing the number of parameters via symmetries [18]. We propose a new approach to this problem based on the algebro-geometric concept of a complete collineation. The idea is that if an MLE does not exist, or is not unique, given observed data, the data may be perturbed so that a unique MLE can be found. This unique MLE can then be used to single out an MLE given the initial data, if one exists, or to assign a statistically meaningful MLE to it. The key to proving these results is to have the right notion of perturbation. We propose that perturbations arising from complete collineations are a natural choice.

*Received by the editors October 2, 2024; accepted for publication (in revised form) January 6, 2026; published electronically June 10, 2026.
https://doi.org/10.1137/24M1694793
**Funding:** The work of the first author was supported by Aarhus University starting grant AUFF-29289. The work of the third author was funded by the European Research Council (ERC) under the European Horizon 2020 research and innovation programme (grant agreement 787840). The work of the fourth author was supported by National Science Foundation grant DMR2011754 and by the Sloan Foundation.
†Department of Mathematics, Aarhus University, DK-8000 Aarhus, Denmark (gergely.berczi@math.au.dk).
‡Newnham College, University of Cambridge, CB3 9DF Cambridge, UK (eloise.hamilton@newn.cam.ac.uk).
§Institut für Mathematik, Technische Universität Berlin, 10587 Berlin, Germany (ph-reichenbach@t-online.de).
¶School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (aseigal@seas.harvard.edu).

The distributions we consider are mean-centered $m$-dimensional Gaussians, for some dimension $m$. Our models are parametrized by certain subsets of the cone of $m \times m$ positive definite matrices. Sample data can be collected into a matrix $Y$ of size $n \times m$, where $n$ is the number of observations. The existence or uniqueness of the MLE given $Y$ depends on the model and the properties of the matrix $Y$. For example, if the model is the full cone of positive definite matrices, the MLE given $Y$ exists and is unique if and only if $Y$ has full column rank. This cannot occur for $n < m$, but it occurs generically once $n \geq m$.

In this paper, we think of a sample $Y \in \mathbb{R}^{n\times m}$ as a linear map $\mathbb{R}^m \to \mathbb{R}^n$. If the MLE does not exist or is not unique given $Y$, then $Y$ is a degenerate linear map, i.e., it does not have maximal rank. We adopt the view that $Y$ should be considered not on its own but together with the additional data that a complete collineation provides. This additional data can be packaged into a new sample $\widetilde{Y}$, which we call a stabilization of $Y$, such that the MLE is unique given $\widetilde{Y}$. We think of $\widetilde{Y}$ as a "well-behaved" degeneration of a sample corresponding to a nondegenerate linear map, i.e., a map of maximal rank.

**1.1. Why complete collineations?** While degenerate linear maps are the most obvious candidates for degenerations of nondegenerate linear maps, an important lesson originating in the work of late 19th century geometers is that they are not the right notion of degeneration from the point of view of enumerative geometry [29]. The key insight from this line of work is that degenerations should carry more information than just that of a degenerate linear map; the key contribution lies in identifying exactly what this information should be. Complete collineations encode the necessary information.

A collineation between two projective spaces $\mathbb{P}(V)$ and $\mathbb{P}(W)$ is the scalar equivalence class $[f]$ of a nondegenerate linear map $f : V \to W$. By convention, we map from the smaller projective space to the larger one, so we assume that $\dim V \leq \dim W$. The term collineation originates in the fact that the map $[f]$ sends collinear points in $\mathbb{P}(V)$ to collinear points in $\mathbb{P}(\operatorname{im} f)$ bijectively. In fact, the map $[f]$ not only maps lines to lines but also maps $i$-planes to $i$-planes, via associated maps $[\wedge^i f] : \mathbb{P}(\wedge^i V) \to \mathbb{P}(\wedge^i W)$ for $i$ from 1 to $\dim V$. By contrast, if $f$ is a degenerate map from $V$ to $W$, then while the equivalence class $[f]$ is well-defined, the equivalence classes $[\wedge^i f]$ may no longer be well-defined, since $i$-planes may collapse to $j$-planes for some $j < i$. Complete collineations are degenerations of collineations that preserve the higher-order information of the $i$-plane to $i$-plane correspondences $[\wedge^i f]$ [29, p. 254]. Concretely, a complete collineation from $\mathbb{P}(V)$ to $\mathbb{P}(W)$ with $\dim V \leq \dim W$ is a finite sequence $([f_1], \ldots, [f_t])$ of equivalence classes of linear maps, where $f_1 : V \to W$, $f_i : \ker f_{i-1} \to \operatorname{coker} f_{i-1}$ for $i \geq 2$ and $f_t$ is the first nondegenerate map; see section 2 for details.

Given a degenerate sample, or equivalently a linear map $f : \mathbb{R}^m \to \mathbb{R}^n$, we define a stabilization of $f$, or $f$-stabilization, to be $\widetilde{f} = f + f'$ where the perturbation $f'$ comes from a complete collineation between $\mathbb{P}(\mathbb{R}^m)$ and $\mathbb{P}(\mathbb{R}^n)$ with first term $[f]$. We will always reduce to the case where $m \leq n$ (see section 4.1). A precise definition of sample stabilizations is in section 5.1.

Properties of complete collineations ensure that the MLE is unique given $\widetilde{f}$. While various conditions could be placed on $f'$ to ensure that the MLE given $\widetilde{f}$ is unique, our conditions have the advantage that the MLE given $\widetilde{f}$ and the MLEs given $f$, if they exist, are closely related. In this paper, we use complete collineations to resolve nonidentifiability of the MLE.

**1.2. Main results.** Fix a directed acyclic graph (DAG) $\mathcal{G}$ on vertices $\{1,\ldots,m\}$ with edge set $E$. A directed edge from $j$ to $i$ is denoted by $j \to i$. The acyclicity rules out directed cycles $j \to i \to \cdots \to k \to j$. The statistical models we consider are directed Gaussian graphical models on DAGs. We call these DAG models, for short. They have $m+|E|$ parameters, one for each edge and one for each vertex. The MLE given data $Y$ consists of estimates for all of these parameters; see section 3 for details. We work throughout over a field $\mathbb{K}$ which can be taken to be either $\mathbb{R}$ or $\mathbb{C}$. Our results hold over both fields.

Our first main result relates the MLE given a stabilization to an MLE given an original sample. We denote the span of a set of vectors $\{v_1,\ldots,v_k\}$ by $\langle v_1,\ldots,v_k\rangle$ and the projection of a vector $v$ onto a linear space $L$ by $\pi_L(v)$. A child vertex in $\mathcal{G}$ is a vertex $i$ in a DAG $\mathcal{G}$ with a parent in $\mathcal{G}$, i.e., with an edge $j \to i$ in $\mathcal{G}$ for some vertex $j$.

Theorem 1.1. *Fix a DAG $\mathcal{G}$ and a sample $f \in \mathbb{K}^{n\times m}$. Let $\widetilde{f} = f + f'$ denote a stabilization of $f$. Let $f_i$ and $v_i$ denote the columns of $f$ and $f'$, respectively. We have the following results concerning maximum likelihood estimation in the DAG model on $\mathcal{G}$:*

(a) *the MLE given $\widetilde{f}$ is unique;*

(b) *the MLE given $\widetilde{f}$ is an MLE given $f$ if and only if*

$$v_i \in \langle v_j : j \to i\rangle \text{ and } \overline{f_i} + v_i \in \langle f_j + v_j : j \to i\rangle,$$

*where $\overline{f_i} := \pi_{\langle f_j : j \to i\rangle}(f_i)$, for all child vertices $i$;*

(c) *the MLE given $\widetilde{f}(\epsilon) := f + \epsilon f'$ is unique for all $\epsilon \neq 0$, and has a well-defined limit as $\epsilon$ tends to zero, called the* limit MLE given $\widetilde{f}$, *which can be described explicitly (see Theorem 8.3(c));*

(d) *the limit MLE given $\widetilde{f}$ is an MLE given $f$, if one exists.*

Our second main result addresses when an MLE given $f$ is the MLE or limit MLE given a stabilization of $f$.

Theorem 1.2. *Fix a DAG $\mathcal{G}$ and a sample $f \in \mathbb{K}^{n\times m}$. Let $\alpha$ denote an MLE given $f$ in the DAG model on $\mathcal{G}$. Then we have the following:*

(a) *there is a locally Zariski closed subvariety $X_f \subseteq \mathbb{K}^{n\times m}$ parametrizing stabilizations of $f$;*

(b) *there is a Zariski closed subvariety $X_{f,\alpha} \subseteq X_f$, with defining equations given in (6.5), parameterizing stabilizations $\widetilde{f}$ of $f$ such that the MLE given $\widetilde{f}$ is $\alpha$, so that*

$$X_{f,\alpha} \neq \emptyset$$

*if and only if $\alpha$ is the MLE given an $f$-stabilization;*

(c) *there is a Zariski closed subvariety $X^{\lim}_{f,\alpha} \subseteq X_f$, with defining equations given in (8.2), parameterizing stabilizations $\widetilde{f}$ of $f$ such that the limit MLE given $\widetilde{f}$ is $\alpha$, so that*

$$X^{\lim}_{f,\alpha} \neq \emptyset$$

*if and only if $\alpha$ is the limit MLE given an $f$-stabilization.*

We apply the above results to DAG models on a star-shaped graph, which in this paper refers to a connected DAG with a unique child vertex, as in Figure 1. Such models study the linear dependence of one variable on all others.

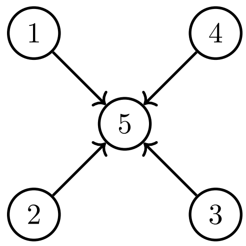



**Figure 1.** *Star-shaped graph with $m=5$ vertices.*

Theorem 1.3. *Consider a star-shaped DAG $\mathcal{G}$ and a sample $f$. If the MLE given $f$ exists in the DAG model on $\mathcal{G}$, then the MLE given any stabilization $\widetilde{f}$ of $f$ is the unique MLE given $f$ of minimal 2-norm.*

Theorem 1.3 exhibits a model where exactly one of the MLEs given a sample $f$ can be obtained from the MLE or limit MLE given a stabilization of $f$. The unique MLE singled out should be viewed as the "preferred" one, thus resolving the problem of nonidentifiability of the MLE given $f$. For other DAGs, different stabilizations may give different MLEs, and resolving nonidentifiability of the MLE relies on a choice of stabilization. We describe a sampling algorithm which constructs an $f$-stabilization from any sample $f$ via a finite sequence of samples in section 5.3. Each sample is obtained by sampling linear combinations of the nodes of $\mathcal{G}$, with the number of samples needed strictly decreasing at each step.

**1.3. Related work.** We are not aware of any existing work connecting complete collineations to algebraic statistics. Nevertheless, in a different direction, the closely related concept of complete quadrics has recently been used in algebraic statistics to study particular classes of Gaussian models [21, 10, 19]. Complete quadrics are defined analogously to complete collineations, with the additional constraint that $\dim V = \dim W$ and that $f$ is symmetric; their moduli space enjoys the same features as the moduli space of complete collineation. References [21, 19] study generic linear concentration models. These are Gaussian models whose concentration matrices (i.e., the inverses of the covariance matrices) are $m \times m$ positive definite matrices lying in a fixed $d$-dimensional generic linear subspace $L$ of $m \times m$ symmetric matrices. The maximum likelihood (ML) degree of such a model is the number of complex critical points of the log-likelihood function for a generic sample covariance matrix, which depends only on $m$ and $d$ by genericity of $L$ and is denoted by $\phi(m,d)$. References [21, 19] connect intersection theory on the space of complete quadrics to the computation of $\phi(m,d)$, leading to a proof that $\phi(m,d)$ is polynomial in $m$ for fixed $d>0$ in [19], as conjectured in [26]. By contrast, [10] considers Gaussian graphical models, which are examples of nongeneric linear concentration models, and uses intersection theory on the space of complete quadrics to compute the degree of the projective variety associated to Gaussian graphical models on cyclical graphs, answering another conjecture of [26].

Our proof of Theorem 1.2, the main result of our paper, is closely related to Tikhonov regularization [30, 14, 16, 20, 23]. Indeed, we prove Theorem 1.2 by first establishing a result about solutions of underdetermined linear systems (Theorem 7.1). Given an underdetermined linear system $Ax = \pi_A(b)$, we show that if $E$ has the property that $A+E$ has full rank so that

$(A+\epsilon E)x = \pi_{A+\epsilon E}(b)$ has a unique solution $x(\epsilon)$ for any $\epsilon \neq 0$, and if the columns of $E$ are orthogonal to the columns of $A$ and to $b$, then $\lim_{\epsilon\to 0} x(\epsilon)$ exists and is a solution to the initial linear system. In the context of Tikhonov regularization, this result can be viewed as providing a new class of regularization matrices with the following two properties: first, the solution to the regularized minimization problem has a well-defined limit as the deformation parameter tends to zero, and second, this limit coincides with a solution to the original minimization problem. Importantly, while in special cases this limit solution coincides with the minimal norm solution, as is always the case in standard Tikhonov regularization, in general the limit solution need not coincide with the minimal norm solution. We give examples in section 7.

### 1.4. Organization.

We give preliminaries from algebraic geometry and algebraic statistics in sections 2 and 3, respectively. We review, for different DAG models, which MLE properties can occur in section 4. We introduce sample stabilizations and their parameter spaces in section 5 (Theorem 1.2(a)), and show how a sample stabilization is constructed from a complete collineation. Sections 6–9 prove the main results. Section 6 focuses on the MLE given a sample stabilization (Theorem 1.1(a) and (b), and Theorem 1.2 (b)). Section 7 constructs unique solutions to underdetermined linear systems as the limit of a solution to a perturbation of the linear system. This result is applied in section 8 to study the limit MLE given sample stabilizations (Theorems 1.1(c) and (d), and 1.2(c)). In section 9 we apply the results to star-shaped graphs (Theorem 1.3). Finally, we discuss directions for future work in section 10.

## 2. Algebraic geometry preliminaries.

We review the construction of the moduli space of complete collineations and the definition of a complete collineation that we will work with in this paper.

### 2.1. The moduli space of complete collineations.

We start by defining the moduli space of complete collineations. A complete collineation is an element of this moduli space. We will give another definition of a complete collineation that is easier to work with in section 2.2.

Definition 2.1 (the moduli space of complete collineations). *Fix two vector spaces $V$ and $W$ with $\dim V \leq \dim W$. The* moduli space of complete collineations from $\mathbb{P}(V)$ to $\mathbb{P}(W)$ *is the closure of the graph of the rational map*

$$\phi : \mathbb{P}(\mathrm{Hom}(V,W)) \dashrightarrow \mathbb{P}(\mathrm{Hom}(\wedge^2 V, \wedge^2 W)) \times \cdots \times \mathbb{P}(\mathrm{Hom}(\wedge^r V, \wedge^r W))$$
$$[M] \mapsto ([\wedge^2 M], \ldots, [\wedge^r M]),$$

*where $r = \dim V$.*

Note that $\phi$ is only well-defined on the locus inside $\mathbb{P}(\mathrm{Hom}(V,W))$ parametrizing collineations, i.e., maps of maximal rank.

By construction, the moduli space of complete collineations contains as an open dense subset the space of maximal rank linear maps up to scaling. It can therefore be viewed as a compactification of the space of maps of maximal rank in $\mathbb{P}(\mathrm{Hom}(V,W))$. This is an alternative compactification to the "obvious" one given by $\mathbb{P}(\mathrm{Hom}(V,W))$, and has the advantage of having nicer geometric properties: its boundary is a normal crossing divisor, by contrast with the compactification given by $\mathbb{P}(\mathrm{Hom}(V,W))$ whose boundary is highly singular.

This geometric property makes the moduli space of complete collineations useful for tackling enumerative geometry problems related to linear maps [17, 28].

**2.2. Points of the moduli space.** Despite the simple construction of the moduli space of complete collineations, describing points in the boundary is difficult. In other words, given an element of

$$\mathbb{P}(\mathrm{Hom}(V,W)) \times \mathbb{P}(\mathrm{Hom}(\wedge^2 V, \wedge^2 W)) \times \cdots \times \mathbb{P}(\mathrm{Hom}(\wedge^r V, \wedge^r W))$$

with first term not of maximal rank, it is not obvious which properties the remaining terms need to satisfy for the element to lie in the moduli space of complete collineations. Thankfully, there is an alternative construction of the moduli space of complete collineations from which a description of points in the boundary can more readily be extracted.

This construction is obtained via a sequence of blow-ups of $\overline{C} := \mathbb{P}(\mathrm{Hom}(V,W))$, as shown by Vaisencher in [31]. The sequence can be described inductively as follows: set $\overline{C}_0 = \overline{C}$ and for $i \geq 1$ let $\overline{C}_i$ denote the blow-up of $\overline{C}_{i-1}$ along the proper transform in $\overline{C}_{i-1}$ of the locus of maps $[f] \in \overline{C}$ of rank less than or equal to $i$. Then the moduli space of complete collineations from $V$ to $W$ is isomorphic to $\overline{C}_r$. In particular, points of the blow-up are in one-to-one correspondence with complete collineations from $V$ to $W$. Moreover, points of the blow-up can be described explicitly by analyzing the exceptional divisors at each stage of the blow-up. Doing so yields the following definition, which we will use for the rest of this paper.

Definition 2.2 (complete collineations). *Fix two vector spaces $V$ and $W$ with $\dim V \leq \dim W$. A* complete collineation *from $\mathbb{P}(V)$ to $\mathbb{P}(W)$ is a finite sequence $([f_1], \ldots, [f_t])$ of scalar equivalence classes of maps:*

$$\begin{aligned} f_1 &: V \to W \\ f_2 &: \ker f_1 \to \operatorname{coker} f_1 \\ &\vdots \\ f_t &: \ker f_{t-1} \to \operatorname{coker} f_{t-1} \end{aligned}$$

*where each $f_i$ is degenerate except for $f_t$. An* affine lift *of a complete collineation $([f_i])_{i=1}^t = ([f_1], \ldots, [f_t])$ from $\mathbb{P}(V)$ to $\mathbb{P}(W)$ is a sequence $(g_1, \ldots, g_t)$ where $[g_i] = [f_i]$ for each $i$.*

*Example* 2.3. Consider the map $f_1 \colon \mathbb{C}^3 \to \mathbb{C}^4$ given by

$$f_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with respect to the standard bases $\{b_1, b_2, b_3\}$ for $\mathbb{C}^3$ and $\{e_1, e_2, e_3, e_4\}$ for $\mathbb{C}^4$. Then $\ker f_1 = \langle b_2, b_3 \rangle$ and $\operatorname{coker} f_1$ can be identified with the orthogonal complement $\langle e_2, e_3, e_4 \rangle$ of $\operatorname{im} f_1$. Consider the map $f_2 \colon \ker f_1 \to \operatorname{coker} f_1$ given by $b_2 \mapsto e_2$ and $b_3 \mapsto 0$. Then $\ker f_2 = \langle b_3 \rangle$ and $\operatorname{coker} f_2 \cong \langle e_3, e_4 \rangle$. Consider $f_3 \colon \ker f_2 \to \operatorname{coker} f_2$ given by $b_3 \mapsto e_3$. Then $([f_1], [f_2], [f_3])$ is a complete collineation from $\mathbb{C}^3$ to $\mathbb{C}^4$. Then $\ker f_1 = \langle b_2, b_3 \rangle$ and $\operatorname{coker} f_1$ can be identified with the orthogonal complement $\langle e_2, e_3, e_4\rangle$ of $\operatorname{im} f_1$. Consider the map $f_2\colon \ker f_1 \to \operatorname{coker} f_1$ given by $b_2 \mapsto e_2$ and $b_3 \mapsto 0$. Then $\ker f_2 = \langle b_3\rangle$ and $\operatorname{coker} f_2 \cong \langle e_3, e_4\rangle$. Consider $f_3\colon \ker f_2 \to \operatorname{coker} f_2$ given by $b_3 \mapsto e_3$. Then $([f_1],[f_2],[f_3])$ is a complete collineation from $\mathbb{C}^3$ to $\mathbb{C}^4$.

The blow-up construction of the moduli space $\mathcal{M}$ of complete collineations from $\mathbb{P}(V)$ to $\mathbb{P}(W)$ can be used to show that $\mathcal{M}$ admits a stratification $\mathcal{M} = \bigsqcup_{t=1}^m \mathcal{M}_t$ by locally closed subvarieties $\mathcal{M}_t$ parametrizing complete collineations with $t$ terms. Moreover, the composition of the blow-down maps gives a surjective morphism $\pi : \mathcal{M} \to \mathbb{P}(\operatorname{Hom}(V,W))$, which maps $([f_i])_{i=1}^t$ to $[f_1]$. Given a map $[f] \in \mathbb{P}(\operatorname{Hom}(V,W))$, define $\mathcal{M}_{[f]} := \pi^{-1}([f])$, and let $\mathcal{M}_{[f]}^t$ denote the intersection of this fibre with the stratum $\mathcal{M}_t$, so that $\mathcal{M}_{[f]}^t = \{([f_i])_{i=1}^t \in \mathcal{M} : [f_1] = [f]\}$. On each $\mathcal{M}_{[f]}^t$ there is a torus fibration

$$\mathcal{A}_{[f]}^t \to \mathcal{M}_{[f]}^t \tag{2.1}$$

such that the fibre over any complete collineation is the set of its affine lifts.

## 3. Algebraic statistics preliminaries.

We give background on maximum likelihood estimation and DAG models.

### 3.1. Maximum likelihood estimation.

An $m$-dimensional Gaussian with mean zero has density

$$f_\Sigma(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2} y^\mathsf{T} \Sigma^{-1} y\right),$$

where $y \in \mathbb{R}^m$ and the covariance $\Sigma$ lies in the cone of $m \times m$ positive definite matrices $\mathrm{PD}_m$. We refer to a multivariate Gaussian model by its set $\mathcal{M} \subset \mathrm{PD}_m$ of covariance matrices. The elements $\Sigma \in \mathcal{M}$ are parameters for the model. A maximum likelihood estimate (MLE) given sample data consists of parameters that maximize the likelihood of observing that sample.

We collect independent samples $Y_1, \dots, Y_n \in \mathbb{R}^m$ as the rows of a matrix $Y \in \mathbb{R}^{n\times m}$. Our convention that the rows are indexed by samples and the columns by variables is the transpose of that used in related work [18, 1, 22]. A maximum likelihood estimate (MLE) given data $Y$ is a point $\hat{\Sigma} \in \mathcal{M}$ that maximizes the likelihood of observing $Y$. The likelihood function is $L_Y(\Sigma) = \prod_{i=1}^n f_\Sigma(Y_i)$. We work with the function

$$\ell_Y(\Sigma) = -\log\det(\Sigma) - \operatorname{tr}(\Sigma^{-1} S_Y), \tag{3.1}$$

where $S_Y = \frac{1}{n} Y^\mathsf{T} Y$. This is the log-likelihood function, up to additive and positive multiplicative constants, hence has the same maximizers. An MLE given $Y$ in $\mathcal{M}$ is therefore

$$\hat{\Sigma} := \arg\max_{\Sigma\in\mathcal{M}} \ell_Y(\Sigma)$$

if such a maximizing $\Sigma \in \mathcal{M}$ exists. We consider the following four properties which can occur when maximizing $\ell_Y(\Sigma)$ over $\Sigma \in \mathcal{M}$:

(a) $\ell_Y$ is unbounded from above,
(b) $\ell_Y$ is bounded from above,
(c) the MLE exists (i.e., $\ell_Y$ is bounded from above and attains its supremum),
(d) the MLE exists and is unique.

*Example* 3.1. Let $\mathcal{M} = \mathrm{PD}_m$ and fix a sample $Y \in \mathbb{R}^{n\times m}$. The MLE given $Y$ is $S_Y = \frac{1}{n} Y^\mathsf{T} Y$ if it is invertible; see, e.g., [27, Proposition 5.3.7]. The matrix $S_Y$ lies in the model $\mathrm{PD}_m$ if and only if it is invertible. If it is not invertible, then $\ell_Y$ is unbounded and the MLE does not exist. Put differently, the MLE given $Y$ exists if and only if $Y$ has full column rank.

We define the *maximum likelihood threshold* (mlt) of a multivariate Gaussian model to be the minimal number of samples needed for the MLE to generically exist and be unique. Example 3.1 has $\mathrm{mlt} = m$. The study of mlts is an active area of study, with recent developments, including [4, 11, 15, 5, 12, 8, 9].

*Remark* 3.2.

(i) We assume that the mean is known to be zero. Alternatively, one could estimate the mean in addition to the covariance matrix, i.e., consider a model $\mathbb{R}^m \times \mathcal{M}$ with $\mathcal{M} \subseteq \mathrm{PD}_m$. The MLE for the mean parameter is then the sample mean. Thus, after shifting to the sample mean, one can translate to the mean zero setting. This process shifts the mlt by one; see [22, Remark 6.3.7].

(ii) For $m$-dimensional complex multivariate Gaussian distributions [33], one can do maximum likelihood estimation similarly to the above. The covariance matrix $\Sigma$ is Hermitian positive-definite and the sample matrix $Y$ lies in $\mathbb{C}^{n\times m}$. The log-likelihood function is, up to additive and positive multiplicative constants, as in (3.1) with $S_Y$ now formed using the conjugate transpose; see [8, section 1.2] and [22, section 6.3]. From here on, we will work over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, as in [22].

**3.2. Directed Gaussian graphical models.** Linear structural equation models study linear relationships between noisy variables of interest. Directed Gaussian graphical models are a special case. Let $\mathcal{G} = (V, E)$ be a DAG on vertices $V = \{1, 2, \ldots, m\}$ and directed edges $E$. A directed edge from $j$ to $i$ is denoted by $j \to i$ and the absence of such an edge by $j \not\to i$. The *parents* of $i$ in $\mathcal{G}$ is the subset of vertices

$$\mathrm{pa}(i) = \{j \in V \mid (j \to i) \in E\}.$$

A *directed Gaussian graphical model on* $\mathcal{G}$ is defined by the linear structural equation

$$y = \Lambda y + \varepsilon, \qquad \text{i.e.,} \qquad y_i = \sum_{j\in \mathrm{pa}(i)} \lambda_{ij} y_j + \epsilon_i, \tag{3.2}$$

where $y \in \mathbb{K}^m$, and $\lambda_{ij} = 0$ for $j \not\to i$ in $\mathcal{G}$. Directed Gaussian graphical models assume normally distributed noise $\epsilon \sim N(\mu, \Omega)$ with $\Omega$ diagonal. We assume that the variables are mean-centered, so that $\mu = 0$. The linear relationships are recorded in the term $\Lambda y$ while the noise term is $\epsilon$. We refer to a directed Gaussian graphical model on a DAG as a DAG model, for short.

The vector $y$ follows a multivariate normal distribution with mean 0 and covariance

$$\Sigma = (I - \Lambda)^{-1} \Omega (I - \Lambda)^{-*} \tag{3.3}$$

by (3.2), where $\Lambda$ has entries $\lambda_{ij}$ and $(\cdot)^{-*}$ denotes inverse conjugate transpose (which is the inverse transpose if $\mathbb{K} = \mathbb{R}$). The DAG model on $\mathcal{G}$ is

$$\mathcal{M} = \{\Sigma \in \mathrm{PD}_m \mid \Sigma = (I - \Lambda)^{-1}\Omega(I - \Lambda)^{-*},\ \lambda_{ij} = 0 \text{ unless } j \to i \text{ in } \mathcal{G},\ \Omega \text{ diagonal}\}.$$

An MLE given $Y$ in the DAG model on $\mathcal{G}$ consists of edge weights $\Lambda$ and variance $\Omega$.

Denote the entries of $\Omega$ by $\omega_i$, and recall that $\lambda_{ij}$ are the entries of $\Lambda$. The function $\ell_Y$ from (3.1) can be written in terms of the parameters $\omega_i$ and $\lambda_{ij}$. Its negation $-\ell_Y$ is

$$\sum_{i=1}^{m}\left(\log\omega_i + \frac{1}{n\omega_i}\|Y^{(i)} - \sum_{j\in \mathrm{pa}(i)} \lambda_{ij}Y^{(j)}\|^2\right), \tag{3.4}$$

where $Y^{(k)}$ denotes the $k$th column of the sample matrix $Y$ for $k\in\{1,\dots,m\}$; see [18, Theorem 4.9]. An MLE given $Y$ consists of $\hat\lambda_{ij}$ and $\hat\omega_i$ that minimize the above expression. The $\hat\lambda_{ij}$ are therefore coefficients of each $Y^{(j)}$ in the orthogonal projection of $Y^{(i)}$ onto $\langle Y^{(j)} : j\in \mathrm{pa}(i)\rangle$. The $\hat\omega_i$ are the residuals $\frac{1}{n}\|Y^{(i)} - \sum_{j\in\mathrm{pa}(i)}\hat\lambda_{ij}Y^{(j)}\|^2$, provided that the residual is strictly positive; see the proof of [18, Theorem 4.9] or of [22, Theorem 6.3.16].

We can consider maximum likelihood estimation of just the $\Lambda$ parameters or just the $\Omega$ parameters. We refer to these as the $\Lambda$-MLE and $\Omega$-MLE given $Y$, respectively.

*Example* 3.3. Let $\mathcal{G}$ be the DAG $1\to 3\leftarrow 2$. The DAG model on $\mathcal{G}$ is parametrized by $\lambda=(\lambda_{31},\lambda_{32})$ and $\omega=(\omega_1,\omega_2,\omega_3)$. Fix sample matrices

$$Y=\begin{pmatrix}1&0&1\\0&1&1\end{pmatrix},\qquad Y'=\begin{pmatrix}1&1&0\\0&0&1\end{pmatrix},\qquad Y''=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}.$$

The $\Lambda$-MLE given $Y$ is $(1,1)$ and the $\Omega$-MLE given $Y$ does not exist. Hence, the MLE given $Y$ does not exist. The $\Omega$-MLE given $Y'$ is $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$, while the $\Lambda$-MLEs are $\{(t,-t): t\in\mathbb{K}\}$. Finally, the $\Omega$-MLE given $Y''$ is $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ and the $\Lambda$-MLE is $\alpha=(0,0)$.

Proposition 3.4. *A $\Lambda$-MLE always exists, but may not be unique. An $\Omega$-MLE may not exist, but is unique whenever it does.*

*Proof.* Coefficients of each $Y^{(j)}$ in the projection of $Y^{(i)}$ onto $\langle Y^{(j)} : j\in\mathrm{pa}(i)\rangle$ always exist, hence a $\Lambda$-MLE always exists. The $\hat\lambda_{ij}$ are unique if and only if the submatrix of $Y$ with columns indexed by $\mathrm{pa}(i)$ has full column rank. The above residual formula for $\hat\omega_i$ shows that they are unique whenever they exist. ∎

The existence and uniqueness of the MLE given a sample matrix $Y\in\mathbb{K}^{n\times m}$ can be described by linear dependence conditions on $Y$. For a vertex $i$ in $\mathcal{G}$ we write $Y^{(\mathrm{pa}(i))}$ for the submatrix of $Y$ with columns indexed by the parents of $i$ in $\mathcal{G}$, and by $Y^{(\mathrm{pa}(i)\cup i)}$ the submatrix of $Y$ with columns indexed by $\{i\}\cup\mathrm{pa}(i)$.

Theorem 3.5 (see [18, Theorem 4.9] and [22, Theorem 6.3.16]). *Consider the DAG model on $\mathcal{G}$ with $m$ vertices, and fix a sample matrix $Y\in\mathbb{K}^{n\times m}$. The following possibilities characterize maximum likelihood estimation given $Y$:*

| | | | |
|---|---|---|---|
| (a) | *$\ell_Y$ unbounded from above* | $\Leftrightarrow$ | $\exists\, i\in\{1,\dots m\}\colon Y^{(i)}\in\langle Y^{(j)} : j\in\mathrm{pa}(i)\rangle$, |
| (b) | *MLE exists* | $\Leftrightarrow$ | $\forall\, i\in\{1,\dots m\}\colon Y^{(i)}\notin\langle Y^{(j)} : j\in\mathrm{pa}(i)\rangle$, |
| (c) | *MLE exists uniquely* | $\Leftrightarrow$ | $\forall\, i\in\{1,\dots m\}\colon Y^{(\mathrm{pa}(i)\cup i)}$ *has full column rank.* |

The above theorem uses the convention that the linear hull of the empty set is the zero vector space. In particular, if a sample matrix $Y$ has a column of zeros, then $\ell_Y$ is unbounded from above, regardless of whether the corresponding vertex has parents in $\mathcal{G}$. For a DAG model on $\mathcal{G}$ the mlt is

$$\text{mlt}(\mathcal{G}) := \max_{i\in\{1,\dots,m\}} |\text{pa}(i)| + 1 \tag{3.5}$$

by Theorem 3.5; see also [11, Theorem 1].

*Remark* 3.6. There is a correspondence between the existence and uniqueness of the MLE and notions of stability from Geometric Invariant Theory; see [18, Theorem A.2] and [22, Theorem 10.6.4]. For a DAG model, there are three equivalences:

$$\begin{array}{cccc} & Y \text{ unstable} & \Leftrightarrow & \text{MLE does not exist,} \\ Y \text{ semistable} \Leftrightarrow & Y \text{ polystable} & \Leftrightarrow & \text{MLE exists,} \\ & Y \text{ stable} & \Leftrightarrow & \text{MLE exists uniquely.} \end{array} \tag{3.6}$$

Stability is under right multiplication by the set of invertible matrices $g$ with $\det g = 1$ and $g_{ij} = 0$ for all $i \neq j$ with $j \not\to i$ in $\mathcal{G}$; see [18, Definition A.1]. This is a group if and only if the DAG $\mathcal{G}$ is transitive; see [1, Proposition 5.1]. A DAG is transitive if it has the property that a path $k \to j \to i$ implies the presence of an edge $k \to i$.

## 4. Samples with nonunique MLE.

The MLE does not exist given $Y$ in a directed Gaussian graphical model if certain submatrices of $Y$ have deficient column rank, as described in section 3. There are two ways this can happen. The first is that the number of samples $n$ is too small, the second is that the columns of $Y$ are not generic. We relate these two possibilities in section 4.1. This enables us to assume without loss of generality that $n \geq m$.

With too few samples, the MLE will not exist, and with sufficiently many generic samples, the MLE will exist and be unique. Between these extremes, different possibilities occur, which we characterize in section 4.2. Our result holds in the setting of transitive DAGs.

### 4.1. Relating too few samples to nongeneric samples.

We relate maximum likelihood estimation when $n \leq m$ to the setting $n \geq m$.

Proposition 4.1. *Fix sample data $Y \in \mathbb{K}^{n\times m}$. Then the MLEs given $Y$ equal the MLEs given $Z$, where $Z \in \mathbb{K}^{kn\times m}$ is the matrix obtained from $Y$ by taking $k$ copies of $Y$ and stacking them vertically.*

*Proof.* The $\Lambda$-MLEs given $Z$ are $\hat{\lambda}_{ij}$ that minimize each $\|Z^{(i)} - \sum_{j\in\text{pa}(i)} \lambda_{ij} Z^{(j)}\|^2$. Since $\|Z^{(i)} - \sum_{j\in\text{pa}(i)} \lambda_{ij} Z^{(j)}\|^2 = k\|Y^{(i)} - \sum_{j\in\text{pa}(i)} \lambda_{ij} Y^{(j)}\|^2$, both norms are minimized at $\hat{\lambda}_{ij}$. Hence, the $\Lambda$-MLEs given $Y$ and $Z$ agree. The $\Omega$-MLE components $\hat{\omega}_i$ given $Z$ are the residuals $\frac{1}{kn}\|Z^{(i)} - \sum_{j\in\text{pa}(i)} \hat{\lambda}_{ij} Z^{(j)}\|^2$. We have $\frac{1}{nk}\|Z^{(i)}\|^2 = \frac{k}{nk}\|Y^{(i)}\|^2 = \frac{1}{n}\|Y^{(i)}\|^2$. The same norm computations hold for $Z^{(i)} - \sum_{j\in\text{pa}(i)} \hat{\lambda}_{ij} Z^{(j)}$. Hence, the $\Omega$-MLEs given $Y$ and $Z$ agree. ∎

Proposition 4.1 allows us to assume without loss of generality that $n \geq m$. Indeed, if $n < m$, we let $k$ be minimal such that $kn \geq m$ and replace $Y$ by $Z \in \mathbb{K}^{kn\times m}$.

### 4.2. Possibilities for MLE existence and uniqueness.

We study all possibilities that can arise for ML estimation in transitive DAG models. The following theorem characterizes which MLE properties can occur. An unshielded collider is an induced subgraph $i \to j \leftarrow$

**Table 1**
*Possible MLE properties for transitive DAG models.*

| | Does not exist | Exists but not unique | Unique |
|---|---|---|---|
| $n \leq d(\mathcal{G})$ | ✓ | | |
| $d(\mathcal{G}) < n < \text{mlt}(\mathcal{G})$ | ✓ | ✓ | |
| $n \geq \text{mlt}(\mathcal{G})$, unshielded colliders | ✓ | ✓ | ✓ |
| $n \geq \text{mlt}(\mathcal{G})$, no unshielded colliders | ✓ | | ✓ |

$k$ with no edge connecting $i$ and $k$. Recall from (3.5) that the mlt $\text{mlt}(\mathcal{G})$ of a DAG is $\max_{i\in\{1,\dots,m\}} |\text{pa}(i)| + 1$. The depth $d(\mathcal{G})$ of a DAG is the number of arrows in a longest path in $\mathcal{G}$. If $\mathcal{G}$ is transitive, then $d(\mathcal{G}) \leq \text{mlt}(\mathcal{G}) - 1$.

Theorem 4.2. *Let $\mathcal{G}$ be a transitive DAG, and let $n$ denote the number of samples. The MLE properties that can occur in the DAG model on $\mathcal{G}$ are as per Table 1.*

*Proof.* We use the characterization of the existence and uniqueness of the MLE from Theorem 3.5. Define $d := d(\mathcal{G})$ and $\text{mlt} := \text{mlt}(\mathcal{G})$. By definition, there is a directed path

$$p_0 \longleftarrow p_1 \longleftarrow p_2 \longleftarrow \cdots \longleftarrow p_d$$

in $\mathcal{G}$. The transitivity of $\mathcal{G}$ implies that $p_{j+1}, \dots, p_d$ are parents of $p_j$ for all $j = 0, 1, \dots, d$.

Assume $n \leq d$. Then for any $Y \in \mathbb{K}^{n\times m}$ the vectors $Y^{(p_j)} \in \mathbb{K}^n$ for $j = 0, 1, \dots, d$ are linearly dependent, since $n < d + 1$. Therefore, there is some nontrivial linear combination $\sum_j \lambda_j Y^{(p_j)} = 0$. Let $k$ be minimal such that $\lambda_k \neq 0$. Then $Y^{(p_k)}$ is a linear combination of (some of) its parent columns. Hence, the MLE given $Y$ does not exist.

Next, assume $d < n < \text{mlt}$. The MLE does not exist given almost all $Y$, by the definition of mlt. However, the MLE does exist given special samples, as follows. Fix linear independent vectors $f_0, f_1, \dots, f_d \in \mathbb{K}^n$ using $n \geq d + 1$ and denote by $d(i)$ the number of arrows of a longest directed path in $\mathcal{G}$ starting at $i$. Then $0 \leq d(i) \leq d$. We have $d(i) = 0$ if and only if vertex $i$ is not in $\text{pa}(j)$ for any $j$. Moreover, if $p \to i$, then $d(p) > d(i)$. Define $Y \in \mathbb{K}^{n\times m}$ by setting $Y^{(i)} := f_{d(i)}$ for all $i \in \{1, \dots, m\}$. The parent columns of $Y^{(i)} = f_{d(i)}$ are all contained in $\{f_{d(i)+1}, \dots, f_{d(\mathcal{G})}\}$, by construction. Thus, $Y^{(i)}$ is not in the linear span of its parent columns and hence the MLE given $Y$ exists. Observe that there is a vertex $i$ in $\mathcal{G}$ such that $n < 1 + |\text{pa}(i)|$, since $n < \text{mlt}$. Therefore, for any $Y \in \mathbb{K}^{n\times m}$ the submatrix $Y^{(i\cup\text{pa}(i))}$ does not have full column rank, so the MLE given $Y$ is not unique.

Finally, assume $n \geq \text{mlt}$. The MLE is unique given generic samples $Y \in \mathbb{K}^{n\times m}$, by the definition of mlt. The MLE does not exist for a matrix with a column of zeros, for example. It remains to see whether the MLE given $Y$ can exist but not be unique. If there is an unshielded collider $j \to i \leftarrow k$ in $\mathcal{G}$, we create such a $Y$ by taking a generic $Y$ and replacing $Y^{(k)}$ by $Y^{(j)}$. Since $j \notin \text{pa}(k)$ and $k \notin \text{pa}(j)$, the MLE exists, but since two rows indexed by parents of $i$ are equal, it is not unique. We conclude with the case where there is no unshielded collider in $\mathcal{G}$. Assume there is some sample matrix $Y$ such that the MLE is not unique given $Y$. By Theorem 3.5(b) and (c) there is some $i$ such that $Y^{(\text{pa}(i))}$ does not have full column rank. Let $J \subset \text{pa}(i)$ denote the indexing set for those columns that appear with nonzero coefficient in a linear dependence relation among the columns of $Y^{(\text{pa}(i))}$.

Since there are no unshielded colliders in $\mathcal{G}$, there is some $k \in J$ with $J\backslash\{k\} \subset \mathrm{pa}(k)$. But then $Y^{(k)} \in \mathrm{span}\{Y^{(j)} : j \in J \setminus \{k\}\} \subseteq \mathrm{span}\{Y^{(j)} : j \in \mathrm{pa}(k)\}$, which contradicts existence of the MLE. ∎

Section 4.1 implies that we can always assume that we are in the situation where $n \geq m$, by duplicating samples enough times. So we may restrict our attention to the bottom two rows of Table 1. Given a sample $Y$ with nonunique MLE given $Y$, we will see in section 5 how to construct using a complete collineation a new sample $\widetilde{Y}$ with unique MLE given $\widetilde{Y}$. Then in sections 6 and 8 we will relate the MLE given $\widetilde{Y}$ to the MLE(s) given $Y$, and show how $\widetilde{Y}$ can be used to resolve nonidentifiability of the MLE given $Y$.

## 5. From complete collineations to sample stabilizations.

In this section, we introduce the stabilization of a sample. We call it a stabilization because, as we will see, the MLE given any stabilization of a sample is unique; cf. (3.6). There are many ways we could obtain from a sample a new sample with unique MLE. The notion of stabilization that we introduce here is based on complete collineations, and has the advantage that we can relate the MLE given a stabilization to MLEs given the original sample if they exist. We define sample stabilizations in section 5.1 and explain how a sample stabilization can be constructed from a complete collineation. We construct a parameter space for sample stabilizations as an algebraic variety in section 5.2. We describe an algorithm for obtaining complete collineations via sampling in section 5.3.

Convention 5.1. *We assume $n \geq m$. This is without loss of generality by section 4.1.*

### 5.1. Sample stabilizations from complete collineations.

Defining sample stabilizations requires taking orthogonal complements in $\mathbb{K}^n$ and $\mathbb{K}^m$. To this end, we fix the standard inner products on $\mathbb{K}^n$ and $\mathbb{K}^m$, with $v \cdot w = \sum_{i=1}^n v_i w_i^*$ where $w_i^*$ denotes the complex conjugate.

Definition 5.2 (sample perturbations and stabilizations). *Fix a sample $f : \mathbb{K}^m \to \mathbb{K}^n$. A linear map $f' : \mathbb{K}^m \to \mathbb{K}^n$ is a* perturbation *of $f$, or $f$-*perturbation*, if it satisfies the following conditions:*

(i) $\mathrm{im}\, f' \subseteq (\mathrm{im}\, f)^\perp$;

(ii) $(\ker f')^\perp = \ker f$.

*A* stabilization *of $f$, or $f$-*stabilization*, is a sum $\widetilde{f} = f + f'$, where $f'$ is an $f$-perturbation.*

Equivalently, a linear map $f' : \mathbb{K}^m \to \mathbb{K}^n$ is an $f$-perturbation if and only if its rows and columns are orthogonal to the rows and columns of $f$, respectively, and $\dim \ker f' = \dim \mathrm{im}\, f$.

Lemma 5.3 (Theorem 1.1(a)). *Any stabilization $\widetilde{f}$ of a sample $f$ has maximal rank. In particular, the MLE given $\widetilde{f}$ is unique in any DAG $\mathcal{G}$ on $m$ vertices.*

*Proof.* Write $\widetilde{f}$ as $f + f'$ where $f'$ is an $f$-perturbation. We wish to show that $\widetilde{f}$ has trivial kernel. To this end, suppose that $\widetilde{f}(v) = 0$ for some $v \in \mathbb{K}^m$. Write $v = v_1 + v_2$ where $v_1 \in \ker f$ and $v_2 \in (\ker f)^\perp$. Then $\widetilde{f}(v) = f'(v_1) + f(v_2)$. By (i) we know that $f'(v_1) \in (\mathrm{im}\, f)^\perp$; therefore, $\widetilde{f}(v) = 0$ if and only if $f'(v_1) = f(v_2) = 0$. By (ii) we have $v_1 \in (\ker f')^\perp$; therefore, $v_1 = 0$. Since $v_2 \in (\ker f)^\perp$, we also have $v_2 = 0$. Therefore, $v = 0$ as required. Moreover, since by Convention 5.1 we are assuming $n \geq m$, the MLE given $\widetilde{f}$ is unique in the DAG model on any DAG $\mathcal{G}$ on $m$ vertices, by Theorem 3.5. ∎

We now show how sample stabilizations can be constructed from complete collineations.

*Construction* 5.4 (an $f$-stabilization from a complete collineation). *Fix a sample $f$ and consider a complete collineation $([f_1],\ldots,[f_t])$ from $\mathbb{P}(\mathbb{K}^m)$ to $\mathbb{P}(\mathbb{K}^n)$ with $[f_1]=[f]$. Choose an affine lift $(f_1,f_2,\ldots,f_t)$ with $f_1=f$. Each $f_i$ is a nonzero map $\ker f_{i-1}\to \operatorname{coker} f_{i-1}$, with $f_t$ the first nondegenerate map (which must be injective since we are assuming $m\le n$).*

*We first explain how to turn each map $f_i$ into a map to $\mathbb{K}^n$. Using the standard inner product on $\mathbb{K}^n$, we identify $\operatorname{coker} f_1$ with $(\operatorname{im} f_1)^{\perp}$, a subspace of $\mathbb{K}^n$. In this way, we view $f_2$ as a map $\ker f_1\to\mathbb{K}^n$. The standard inner product on $\mathbb{K}^n$ restricts to one on $(\operatorname{im} f_1)^{\perp}$, which enables us to identify $\operatorname{coker} f_2$ with the orthogonal complement of $\operatorname{im} f_2$ inside $(\operatorname{im} f_1)^{\perp}$:*

$$\operatorname{coker} f_2=\{x\in(\operatorname{im} f_1)^{\perp}\mid \langle x,y\rangle=0 \text{ for all } y\in \operatorname{im} f_2\}.$$

*Thus, we can view $f_3$ as a map $\ker f_2\to\mathbb{K}^n$. Proceeding in this way, we identify each $\operatorname{coker} f_i$ as the orthogonal complement of $\operatorname{im} f_i$ in $(\operatorname{im} f_{i-1})^{\perp}$, and thus view $f_{i+1}$ as a map $\ker f_i\to\mathbb{K}^n$. Note that the images of each $f_i$ have pairwise trivial intersection.*

*Next, we explain how to turn each map $f_i$ into a map with domain $\mathbb{K}^m$. Set $f_2'=f_2$. It is a map with domain $\ker f_1$. Let*

$$f_3' : \ker f_1=\ker f_2\oplus(\ker f_2)^{\perp}\to\mathbb{K}^n$$

*denote the precomposition of $f_3$ with the projection from $\ker f_1$ to $\ker f_2$. In the above equation, the orthogonal complement $(\ker f_2)^{\perp}$ is taken inside $\ker f_1$. Let*

$$f_{i+1}' : \ker f_1\to\mathbb{K}^n$$

*denote the precomposition of $f_{i+1}$ with the sequence of projections $\ker f_1\twoheadrightarrow\cdots\twoheadrightarrow\ker f_i$. The process ends when we reach $f_t' : \ker f_1\to\mathbb{K}^n$, whose restriction to $\ker f_{t-1}$ has trivial kernel. Since the images of each $f_i'$ have pairwise trivial intersection, we obtain an injective map*

$$f_2'+\cdots+f_t' : \ker f_1\to\mathbb{K}^n$$

*with image contained in $(\operatorname{im} f_1)^{\perp}$. Precomposing with the projection $\mathbb{K}^m=\ker f_1\oplus(\ker f_1)^{\perp}\to \ker f_1$ gives a map $f' : \mathbb{K}^m\to\mathbb{K}^n$ with kernel $(\ker f_1)^{\perp}$ and image contained in $(\operatorname{im} f_1)^{\perp}$.*

Lemma 5.5. *The map $f' : \mathbb{K}^m\to\mathbb{K}^n$, obtained by precomposing $f_2'+\cdots+f_t'$ with the projection $\mathbb{K}^m\to\ker f_1$, is an $f$-perturbation.*

*Proof.* By construction we have $(\ker f')^{\perp}=\ker f$ and $\operatorname{im} f'\subseteq(\operatorname{im} f)^{\perp}$. Hence, both conditions of Definition 5.2 required for $f'$ to be an $f$-perturbation are satisfied. ∎

By Lemma 5.5, we set $\widetilde{f}:=f+f'$ to obtain an $f$-stabilization.

There are no choices involved in this construction, beyond the standard bases and inner products on $\mathbb{K}^n$ and $\mathbb{K}^m$, which are set once and for all. Thus, we have a canonical way of obtaining an $f$-stabilization given a sample $f$ and an affine lift of a complete collineation with first term $[f]$.

Proposition 5.6. *Given a sample $f$, an affine lift $(f_1,\ldots,f_t)$ of a complete collineation from $\mathbb{P}(\mathbb{K}^m)$ to $\mathbb{P}(\mathbb{K}^n)$ with first term $f_1=f$ uniquely determines an $f$-perturbation $f'$ and an $f$-stabilization $\widetilde{f}=f+f'$, via Construction 5.4.*

*Example* 5.7 (illustration of Construction 5.4). Let

$$f = f_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

so that $m = 3$ and $n = 4$. Let $\{b_1, b_2, b_3\}$ and $\{e_1, e_2, e_3, e_4\}$ denote the standard bases for $\mathbb{K}^3$ and $\mathbb{K}^4$, respectively. Then $\ker f = \langle b_3 \rangle$ while $(\operatorname{im} f)^\perp = \langle e_3, e_4 \rangle$. A nonzero map $f_2 : \ker f \to (\operatorname{im} f)^\perp$ is of the form $b_3 \mapsto c_1 e_3 + c_2 e_4$ for some $c_1, c_2 \in \mathbb{K}$ not both zero. This map is necessarily injective, so $(f_1, f_2)$ is an affine lift of a complete collineation from $\mathbb{P}(\mathbb{K}^3)$ to $\mathbb{P}(\mathbb{K}^4)$. Then

$$\widetilde{f} = f + f' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & c_1 \\ 0 & 0 & c_2 \end{pmatrix}.$$

**5.2. The parameter space of sample stabilizations.** A perturbation of a sample $f$ is a linear map from $\mathbb{K}^m$ to $\mathbb{K}^n$, or alternatively an element in $X := \mathbb{K}^{n\times m}$. We describe the subvariety $X_f$ of $X$ parametrizing $f$-perturbations. This is a parameter space for $f$-stabilizations.

Fix a sample $f$, and let $r := \dim \operatorname{im} f$. Let $Y_f \subseteq X$ be the subspace of maps $f' : \mathbb{K}^m \to \mathbb{K}^n$ that descend to a map

$$\mathbb{K}^m/(\ker f)^\perp \cong \ker f \to (\operatorname{im} f)^\perp.$$

In other words, $f' \in Y_f$ if and only if the columns of $f'$ are orthogonal to the columns of $f$ and the rows of $f'$ are orthogonal to the rows of $f$. The space $Y_f$ is cut out by linear equations in $X$. Let $X_{m-r}$ be the rank $m - r$ matrices in $X$. This is a locally closed subvariety of $X$, as it is closed inside the open subvariety given by matrices of rank less than or equal to $m - r$. Set

$$X_f := X_{m-r} \cap Y_f. \tag{5.1}$$

Proposition 5.8 (Theorem 1.2(a)). *Fix $f \in \mathbb{K}^{n\times m}$. Then $f' \in \mathbb{K}^{n\times m}$ is an $f$-perturbation if and only if $f' \in X_f$.*

*Proof.* If $f'$ is an $f$-perturbation, then it descends to a map from $\mathbb{K}^m/(\ker f)^\perp \cong \ker f$ to $(\operatorname{im} f)^\perp$, by definition. Thus, it lies in $Y_f$. If $f' \in Y_f$, then it is an $f$-perturbation if and only if $\dim \ker f' = \dim \operatorname{im} f = r$, or equivalently, if and only if $\dim \operatorname{im} f' = m - r$. Hence, $f' \in Y_f$ lies in $X_f$ if and only if it has rank $m - r$. ∎

Definition 5.9 (parameter space of $f$-stabilizations). *Given a sample $f$, the subvariety $X_f \subseteq X = \mathbb{K}^{n\times m}$ defined in (5.1) is the* parameter space of $f$-stabilizations.

Note that Proposition 5.6 gives a map from the space $\mathcal{A}^t_{[f]}$ defined in (2.1) to the parameter space $X_f$ of $f$-stabilizations for each $t = 1, \ldots, m$.

**5.3. Complete collineations via sampling.** In section 5.1 we have seen that, given a sample $f$, an affine lift $(f, f_2, \dots, f)$ of a complete collineation determines an $f$-stabilization and the MLE given an $f$-stabilization is always unique. In section 8, we will see that this uniquely determines an MLE given $f$ if one exists (otherwise a $\Lambda$-MLE given $f$), by taking the limit of the MLE given the stabilization, as the stabilization tends to the initial sample. Thus, an $f$-stabilization can be viewed as a way of resolving the nonidentifiability of the MLE given $f$. A natural question is whether the $f$-stabilization can be thought of as adding additional data samples to the data matrix $f$. We have seen in section 5.1 that an $f$-stabilization can be constructed from an affine lift of a complete collineation with first term $f$. Therefore, we may instead ask whether an affine lift of such a complete collineation can be obtained via sampling. We outline such a procedure below, noting its limitations.

The following is a procedure for obtaining an affine lift $(f_1, \dots, f_t)$ of a complete collineation from $\mathbb{P}(\mathbb{K}^m)$ to $\mathbb{P}(\mathbb{K}^n)$ with first term $f : \mathbb{K}^m \to \mathbb{K}^n$. We assume $n \geq m$, which is without loss of generality by section 4.1. If $f$ has full rank, then $([f])$ is a complete collineation. If not, compute a basis for $\ker f$. We think of it as consisting of $\dim \ker f$ new variables, each of which is a linear combination of the original $m$ variables. Sampling along a linear combination of variables appears in data analysis contexts such as [24]. Sample these new variables $\dim \operatorname{coker} f = n - \dim f$ times. This gives a $\dim \operatorname{coker} f \times \dim \ker f$ matrix. By identifying $\operatorname{coker} f$ with $(\operatorname{im} f)^\perp$ via the standard inner product on $\mathbb{K}^n$, and choosing a basis for $(\operatorname{im} f)^\perp$, this matrix determines a map $f_2 : \ker f \to (\operatorname{im} f)^\perp \cong \operatorname{coker} f$. If $f_2$ has maximal rank, then $([f], [f_2])$ is a complete collineation, and we stop. If not, we follow the same procedure, replacing $f$ by $f_2$. Eventually, we reach $f_t$ of maximal rank, thus giving the desired affine lift $(f, f_2, \dots, f_t)$.

The above is a prodecure that takes as input a sample matrix $f$ and outputs a stabilization. This procedure falls short of being deterministic, because it depends on choices of bases: the basis for $\ker f_i$ and for $(\operatorname{im} f_i)^\perp$. While the choice of basis for $\ker f_i$ determines how the additional sampling must be done, the choice of basis for $(\operatorname{im} f_i)^\perp$ has no statistical significance—it only determines how to turn the additional samples into the affine lift of a complete collination. It is unclear to us how the procedure could be modified so that there is a canonical way from passing from the sample to such an affine lift. This question may be better answered from a different perspective, by thinking about what statistical model might have the moduli space of complete collineations as its space of samples. Section 10.2 explores this perspective.

**6. MLEs given stabilizations.** Let $\mathcal{G}$ be a connected DAG on $m$ vertices. We study the MLE given a stabilization $\widetilde{f} := f + f'$ in the DAG model on $\mathcal{G}$. We obtain necessary and sufficient conditions for the MLE given an $f$-stabilization $\widetilde{f}$ to be an MLE given $f$ in section 6.1. If an MLE given $f$ does not exist, we study the analogous question for the $\Lambda$-MLE, which always exists by Proposition 3.4. We study which MLEs can be obtained as the MLE given a stabilization in section 6.2.

**6.1. When is the MLE given an $f$-stabilization an MLE given $f$?**

Proposition 6.1 (When is the $\Lambda$-MLE given an $f$-stabilization a $\Lambda$-MLE given $f$?). *Fix a DAG $\mathcal{G}$, a sample $f$, and an $f$-stabilization $\widetilde{f} = f + f'$. Let $f_i$ be the columns of $f$ and $v_i$ the columns of $f'$. The $\Lambda$-MLE given $\widetilde{f}$ is a $\Lambda$-MLE given $f$ in the DAG model on $\mathcal{G}$ if and only if*

$$\overline{f_i} + \overline{v_i} \in \langle f_j + v_j : j \to i \rangle \tag{6.1}$$

*for all child vertices* $i$*, where* $\overline{f_i} := \pi_{\langle f_j : j\to i\rangle}(f_i)$ *and* $\overline{v_i} := \pi_{\langle v_j : j \to i\rangle}(v_i)$.

*Proof.* For a vertex $i$ of $\mathcal{G}$, we call the components $\lambda_{ij}$ of the $\Lambda$-MLE indexed by arrows $j \to i$ the $\Lambda_i$-MLE. We show that the $\Lambda_i$-MLE given $\widetilde{f}$ is a $\Lambda_i$-MLE given $f$ if and only if (6.1) holds. The $\Lambda_i$-MLE for $\widetilde{f}$ consists of coefficients $\{\lambda_{ij}\}_{j\to i}$ such that

$$\pi_{\langle f_j+v_j : j \to i\rangle}(f_i + v_i) = \sum_{j\to i} \lambda_{ij}(f_j + v_j). \tag{6.2}$$

Using the containment $\langle f_j + v_j : j \to i\rangle \subseteq \langle f_j : j \to i\rangle \oplus \langle v_j : j \to i\rangle$, we obtain

$$\begin{aligned}\pi_{\langle f_j+v_j:j\to i\rangle}(f_i+v_i) &= \pi_{\langle f_j+v_j:j\to i\rangle}(\pi_{\langle f_j:j\to i\rangle\oplus\langle v_j:j\to i\rangle}(f_i+v_i))\\ &= \pi_{\langle f_j+v_j:j\to i\rangle}(\pi_{\langle f_j:j\to i\rangle\oplus\langle v_j:j\to i\rangle}(f_i)) + \pi_{\langle f_j+v_j:j\to i\rangle}(\pi_{\langle f_j:j\to i\rangle\oplus\langle v_j:j\to i\rangle}(v_i))\\ &= \pi_{\langle f_j+v_j:j\to i\rangle}(\overline{f_i}+\overline{v_i}),\end{aligned}$$

since $\langle v_1,\dots,v_m\rangle$ and $\langle f_1,\dots,f_m\rangle$ are orthogonal. If $\overline{f_i}+\overline{v_i} \in \langle f_j + v_j : j\to i\rangle$, then

$$\pi_{\langle f_j+v_j:j\to i\rangle}(f_i+v_i) = \overline{f_i}+\overline{v_i}.$$

This is $\sum_{j\to i}\lambda_{ij}f_j + \sum_{j\to i}\lambda_{ij}v_j$, by (6.2). We conclude that $\overline{f_i} = \sum_{j\to i}\lambda_{ij}f_j$, again by the orthogonality of $\langle f_1,\dots,f_m\rangle$ and $\langle v_1,\dots,v_m\rangle$. Hence, the $\Lambda_i$-MLE for $\widetilde{f}$ is a $\Lambda_i$-MLE for $f$.

Conversely, assume that the $\Lambda_i$-MLE for $\widetilde{f}$ is a $\Lambda_i$-MLE for $f$. This means $\overline{f_i} = \sum_{j\to i}\lambda_{ij}f_j$ for the same coefficients $\lambda_{ij}$ as in (6.2). Define $\overline{v_i} = \sum_{j\to i}\nu_{ij}v_j$ and $x = \sum_{j\to i}(\nu_{ij}-\lambda_{ij})v_j$. Then

$$\pi_{\langle f_j+v_j:j\to i\rangle}(\overline{f_i}+\overline{v_i}) = \sum_{j\to i}\lambda_{ij}(f_j+v_j) = \sum_{j\to i}\lambda_{ij}f_j + \sum_{j\to i}\lambda_{ij}v_j = \overline{f_i}+\overline{v_i} - x.$$

We have $x \in \langle v_j : j\to i\rangle$, by definition. Moreover, $x \in \langle f_j + v_j : j \to j\rangle^\perp$, since the projection of $\overline{f_i}+\overline{v_i}$ onto $\langle f_j+v_j : j\to i\rangle$ differs from $\overline{f_i}+\overline{v_i}$ by $x$. Hence, $x\cdot(f_j+v_j) = 0$ for all $j\to i$. But $x\cdot(f_j+v_j) = x\cdot v_j$, since $f_j$ and $x$ are orthogonal for all $j \to i$. Hence, $x \in \langle v_j : j\to i\rangle^\perp \cap \langle v_j : j \to i\rangle$ and we conclude that $x = 0$. ∎

If an MLE given $f$ exists, we can ask when the MLE given an $f$-stabilization $\widetilde{f}$ is an MLE given $f$. Corollary 6.2 below gives a complete answer—this is Theorem 1.1(b).

**Corollary 6.2** (When is the MLE given an $f$-stabilization an MLE given $f$?). *Fix a DAG* $\mathcal{G}$*, sample* $f$*, and* $f$*-stabilization* $\widetilde{f} = f + f'$*. Let* $f_i$ *denote the columns of* $f$ *and* $v_i$ *the columns of* $f'$*. Then the MLE given* $\widetilde{f}$ *in the DAG model on* $\mathcal{G}$ *is an MLE given* $f$ *if and only if*

$$v_i \in \langle v_j : j\to i\rangle \quad \text{and} \quad \overline{f_i} + v_i \in \langle f_j + v_j : j \to i\rangle \tag{6.3}$$

*for all child vertices* $i$ *of* $\mathcal{G}$*, where* $\overline{f_i} := \pi_{\langle f_j : j\to i\rangle}(f_i)$.

*Remark* 6.3 (sanity check). It follows from Corollary 6.2 that if the MLE given $f$ does not exist, then the condition given in (6.3) can never be satisfied by an $f$-stabilization $\widetilde{f}$.

This can be seen directly as follows. Suppose that the MLE given $f$ does not exist. Then there is some $i$ with $f_i \in \langle f_j : j \to i \rangle$, so that $\overline{f_i} = f_i$. Suppose $\widetilde{f}$ is an $f$-stabilization satisfying (6.3). Then $f_i + v_i \in \langle f_j + v_j : j \to i \rangle$, contradicting the existence of the MLE given $\widetilde{f}$, by Theorem 3.5.

*Example* 6.4. Consider the DAG $2 \leftarrow 1 \to 3$ and the sample $f$ with first column $(1,0,0)$ and second and third columns $(0,1,1)$. This sample has a nonunique MLE by Theorem 3.5. Any $f$-perturbation $f'$ must have as its first column $(0,0,0)$, as its second column $(0,0,a)$ for some nonzero $a \in \mathbb{K}$, and as its third column $(0,0,-a)$. Then the equations from (6.3) reduce to $v_2, v_3 \in \langle (1,0,0) \rangle$, which can never hold given that $a \neq 0$. Thus the MLE given an $f$-stabilization can never be an MLE given $f$.

By contrast with Example 6.4, we will see in section 9 that for star-shaped graphs with a single child vertex, and samples for which an MLE given the sample exists, the MLE given any stabilization of the sample is an MLE given the original sample (see Proposition 9.1). It follows that there are examples of graphs and samples for which the MLE given a stabilization of the sample is *always* an MLE given the original sample, and examples where the MLE given a stabilization of the sample is *never* an MLE given the original sample. The question of whether there are examples for which the MLE given a stabilization of the sample is only *sometimes* an MLE given the original sample remains open.

Question 6.5. *Is there a DAG $\mathcal{G}$ and sample $f$ such that an MLE given $f$ exists, and for some $f$-stabilizations the MLE given the stabilization is an MLE given $f$, yet for other it isn't?*

*Proof of Corollary* 6.2. Define the $\Lambda_i$-MLE as in the proof of Proposition 6.1 and similarly define the $\Omega_i$-MLE to be the component $\omega_i$ of the $\Omega$-MLE indexed by $i$. Suppose that the $\Lambda_i$-MLE and $\Omega_i$-MLE given $\widetilde{f}$ are a $\Lambda_i$-MLE and $\Omega_i$-MLE given $f$. Then $\overline{f_i} + \overline{v_i} \in \langle f_j + v_j : j \to i \rangle$, by Proposition 6.1. The $\Omega_i$-MLE given $\widetilde{f}$ is the norm of $\pi_{\langle f_j + v_j : j \to i \rangle}(f_i + v_i) - f_i - v_i$. This is, equivalently, the norm of $\overline{f_i} - f_i + \overline{v_i} - v_i$ since

$$\pi_{\langle f_j + v_j : j \to i \rangle}(f_i + v_i) = \pi_{\langle f_j + v_j : j \to i \rangle}(\overline{f_i} + \overline{v_i}) = \overline{f_i} + \overline{v_i}.$$

The $\Omega_i$-MLE given $f$ is the norm of $\overline{f_i} - f_i$. For the MLEs to coincide, the vectors $\overline{f_i} - f_i + \overline{v_i} - v_i$ and $\overline{f_i} - f_i$ must have the same norm. But given that $\overline{v_i} - v_i$ lies in $\langle \overline{f_i} - f_i \rangle^\perp$, this means the vectors must be equal, so that $v_i = \overline{v_i}$. Hence, $v_i \in \langle v_j : j \to i \rangle$.

For the other direction, suppose that $\overline{f_i} + \overline{v_i} \in \langle f_j + v_j : j \to i \rangle$ and $v_i \in \langle v_j : j \to i \rangle$. The first condition ensures that the $\Lambda_i$-MLE given $\widetilde{f}$ is a $\Lambda_i$-MLE given $f$, by Proposition 6.1. It remains to show that the $\Omega_i$-MLE given $\widetilde{f}$ is the $\Omega_i$-MLE given $f$. But this follows from the fact that $v_i \in \langle v_j : j \to i \rangle$, by the same calculations as in the previous paragraph. ∎

*Remark* 6.6. It is reasonable to wonder whether there always exists a stabilization of $f$ whose MLE is an MLE given $f$. Proposition 9.6 will show that this is not necessarily the case.

**6.2. When is an MLE given $f$ the MLE given an $f$-stabilization?** Corollary 6.2 gives necessary and sufficient conditions for the MLE given an $f$-stabilization to be an MLE given $f$. It is natural to ask which MLEs given $f$ are the MLE given some $f$-stabilization. We reformulate this question geometrically, showing that it reduces to asking whether a locally closed subvariety of the parameter space $X_f$ from Definition 5.9 is nonempty. As a first step, we characterise when, for a fixed MLE $\alpha$ given $f$, the MLE given an $f$-stabilization is also $\alpha$.

Proposition 6.7. *Let $\alpha$ be an MLE given $f$ in a DAG model on $\mathcal{G}$. Let $\lambda$ denote the $\Lambda$-MLE part of $\alpha$. Fix an $f$-perturbation $f'$, and let $v_i$ denote its columns. Then the MLE given $\widetilde{f} := f + f'$ is $\alpha$ if and only if, for every child vertex $i$,*

$$v_i = \sum_{j \to i} \lambda_{ij} v_j. \tag{6.4}$$

*Proof.* Suppose that $\alpha$ is the MLE given $\widetilde{f}$. Since $\alpha$ is also an MLE given $f$, we have

$$v_i \in \langle v_j : j \to i \rangle \quad \text{and} \quad \overline{f_i} + v_i \in \langle f_j + v_j : j \to i \rangle$$

for all child vertices $i$, where $\overline{f_i} := \pi_{\langle f_j : j \to i \rangle}(f_i)$, by Corollary 6.2. Since $\lambda$ is a $\Lambda$-MLE given $f$, we know that $\overline{f_i} = \sum_{j \to i} \lambda_{ij} f_j$. Since $\lambda$ is also an $\Lambda$-MLE given $\widetilde{f}$, we have

$$\pi_{\langle f_j + v_j : j \to i \rangle}(f_i + v_i) = \sum_{j \to i} \lambda_{ij}(f_j + v_j) = \pi_{\langle f_j + v_j : j \to i \rangle}(\overline{f_i} + \overline{v_i}) = \overline{f_i} + \overline{v_i} = \overline{f_i} + v_i.$$

It follows from orthogonality of the $f_i$ and $v_i$ that $v_i = \sum_{j \to i} \lambda_{ij} v_j$, as required.

Conversely, suppose that $v_i = \sum_{j \to i} \lambda_{ij} v_j$ for all child vertices $i$. Then $\overline{v_i} = v_i$ and, since $\overline{f_i} = \sum_{j \to i} \lambda_{ij} f_j$, it follows that $\overline{f_i} + \overline{v_i} = \sum_{j \to i} \lambda_{ij}(f_j + v_j) \in \langle f_j + v_j : j \to i \rangle$. Hence,

$$\pi_{\langle f_j + v_j : j \to i \rangle}(f_i + v_i) = \pi_{\langle f_j + v_j : j \to i \rangle}(\overline{f_i} + \overline{v_i}) = \overline{f_i} + \overline{v_i} = \sum_{j \to i} \lambda_{ij}(f_j + v_j),$$

so that $\lambda$ is the $\Lambda$-MLE given $\widetilde{f}$. Since $\overline{v_i} = v_i$ and $\overline{f_i} + \overline{v_i} = \sum_{j \to i} \lambda_{ij}(f_j + v_j) \in \langle f_j + v_j : j \to i \rangle$ for all $i$, by Corollary 6.2 we know that the $\Omega$-MLE given $\widetilde{f}$ is also an $\Omega$-MLE given $f$. The latter is unique. Therefore, $\alpha$ is the MLE given $\widetilde{f}$. ∎

We can use Proposition 6.7 to characterize geometrically when an MLE given $f$ is the MLE given an $f$-stabilization. Fix $\alpha$ an MLE given $f$, with $\lambda$ its $\Lambda$-MLE component. For every child vertex $i$ let $Y_{\alpha,i} \subseteq X = \mathbb{K}^{n \times m}$ denote the linear subspace defined by

$$v_i - \sum_{j \to i} \lambda_{ij} v_j = 0. \tag{6.5}$$

Define $Y_\alpha = \bigcap_i Y_{\alpha,i} \subseteq X$ where the intersection is over all child vertices $i$, and

$$X_{f,\alpha} = Y_\alpha \cap X_f. \tag{6.6}$$

By construction, $X_{f,\alpha}$ is a closed subvariety of $X_f$. Moreover, by Proposition 6.7 we have that $\widetilde{f} \in X_{f,\alpha}$ if and only if the MLE given $\widetilde{f}$ is $\alpha$. We obtain the following, which is Theorem 1.2(b).

Corollary 6.8 (When is an MLE given $f$ the MLE given an $f$-stabilization?). *Let $\alpha$ be an MLE given a sample $f$. Then $\alpha$ is the MLE given an $f$-stabilization if and only if $X_{f,\alpha} \neq \emptyset$.*

Definition 6.9 (parameter space of $f$-stabilizations with MLE $\alpha$). *Let $f$ denote a sample, and let $\alpha$ be an MLE given $f$. Then the closed subvariety $X_{f,\alpha} \subseteq X_f$ is the* parameter space of $f$-stabilizations $\widetilde{f}$ such that $\alpha$ is the MLE given $\widetilde{f}$.

The question of which MLEs given $f$ can be obtained as MLEs given an $f$-stabilization amounts, therefore, to determining whether $X_{f,\alpha}$ is nonempty. We have given defining equations for $X_{f,\alpha}$ in (6.5). This means that, in cases where the subvariety cannot be described explicitly, we can apply techniques from algebraic geometry to determine whether the subvariety $X_{f,\alpha}$ is nonempty; see [13] for $\mathbb{K} = \mathbb{C}$, and [2] for $\mathbb{K} = \mathbb{R}$.

*Example* 6.10 (example such that $X_{f,\alpha} = \emptyset$ for any $\alpha$). Consider again the DAG and sample from Example 6.4. Since the MLE given any $f$-stabilization is never an MLE given $f$, as shown in Example 6.4, it follows that $X_{f,\alpha} = \emptyset$ for any MLE $\alpha$ given $f$. This can also be checked via Corollary 6.8. Indeed, given an MLE $\alpha$ given $f$, the equations in (6.5) reduce to $v_2 = v_3 = 0$, so that $Y_\alpha = \emptyset \subseteq \mathbb{K}^{3\times 3}$, and therefore, $X_{f,\alpha} = Y_\alpha \cap X_f = \emptyset$.

*Example* 6.11 (example such that $X_{f,\alpha} \neq \emptyset$ for a unique $\alpha$). Consider the DAG $1 \to 3 \leftarrow 2$ and sample

$$f = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then any $\Lambda$-MLE given $f$ is of the form $(0, \lambda_{32})$ for some $\lambda_{32} \in \mathbb{K}$. The equations from (6.5) yield $v_3 - \lambda_{32} v_2 = 0$ as the defining equation of the variety $Y_\alpha$. Moreover, the variety $X_f$ in this case can be described as $3 \times 3$ matrices with zeroes everywhere except for the middle of the bottom row. Thus $Y_\alpha \cap X_f$ is nonempty if and only if $\lambda_{32} = 0$, which uniquely determines the MLE $\alpha$ given $f$. We note that in this case the $\Lambda$-component of $\alpha$ is the minimal norm $\Lambda$-MLE.

Example 6.11 will be generalized in section 9, which will show that for star-shaped graphs, sample $f$ and MLE $\alpha$ given $f$, the variety $X_{f,\alpha}$ is nonempty if and only if the $\Lambda$-component of $\alpha$ is the minimal norm $\Lambda$-MLE (see Proposition 9.2). Examples 6.10 and 6.11 give examples of DAGs and samples $f$ such that $X_{f,\alpha}$ is either empty for all MLEs $\alpha$ given $f$, or empty only for one MLE $\alpha$ given $f$. The following is an open question.

Question 6.12. *Is there a DAG, a sample $f$, and two distinct MLEs $\alpha$ and $\alpha'$ given $f$ such that $X_{f,\alpha}$ and $X_{f,\alpha'}$ are both nonempty?*

**7. Solutions of underdetermined linear systems using stabilizations.** We investigate how solutions to underdetermined linear systems of a particular form can be obtained from limits of solutions to related full rank systems. In section 8 we will think of this limit solution as a way to choose a unique $\Lambda$-MLE from a choice of infinitely many. We will see through examples that the limit solution does not necessarily coincide with the minimal norm solution.

As in section 6, we fix the standard inner product on $\mathbb{K}^n$. Fix a matrix $A \in \mathbb{K}^{n\times p}$ and a vector $b \in \mathbb{K}^n$, with $n \geq p$. Let $\pi_A(b)$ denote the projection of $b$ onto the column space of $A$. We consider linear systems of the form

$$Ax = \pi_A(b).$$

One solution is given by the pseudo-inverse $x = A^+\pi_A(b)$. It is the unique solution if and only if the matrix $A$ has full column rank, in which case it is $x = (A^\dagger A)^{-1}A^\dagger\pi_A(b)$. Note that $A^+\pi_A(b) = A^+b$, by properties of the pseudoinverse.

In this section, we establish the following result.

Theorem 7.1. *Let $A(\epsilon) = A + \epsilon E$ and $b(\epsilon) = b + \epsilon v$, where $A(\epsilon)$ has full column rank for each $\epsilon \neq 0$ and the columns of $A$ and $b$ are orthogonal to the columns of $E$ and to $v$. Let $x(\epsilon) = A(\epsilon)^+\pi_{A(\epsilon)}(b(\epsilon))$. Then the limit*

$$x := \lim_{\epsilon\to 0} x(\epsilon)$$

*exists and it is a solution to $Ax = \pi_A(b)$, with explicit description given in Corollary A.5.*

The challenge in proving Theorem 7.1 is that the pseudo-inverse is not necessarily a continuous function in the elements of the matrix. It is continuous if and only if $A(\epsilon)$ and $A$ have the same rank for sufficiently small $\epsilon$; see [25]. When $A(\epsilon)$ and $A$ do not have the same rank, the limit $\lim_{\epsilon\to 0}(A(\epsilon))^+$ does not exist; consider, for example,

$$A(\epsilon) = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \end{pmatrix}.$$

Luckily, we are not interested in $A(\epsilon)^+$ and its limit, but rather in $A(\epsilon)^+\pi_{A(\epsilon)}(b(\epsilon))$ and its limit. As we will see, multiplying by $\pi_{A(\epsilon)}(b(\epsilon))$ resolves the discontinuity to give a well-defined limit solution.

*Remark* 7.2 (connection to Tikhonov regularization). Given a matrix $A \in \mathbb{K}^{n\times p}$ and a vector $b \in \mathbb{K}^n$ with $n \geq p$, ordinary least squares seeks to minimize $||Ax-b||_2$. The problem is ill-posed if the solution is not unique. In this case Tikhonov regularization (see [30, 14, 16, 20, 23]) is a method for obtaining a unique solution with desirable properties, by adding a regularization term:

$$||Ax - b||_2 + ||\Gamma x||_2.$$

The matrix $\Gamma$, called the Tikhonov matrix, is chosen such that the regularized minimization problem has a unique solution with desirable properties. The corresponding solution is

$$x(\Gamma) = (A^\dagger A + \Gamma^\dagger\Gamma)^{-1}A^\dagger b. \tag{7.1}$$

Standard Tikhonov regularization refers to the situation where $\Gamma^\dagger\Gamma$ is a scalar multiple of the identity matrix. In this case, the solution $x(\Gamma)$ tends to the minimal norm solution $x = A^+b$ as the scalar tends to zero. Generalised Tikhonov regularization is the case where $\Gamma^\dagger\Gamma$ is not the identity matrix. In this case, depending on $\Gamma$ and $A$, the solution $x(\Gamma)$ may not have a limit, as $\Gamma$ tends to the zero matrix and, even if it does, this limit may not be a solution of the original minimization problem.

Theorem 7.1 relates to Tikhonov regularization, as follows. If we assume in the statement of Theorem 7.1 that $v = 0$ and set $\Gamma = \epsilon E$, then the solution $x(\epsilon)$ in Theorem 7.1 is the solution to the Tikhonov regularized minimization problem in (7.1). If we assume, in addition, that the columns of $E$ are orthonormal, then $E^\dagger E$ is the identity, and we are in the setting of standard

Tikhonov regularization. In general, this condition will not be satisfied and Theorem 7.1 can be viewed as introducing a new class of Tikhonov matrices for which the solution $x(\epsilon) := x(\Gamma)$ to the regularized minimization problem has a well-defined limit, as $\epsilon$ tends to zero, and this limit is a solution to the original minimization problem. In special cases this limit solution coincides with the minimal norm solution, as is the case in standard Tikhonov regularization, but, in general, the limit solution need not be the minimal norm solution; see Examples 7.3 and 9.4.

*Example* 7.3 (the limit of $x(\epsilon)$ may not be the minimal norm solution). As mentioned in Remark 7.2, in standard Tikhonov regularization the limit of the solution to the regularized optimization problem is the minimal norm solution. For the generalized Tikhonov matrices introduced in Theorem 7.1, this is not necessarily the case, as follows. Let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad E = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then $Ax = \pi_A(b)$ has solutions $x = (0, \beta)^\mathsf{T}$ for any $\beta \in \mathbb{K}$. The minimal norm solution is $x_0 = (0,0)^\mathsf{T}$. If $v = (0,0,0)^\mathsf{T}$, then $(A + \epsilon E)x(\epsilon) = \pi_{A(\epsilon)}(b + \epsilon v)$ has solution $x(\epsilon) = (0,0)^\mathsf{T}$ for $\epsilon \neq 0$. Its limit as $\epsilon$ tends to zero is the minimal norm solution. However, if instead $v = (0,0,1)^\mathsf{T}$, then $(A + \epsilon E)x(\epsilon) = \pi_{A(\epsilon)}(b + \epsilon v)$ has the solution $x(\epsilon) = (0,1)^\mathsf{T}$. Its limit as $\epsilon$ tends to zero is $(0,1)^\mathsf{T}$, which is not the minimal norm solution.

In the above example, the solution to the regularized optimization problem is itself a solution to the original problem, but this need not always be the case, as the next example shows. Let

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \text{and} \quad v = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Then

$$x(\epsilon) = \frac{1}{1 + 2\epsilon^2} \left( \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} + \epsilon^2 \begin{pmatrix} 2 \\ -1 \\ 2 \\ 0 \end{pmatrix} \right),$$

which tends to $(1,0,0,1)^\mathsf{T}$ as $\epsilon$ tends to zero. Note that this is again distinct from the minimal norm solution $(1,0,0,0)^\mathsf{T}$.

Theorem 7.1 has the following geometric interpretation. The vector $x(\epsilon) = A(\epsilon)^+ \pi_{A(\epsilon)}(b(\epsilon))$ gives the coefficients for the projection of $b(\epsilon)$ onto the column space of $A(\epsilon)$:

$$\pi_{A(\epsilon)}(b(\epsilon)) = \sum_{i=1}^{p} x_i(\epsilon) A(\epsilon)_i,$$

where $x_i(\epsilon)$ denotes the $i$th entry of $x$ and $A(\epsilon)_i$ the $i$th column of $A(\epsilon)$. Theorem 7.1 implies that these coefficients do not go off to infinity. Now we also have

$$\pi_{A(\epsilon)}(b(\epsilon)) = \pi_{A(\epsilon)}(\overline{b} + \epsilon\overline{v}), \tag{7.2}$$

where $\overline{b} = \pi_A(b)$ and $\overline{v} = \pi_E(v)$. This follows from the proof of Proposition 6.1, since the columns of $A$ and $b$ are orthogonal to the columns of $E$ and to $v$. So instead of projecting $b(\epsilon)$ we can project $\overline{b} + \epsilon\overline{v}$, which is near the column space of $A$ for small $\epsilon$, and hence also near the column space of $A(\epsilon)$ for small $\epsilon$. Therefore, Theorem 7.1 says, roughly, that if we project a vector onto a subspace that is "very close" to it, the coefficients don't go off to infinity. This assumption is important because if we are projecting a vector that is "far away" from our subspace, the limit may not exist. We give two examples below to illustrate the two behaviors.

*Example* 7.4. Fix

$$A(\epsilon) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \epsilon \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad b(\epsilon) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \epsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Then the conditions of Theorem 7.1 are satisfied for $A(\epsilon)$ and $b(\epsilon)$ so $x(\epsilon)$ has a limit as $\epsilon \to 0$. It can be calculated as follows. We have $\pi_A(b) = 0$, so the system $Ax = \pi_A(b)$ has solutions $ce_2$ for any $c \in \mathbb{K}$, where $e_2 = (0,1)^\mathsf{T}$. Moreover, we have $\pi_{A(\epsilon)}(b(\epsilon)) = b(\epsilon)$, so $A(\epsilon)x = \pi_{A(\epsilon)}(b(\epsilon))$ has unique solution $x(\epsilon) = e_2$. Thus, $x(\epsilon)$ has a limit as $\epsilon \to 0$, and this is a solution to $Ax = \pi_A(b)$. Note that the limit $x = e_2$ is not the solution obtained from the pseudo-inverse $A^+b$, which is the minimal norm solution $(0,0)^\mathsf{T}$.

*Example* 7.5. Fix

$$A(\epsilon) = A + \epsilon E := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \epsilon \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad b(\epsilon) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \epsilon \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since $(0,1)^\mathsf{T}$ is not orthogonal to the second column $E$, the conditions of Theorem 7.1 are not satisfied. We show that in this case $x(\epsilon)$ does not have a finite limit as $\epsilon \to 0$. Since $\pi_A(b) = 0$, the system $Ax = \pi_A(b)$ has solutions $ce_2$ for any $c \in \mathbb{K}$. Since $\pi_{A(\epsilon)}(b(\epsilon)) = b(\epsilon)$, the system $A(\epsilon)x = \pi_{A(\epsilon)}b(\epsilon)$ has unique solution $x(\epsilon) = \frac{1}{\epsilon}e_2$. This does not have a finite limit as $\epsilon \to 0$.

The existence of the limit $x = \lim_{\epsilon\to 0} x(\epsilon)$ is proven in Appendix A, where we give a geometric and an algebraic argument. We show that $\lim_{\epsilon\to 0} x(\epsilon)$ is a solution to $Ax = \pi_A(b)$, in Proposition 7.6 below.

Proposition 7.6 (the limit is a solution). *Let $A(\epsilon) = A + \epsilon E$ and $b(\epsilon) = b + \epsilon v$ be as in Theorem 7.1. Let $x(\epsilon)$ denote the unique solution to $A(\epsilon)x(\epsilon) = \pi_{A(\epsilon)}(b(\epsilon))$ for $\epsilon \neq 0$. Assume that $x := \lim_{\epsilon\to 0} x(\epsilon)$ exists. Then $x$ is a solution to $Ax = \pi_A(b)$.*

*Proof.* We show that $\pi_{A(\epsilon)}(b(\epsilon))$ tends to $\pi_A(b)$ as $\epsilon \to 0$. Let $f_i$ and $v_i$ denote the columns of $A$ and $E$, respectively. For $\epsilon > 0$, each $\langle f_i + \epsilon v_i : 1 \le i \le p\rangle$ determines a point $L_\epsilon$ in the Grassmannian $G(p,n)$ of $p$-dimensional subspaces of $\mathbb{K}^n$. Since $G(p,n)$ is compact, there is a limit subspace $L_0 \in G(p,n)$ as $\epsilon \to 0$. Choose a basis $b_1^0, \dots, b_p^0$ for $L_0$. By the geometric version of Nakayama's lemma, this basis can be lifted to a basis $b_1^\epsilon, \dots, b_p^\epsilon$ of $L_\epsilon$ for small $\epsilon$.

Let $M_\epsilon$ be the $n \times p$ matrix with columns $b_1^\epsilon, \ldots, b_p^\epsilon$. Then

$$\pi_{A(\epsilon)}(b(\epsilon)) = \pi_{A(\epsilon)}(\overline{b} + \epsilon\overline{v}) = \pi_{L_\epsilon}(\overline{b} + \epsilon\overline{v}) = M_\epsilon M_\epsilon^+(\overline{b} + \epsilon\overline{v}).$$

Recall that the second equality follows from orthogonality of the columns of $A$ and $b$ with the columns of $E$ and $v$. Now $\lim_{\epsilon\to 0} M_\epsilon = M_0$, and $M_0$ has the same rank as $M_\epsilon$ for $\epsilon \neq 0$. Therefore, we have $\lim_{\epsilon\to 0} M_\epsilon^+ = M_0^+$, by [3] (see also [25]). Therefore,

$$\begin{aligned}\lim_{\epsilon\to 0} \pi_{A(\epsilon)}(b(\epsilon)) &= \lim_{\epsilon\to 0} M_\epsilon M_\epsilon^+(\overline{b} + \epsilon\overline{v}) = M_0 M_0^+(\overline{b}) \\ &= \pi_{L_0}(\overline{b}) = \pi_A(\overline{b}) = \overline{b} = \pi_A(b).\end{aligned}$$ ∎

*Remark* 7.7 (Examples 7.4 and 7.5, revisited). The choices involved in the proof of Proposition 7.6 can be made explicit if we work with Example 7.4. In this example, we have $n, p = 2$, so $G(p,n)$ consists of a single point, namely $\mathbb{K}^2$. Therefore, $L_\epsilon = L_0 = \mathbb{K}^2$ for each $\epsilon \neq 0$. We can take the standard basis $\{b_1^0 = e_1, b_2^0 = e^2\}$ for $L_0$. This same basis is a lift to a basis of $L_\epsilon = \mathbb{K}^2$ for any $\epsilon$, i.e., we take $b_1^\epsilon = e_1$ and $b_2^\epsilon = e_2$. Then $M_\epsilon$ is the two by two identity matrix.

Example 7.5 shows that if the columns of $A$ and $b$ are not orthogonal to the columns of $E$ and $v$ in Proposition 7.6, then the limit may not exist. Indeed, in this example it is not the case that $b$ is orthogonal to the columns of $E$, and the limit does not exist. We can see explicitly where the proof of Proposition 7.6 fails if the orthogonality assumption does not hold: orthogonality is used to show that $\pi_{A(\epsilon)}(b(\epsilon)) = \pi_{A(\epsilon)}(\overline{b} + \epsilon\overline{v})$. This equality fails in the case of Example 7.5: the left-hand side is the vector $(0,\epsilon)^\mathsf{T}$, while the right-hand side is the zero vector.

## 8. MLEs given sample stabilizations in the limit.

We gave necessary and sufficient conditions for the MLE given an $f$-stabilization to be an MLE given $f$ in section 6. In this section, we consider the limit of the $\Lambda$-MLE or MLE given $\widetilde{f}(\epsilon) := f + \epsilon f'$ as $\epsilon \to 0$. Our intuition is that the original data matrix was insufficient, in that its MLE did not exist or was not unique. But the the stabilization is well-behaved, in that there is a unique MLE given the stabilization. This motivates us to define a family of good matrices that tend to the original data matrix, and to assess what happens to the MLE in the limit. We show that we always obtain an MLE given $f$ (if one exists, otherwise a $\Lambda$-MLE), in section 8.1. We study which MLEs given $f$ can be obtained as MLEs given $f$-stabilizations under such a limit, in section 8.2.

### 8.1. The limit MLE given $\widetilde{f}$ exists and is an MLE given $f$.

We prove that if $\widetilde{f} = f + f'$ is an $f$-stabilization, then the MLE given $\widetilde{f}(\epsilon) := f + \epsilon f'$ has a well-defined limit as $\epsilon$ tends to zero, and moreover, that this limit is an MLE given $f$ if one exists. If the MLE given $f$ does not exist, then the previous statement remains true by considering the $\Lambda$-MLE. We also describe the $\Lambda$-MLE and MLE given $f$ that is picked out by this process. We start by proving the result about $\Lambda$-MLEs, before turning to MLEs in Theorem 8.3.

Proposition 8.1 (limit $\Lambda$-MLE given an $f$-stabilization). *Fix a DAG $\mathcal{G}$, a sample $f$, and an $f$-stabilization $\widetilde{f} = f + f'$. Let $f_i$ denote the columns of $f$ and $v_i$ the columns of $f'$. For each* *child vertex $i$, let $C_i'(\epsilon) := A_i^\top A_i + \epsilon E_i^\top E_i$, where $A_i$ (respectively, $E_i$) is the $n \times |\mathrm{pa}(i)|$ matrix with columns the subset of the $f_j$ (respectively, $v_j$) such that $j \to i$. Let $\overline{f_i} = \pi_{\langle f_j : j \to i\rangle}(f_i)$, and let $\overline{v_i} = \pi_{\langle v_j : j\to i\rangle}(v_i)$. Let $\widetilde{f}(\epsilon) = f + \epsilon f'$ for $\epsilon \neq 0$. Then we have the following results about MLEs in the DAG model on $\mathcal{G}$:*

(a) *a unique $\Lambda$-MLE exists given $\widetilde{f}(\epsilon)$ for any $\epsilon \neq 0$;*
(b) *fix a vertex $i$ and suppose for simplicity that $f_1, \ldots, f_p$ and $v_1, \ldots, v_p$ are the columns of $f$ and $f'$, respectively indexed by edges $j \to i$. Then the $\Lambda_i$-MLE given $\widetilde{f}(\epsilon)$ has a well-defined limit as $\epsilon$ tends to zero, given by*

$$\frac{1}{\mathrm{tr}\left(\frac{d^{l-1}}{\epsilon^{l-1}}\Big|_{\epsilon=0} \mathrm{adj}\, C_i'(\epsilon) E^\top E\right)} \left( \frac{d^l}{d\epsilon^l}\Big|_{\epsilon=0} \mathrm{adj}\, C_i'(\epsilon) \begin{pmatrix} f_1 \cdot \overline{f_i} \\ \vdots \\ f_p \cdot \overline{f_i} \end{pmatrix} + l\, \frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \mathrm{adj}\, C_i'(\epsilon) \begin{pmatrix} v_1 \cdot \overline{v_i} \\ \vdots \\ v_p \cdot \overline{v_i} \end{pmatrix} \right),$$

*where $l \in \{0, \ldots, p\}$ denotes the smallest integer such that*

$$\mathrm{tr}\left( \frac{d^{l-1}}{\epsilon^{l-1}}\Big|_{\epsilon=0} \mathrm{adj}\, C_i'(\epsilon) E^\top E \right) \neq 0;$$

(c) *the limit of the $\Lambda$-MLE given $\widetilde{f}(\epsilon)$ as $\epsilon$ tends to zero is a $\Lambda$-MLE given $f$;*
(d) *if $\overline{f_i} + \overline{v_i} \in \langle f_j + v_j : j \to i\rangle$ for all vertices $i$, then the $\Lambda$-MLE given $\widetilde{f}(\epsilon)$ is independent of $\epsilon$ and is a $\Lambda$-MLE given $f$.*

*Proof.* If $f'$ is an $f$-perturbation, then $\epsilon f'$ is also an $f$-perturbation for any $\epsilon \neq 0$. Therefore, $\widetilde{f}(\epsilon)$ is an $f$-stabilization for any $\epsilon \neq 0$ and so by Lemma 5.3 there is a unique MLE given $\widetilde{f}$. In particular, there is a unique $\Lambda$-MLE given $\widetilde{f}$. This proves (a).

We now turn to (b) and (c). We can find the $\Lambda$-MLE by finding each $\Lambda_i$-MLE independently. The $\Lambda_i$-MLE given a sample $Y$ are the coefficients $\lambda_{ij}$ in front of each $Y^{(j)}$ in the orthogonal projection of $Y^{(i)}$ onto the span of $\{Y^{(j)} : j \to i\}$. Hence, they are the entries of $x$ in a linear system of the form $Ax = \pi_A(b)$, where $A$ has the vectors $Y^{(j)}$ for $j \to i$ as its columns and $b = Y^{(i)}$. Therefore, the $\Lambda_i$-MLE is not unique if and only if the linear system is underdetermined.

By definition, the $\Lambda_i$-MLE given $\widetilde{f}(\epsilon)$ is the unique solution $x(\epsilon)$ to the linear system

$$(A_i + \epsilon E_i) x(\epsilon) = \pi_{\langle f_j + \epsilon v_j : j \to i\rangle}(f_i + v_i).$$

Recall that the matrix $A_i + \epsilon E_i$ is the matrix obtained from $f + \epsilon f'$ by picking out those columns indexed by vertices $j$ such that $j \to i$. Since the columns of $f$ are orthogonal to the columns of $f'$ by the definition of a sample stabilization, it follows that the columns of $A_i$ and $f_i$ are orthogonal to the columns of $E_i$ and $v_i$. We also know that $f + \epsilon f'$ has full column rank, since $f + \epsilon f'$ is an $f$-stabilization. Therefore, $A_i + \epsilon E_i$ has full column rank for each $\epsilon \neq 0$. We have thus shown that $A_i(\epsilon) = A_i + \epsilon E_i$ and $f_i(\epsilon) = f_i + \epsilon v_i$ satisfy the assumptions of Theorem 7.1. It follows that $x(\epsilon)$ has a well-defined limit as $\epsilon$ tends to zero, and moreover, that the limit $x(0)$ is a solution to $A_i x = \overline{f_i}$. This ensures that the limit $x(0)$ is a $\Lambda_i$-MLE given $f$. The formula is obtained from Corollary A.5.

It remains to show (d). If $\overline{f_i} + \overline{v_i} \in \langle f_j + v_j : j \to i\rangle$, so that $\overline{f_i} + \overline{v_i} = \sum_{j \to i} \mu_j (f_j + v_j)$ for some $\mu_j \in \mathbb{K}$, then the $\Lambda_i$-MLE given $\widetilde{f}(\epsilon)$ is $\mu_i(\epsilon) = \sum_{j\to i} \mu_j e_j$, which is independent of $\epsilon$, by

Lemma A.3. Hence, the limit Λ-MLE, which is a Λ-MLE given $f$, is also a Λ-MLE given $\widetilde{f}(\epsilon)$ for any $\epsilon \neq 0$. ■

*Remark* 8.2 (*connection to Proposition* 6.1). Proposition 8.1(d) is the reverse implication of Proposition 6.1. We included it above because we prove it using a different method.

We build on Proposition 8.1 to obtain an analogous result about MLEs.

Theorem 8.3 (limit MLE given a sample stabilization). *Fix a DAG $\mathcal{G}$, a sample $f$, and an $f$-stabilization $\widetilde{f} := f + f'$. Let $\widetilde{f}(\epsilon) := f + \epsilon f'$ for $\epsilon \neq 0$. Then we have the following results about MLEs in the DAG model on $\mathcal{G}$:*
(a) *$\widetilde{f}(\epsilon)$ has a unique MLE for any $\epsilon \neq 0$;*
(b) *the MLE given $\widetilde{f}(\epsilon)$ has a well-defined limit as $\epsilon \to 0$;*
(c) *if $f$ has at least one MLE, then the limit is an MLE given $f$, more precisely the unique MLE with Λ-MLE component given in Proposition 8.1(b);*
(d) *the MLE given $\widetilde{f}(\epsilon)$ is independent of $\epsilon$ and an MLE given $f$ if and only if*

$$v_i \in \langle v_j : j \to i \rangle \quad \text{and} \quad \overline{f_i} + v_i \in \langle f_j + v_j : j \to i \rangle, \tag{8.1}$$

*for all child vertices $i$, where $\overline{f_i} := \pi_{\langle f_j : j \to i \rangle}(f_i)$.*

Theorem 8.3(a) and (c) is Theorem 1.1(b), while Theorem 8.3(c) is Theorem 1.1(d).

*Proof.* For (a), see the proof of Proposition 8.1(a). By Proposition 8.1(b), the Λ-MLE given $\widetilde{f}(\epsilon)$ has a well-defined limit as $\epsilon \to 0$. It remains to show that the Ω-MLE also has a well-defined limit. The Ω-MLE $\omega(\epsilon)$ given $\widetilde{f}(\epsilon)$ has components

$$\omega_i(\epsilon) = ||\pi_{\langle f_j + \epsilon v_j : j \to i \rangle}(f_i + \epsilon v_i) - f_i - \epsilon v_i||.$$

We have

$$\pi_{\langle f_j + \epsilon v_j : j \to i \rangle}(f_i + \epsilon v_i) \to \pi_{\langle f_j : j \to i \rangle}(f_i)$$

as $\epsilon \to 0$, by the proof of Proposition 7.6. Since the limit commutes with taking the norm, it follows that $\omega_i(\epsilon)$ tends to $||\pi_{\langle f_j : j \to i \rangle}(f_i) - f_i||$ as $\epsilon$ tends to zero. Therefore, the Ω-MLE given $\widetilde{f}(\epsilon)$ has a limit as $\epsilon$ tends to zero, which proves (b). Moreover, if the Ω-MLE given $f$ exists, then the limits $||\pi_{\langle f_j : j \to i \rangle}(f_i) - f_i||$ for all $j \to i$ make up the Ω-MLE given $f$. Together with Proposition 8.1(b), this establishes (c).

To prove (d), suppose first that the MLE given $\widetilde{f}(\epsilon)$ is independent of $\epsilon$ and an MLE given $f$. The second assumption ensures by Corollary 6.2 that the equations in (8.1) are satisfied for all child vertices $j$. Conversely, if these equations are satisfied, then they are also satisfied if the $v_i$ and $v_j$ are replaced by $\epsilon v_i$ and $\epsilon v_j$ for $\epsilon \neq 0$. Therefore, the MLE given $\widetilde{f}(\epsilon)$ is an MLE given $f$, by Corollary 6.2. In particular, the Ω-MLE given $\widetilde{f}(\epsilon)$ is the unique Ω-MLE given $f$, which is independent of $\epsilon$. Moreover, if these equations are satisfied, then by Proposition 8.1(d) the Λ-MLE given $\widetilde{f}(\epsilon)$ is independent of $\epsilon$ and also a Λ-MLE given $f$. This shows that the MLE given $\widetilde{f}(\epsilon)$ is independent of $\epsilon$ and an MLE given $f$. ■

*Remark* 8.4 (*strengthening Theorem* 8.3(c)). Our proof of Theorem 8.3(c) proves a stronger statement, which does not require that an MLE given $f$ exists: if an MLE exists on a subset of vertices, then the limit of the partial MLE given $\widetilde{f}(\epsilon)$ on these vertices is a partial MLE given $f$.

We conclude this section by giving a name to the MLEs and $\Lambda$-MLEs obtained in the limit.

Definition 8.5 (limit MLE given a sample stabilization). *Given a sample $f$ and an $f$-stabilization $\widetilde{f} = f + f'$, the* limit $\Lambda$-MLE given $\widetilde{f}$ *is the limit as $\epsilon$ tends to zero of the $\Lambda$-MLE given $\widetilde{f}(\epsilon) = f + \epsilon f'$. If the MLE given $f$ exists, then the* limit MLE given $\widetilde{f}$ *is defined analogously.*

### 8.2. When is an MLE given $f$ the limit MLE given an $f$-stabilization?

We know that for a sample $f$, the limit MLE given any $f$-stabilization is an MLE given $f$ if $f$ admits at least one MLE, by Theorem 8.3. In this section, we address the following question: which MLEs given $f$ are limit MLEs given $f$-stabilizations? This question should be viewed as an extension of the question posed in section 6.2 regarding which MLEs given $f$ coincide with the MLE given an $f$-stabilization. We approach the question geometrically, giving an analogue of Corollary 6.8. We start first by answering the question for $\Lambda$-MLEs in Proposition 8.6 below. The solution to the problem for MLEs will follow immediately; see Corollary 8.7.

The statement of Proposition 8.6 requires defining for a $\Lambda$-MLE $\lambda$ given $f$ an associated locally closed subvariety $X_{f,\lambda}^{\lim}$ of the parameter space $X_f \subseteq X = \mathbb{K}^{n\times m}$ of $f$-stabilizations defined in section 5.2. This subvariety will parametrize $f$-stabilizations $\widetilde{f}$ such that the limit $\Lambda$-MLE given $\widetilde{f}$ is $\lambda$. To define $X_{f,\lambda}^{\lim}$, fix a sample $f$ and $\lambda$ a $\Lambda$-MLE given $f$. Let $\lambda_i$ denote the $\Lambda_i$-MLE for each child vertex $i$. We represent $\lambda_i$ as a column vector of length $|\operatorname{pa}(i)|$.

By Proposition 8.1(b) and Corollary A.5 we know that for any $f$-perturbation $\widetilde{f}$ and any vertex $i$, the limit of the $\Lambda_i$-MLE given $\widetilde{f}(\epsilon) = f + \epsilon f'$ as $\epsilon$ tends to zero equals $D_l/c_l$, where $l$ is defined in Corollary A.5, and $D_l$ and $c_l$ are defined in (A.11) and (A.12), respectively. We are therefore interested in whether or not there exists an $f$-stabilization $\widetilde{f} = f + f'$ in the parameter space $X_f$ such that the following equation is satisfied for each vertex $i$:

$$c_l \lambda_i - D_l = 0. \tag{8.2}$$

Each entry in the above vector can be viewed as a polynomial in the entries of $f'$. Therefore, (8.2) cuts out a closed subvariety $X_{f,\alpha}^{i,\lim}$ of $X_f$ defined by the vanishing of the polynomial equations appearing in the entries of the vector in the left-hand side of (8.2). Let

$$X_{f,\alpha}^{\lim} = \bigcap_j X_{f,\alpha}^{i,\lim} \subseteq X_f,$$

where the intersection ranges over all child vertices $j$ of $\mathcal{G}$. This is a closed subvariety of the parameter space $X_f$ of $f$-stabilizations, with defining equations given explicitly by (8.2). We have thus proved the following.

Proposition 8.6 (When is a $\Lambda$-MLE given $f$ the limit $\Lambda$-MLE given an $f$-stabilization?). *Let $f$ denote a sample, and let $\lambda$ be a $\Lambda$-MLE given $f$. Then*

$$X_{f,\lambda}^{\lim} \subseteq X_f$$

*parametrizes those $f$-stabilizations such that the $\Lambda$-MLE given $\widetilde{f}(\epsilon) := f + \epsilon f'$ tends to $\lambda$ as $\epsilon$ tends to zero. In particular, the $\Lambda$-MLE $\lambda$ given $f$ is the limit $\Lambda$-MLE given an $f$-stabilization if and only if*

$$X_{f,\lambda}^{\lim} \neq \emptyset.$$

We can use Proposition 8.6 to answer the analogous question for MLEs rather than $\Lambda$-MLEs, as per Corollary 8.7 below which corresponds to Theorem 1.2(c).

Corollary 8.7 (When is an MLE given $f$ the limit MLE given an $f$-stabilization?). *Assume an MLE $\alpha$ exists given sample $f$. Then $X_{f,\alpha}^{\lim} \subseteq X_f$ parameterizes the $f$-stabilizations $\widetilde{f}$ such that the limit MLE given $\widetilde{f}$ is $\alpha$. In particular, $\alpha$ is a limit MLE given an $f$-stabilization if and only if $X_{f,\alpha}^{\lim} \neq \emptyset$.*

*Proof.* Let $\lambda$ denote the $\Lambda$-MLE component of $\alpha$. By Proposition 8.6 we know that $\widetilde{f}$ lies in $X_{f,\lambda}^{\lim}$ if and only if its limit $\Lambda$-MLE is $\lambda$. But by Theorem 8.3(c) we also know that the limit MLE is an MLE given $f$, as by assumption $f$ has at least one MLE. Since $\Omega$-MLEs are unique, there is a unique MLE given $f$ with a fixed $\Lambda$-MLE component. The MLE $\alpha$ has this property; therefore, the limit MLE given $\widetilde{f}$ is $\alpha$ as required. ■

Corollary 8.7 shows that $X_{f,\alpha}^{\lim}$ parametrizes those $f$-stabilizations $\widetilde{f}$ in $X_f$ satisfying the property that the limit MLE given $\widetilde{f}$ is $\alpha$, which leads us naturally to the following.

Definition 8.8 (parameter space of $f$-stabilizations with limit MLE $\alpha$). *Let $f$ denote a sample, and let $\alpha$ be an MLE given $f$. Then the closed subvariety*

$$X_{f,\alpha}^{\lim} \subseteq X_f$$

*of the parameter space of $f$-stabilizations is the* parameter space of $f$-stabilizations $\widetilde{f}$ such that $\alpha$ is the limit MLE given $\widetilde{f}$.

## 9. Star-shaped graphs.

Maximum likelihood estimation in a DAG model on a general DAG $\mathcal{G}$ defines a coupled collection of linear regression problems, one for each vertex with parents. The building block of such a process is the regression of a single vertex onto its parents. This is the special case of a star-shaped graph $\mathcal{G}$. These are connected graphs with a single child vertex; see Figure 1. Statistical models determined by graphs of this type express the child node as a linear combination of the parent nodes plus noise, via a single linear regression. In section 9.1, we consider the case where the MLE exists. In section 9.2, we consider the case where the MLE does not exist.

### 9.1. When the MLE exists.

We show that for a star-shaped $\mathcal{G}$, if the MLE exists given a sample $f$, then the MLE given any $f$-stabilization is the same: the minimal norm MLE given $f$. We apply results from section 6 to prove this. First, we show that the conditions given in Corollary 6.2 are satisfied for all $f$-stabilizations. These characterize when the MLE given an $f$-stabilization is an MLE given $f$. We prove this in Proposition 9.1. Second, we show that only one MLE given $f$ can be obtained in this way and describe it explicitly; see Proposition 9.2 below. This gives an explicit description of the parameter spaces $X_{f,\alpha}$ from Definition 6.9, for all samples $f$ and MLEs $\alpha$ given $f$; see Corollary 9.3.

Proposition 9.1 (the MLE given any $f$-stabilization is an MLE given $f$). *Fix a star-shaped graph $\mathcal{G}$ on $m$ vertices. Let $f$ be a sample, and let $\widetilde{f}$ be a stabilization of $f$. Assume the MLE given $f$ exists. Then the MLE given $\widetilde{f}$ is an MLE given $f$.*

*Proof.* Without loss of generality, the unique child vertex is vertex $m$. Let $f'$ be any $f$-perturbation. To show that the MLE given $\widetilde{f} = f + f'$ is an MLE given $f$, by applying Corollary 6.2 to a star-shaped graph it suffices to show that

$$v_m \in \langle v_i : i \to m\rangle \quad \text{and} \quad \overline{f_m} + \overline{v_m} \in \langle f_i + v_i : i \to m\rangle.$$

We will starting by proving

$$\langle f_i + v_i : i \to m\rangle = \langle f_i : i \to m\rangle \oplus \langle v_i : i \to m\rangle, \tag{9.1}$$

which implies that $\overline{f_m} + \overline{v_m} \in \langle f_i + v_i : i \to m\rangle$, since $\overline{f_m}$ and $\overline{v_m}$ lie in $\langle f_i : i \to m\rangle$ and $\langle v_i : i \to m\rangle$, respectively.

We have

$$\langle f_i + v_i : i \to m\rangle \subseteq \langle f_i : i \to m\rangle \oplus \langle v_i : i \to m\rangle \subseteq \langle f_i : i \to m\rangle \oplus \langle v_1, \ldots, v_m\rangle. \tag{9.2}$$

The left-hand side has dimension equal to the number of parents of $m$, namely $m-1$, since the rows of $\widetilde{f}$ are linearly independent. We now show that the right-hand side has dimension less than or equal to $m-1$.

By definition of an $f$-perturbation, the map $f' : \mathbb{K}^m \to \mathbb{K}^n$ has kernel of dimension $r := \dim \operatorname{im} f$. Then the span of the columns $v_1, \ldots, v_m$ of $f'$ has dimension $m - r$, and the span of the columns $f_1, \ldots, f_m$ of $f$ has dimension $r$. Since the MLE given $f$ exists, we know that $f_m \notin \langle f_i : i \to m\rangle$. Therefore, $\dim\langle f_i : i \to m\rangle = r - 1$. It follows that the right-hand side of (9.2) has dimension $m - 1$. As a result, the inclusions in (9.2) above must all be equalities, giving (9.1).

It remains to show that $v_m \in \langle v_i : i \to m\rangle$. In fact, we will show the stronger statement that $v_m = 0$. Recall that $f'$ has kernel equal to $(\ker f)^\perp$. Therefore, to show that $v_m = 0$, it suffices to show that the standard basis vector $e_m := (0, \ldots, 0, 1) \in \mathbb{K}^m$ lies in $(\ker f)^\perp = \operatorname{im} f^\mathsf{T}$, as $v_m = f'(e_m)$. Since the MLE given $f$ exists, we know that $f_m \notin \langle f_i : i \to m\rangle$, so that $x := f_m - \overline{f_m} \neq 0$. Note that $x \in \langle f_1, \ldots, f_m\rangle$. By construction $x \in \langle f_i : i \to m\rangle^\perp$; therefore, $f_i \cdot x = 0$ for all $i \to m$. Note also that $f_m \cdot x \neq 0$, since otherwise $x \in \langle f_1, \ldots, f_m\rangle^\perp \cap \langle f_1, \ldots, f_m\rangle = \{0\}$, which contradicts $x \neq 0$. Therefore, $f^\mathsf{T}(x) = e_m$, so that $e_m \in \operatorname{im} f^\mathsf{T}$ as required. ∎

We now strengthen Proposition 9.1. That is, in Proposition 9.2 below we show that for a sample $f$ such that an MLE given $f$ exists, not only do we have that the MLEs given $\widetilde{f}$ are MLEs given $f$ for any $f$-stabilization $\widetilde{f}$, but also that only one MLE given $f$ can be obtained in this way, namely the minimal norm MLE given $f$. This is Theorem 1.3.

Proposition 9.2 (the MLE given any $f$-stabilization is the minimal norm MLE given $f$). *Fix a star-shaped graph $\mathcal{G}$ on $m$ vertices, and let $f$ denote a sample such that an MLE given $f$ exists. Then the $\Lambda$-MLE given any stabilization $\widetilde{f}$ of $f$ is the minimal norm $\Lambda$-MLE given $f$.*

Corollary 9.3 below gives an explicit description of the parameter space $X_{f,\alpha}$ from section 6.2, for any sample $f$ for which $\alpha$ is an MLE given $f$.

Corollary 9.3. *Fix a connected DAG $\mathcal{G}$ on $m$ vertices with a unique child vertex, and let $f$ denote a sample such that an MLE given $f$ exists. Let $\alpha$ denote any MLE given $f$. Then*

$$X_f \supseteq X_{f,\alpha} = \begin{cases} \emptyset & \text{if } \alpha \text{ is not the minimal norm MLE given } f; \\ X_f & \text{if } \alpha \text{ is the minimal norm MLE given } f. \end{cases}$$

*Proof of Proposition* 9.2. By relabeling the vertices of $\mathcal{G}$ if necessary we can assume that $m$ is the unique child vertex. Let $\widetilde{f} = f + f'$ denote a stabilization of $f$.

The MLE given $\widetilde{f}$ is an MLE given $f$, by Proposition 9.1. Moreover, in the proof of Proposition 9.1 we have also shown that $v_m = 0$, where $v_m$ is the last column of $f'$. To show that the MLE given $\widetilde{f}$ is the minimal norm MLE given $f$, recall that the $\Lambda$-MLE $\{\lambda_{im}\}_{i\to m}$ given $\widetilde{f}$ is determined by the equation

$$\overline{f_m} + \overline{v_m} = \sum_{i\to m} \lambda_{im} f_i + \sum_{i\to m} \lambda_{im} v_i.$$

Since $\overline{v_m} = \overline{0} = 0$, the coefficients $\lambda_{im}$ satisfy $\sum_{i\to m} \lambda_{im} f_i = \overline{f_m}$ and $\sum_{i\to m} \lambda_{im} v_i = 0$, using the fact that the $v_i$ and $f_i$ are orthogonal to each other. The latter equation is equivalent to asking that the vector $\lambda_m = (\lambda_{1m}, \lambda_{2m}, \ldots, \lambda_{m-1,m})$ lies in $\ker f'_m$ where $f'_m$ is obtained from $f'$ by removing the last column. Let $f_m$ denote the matrix obtained by removing the last column of $f$. Then the minimal norm MLE given $f$ has as its $\Lambda$-MLE the solution to the system $f_m x = \overline{f_m}$ which lies in $(\ker f_m)^\perp$. We claim now that $(\ker f_m)^\perp = \ker f'_m$.

By definition of a sample perturbation, we know that $(\ker f)^\perp = \ker f'$. Since $v_m = 0$, the rows of $f'_m$ and of $f_m$ are also orthogonal to each other; therefore, $(\ker f_m)^\perp \subseteq \ker f'_m$. To show that equality holds, we calculate the dimension of each side. Since $f_m \notin \langle f_i : i \to m\rangle$ by semistability of $f$, on the left-hand side we have

$$\dim(\ker f_m)^\perp = \dim \operatorname{im} f_m = \dim \operatorname{im} f - 1.$$

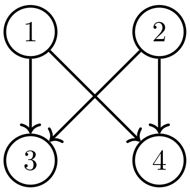
Since $v_m = 0$, we also have that $\dim\ker f'_m = \dim\ker f' - 1$. So on the right-hand side we have

$$\dim\ker f'_m = \dim\ker f' - 1 = \dim(\ker f)^\perp - 1 = \dim\operatorname{im} f - 1.$$

Thus, $(\ker f_m)^\perp = \ker f'_m$. Hence, the MLE given $\widetilde{f}$ is the minimal norm MLE given $f$. ■

For general DAG models, we do not expect the above results to continue to hold for all samples $f$ whose MLE exists. Below is an explicit counterexample.

*Example* 9.4. Let $\mathcal{G}$ be the graph




and let

$$f = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad f' = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 2 & -3 & 3 \end{pmatrix}.$$

The sum $f+f'$ is a stabilization, since the rows of $f$ are orthogonal to those of $f'$, and likewise the columns of $f$ are orthogonal to those of $f'$, and the matrix $f+f'$ is full rank. The MLEs given $f$ in the model on $\mathcal{G}$ are the scalars $\lambda_{ij}$ for all edges $j \to i$ that satisfy $\lambda_{31}+\lambda_{32}=1$, $\lambda_{41}+\lambda_{42}=0$. The MLEs given $f+\epsilon f'$ are

$$\hat{\lambda}_{31}=5, \quad \hat{\lambda}_{32}=-4, \quad \hat{\lambda}_{41}=3, \quad \hat{\lambda}_{42}=-3.$$

This is not the mimimal norm solution.

**9.2. When the MLE does not exist.** Since the $\Lambda$-MLE always exists, we study which $\Lambda$-MLEs can be achieved as the $\Lambda$-MLE given a stabilization. We may also ask which $\Lambda$-MLEs can be achieved as the limit $\Lambda$-MLE given a stabilization. We address both these questions through specific examples.

Proposition 9.5 below gives an example of a DAG $\mathcal{G}$ and sample $f$ such that the $\Lambda$-MLE given any $f$-stabilization is a $\Lambda$-MLE given $f$. It also shows that any $\Lambda$-MLE given $f$ can be obtained as the $\Lambda$-MLE given an $f$-stabilization. This is in contrast with Proposition 9.2 in section 9.1 above, where only one $\Lambda$-MLE can be obtained. Finally, it provides an explicit description of the varieties $X^{\lim}_{f,\lambda}$ appearing in Proposition 8.6.

Proposition 9.5. *Let $\mathcal{G}$ denote the DAG $1 \to 3 \leftarrow 2$, and let $f$ denote the sample with first column $f_1=(1,0,\dots,0)$ and zero second and third column. Then the $\Lambda$-MLE given any $f$-stabilization is an MLE given $f$, and moreover, any $\Lambda$-MLE $\lambda$ of $f$ can be achieved as the $\Lambda$-MLE given a suitable $f$-stabilization. In addition, given a $\Lambda$-MLE $\lambda=(0,b)$ of $f$, we have*

$$X^{\lim}_{f,\lambda}=\{f' \in X_f : v_2 \cdot v_3 = b(v_2 \cdot v_2)\},$$

*where $v_2$ and $v_3$ are the second and third columns of $f'$, respectively.*

*Proof.* The $\Lambda$-MLEs given $f$ are pairs of the form $(0,b)$ for $b \in \mathbb{K}$. We now show that the $\Lambda$-MLE given any $f$-stabilization has this form. A map $f' : \mathbb{K}^m \to \mathbb{K}^n$ with columns $v_1, v_2, v_3$ is an $f$-perturbation if and only if $v_1=0$ (to ensure $\operatorname{im} f' \subseteq (\operatorname{im} f)^\perp$), $v_2, v_3$ have zero first entry (to ensure that $(\ker f')^\perp \subseteq \ker f$) and are linearly independent (to give equality $(\ker f')^\perp = \ker f$). Choose such an $f$-perturbation $f'$.

Then the $\Lambda$-MLE given $\widetilde{f}=f+f'$ is the pair $(c,d)$ such that

$$\pi_{\langle f_1+v_1, f_2+v_2\rangle}(f_3+v_3)=cf_1+dv_2.$$

But

$$\pi_{\langle f_1+v_1, f_2+v_2\rangle}(\overline{f_3}+\overline{v_3})=\pi_{\langle f_1, v_2\rangle}(\overline{v_3})=\overline{v_3},$$

since $\overline{v_3} \in \langle v_2\rangle$. Here $\overline{f_3}=\pi_{\langle f_1,f_2\rangle}(f_3)$ and $\overline{v_3}=\pi_{\langle v_1,v_2\rangle}(v_3)$. Therefore, $cf_1+dv_2=\overline{v_3}\in\langle v_2\rangle$. Since $v_2$ and $f_1$ are orthogonal, it follows that $c=0$ and that $dv_2=\overline{v_3}$. Therefore, $d = v_2 \cdot v_3 / v_2 \cdot v_2$. In other words, the $\Lambda$-MLE given $\widetilde{f}$ is the pair $(0, v_2\cdot v_3/v_2\cdot v_2)$, which is a well-defined $\Lambda$-MLE given $f$. Note that we could also obtain this result by showing instead that the condition of Proposition 6.1 holds, but the direct proof we have given also proves the second part of Proposition 9.5. Indeed, given any $\Lambda$-MLE $(0,b)$ of $f$, we can always find an $f$-perturbation $f'$ such that $v_2 \cdot v_3 = b(v_2 \cdot v_2)$.

The equation above defines a quadratic $Q_b$ in $X_f$, and the $\Lambda$-MLE given $\widetilde{f}$ is $(0,b)$ (which coincides with the limit $\Lambda$-MLE given $\widetilde{f}$) if and only if $\widetilde{f} \in Q_b$. Therefore, setting $\lambda = (0,b)$ we have

$$X_{f,\lambda}^{\lim} = X_f \cap Q_b = \{f' \in X_f : \langle v_2, v_3\rangle = b\langle v_2, v_2\rangle\}. \qquad \blacksquare$$

We now give an example of a sample $f$ such that the $\Lambda$-MLE given any $f$-stabilization is never a $\Lambda$-MLE given $f$. We use this example to illustrate Proposition 8.1, by describing the $\Lambda$-MLE given $\widetilde{f}(\epsilon) = f + \epsilon f'$ for any $f$-stabilization $\widetilde{f} = f + f'$ and its limit as $\epsilon \to 0$.

Proposition 9.6. *Let $\mathcal{G}$ denote the DAG $1 \to 3 \leftarrow 2$. Let $f$ denote the sample with first column $f_1 = (1,0,\dots,0)$, second column $f_2 = (1,1,0,\dots,0)$, and third column $f_3 = (2,1,0,\dots,0)$, with unique $\Lambda$-MLE given $f$ equal to $(1,1)$. Then the $\Lambda$-MLE given any $f$-stabilization is not $(1,1)$. Moreover, the $\Lambda$-MLE given $\widetilde{f}(\epsilon) = f + \epsilon f'$ for any $f$-stabilization $\widetilde{f} = f + f'$ and $\epsilon \neq 0$ is*

$$\left(\frac{1 - 2\epsilon^2(v\cdot v)}{1+\epsilon^2(v\cdot v)}, 1\right),$$

*which tends to $(1,1)$ as $\epsilon \to 0$.*

*Proof.* Let $f'$ denote an $f$-perturbation. Then $f'$ has columns $-v, -v, v$ for some nonzero $v$ with zero first and second entries. The $\Lambda$-MLE given $\widetilde{f} = f + f'$ is the pair $(\alpha,\beta)$ such that

$$\pi_{\langle f_1 - v, f_2 - v\rangle}(f_3 + v) = \alpha(f_1 - v) + \beta(f_2 - v).$$

We claim that the $\Lambda$-MLE given $\widetilde{f}$ is not $(1,1)$, the unique $\Lambda$-MLE given $f$. By Proposition 6.1, this follows from the fact that $f_3 + v$ does not lie in $\langle f_1 - v, f_2 - v\rangle$.

We can check directly that the $\Lambda$-MLE given $\widetilde{f}$ is not a $\Lambda$-MLE given $f$. To calculate the $\Lambda$-MLE given $\widetilde{f}$, observe that $\langle f_1 - v, f_2 - v\rangle = \langle f_1 - v, e_2\rangle$, where $e_2$ is the second standard basis vector in $\mathbb{K}^n$. Since the vectors in the latter span are orthogonal, we have

$$\begin{aligned}
\pi_{\langle f_1 - v, f_2 - v\rangle}(f_3 + v) &= \pi_{\langle f_1 - v, e_2\rangle}(f_3 + v)\\
&= \frac{(f_1 - v)\cdot(f_3+v)}{(f_1 - v)\cdot(f_1 - v)}(f_1 - v) + \frac{e_2\cdot(f_3+v)}{e_2\cdot e_2}e_2\\
&= \frac{2 - v\cdot v}{1 + v\cdot v}(e_1 - v) + e_2\\
&= \frac{1 - 2v\cdot v}{1 + v\cdot v}(e_1 - v) + (e_1 + e_2 - v).
\end{aligned}$$

Therefore, the $\Lambda$-MLE given $\widetilde{f}$ is

$$(\alpha,\beta) = \left(\frac{1 - 2v\cdot v}{1 + v\cdot v}, 1\right).$$

Since $v$ is nonzero, we have that $(\alpha,\beta) \neq (1,1)$ for any $\widetilde{f}$. An analogous calculation to the one above shows that the $\Lambda$-MLE given $\widetilde{f}(\epsilon)$ is

$$\left(\frac{1 - 2\epsilon^2(v\cdot v)}{1+\epsilon^2(v\cdot v)}, 1\right).$$

This expression, while never equal to the unique $\Lambda$-MLE $(1,1)$ given $f$ for $\epsilon \neq 0$, tends to $(1,1)$ as $\epsilon$ tends to zero. ■

The above results suggest the following open questions, which are open even for DAG models on star-shaped graphs.

Question 9.7. *Given a DAG $\mathcal{G}$, can we characterize those samples $f$ such that the $\Lambda$-MLE given any $f$-stabilization is a $\Lambda$-MLE given $f$? Is there a sample $f$ such that some $f$-stabilizations have as their $\Lambda$-MLE a $\Lambda$-MLE given $f$, but others do not? Is there an unstable sample $f$ and $\Lambda$-MLE $\lambda$ of $f$ such that $X^{\lim}_{f,\lambda}$ is empty or all of $X_f$?*

Regarding the first question, Proposition 9.1 shows that for DAG models on star-shaped graphs all samples $f$ such that an MLE given $f$ exists have this property, whilst Propositions 9.5 and 9.6 show that unstable samples may or may not have this property. We conjecture, based on these results, that for star-shaped graphs the $\Lambda$-MLE given any $f$-stabilization of a sample $f$ is a $\Lambda$-MLE given $f$ either if an MLE given $f$ exists, or if $f$ does not admit any linear dependencies amongst the unique set of parents.

## 10. Outlook.

We explore in section 10.1 the statistical implications of our results and conclude in section 10.2 with an open problem.

### 10.1. Statistical implications.

Our results can be interpreted as providing a method for resolving nonidentifiability of the MLE in DAG models. Consider a DAG model on a DAG $\mathcal{G}$ and a sample matrix $Y \in \mathbb{K}^{n\times m}$. If the MLE given $Y$ is not unique, then section 5.3 describes a way of performing some additional sampling (involving a strictly smaller number of samples and measurements than those needed to obtain $Y$) to obtain an affine lift $(Y_1,\dots,Y_t)$ of a complete collineation with first term $Y_1 = Y$, which can be packaged into a $Y$-stabilization $\widetilde{Y} = Y + Y'$ using the construction given in section 5.1. Any $\widetilde{Y}$ obtained in this way has the property that the MLE given $\widetilde{Y}(\epsilon) := Y + \epsilon Y'$ is unique for any $\epsilon \neq 0$ and that, as $\epsilon$ tends to zero, the MLE tends to an MLE given $Y$ if one exists (if not, then the result holds for the $\Lambda$-MLE). Thus, choosing a stabilization $\widetilde{Y}$ is a way of singling out a unique MLE given $Y$, which, moreover, has a statistical interpretation in that it is the MLE given a perturbation of $Y$. The unique MLE given $Y$ that is picked out by a stabilization will depend on the stabilization: different choices of stabilization may pick out different MLEs given $Y$. In special cases, the MLE given $Y$ obtained from a $Y$-stabilization is independendent of the stabilization and coincides with the minimal norm MLE.

### 10.2. Complete collineations as the sample space for a statistical model.

This paper gives a way to package an affine lift of a complete collineation $\mathbb{P}(\mathbb{K}^m) \to \mathbb{P}(\mathbb{K}^n)$ into a sample matrix for a DAG model on $m$ vertices. Samples constructed in this way have the advantage that the associated MLE always exists and is unique. In light of this, it is reasonable to ask whether there exists a statistical model such that samples for this model are from the outset affine lifts of complete collineations from $\mathbb{P}(\mathbb{K}^m)$ to $\mathbb{P}(\mathbb{K}^n)$ for some $n$. If such a model were to exist, our expectation based on the results proved in this paper is that the MLE given any sample for this model would always exist and be unique.

Gaussian group models [1] may offer a promising starting point in the search of such a model, if we assume that $\mathcal{G}$ is transitive. In this case the DAG model on $\mathcal{G}$ coincides with the Gaussian group model determined by the representation of the group $G(\mathcal{G})$ on $\mathbb{K}^m$, where

$$G(\mathcal{G}) := \{a \in \mathrm{GL}_m(\mathbb{K}) \mid a_{ij} = 0 \text{ for all } i \neq j \text{ such that } j \not\to i \text{ in } \mathcal{G}\}.$$

The representation of $G(\mathcal{G})$ on $\mathbb{K}^m$ naturally extends to a representation on $(\mathbb{K}^m)^n$ for any $n$. This is the right multiplication action of $G(\mathcal{G})$ on $\mathrm{Mat}_{n\times m}(\mathbb{K})$, which induces a right multiplication action on $\mathbb{P}(\mathrm{Hom}(\mathbb{K}^m,\mathbb{K}^n))$. Since $\mathcal{M}$ is a blow-up of $\mathbb{P}(\mathrm{Hom}(\mathbb{K}^m,\mathbb{K}^n))$ along $G(\mathcal{G})$-invariant centers, it has an induced action of $G(\mathcal{G})$. By contrast to $\mathbb{P}(\mathrm{Hom}(\mathbb{K}^m,\mathbb{K}^n))$, the moduli space $\mathcal{M}$ is not of the form $\mathbb{P}(U)$ with the $G(\mathcal{G})$ action induced by a representation $G(\mathcal{G}) \to \mathrm{GL}(U)$, so it is not obvious how to associate to the action of $G(\mathcal{G})$ on $\mathcal{M}$ a Gaussian group model.

One approach is to use the fact that the action of $G(\mathcal{G})$ on $\mathcal{M}$ is linear, so that $\mathcal{M}$ can be embedded $G(\mathcal{G})$-equivariantly inside a larger projective space $\mathbb{P}(U)$, with $G(\mathcal{G})$ acting linearly on $\mathbb{P}(U)$ via a representation $G(\mathcal{G}) \to \mathrm{GL}(U)$. This representation does indeed gives rise to a Gaussian group model. Unfortunately, this is not quite the model we are after. The sample space is too big: we are interested only in the subvariety of those samples corresponding to complete collineations, and it is unclear how to interpret the condition that $[f] \in \mathbb{P}(U)$ lies in $\mathcal{M}$ in a statistically meaningful way. Moreover, it is unclear how to relate MLEs for this new model to MLEs for the original model.

A remaining open problem then is whether there is another statistical model that can be constructed from the action of $G(\mathcal{G})$ on $\mathcal{M}$, one in which samples are affine lifts of complete collineations, MLEs given samples are always unique, and MLEs can be more easily related to those of the DAG model on $\mathcal{G}$.

## Appendix A. Existence of the limit in Theorem 7.1.

In this appendix, we give two proofs that the limit in Theorem 7.1 exists.

### A.1. Geometric proof.

We prove the first part of Theorem 7.1, that the limit $\lim_{\epsilon\to 0} x(\epsilon)$ exists, via a geometric argument. Recall that $x(\epsilon)$ for $\epsilon \neq 0$ is defined by the equation

$$A(\epsilon)x(\epsilon) = \pi_{A(\epsilon)}(b(\epsilon)),$$

where $A(\epsilon) = A + \epsilon E$ and $b(\epsilon) = b + \epsilon v$ with the columns of $A$ and $b$ orthogonal to the columns of $E$ and $v$. Let $f_1,\dots,f_p$ and $v_1,\dots,v_p$ denote the columns of $A$ and $E$, respectively. Then the coefficients $x_i(\epsilon)$ of the vector $x(\epsilon)$ satisfy

$$\pi_{A(\epsilon)}(b+\epsilon v) = \sum_{i=1}^{p} x_i(\epsilon)(f_i + \epsilon v_i). \tag{A.1}$$

The orthogonality assumptions ensure that $\pi_{A(\epsilon)}(b+\epsilon v) = \pi_{A(\epsilon)}(\overline{b} + \epsilon\overline{v})$, where $\overline{b} = \pi_A(b)$ and $\overline{v} = \pi_E(v)$, by (7.2). We can therefore assume without loss of generality that $b$ and $v$ lie in the column spaces of $A$ and $E$, respectively (if not, then we just work with $\overline{b}$ and $\overline{v}$ instead).

By replacing some of the $f_i$ and $v_i$ by their negatives if necessary, we can assume that $b+\epsilon v$ sits inside the positive orthant, i.e., that $x_i(\epsilon) \geq 0$ for $1 \leq i \leq p$. Our aim is to show that the coefficients $x_i(\epsilon)$ are bounded. This is enough to conclude that $\lim_{\epsilon\to 0}(x_i(\epsilon))$ exists for each $i$, by the following argument. As $A(\epsilon)$ has full rank for each $\epsilon \neq 0$, we have the following formula for $x(\epsilon)$ when $\epsilon \neq 0$:

$$x(\epsilon) = (A(\epsilon)^\dagger A(\epsilon))^{-1} A(\epsilon)^\dagger \pi_{A(\epsilon)}(b(\epsilon)).$$

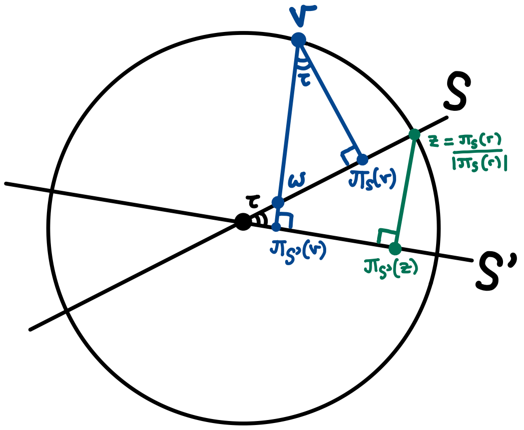



**Figure 2.** *Sketch of the setup of Lemma A.2.*

Since the entries of the vector on the right-hand side are rational functions in $\epsilon$, so are the coefficients of $x(\epsilon)$. But a bounded rational function in $\epsilon$ has a finite limit as $\epsilon \to 0$.

To prove that the coefficients $x_i(\epsilon)$ are bounded above, we start by applying the projection $\pi_A^\perp := \pi_{\langle f_1,\dots,f_p\rangle^\perp}$ to both sides of (A.1). This yields

$$\pi_A^\perp(\pi_{A(\epsilon)}(b+\epsilon v)) = \sum_{i=1}^{p} x_i(\epsilon)\pi_A^\perp(f_i+\epsilon v_i) = \epsilon \sum_{i=1}^{p} x_i(\epsilon)\pi_A^\perp(v_i). \tag{A.2}$$

Our aim is to show that the left-hand side of (A.2) is bounded above by a quantity proportional to $\epsilon$ for small enough $\epsilon$. We use the following two lemmas.

Lemma A.1. *Let $P,S$ be subspaces of $\mathbb{K}^n$ with $P \subseteq S$. Then projections $\pi_S$ and $\pi_{P^\perp}$ commute.*

*Proof.* If $U,V$ are subspaces of $\mathbb{K}^n$ satisfying $V^\perp \subset U$, then

$$\pi_U \circ \pi_V = \pi_{(U\cap V)\oplus V^\perp} \circ \pi_V = \pi_{U\cap V} \circ \pi_V + \pi_{V^\perp} \circ \pi_V = \pi_{U\cap V}.$$

Therefore, the composition of projections is a projection itself (onto $U \cap V$), so the projections commute. Applying this result with $U = S$ and $V = P^\perp$ yields the desired result. ∎

Lemma A.2. *Let $S,S' \in G(p,n)$ denote two subspaces of $\mathbb{K}^n$ of dimension $p$. Let*

$$d(S,S') = \sup_{u\in S, |u|=1} d(u,S'),$$

*where $d(u,S')$ is the distance between $u$ and $S'$. Then if $d(S,S') < \epsilon < 1/4$, for any unit vector $v \in \mathbb{K}^n$, we have $|\pi_S(v) - \pi_{S'}(v)| < 4\epsilon$.*

*Proof.* Assume that $d(S,S') < \epsilon < 1/4$. Let $|v| = 1$ be a unit vector, and let $S'^\perp(v) \subset \mathbb{K}^n$ be the affine space orthogonal to $S'$ and passing through $v$. First, assume that $\pi_{S'}(v)$ and $\pi_S(v)$ are both nonzero, and let $w \in S \cap S'^\perp(v)$ be any point in the intersection $S \cap S'^\perp(v)$; see Figure 2. Observe that $\pi_{S'}(v) = \pi_{S'}(w)$. Then

$$|\pi_S(v) - \pi_{S'}(v)| \le |\pi_{S'}(v) - w| + |\pi_S(v) - w| = d(w,S') + |\pi_S(v) - w|. \tag{A.3}$$

We wish to bound the two terms on the right-hand side of (A.3). First, we show that $d(w, S') < 2\epsilon$. Let $\tau$ denote the angle formed by the subspaces $S$ and $S'$. For $\epsilon$ small enough we have $\cos(\tau) \neq 0$. Then

$$|w| = \frac{|\pi_{S'}(w)|}{\cos(\tau)} = \frac{|\pi_{S'}(v)|}{\cos(\tau)} \leq \frac{1}{\cos(\tau)}.$$

So by choosing $\epsilon$ sufficiently small we can ensure that $|w| < 2$. By definition of $d(S, S')$,

$$d(S, S') \geq d\left(\frac{w}{|w|}, S'\right) = \frac{d(w, S')}{|w|}.$$

Therefore, $d(w, S') = |w| d(S, S') < 2d(S, S') < 2\epsilon$. Next, we show that $|\pi_S(v) - w| < 2\epsilon$. Let $z := \frac{\pi_S(v)}{|\pi_S(v)|}$ be the unit vector in direction $\pi_S(v)$. The two triangles

$$(v, w, \pi_S(v)) \quad \text{and} \quad (0, z, \pi_{S'}(z))$$

are similar as they have two equal angles; see Figure 2. Their scaling ratio is

$$\frac{|v - w|}{|z|} = |v - w| \leq |v| + |w| = |w| < 2.$$

By similarity of the triangles,

$$\frac{|\pi_S(v) - w|}{|\pi_{S'}(z) - z|} = |v - w|,$$

and hence $|\pi_S(v) - w| < 2|z - \pi_{S'}(z)| = 2d(z, S) \leq 2d(S, S') = 2\epsilon$. Hence, from (A.3) we have

$$|\pi_S(v) - \pi_{S'}(v)| \leq d(w, S') + |\pi_S(v) - w| < 2\epsilon + 2\epsilon = 4\epsilon.$$

Finally, we consider the case where $\pi_{S'}(v) = 0$ (the case $\pi_S(v) = 0$ is similar). Assume first that $\pi_{S'}(v) = 0$. Then $|\pi_S(v) - \pi_{S'}(v)| = |\pi_S(v)|$, and the triangles $(0, v, \pi_S(v))$ and $(0, z, \pi_{S'}(z))$ are congruent (isometric). Hence,

$$|\pi_S(v)| = |z - \pi_{S'}(z)| = d(z, S') < \epsilon.$$ ∎

We use Lemmas A.1 and A.2 to show that the left-hand side of (A.2) is bounded above by a quantity proportional to $\epsilon$. Let $L_\epsilon := \langle f_1 + \epsilon v_1, \ldots, f_p + \epsilon v_p \rangle$, and let $L_0$ denote the limit of $L_\epsilon$ as $\epsilon \to 0$ in the Grassmannian $G(p, n)$ of $p$-dimensionsal subspaces of $\mathbb{R}^n$. Since $\langle f_1, \ldots, f_p \rangle \subseteq L_0$, we can apply Lemma A.1 with $P = \langle f_1, \ldots, f_p \rangle$ and $S = L_0$ to obtain

$$\pi_A^\perp(\pi_{L_0}(b + \epsilon v)) = \pi_{L_0}(\pi_A^\perp(b + \epsilon v)) = \epsilon \pi_{L_0}(\pi_A^\perp(v)), \tag{A.4}$$

where the second equality follows from $\pi_A^\perp(b) = 0$ (since $b$ lies in the column space of $A$ by assumption). Therefore, by (A.2) we have

$$\begin{aligned}
\epsilon \sum_{i=1}^{p} x_i(\epsilon) \pi_A^\perp(v_i) &= \pi_A^\perp(\pi_{A(\epsilon)}(b + \epsilon v)) \\
&= \pi_A^\perp(\pi_{A(\epsilon)}(b + \epsilon v)) + \pi_A^\perp(\pi_{L_0}(b + \epsilon v)) - \pi_A^\perp(\pi_{L_0}(b + \epsilon v)) \\
&= \pi_A^\perp(\pi_{A(\epsilon)}(b + \epsilon v)) + \epsilon \pi_{L_0}\left(\pi_A^\perp(v)\right) - \pi_A^\perp(\pi_{L_0}(b + \epsilon v)) \text{ by (A.4)} \\
&= \epsilon \pi_{L_0}\left(\pi_A^\perp(v)\right) + \pi_A^\perp((\pi_{A(\epsilon)}(b + \epsilon v)) - \pi_{L_0}(b + \epsilon v)).
\end{aligned} \tag{A.5}$$

We can use Lemma A.2 to bound the norm of the difference $\pi_{A(\epsilon)}(b+\epsilon v)-\pi_{L_0}(b+\epsilon v)=\pi_{L_\epsilon}(b+\epsilon v)-\pi_{L_0}(b+\epsilon v)$, by setting $S=L_\epsilon$ and $S'=L_0$. Note that by definition

$$d(L_\epsilon,L_0)<d(f_i+\epsilon v_i,f_i)=\epsilon|v_i|$$

for $1\le i\le p$ and hence

$$d(L_\epsilon,L_0)<\epsilon\max_i|v_i|.$$

Then, provided that $\epsilon\max_i|v_i|<\frac{1}{4}$, we obtain

$$\left|\pi_{L_\epsilon}\left(\frac{b+\epsilon v}{|b+\epsilon v|}\right)-\pi_{L_0}\left(\frac{b+\epsilon v}{|b+\epsilon v|}\right)\right|<4\epsilon\max_i|v_i|$$

by Lemma A.2. Hence,

$$|\pi_{A(\epsilon)}(b+\epsilon v)-\pi_{L_0}(b+\epsilon v)|<4\epsilon|b+\epsilon v|\max_i|v_i|<4\epsilon(|b|+|v|)\max_i|v_i|. \tag{A.6}$$

Projecting onto the orthogonal complement of the column space of $A$ can only decrease the norm, hence we obtain from (A.2), (A.5), and (A.6) that

$$\pi_A^\perp(\pi_{A(\epsilon)}(b+\epsilon v))=\epsilon\sum_{i=1}^p x_i(\epsilon)|\pi_A^\perp(v_i)|<\epsilon|\pi_{L_0}(\pi_A^\perp(v))|+4\epsilon(|b|+|v|)\max_i|v_i|. \tag{A.7}$$

Since $x_i(\epsilon)\ge 0$, it follows that the $x_i(\epsilon)$ are bounded above, with bound

$$x_i(\epsilon)\le\frac{|\pi_{L_0}(\pi_A^\perp(v))|+4(|b|+|v|)\max_i|v_i|}{\max_{1\le i\le p}|\pi_A^\perp(v_i)|}.$$

**A.2. Algebraic proof.** We now prove the first part of Theorem 7.1 again, via an algebraic argument. The algebraic approach has the advantage that it also gives an explicit description of the limit. We assume in this section that $\mathbb{K}=\mathbb{R}$. The proof over $\mathbb{K}=\mathbb{C}$ is identical as long as we replace the transpose by the conjugate transpose.

Let $A(\epsilon)=A+\epsilon E$ and $b(\epsilon)=b+\epsilon v$ be as in Theorem 7.1. Then the unique solution $x(\epsilon)$ to $A(\epsilon)x(\epsilon)=\pi_{A(\epsilon)}(b(\epsilon))$ is given by

$$x(\epsilon)=A(\epsilon)^+\pi_{A(\epsilon)}(b+\epsilon v).$$

Let $f_i$ denote the columns of $A$ and $v_i$ the columns of $E$. Define $\overline{b}=\pi_A(b)$ and $\overline{v}=\pi_E(v)$. Then $\pi_{A(\epsilon)}(b(\epsilon))=\pi_{A(\epsilon)}(\overline{b}+\epsilon\overline{v})$, by (7.2).

Since $A(\epsilon)$ has full column rank by assumption for $\epsilon\ne 0$, its pseudo-inverse is

$$A(\epsilon)^+=(A(\epsilon)^\mathsf{T}(A(\epsilon)))^{-1}A(\epsilon)^\mathsf{T}.$$

Therefore,

$$\begin{aligned}x(\epsilon)&=(A(\epsilon)^\mathsf{T}A(\epsilon))^{-1}A(\epsilon)^\mathsf{T}A(\epsilon)(A(\epsilon)^\mathsf{T}A(\epsilon))^{-1}A(\epsilon)^\mathsf{T}(\overline{b}+\epsilon\overline{v})\\&=(A(\epsilon)^\mathsf{T}A(\epsilon))^{-1}A(\epsilon)^\mathsf{T}(\overline{b}+\epsilon\overline{v}).\end{aligned}$$

Define $C(\epsilon) := A(\epsilon)^{\mathsf{T}} A(\epsilon)$. Note that $C(\epsilon) = A^{\mathsf{T}} A + \epsilon^2 E^{\mathsf{T}} E$, since the columns of $A$ are orthogonal to those of $E$. Then

$$x(\epsilon) = C(\epsilon)^{-1} A(\epsilon)^{\mathsf{T}}(\overline{b} + \epsilon\overline{v}) = \frac{1}{\det C(\epsilon)} \operatorname{adj} C(\epsilon) \begin{pmatrix} f_1 \cdot \overline{b} + \epsilon^2 v_1 \cdot \overline{v} \\ \vdots \\ f_p \cdot \overline{b} + \epsilon^2 v_p \cdot \overline{v} \end{pmatrix}. \tag{A.8}$$

We seek an expression for $x(\epsilon)$ without powers of $\epsilon$ in the denominator. We begin with the case where $\overline{b} + \overline{v}$ lies in $\langle f_i + v_i : 1 \le i \le p\rangle$. This is the same assumption as in Proposition 6.1.

Lemma A.3. *Suppose $\overline{b} + \overline{v} \in \langle f_i + v_i : 1 \le i \le p\rangle$, so that $\overline{b} + \overline{v} = \sum_{i=1}^{p} \mu_i (f_i + v_i)$ for some $\mu_i \in \mathbb{K}$. Let $e_i$ be the $i$th standard basis vector in $\mathbb{K}^p$. Then for all $\epsilon$, we have $x(\epsilon) = \sum_{i=1}^{p} \mu_i e_i$.*

*Proof.* We can assume that $\overline{b} + \overline{v} = f_i + v_i$ for some $1 \le i \le p$. The case where $\overline{b} + \overline{v}$ is a general linear combination follows similarly. We want to prove that $x(\epsilon) = e_i$. We give both an algebraic and a geometric proof of this result. We start with the algebraic proof.

From (A.8) we see that the $i$th entry of $x(\epsilon)$ is $1/\det C(\epsilon)$ times the dot product of the $i$th row of $\operatorname{adj} C(\epsilon)$ with the$i$-th column of $C(\epsilon)$. The latter is $\det C(\epsilon)$, in its cofactor expansion along the $i$th column of $C(\epsilon)$. Hence, the $i$th entry of $x(\epsilon)$ is 1. Now take $i \neq l \in \{1, \ldots, p\}$. The $l$-th entry of $x(\epsilon)$ is $1/\det(\epsilon)$ times the dot product of the $l$th row of $\operatorname{adj} C(\epsilon)$ with the $i$th column of $C(\epsilon)$. Since $l \neq i$, this is a cofactor expansion using a different column, and therefore, the expression vanishes. Hence, $x(\epsilon) = e_i$.

The geometric proof is as follows. The entries of $x(\epsilon)$ are the coefficients in front of $\{f_i + \epsilon v_i : 1 \le i \le p\}$ in the projection of $\overline{b} + \epsilon\overline{v}$ to $\langle f_i + \epsilon v_i : 1 \le i \le p\rangle$. Since we are assuming $\overline{b} + \overline{v} = f_i + v_i$, we have $\overline{b} = f_i$ and $\overline{v} = v_i$ by the orthogonality assumptions. Therefore,

$$\pi_{\langle f_i + \epsilon v_i : i \in \{1,\ldots,p\}\rangle}(\overline{b} + \epsilon\overline{v}) = \pi_{\langle f_i + \epsilon v_i : i \in \{1,\ldots,p\}\rangle}(f_i + \epsilon v_i) = f_i + \epsilon v_i.$$

So $x(\epsilon) = e_i$. ∎

We now turn to the general case. Since $\epsilon$ only appears in $x(\epsilon)$ with even powers, to simplify calculations we let

$$C'(\epsilon) = A^{\mathsf{T}} A + \epsilon E^{\mathsf{T}} E \tag{A.9}$$

and consider the following vector:

$$x'(\epsilon) = \frac{1}{\det C'(\epsilon)} \operatorname{adj} C'(\epsilon) \begin{pmatrix} f_1 \cdot \overline{b} + \epsilon v_1 \cdot \overline{v} \\ \vdots \\ f_p \cdot \overline{b} + \epsilon v_p \cdot \overline{v} \end{pmatrix}. \tag{A.10}$$

Note that $\lim_{\epsilon \to 0} x'(\epsilon)$ exists if and only if $\lim_{\epsilon \to 0} x(\epsilon)$ exists, since $x(\epsilon) = x'(\epsilon^2)$. We expand the polynomial $\det C'(\epsilon)$ as

$$\det C'(\epsilon) = c_0 + c_1\epsilon + c_2\epsilon^2 + \cdots + c_p\epsilon^p \tag{A.11}$$

for some coefficients $c_i \in \mathbb{K}$. Similarly, we write

$$\text{(A.12)} \qquad \operatorname{adj} C'(\epsilon) \begin{pmatrix} f_1 \cdot \overline{b} + \epsilon v_1 \cdot \overline{v} \\ \vdots \\ f_p \cdot \overline{b} + \epsilon v_p \cdot \overline{v} \end{pmatrix} = D_0 + D_1\epsilon + \cdots + D_p \epsilon^p$$

for some column vectors $D_i$, since the entries are polynomials in $\epsilon$. Then

$$\text{(A.13)} \qquad x'(\epsilon) = \frac{D_0 + D_1\epsilon + \cdots + D_p\epsilon^p}{c_0 + c_1\epsilon + \cdots + c_p\epsilon^p}.$$

We see that $\lim_{\epsilon \to 0} x'(\epsilon)$ exists if and only if whenever $c_k = 0$ for all $k \le l$ (for some $0 \le l \le p$), we have $D_k = 0$ for all $k \le l$.

We now describe the coefficients $c_i$. First,

$$c_0 = \det C(0) = \det A^\mathsf{T} A,$$

since to obtain the constant term in $\epsilon$, only the matrix $A^\mathsf{T} A$ need be considered. We use Jacobi's formula for the derivative of a determinant to calculate

$$\begin{aligned} c_1 &= \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \det C'(\epsilon) = \operatorname{tr}\left(\operatorname{adj} C'(\epsilon) \frac{d}{d\epsilon} C'(\epsilon)\right)(0) \\ &= \operatorname{tr}((\operatorname{adj} C'(0)) E^\mathsf{T} E) = \operatorname{tr}((\operatorname{adj} A^\mathsf{T} A) E^\mathsf{T} E). \end{aligned}$$

We apply Jacobi's formula again to compute $c_2$:

$$c_2 = \frac{1}{2} \frac{d^2}{d\epsilon^2}\bigg|_{\epsilon=0} \det C'(\epsilon) = \frac{1}{2} \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \left(\frac{d}{d\epsilon} \det C'(\epsilon)\right) = \frac{1}{2} \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \operatorname{tr}((\operatorname{adj} C'(\epsilon)) E^\mathsf{T} E).$$

Proceeding in this way, we obtain

$$\text{(A.14)} \qquad c_i = \frac{1}{i!} \operatorname{tr}\left(\left(\frac{d^{i-1}}{d\epsilon^{i-1}}\bigg|_{\epsilon=0} \operatorname{adj} C'(\epsilon)\right) E^\mathsf{T} E\right)$$

for all $i \in \{1, \ldots, p\}$. We now turn to the coefficients $D_i$. Expanding (A.13) gives

$$\text{(A.15)} \qquad D_0 + D_1\epsilon + \cdots + D_p\epsilon^p = \operatorname{adj} C'(\epsilon) \begin{pmatrix} f_1 \cdot \overline{b} \\ \vdots \\ f_p \cdot \overline{b} \end{pmatrix} + \epsilon \operatorname{adj} C'(\epsilon) \begin{pmatrix} v_1 \cdot \overline{v} \\ \vdots \\ v_p \cdot \overline{v} \end{pmatrix}.$$

It follows that

$$D_0 = \operatorname{adj} C'(0) \begin{pmatrix} f_1 \cdot \overline{b} \\ \vdots \\ f_p \cdot \overline{b} \end{pmatrix} = \operatorname{adj} A^\mathsf{T} A \begin{pmatrix} f_1 \cdot \overline{b} \\ \vdots \\ f_p \cdot \overline{b} \end{pmatrix}.$$

The coefficient $D_1$ is the sum of the degree 1 part of $\operatorname{adj} C'(\epsilon)$ multiplied by the vector with entries $f_i \cdot \overline{b}$, and the degree 0 part of $\operatorname{adj} C'(\epsilon)$ multiplied by the vector with entries $v_i \cdot \overline{v}$. Therefore, we have

$$D_1 = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \operatorname{adj} C'(\epsilon) \begin{pmatrix} f_1 \cdot \overline{b} \\ \vdots \\ f_p \cdot \overline{b} \end{pmatrix} + \operatorname{adj} C'(0) \begin{pmatrix} v_1 \cdot \overline{v} \\ \vdots \\ v_p \cdot \overline{v} \end{pmatrix}. \tag{A.16}$$

Proceeding in this way, we obtain for all $i \in \{1, \dots, p\}$ that

$$D_i = \frac{1}{i!} \frac{d^i}{d\epsilon^i}\bigg|_{\epsilon=0} \operatorname{adj} C'(\epsilon) \begin{pmatrix} f_1 \cdot \overline{b} \\ \vdots \\ f_p \cdot \overline{b} \end{pmatrix} + \frac{1}{(i-1)!} \frac{d^{i-1}}{d\epsilon^{i-1}}\bigg|_{\epsilon=0} \operatorname{adj} C'(\epsilon) \begin{pmatrix} v_1 \cdot \overline{v} \\ \vdots \\ v_p \cdot \overline{v} \end{pmatrix}. \tag{A.17}$$

We want to show that if $c_i = 0$ for all $i \leq l$ (for some $0 \leq l \leq p$), then $D_i = 0$ for all $i \leq l$. The following lemma achieves this. Indeed, conditions (a) and (b) together ensure that $D_i = 0$ for all $i \leq l$, based on the expression for $D_i$ given in (A.17) above.

**Lemma A.4.** *Fix $0 \leq l \leq p$ and suppose that $c_i = 0$ for all $i \leq l$. Then*
(a) $\frac{d^i}{d\epsilon^i}|_{\epsilon=0} \operatorname{adj} C'(\epsilon) = 0$ *for all $i < l$;*
(b) $\frac{d^l}{d\epsilon^l}|_{\epsilon=0} \operatorname{adj} C'(\epsilon) A^\mathsf{T} A = 0.$

*Proof.* We use strong induction. We start with the base case $l = 0$, so we assume that $c_0 = 0$. For (a) there is nothing to check, since $l = 0$. To show (b), it is enough to show that

$$\operatorname{adj} C'(0) \begin{pmatrix} f_1 \cdot f_k \\ \vdots \\ f_p \cdot f_k \end{pmatrix} = 0$$

for each $k \in \{1, \dots, p\}$. Now $\operatorname{adj} C'(0) = A^\mathsf{T} A$, and we have that

$$\operatorname{adj} A^\mathsf{T} A \begin{pmatrix} f_1 \cdot f_k \\ \vdots \\ f_p \cdot f_k \end{pmatrix} = (0, \dots, 0, \det A^\mathsf{T} A, 0, \dots, 0)^\mathsf{T},$$

where $\det A^\mathsf{T} A$ appears in the $k$th entry, by the same cofactor expansion argument as in the proof of Lemma A.3. Since $c_0 = \det A^\mathsf{T} A = 0$, it follows that the above expression vanishes, which shows (b) when $l = 0$. This establishes the base case.

Fix some $1 \leq l \leq p$ and suppose that $c_i = 0$ for all $i \leq l$. Assume the following:
($a_{l-1}$) $\frac{d^k}{d\epsilon^k}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon) = 0$ for all $k < l - 1$;
($b_{l-1}$) $\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon) A^\mathsf{T} A = 0.$

We wish to show the following:
($a_l$) $\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon) = 0;$
($b_l$) $\frac{d^l}{d\epsilon^l}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon) A^\mathsf{T} A = 0.$

We start by proving $(a_l)$. By $(b_{l-1})$ we know that $\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)A^\mathsf{T}A = 0$. Therefore, it is sufficient to show that

$$\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)E^\mathsf{T}E = 0, \tag{A.18}$$

since $A^\mathsf{T}A + E^\mathsf{T}E = (A+E)^\mathsf{T}(A+E)$ is invertible. We now prove (A.18) using the assumptions $c_l = 0$ and $(a_{l-1})$. Using the expression for $c_l$ given in (A.14) we have

$$\operatorname{tr}\left(\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)E^\mathsf{T}E\right) = 0. \tag{A.19}$$

If $\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)$ is positive semidefinite, then the product inside the trace in (A.19) is zero. This is because $E^\mathsf{T}E$ is positive semidefinite and the trace of a product of positive semidefinite matrices is zero if and only if the product is zero. We can establish that $\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)$ is positive semidefinite using $(b_{l-1})$. By $(b_{l-1})$, we know that

$$\begin{aligned}
\operatorname{adj} C'(\epsilon) &= \epsilon^{l-1}\left(\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)\right) + \epsilon^{l}\left(\frac{d^{l}}{d\epsilon^{l}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)\right) + \cdots \\
&= \epsilon^{l-1}\left(\left(\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)\right) + \epsilon\left(\frac{d^{l}}{d\epsilon^{l}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)\right) + \cdots\right).
\end{aligned}$$

The matrix $C'(\epsilon)$ is positive semidefinite. Hence, $\lim_{\epsilon\to 0} \epsilon^{-l+1} \operatorname{adj} C'(\epsilon)$ is positive semidefinite. But this limit is

$$\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon).$$

Hence, the product inside the trace in (A.19) is zero. This proves $(a_l)$. To show $(b_l)$, it is enough to show that for any $k \in \{1, \dots, p\}$ we have

$$\frac{d^{l}}{d\epsilon^{l}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)\begin{pmatrix} f_1 \cdot f_k \\ \vdots \\ f_p \cdot f_k \end{pmatrix} = 0.$$

Expanding the expression on the left-hand side gives

$$\begin{aligned}
&\frac{d^{l}}{d\epsilon^{l}}\Big|_{\epsilon=0}\left(\operatorname{adj} C'(\epsilon)\begin{pmatrix} f_1 \cdot f_k + \epsilon v_1 \cdot v_k \\ \vdots \\ f_p \cdot f_k + \epsilon v_p \cdot v_k \end{pmatrix}\right) - \epsilon\left(\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)\right)\begin{pmatrix} v_1 \cdot v_k \\ \vdots \\ v_p \cdot v_k \end{pmatrix} \\
&= \left(\frac{d^{l}}{d\epsilon^{l}}\Big|_{\epsilon=0} \det C'(\epsilon)\right) e_k - \epsilon\left(\frac{d^{l-1}}{d\epsilon^{l-1}}\Big|_{\epsilon=0} \operatorname{adj} C'(\epsilon)\right)\begin{pmatrix} v_1 \cdot v_k \\ \vdots \\ v_p \cdot v_k \end{pmatrix}.
\end{aligned}$$

This is zero, because $\frac{d^{l}}{d\epsilon^{l}}\Big|_{\epsilon=0} \det C'(\epsilon) = 0$ and the product inside the trace in (A.19) is zero. This proves $(b_l)$. ■

Corollary A.5. *Let* $A(\epsilon) = A + \epsilon E$ *and* $b(\epsilon) = b + \epsilon v$ *be as in Theorem 7.1. Let* $\{f_1, \ldots, f_p\}$ *be the columns of* $A$*, and let* $\{v_1, \ldots, v_p\}$ *be the columns of* $E$*. Define* $\overline{b} = \pi_A(b)$ *and* $\overline{v} = \pi_E(v)$*. Let* $C'(\epsilon)$ *be defined as at (A.9). Let* $x(\epsilon)$ *be the unique solution to* $A(\epsilon)x(\epsilon) = \pi_{A(\epsilon)}(b(\epsilon))$*. Then the limit* $\lim_{\epsilon \to 0} x(\epsilon)$ *exists and equals* $D_l/c_l$*, where* $l = 0$ *if* $A$ *has full rank, and if not* $l$ *denotes the smallest integer in* $\{1, \ldots, p\}$ *with* $\operatorname{tr}(\frac{d^{l-1}}{\epsilon^{l-1}}|_{\epsilon=0} \operatorname{adj} C'(\epsilon) E^{\top} E) \neq 0$.

*Proof.* Since $c_i = 0$ for all $i < l$, by Lemma A.4 we have $D_i = 0$ for all $i < l$ (based on the expression for $D_i$ given in (A.17)), so that

$$x(\epsilon) = \frac{1}{c_l \epsilon^{2l} + c_{l+1}\epsilon^{2l+2} + \cdots} \left( D_l \epsilon^{2l} + D_{l+1}\epsilon^{2l+2} + \cdots \right) = \frac{D_l + D_{l+1}\epsilon^2 + \cdots}{c_l + c_{l+1}\epsilon^2 + \cdots}.$$

Therefore, the limit of $x(\epsilon)$ as $\epsilon$ tends to zero exists, and equals $D_l/c_l$. ∎

**Acknowledgments.** We thank Visu Makam for helpful discussions. The second author thanks Johan Martens for bringing complete collineations and [31] to her attention in a different context.

## REFERENCES


[1] C. Améndola, K. Kohn, P. Reichenbach, and A. Seigal, *Invariant theory and scaling algorithms for maximum likelihood estimation*, SIAM J. Appl. Algebra Geom., 5 (2021), pp. 304–337, https://doi.org/10.1137/20M1328932.

[2] S. Basu, R. Pollack, and M.-F. Roy, *Algorithms in Real Algebraic Geometry*, Vol. 10, Springer, 2006, https://doi.org/10.1007/3-540-33099-2.

[3] A. Ben-Israel, *On error bounds for generalized inverses*, SIAM J. Numer. Anal., 3 (1966), pp. 585–592, https://doi.org/10.1137/0703050.

[4] D. I. Bernstein, S. Dewar, S. J. Gortler, A. Nixon, M. Sitharam, and L. Theran, *Maximum likelihood thresholds via graph rigidity*, Ann. Appl. Probab., 34 (2024), pp. 3288–3319, https://doi.org/10.1214/23-AAP2039.

[5] G. Blekherman and R. Sinn, *Maximum likelihood threshold and generic completion rank of graphs*, Discrete Comput. Geom., 61 (2019), pp. 303–324, https://doi.org/10.1007/s00454-018-9990-3.

[6] S. L. Buhl, *On the existence of maximum likelihood estimators for graphical Gaussian models*, Scand. J. Stat., 20 (1993), pp. 263–270, https://www.jstor.org/stable/4616281.

[7] P. Danaher, P. Wang, and D. M. Witten, *The joint graphical lasso for inverse covariance estimation across multiple classes*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 76 (2014), pp. 373–397, https://doi.org/10.1111/rssb.12033.

[8] H. Derksen and V. Makam, *Maximum likelihood estimation for matrix normal models via quiver representations*, SIAM J. Appl. Algebra Geom., 5 (2021), pp. 338–365, https://doi.org/10.1137/20M1369348.

[9] H. Derksen, V. Makam, and M. Walter, *Maximum likelihood estimation for tensor normal models via castling transforms*, in Forum of Mathematics, Sigma, Vol. 10, Cambridge University Press, 2022, e50, https://doi.org/10.1017/fms.2022.37.

[10] R. A. Dinu, M. Michałek, and M. Vodička, *Geometry of the Gaussian Graphical Model of the Cycle*, preprint, arXiv:2111.02937, 2021.

[11] M. Drton, C. Fox, A. Käufl, and G. Pouliot, *The maximum likelihood threshold of a path diagram*, Ann. Statist., 47 (2019), pp. 1536–1553, https://doi.org/10.1214/18-AOS1724.

[12] M. Drton, S. Kuriki, and P. Hoff, *Existence and uniqueness of the Kronecker covariance MLE*, Ann. Statist., 49 (2021), pp. 2721–2754, https://doi.org/10.1214/21-AOS2052.

[13] M. Giusti and J. Heintz, *La détermination des points isolés et de la dimension d'une variété algébrique peut se faire en temps polynomial*, Comput. Algebraic Geom. Commutative Algebra, Sympos. Math. 34, Cambridge University Press, 1994, pp. 216–256.

[14] G. H. Golub, P. C. Hansen, and D. P. O'Leary, *Tikhonov regularization and total least squares*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 185–194, https://doi.org/10.1137/S0895479897326432.
[15] E. Gross and S. Sullivant, *The maximum likelihood threshold of a graph*, Bernoulli, 24 (2018), pp. 386–407, https://doi.org/10.3150/16-BEJ881.
[16] P. C. Hansen, *Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion*, SIAM, Philadelphia, 1998, https://doi.org/10.1137/1.9780898719697.
[17] P. Le Barz and Y. Hervier, *Enumerative Geometry and Classical Algebraic Geometry*, Progress in Mathematics, Birkhäuser, 1982, https://doi.org/10.1007/978-1-4684-6726-0.
[18] V. Makam, P. Reichenbach, and A. Seigal, *Symmetries in directed Gaussian graphical models*, Electron. J. Statist., 17 (2023), pp. 3969–4010, https://doi.org/10.1214/23-EJS2192.
[19] L. Manivel, M. Michałek, L. Monin, T. Seynnaeve, and M. Vodička, *Complete quadrics: Schubert calculus for Gaussian models and semidefinite programming*, J. Eur. Math. Soc., 26 (2023), pp. 3091–3135, https://doi.org/10.4171/jems/1330.
[20] G. C. McDonald, *Ridge regression*, Wiley Interdiscip. Rev. Comput. Stat., 1 (2009), pp. 93–100, https://doi.org/10.1002/wics.14.
[21] M. Michałek, L. Monin, and J. A. Wisniewski, *Maximum likelihood degree, complete quadrics, and* $\mathbb{C}^*$*-action*, SIAM J. Appl. Algebra Geom., 5 (2021), pp. 60–85, https://doi.org/10.1137/20M1335960.
[22] P. Reichenbach, *Invariant Theory in Computational Complexity and Algebraic Statistics*, Ph.D. thesis, TU, Berlin, 2023, https://doi.org/10.14279/depositonce-18306.
[23] A. M. E. Saleh, M. Arashi, and B. G. Kibria, *Theory of Ridge Regression Estimation with Applications*, John Wiley & Sons, 2019, https://doi.org/10.1002/9781118644478.
[24] C. Squires, A. Seigal, S. S. Bhate, and C. Uhler, *Linear causal disentanglement via interventions*, in Proceedings of the 40th International Conference on Machine Learning, 2023, 1348.
[25] G. W. Stewart, *On the continuity of the generalized inverse*, SIAM J. Appl. Math., 17 (1969), pp. 33–45, https://doi.org/10.1137/0117004.
[26] B. Sturmfels and C. Uhler, *Multivariate Gaussians, semidefinite matrix completion, and convex algebraic geometry*, Ann. Inst. Statist. Math., 62 (2010), pp. 603–638, https://doi.org/10.1007/s10463-010-0295-4.
[27] S. Sullivant, *Algebraic Statistics*, Grad. Stud. in Math. 194, American Mathematical Society, Providence, RI, 2018, https://doi.org/10.1090/gsm/194.
[28] M. Thaddeus, *Complete collineations revisited*, Math. Ann., 315 (1999), pp. 1432–1807, https://doi.org/10.1007/s002080050324.
[29] A. Thorup and S. Kleiman, *Complete bilinear forms*, in Algebraic Geometry Sundance 1986, Springer, Berlin, Heidelberg, 1988, pp. 253–320.
[30] A. N. Tikhonov, *On the stability of inverse problems*, Dokl. akad. nauk sssr, 39 (1943), pp. 195–198.
[31] I. Vainsencher, *Complete collineations and blowing up determinantal ideals*, Math. Ann., 267 (1984), pp. 417–432, https://doi.org/10.1007/BF01456098.
[32] A. Wille, P. Zimmermann, E. Vranová, A. Fürholz, O. Laule, S. Bleuler, L. Hennig, A. Prelić, P. von Rohr, L. Thiele, E. Zitzler, W. Gruissem, and P. Bühlmann *Sparse graphical Gaussian modeling of the isoprenoid gene network in Arabidopsis thaliana*, Genome Biol., 5 (2004), r92, https://doi.org/10.1186/gb-2004-5-11-r92.
[33] R. A. Wooding, *The multivariate distribution of complex normal variables*, Biometrika, 43 (1956), pp. 212–215, https://doi.org/10.1093/biomet/43.1-2.212.