# The maximum likelihood threshold of a graph

ELIZABETH GROSS[1] and SETH SULLIVANT[2]

[1]*Department of Mathematics and Statistics, One Washington Square, San José State University, San José, CA, 95192-0103, USA. E-mail: elizabeth.gross@sjsu.edu*
[2]*Department of Mathematics, North Carolina State University, Box 8205, Raleigh, NC, 27695-8205, USA. E-mail: smsulli2@ncsu.edu*

The maximum likelihood threshold of a graph is the smallest number of data points that guarantees that maximum likelihood estimates exist almost surely in the Gaussian graphical model associated to the graph. We show that this graph parameter is connected to the theory of combinatorial rigidity. In particular, if the edge set of a graph $G$ is an independent set in the $(n-1)$-dimensional generic rigidity matroid, then the maximum likelihood threshold of $G$ is less than or equal to $n$. This connection allows us to prove many results about the maximum likelihood threshold. We conclude by showing that these methods give exact bounds on the number of observations needed for the score matching estimator to exist with probability one.

*Keywords:* algebraic matroids; Gaussian graphical models; matrix completion; maximum likelihood estimation; maximum likelihood threshold; rigidity theory; score matching estimator; weak maximum likelihood threshold

## 1. Introduction

Let $X = (X_1, \ldots, X_m)$ be a $m$-dimensional random vector distributed according to a multivariate normal distribution, that is, $X \sim \mathcal{N}(\mu, \Sigma)$. In a Gaussian graphical model, an undirected graph $G = (\{1, \ldots, m\}, E)$ encodes the conditional independence structure of the distribution: the edge $(i, j) \notin E$ if and only if $X_i$ and $X_j$ are conditionally independent given the remaining variables. Originally introduced by Dempster [6] under the name of covariance selection models, Gaussian graphical models have found a variety of applications, especially in systems biology and bioinformatics. For example, these models are used to model gene regulatory networks [7,23,30] and to infer pathways in metabolic networks [19]. Lauritzen [21] and Whittaker [29] both give general introductions to Gaussian graphical models.

In this paper, we are concerned with the existence of the maximum likelihood estimator (MLE) of the covariance matrix when the mean vector $\mu = 0$. For Gaussian graphical models, when the number of observations $n$ is larger than the number of random variables $m$, the MLE exists with probability one. But it is often the case, especially in biological applications, that $m \gg n$. In this setting, it is still possible that the MLE exists with probability one, which invites the question: *For a given graph $G$, what is the smallest $n$ such that the maximum likelihood estimator of $\Sigma$ exists with probability one?* We denote the resulting graph invariant by $\mathrm{mlt}(G)$ and call it the *maximum likelihood threshold.*

One motivation for studying the maximum likelihood threshold considers the use of Gaussian graphical models to big data problems with a large number of variates but a small sample size. In this setting, a regularization method (e.g., graphical lasso [9]) might be used to estimate a graph structure. Then with the graph structure fixed, maximum likelihood would be used to compute the estimates of the sparse inverse covariance matrix associated to the graph. However, if the graph has the property that the sample size is smaller than the maximum likelihood threshold, this suggests that one should go back with a larger regularization parameter to re-estimate to try for a sparser graph.

As originally proven in [6] and discussed further in [25], the existence of the MLE for given data set and for a particular Gaussian graphical model is equivalent to the existence of a full rank positive definite matrix completion of the incomplete matrix obtained by keeping only the diagonal entries and entries corresponding to $E$ of the sample covariance matrix $\Sigma_0$. Let $\mathbb{S}^m$ denote the set of $m \times m$ symmetric matrices, $\mathbb{S}^m_{>0}$ the set of $m \times m$ positive definite symmetric matrices, and $\mathbb{S}^m_{\geq 0}$ the set of $m \times m$ positive semidefinite symmetric matrices. Let $\mathrm{Sym}(m, n)$ denote the set of $m \times m$ symmetric matrices of rank $\leq n$. Let

$$\phi_G : \mathbb{S}^m \to \mathbb{R}^{V+E}, \qquad \phi_G(\Sigma) = (\sigma_{ii})_{i \in V} \oplus (\sigma_{ij})_{ij \in E} \tag{1}$$

be the coordinate projection that extracts the diagonal and entries corresponding to edges of $G$ of the symmetric matrix $\Sigma = (\sigma_{ij})_{i,j \in V}$. In the setting of matrix completion problems, the question of determining the maximum likelihood threshold of a graph $G$ is:

**Problem 1.1 (Maximum likelihood threshold).** *Given a graph $G$, what is the smallest $n$ such that for almost all $\Sigma_0 \in \mathrm{Sym}(m, n) \cap \mathbb{S}^m_{\geq 0}$ there exists a $\Sigma \in \mathbb{S}^m_{>0}$ such that $\phi_G(\Sigma_0) = \phi_G(\Sigma)$?*

Since every positive semi-definite matrix of rank $n$ arises as $P^T P$ for some $n \times m$ real matrix $P$ with real columns $\mathbf{p}_i \in \mathbb{R}^n$, Problem 1.1 is equivalent to asking: *Given a graph $G$, what is the smallest $n$ such that for almost all $P = (\mathbf{p}_1, \ldots, \mathbf{p}_m) \in \mathbb{R}^{n \times m}$ there exists a set of linearly independent vectors $Q = (\mathbf{q}_1, \ldots, \mathbf{q}_m) \in \mathbb{R}^{m \times m}$ such that*

$$\|\mathbf{p}_i\|_2 = \|\mathbf{q}_i\|_2 \quad \text{for all } i, \quad \text{and} \quad \mathbf{p}_i \cdot \mathbf{p}_j = \mathbf{q}_i \cdot \mathbf{q}_j \quad \text{for all } ij \in E?$$

This formulation results in a natural connection between the *symmetric minor matroid* and the *generic rigidity matroid*, which we will use to bound the maximum likelihood threshold:

**Theorem 1.2.** *If the edge set of a graph $G$ is an independent set in the $(n-1)$-dimensional generic rigidity matroid, then the maximum likelihood threshold of $G$ is less than or equal to $n$.*

In spite of the seeming importance of the maximum likelihood threshold in applications where $n \ll m$, very little is known about the value of $\mathrm{mlt}(G)$ except in certain special instances. Some of these instances are straightforward:

- $\mathrm{mlt}(G) = 1$ if and only if $G$ has no edges,
- $\mathrm{mlt}(G) = 2$ if and only if $G$ has no cycles, and
- $\mathrm{mlt}(G) = m = \#V$ if and only if $G = K_m$.

For more complicated graphs, Buhl showed in [3] that the $\mathrm{mlt}(G)$ is bounded in terms of the clique number $\omega(G)$ and treewidth $\tau(G)$ of the graph. Recall that the clique number of a graph is the number of vertices of the largest complete subgraph of $G$. The treewidth of a graph is one less than the clique number of the smallest chordal cover of $G$.

**Proposition 1.3 ([3], Corollary 3.3).** *Let G be a graph. Then*

$$\omega(G) \leq \mathrm{mlt}(G) \leq \tau(G) + 1.$$

Proposition 1.3 implies that if $G$ is chordal then $\mathrm{mlt}(G) = \omega(G)$. However, in general, these bounds are far from optimal and far from one another. For instance, there are graphs with $\omega(G) = 2$ and arbitrarily large treewidth.

In this paper, we develop the connection between the maximum likelihood threshold and combinatorial rigidity theory through the *rank* of a graph.

**Definition 1.4.** *The* rank *of a graph G, denoted* $\mathrm{rank}(G)$, *is the smallest n such that* $\dim \phi_G(\mathrm{Sym}(m, n)) = \#V + \#E$.

Uhler [25] showed the following bound relating the maximum likelihood threshold and the rank of $G$.

**Theorem 1.5 ([25], Theorem 3.3).** *Let G be a graph. Then*

$$\mathrm{mlt}(G) \leq \mathrm{rank}(G).$$

The proof of Theorem 1.2 amounts to connecting Theorem 1.5 to the rigidity matroid.

**Remark.** It should be noted that the rank of a graph is different from the *Gaussian rank* of a graph as introduced in [2]. The Gaussian rank is also an upper bound on the maximum likelihood threshold of a graph, and it is discussed in more detail in Section 7.

It is still unknown whether there exists a graph such that $\mathrm{mlt}(G) < \mathrm{rank}(G)$, but this might be because the maximum likelihood threshold is so poorly understood. The main goal of this paper is to develop a better understanding of the notion of the rank of a graph, so that we can develop better bounds on the maximum likelihood threshold. One always has $\mathrm{rank}(G) \leq \tau(G) + 1$, but usually $\mathrm{rank}(G)$ is significantly smaller than $\tau(G) + 1$, which yields substantially improved bounds. For example, for an arbitrary $k_1 \times k_2$ grid with $k_1, k_2 \geq 2$, denoted $Gr_{k_1,k_2}$, $\mathrm{rank}(Gr_{k_1,k_2}) = \mathrm{mlt}(Gr_{k_1,k_2}) = 3$, whereas $\tau(Gr_{k_1,k_2}) + 1 = \min(k_1, k_2) + 1$ is substantially larger (Corollary 3.8).

While the question of whether or not there is a gap between $\mathrm{mlt}(G)$ and $\mathrm{rank}(G)$ remains open, we conclude the paper by turning our attention to another estimator, the *score matching estimator* (SME). The *score matching threshold* is the smallest amount of data such that the SME exists with probability one. Theorem 6.3 states that the score matching threshold of a graph $G$ is equal to its rank. The SME was introduced in [13] and furthered studied in [8]. The score

matching equations are linear, so when the SME exists, computing the estimator is efficient even for large dense graphs. Hence, the SME has promising applications to model selection for high dimensional graphical models.

As Lauritzen and Forbes point out in [8], a simple sufficient condition for the existence of the SME would be advantageous, since it could be used to limit model searches. By the same reasoning, simple sufficient conditions on the existence of the MLE are desirable as well. Corollary 3.3 of Section 3 and Corollary 6.4 of Section 6, give such sufficient conditions for the mlt($G$) and smt($G$) when $n = 3$; the conditions can be checked in $O(\#V(G) \cdot \#E(G))$ time [15].

The remainder of the paper is organized as follows. In Section 2, we discuss algebraic matroids, in particular, the symmetric minor matroid and the combinatorial rigidity matroid. Within this matroidal setting, we show that the rank($G$) is the smallest $n$ for which the set of edges of $E(G)$ are independent in the generic rigidity matroid $\mathcal{A}(n-1)$. In Section 3, we provide a summary of the consequences of this connection for the maximum likelihood threshold. In Section 4, we prove a splitting theorem which allows for the computation of improved bounds on rank($G$) by reducing to smaller graphs, at the expense of calculating the *birank* of bipartite graphs. In Section 5, we introduce the notion of weak maximum likelihood threshold, and we provide a splitting lemma and bounds for the weak maximum likelihood threshold based on the chromatic number. Finally, in Section 6, we show that the rank($G$) is equal to smt($G$) and discuss consequences.

## 2. Combinatorial rigidity theory

In this section, we relate the rank of a graph to combinatorial rigidity theory. This connection is explained via certain algebraic matroids, which we review here. See [22,26] for more background on matroids and [11,28] for background on rigidity theory. Both the rigidity matroid and the symmetric minor matroid are discussed in detail in Section 3.3 and Section 3.2, respectively, of [17] in the context of matroids with graph symmetries with a view towards low rank matrix completion problems.

**Definition 2.1.** *Let $S$ be a set and $\mathcal{I}$ a collection of subsets of $S$ satisfying*:

(1) $\varnothing \in \mathcal{I}$,
(2) *if $X \in \mathcal{I}$ and $Y \subseteq X$ then $Y \in \mathcal{I}$, and*
(3) *if $X, Y \in \mathcal{I}$ with $|X| < |Y|$ then there is a $y \in Y$ such that $X \cup \{y\} \in \mathcal{I}$.*

*The pair $(S, \mathcal{I})$ is called a* matroid *and the elements of $\mathcal{I}$ are called* independent sets.

The protypical example of a matroid comes from linear algebra: if $S$ is a set of vectors and $\mathcal{I}$ consists of all linearly independent subsets then the pair $(S, \mathcal{I})$ is a matroid. Other terminology from matroid theory comes from linear algebra. An independent set of maximal size in $\mathcal{I}$ is called a *basis*. A subset $X \subseteq S$ that contains a basis is said to *span* the matroid. The main type of matroid that we will need in this work comes from algebra.

**Definition 2.2.** *Let $\mathbb{K}$ be a field, and let $S = \{\alpha_1, \ldots, \alpha_d\}$ be elements of a field extension $\mathbb{L}/\mathbb{K}$. The* algebraic matroid *on $S$ is the matroid whose independent sets are the collections of $X \subset S$ that are algebraically independent over $\mathbb{K}$.*

Two typical ways that algebraic matroids arise are via prime ideals and via parametrizations. In particular, let $I \subseteq \mathbb{K}[x] := \mathbb{K}[x_1, \ldots, x_n]$ be a prime ideal, and consider the field extension $K(\mathbb{K}[x]/I)/\mathbb{K}$ where $K(\mathbb{K}[x]/I)$ denotes the field of fractions. The natural algebraic matroid to consider in this context is the matroid on the elements $x_1, \ldots, x_n$.

The algebraic matroid associated to a rational parametrization is described as follows. Let $\mathbb{K}(t) := \mathbb{K}(t_1, \ldots, t_e)$ be the field of fractions of $\mathbb{K}[t] := \mathbb{K}[t_1, \ldots, t_e]$. Consider $d$ rational functions $f_1, \ldots, f_d \in \mathbb{K}(t)$. These determine an algebraic matroid in the obvious way. This is a special case of the prime ideal description because we can take the presentation ideal $I \subseteq \mathbb{K}[x]$ of the $\mathbb{K}$-algebra homomorphism $f : \mathbb{K}[x] \to \mathbb{K}(t)$, $f(x_i) = f_i(t)$. The algebraic matroid on $x_1, \ldots, x_d \in K(\mathbb{K}[x]/I)$ is the same as the algebraic matroid on $f_1, \ldots, f_d$, precisely because $I$ is the ideal of relations among $f_1, \ldots, f_d$.

The generic rigidity matroid $\mathcal{A}(n)$ of dimension $n$ is constructed as follows. Let $P = (p_{ij})_{i,j \in n,m}$ be an $n \times m$ matrix of algebraically independent indeterminates. Let $\mathbf{p}_j$ be the $j$th column of $P$. Consider the algebraic matroid on the set of $\binom{m}{2}$ polynomials

$$f_{ij} = \|\mathbf{p}_i - \mathbf{p}_j\|_2^2 \in \mathbb{R}[p].$$

One should think of this matroid as giving dependence/independence relationships between the set of distances between $m$ generic points in $\mathbb{R}^n$. A graph $G = (V, E)$ with $V = [m] := \{1, \ldots, m\}$ is called *rigid* if, for generic choices of the points $\mathbf{p}_1, \ldots, \mathbf{p}_m \in \mathbb{R}^n$, the set of distances $f_{ij}$ such that $ij \in E$, determine all the other distances $f_{ij}$ with $ij \in \binom{[m]}{2}$. In the approach based on rigidity matroids, one weakens the condition to allow only finitely many possibilities for the other missing distances. In the language of algebraic matroids, this means that the set of polynomials $\{f_{ij} : ij \in E(G)\}$ is a spanning set for the algebraic matroid $\mathcal{A}(n)$. On the other hand, a graph $G$ is *stress-free* precisely when $\{f_{ij} : ij \in E(G)\}$ is an independent set in the algebraic matroid $\mathcal{A}(n)$. When this is the case, we will say $E(G)$ *is an independent set in $\mathcal{A}(n)$.* A graph $G$ that is simultaneously stress-free and rigid in dimension $n$, is called *isostatic*. In the matroid language, this says that $E(G)$ is a basis in the matroid $\mathcal{A}(n)$.

**Remark.** Note that rigidity and being stress-free are properties that hold generically. There are situations where a graph is rigid but there exist non-generic choices of the points $\mathbf{p}_j$ that make the resulting framework flexible. Since we are only interested in generic properties of the graph, we can ignore such issues.

The second algebraic matroid we will study is the symmetric minor matroid $S(m, n)$, which is described at length in [17]. In particular, let $\mathbb{C}[\Sigma] := \mathbb{C}[\sigma_{ij} : 1 \leq i \leq j \leq m]$ and let $I_{n+1}$ be the prime ideal of $(n + 1)$-minors of the generic symmetric matrix $\Sigma$. The symmetric minor matroid $S(m, n)$ is the algebraic matroid of the elements $\sigma_{ij} : 1 \leq i \leq j \leq m$ in the extension $K(\mathbb{C}[\Sigma]/I_{n+1})/\mathbb{C}$. In the language of algebraic matroids, the rank of the graph $G$ can be expressed as follows.

**Proposition 2.3.** *Let G be a graph on vertex set $[m]$. The rank of G is the smallest n such that $\{\sigma_{ii} : i \in [m]\} \cup \{\sigma_{ij} : ij \in E(G)\}$ is an independent set of the symmetric minor matroid $S(m, n)$.*

The ideal $I_{n+1}$ is the vanishing ideal of a parametrization, a fact that we will use in connecting the matroid $S(m, n)$ to the matroid $\mathcal{A}(n - 1)$. Indeed, every symmetric matrix of rank $\leq n$ can be realized as $P^T P$ for some $n \times m$ matrix $P$ (over $\mathbb{C}$). Hence, if we let $g_{ij} = \mathbf{p}_i \cdot \mathbf{p}_j$ for $1 \leq i \leq j \leq m$ then the algebraic matroid on these elements is the same as the algebraic matroid $S(m, n)$.

For the rank problem of interest, that is the rank problem for studying the maximum likelihood threshold, we are always looking at sets that contain all of the diagonal elements $\sigma_{ii}$. Hence, we can look at independent sets in the matroid contraction by that collection of elements. From the standpoint of algebraic matroids, that amounts to studying the algebraic matroid $S(m, n)/\text{diag}$, of the field extension $K(\mathbb{C}[\Sigma]/I_{n+1})/\mathbb{C}(\sigma_{ii} : i \in [m])$ with ground set consisting of the elements $\sigma_{ij} : i < j$.

**Theorem 2.4.** *The algebraic matroids $S(m, n)/\text{diag}$ and $\mathcal{A}(n - 1)$ are isomorphic.*

Theorem 2.4 is a Corollary to [4], Theorem 8, which explores the question in terms of the *coning of a graph* and *spherical rigidity*. Spherical rigidity was first discussed in [27]. Here, we provide a proof of Theorem 2.4, not only for completeness, but also to set up the terminology and techniques needed for results in later sections.

To prove Theorem 2.4 we use the fact that the algebraic matroid over a field of characteristic zero is isomorphic to the representable matroid obtained from evaluating the Jacobian at a generic point of the parameter space. We will make these evaluations for both of the matroids $S(m, n)$ and $\mathcal{A}(n - 1)$ and compare the results. These particular Jacobian matrices will appear at other points in the paper so we introduce them outside of the proof.

First, consider the map $f : \mathbb{R}^{n \times m} \to \mathbb{R}^{m(m-1)/2}$ with

$$f(P) = \left(\|\mathbf{p}_i - \mathbf{p}_j\|_2^2\right)_{1 \leq i < j \leq m},$$

where we let $P = \begin{pmatrix} \mathbf{p}_1 & \cdots & \mathbf{p}_m \end{pmatrix}$.

The Jacobian $J(f, P)$ of this map is an $mn \times \binom{m}{2}$ matrix. The rows of $J(f, P)$ should be grouped into $m$ blocks of size $n$ corresponding to the $m$ points $\mathbf{p}_1, \ldots, \mathbf{p}_m$. The $ij$ column of the Jacobian matrix is the vector with zeros in all blocks except the $i$th and $j$th blocks which have $\mathbf{p}_i - \mathbf{p}_j$ and $\mathbf{p}_j - \mathbf{p}_i$, respectively (we have ignored the extra factor of 2 that appears in all entries). For example, for $m = 4$ the matrix $J(f, P)$ is

$$\begin{pmatrix} \mathbf{p}_1 - \mathbf{p}_2 & \mathbf{p}_1 - \mathbf{p}_3 & \mathbf{p}_1 - \mathbf{p}_4 & 0 & 0 & 0 \\ \mathbf{p}_2 - \mathbf{p}_1 & 0 & 0 & \mathbf{p}_2 - \mathbf{p}_3 & \mathbf{p}_2 - \mathbf{p}_4 & 0 \\ 0 & \mathbf{p}_3 - \mathbf{p}_1 & 0 & \mathbf{p}_3 - \mathbf{p}_2 & 0 & \mathbf{p}_3 - \mathbf{p}_4 \\ 0 & 0 & \mathbf{p}_4 - \mathbf{p}_1 & 0 & \mathbf{p}_4 - \mathbf{p}_2 & \mathbf{p}_4 - \mathbf{p}_3 \end{pmatrix}.$$

On the other hand, consider the parameterization map $g : \mathbb{R}^{n \times m} \to \mathbb{R}^{m(m+1)/2}$

$$g(P) = (\mathbf{p}_i \cdot \mathbf{p}_j)_{1 \leq i \leq j \leq m}.$$

The Jacobian $J(g, P)$ of this map is an $mn \times \binom{m+1}{2}$ matrix. The rows of $J(g, P)$ should be grouped into $m$ blocks of size $n$ corresponding to the $m$ points $\mathbf{p}_1, \ldots, \mathbf{p}_m$. When $i \neq j$, the $ij$ column in $J(g, P)$ has zero vectors in all blocks except for the $i$th and $j$th blocks, which have $\mathbf{p}_j$ and $\mathbf{p}_i$, respectively. When $i = j$, there is only one nonzero block, which is $2\mathbf{p}_i$. For example, for $m = 4$ the matrix $J(g, P)$ is

$$\begin{pmatrix} 2\mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{p}_1 & 0 & 0 & 2\mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{p}_1 & 0 & 0 & \mathbf{p}_2 & 0 & 2\mathbf{p}_3 & \mathbf{p}_4 & 0 \\ 0 & 0 & 0 & \mathbf{p}_1 & 0 & 0 & \mathbf{p}_2 & 0 & \mathbf{p}_3 & \mathbf{p}_4 \end{pmatrix}.$$

To make a connection to recent investigations regarding the uniqueness of low rank matrices [14,24], note that the Jacobian $J(g, P)$ is the transpose of the *completability matrix* of the complete graph on $m$ vertices (including self-loops) introduced in [24] and discussed further in [14]. Thus, the symmetric minor matroid and the completeability matroid from [14] are isomorphic.

**Proof of Theorem 2.4.** Note that since all polynomials involved are defined over the integers, the underlying ground field can be changed to be $\mathbb{C}$, $\mathbb{R}$, or $\mathbb{Q}$ without changing the matroid in any of the matroids in question.

We also note that the rank of the Jacobian $J(g, P)$ does not change if we scale each point $\mathbf{p}_i$ by a nonzero constant $\lambda_i$. Indeed, performing such a scaling is equivalent to multiplying the rows corresponding the row block indexed by $i$ by $\lambda_i^{-1}$ and multiplying the column indexed by $ij$ by $\lambda_i \lambda_j$, and row and column operations do not change the rank of a matrix.

Since the matrix $P$ is generic, we can assume that all of the coordinates of each $\mathbf{p}_i$ are nonzero. By choosing an appropriate scaling, we may assume that the $n$th coordinate of each $\mathbf{p}_i$ is equal to 1. We write this formally as $\mathbf{p}_i = \binom{\mathbf{p}_i'}{1}$, with $\mathbf{p}_i' \in \mathbb{R}^{n-1}$ which we can assume to be generic. Let $P' = \begin{pmatrix} \mathbf{p}_1' & \cdots & \mathbf{p}_m' \end{pmatrix}$.

Divide the columns corresponding to pairs $ii$ by 2. Then subtract the column corresponding to pair $ii$ from each column corresponding to $ij$. Let $M$ be the resulting matrix. The columns of $M$ corresponding to the tuples $ii$ are clearly linearly independent of all other columns, because they are the only columns that contain nonzero entries in the last position in each block. Hence, when we contract by these diagonal elements we can delete the last row from each block (since we get all zeros). The resulting matrix, the matrix that represents the matroid $S(m, n)/\operatorname{diag}$, is $J(f, P')$. Hence, $S(m, n)/\operatorname{diag}$ is isomorphic to $\mathcal{A}(n-1)$ as claimed. $\qquad\square$

Theorem 2.4 means that the rank of a graph can be precisely characterized in terms of the independence condition in the matroid $\mathcal{A}(n-1)$.

**Theorem 2.5.** *Let $G = (V, E)$ be a graph. Then $\operatorname{rank}(G) = n$ if and only if $E$ is an independent set in $\mathcal{A}(n-1)$ and is not an independent set in $\mathcal{A}(n-2)$.*

Combined with Theorem 1.5, which bounds the $\operatorname{mlt}(G)$ by $\operatorname{rank}(G)$, Theorem 2.5 implies Theorem 1.2, the main theorem stated in the Introduction.

# 3. Basic results on rank(*G*)

In this section, we catalogue some basic results about the rank of a graph *G* that follow immediately from the connection to combinatorial rigidity, including bounds on the number of edges that can be involved and graph constructions that preserve rank. While these do not represent new results for rigidity theory, they provide novel consequences for the maximum likelihood threshold.

**Proposition 3.1 ([11], Lemma 2.5.5).** *Let $G' = (V', E')$ and $G = (V, E)$ such that $G'$ is a subgraph of $G$. Then if $E$ is independent in $\mathcal{A}(n-1)$, the set $E'$ is independent in $\mathcal{A}(n-1)$. Consequently*, $\mathrm{rank}(G') \leq \mathrm{rank}(G)$.

The condition in the next theorem is called *Laman's condition* in the combinatorial rigidity literature and is a necessary condition for a set to be independent in the rigidity matroid $\mathcal{A}(n-1)$. We state the theorem in terms of the rank of *G*, using the equivalence established in Theorem 2.5.

**Theorem 3.2 ([11], Theorem 2.5.4).** *Let $G = (V, E)$ be a graph, and suppose that* $\mathrm{rank}(G) \leq n$. *Then, for all subgraphs $G' = (V', E')$ of G such that $\#V' \geq n-1$ we must have*

$$\#E' \leq \#V'(n-1) - \binom{n}{2}. \tag{2}$$

Laman's theorem [20] states that the condition of Theorem 3.2 is both necessary and sufficient for a set to be independent in $\mathcal{A}(2)$, which combined with Theorem 2.4 and Theorem 1.5 gives us the following corollary in regards to the maximum likelihood threshold.

**Corollary 3.3.** *Let $G = (V, E)$ be a graph, if for all subgraphs $G' = (V', E')$ of G*

$$\#E' \leq 2(\#V') - 3,$$

*then* $\mathrm{mlt}(G) \leq 3$.

Naively checking the conditions of Corollary 3.3 is inefficient. However, there is a polynomial times algorithm that can check if an edge set is independent in $\mathcal{A}(2)$ [15].

***Example 3.4.*** Let *G* be the complete bipartite graph $K_{3,3}$. The treewidth of *G* is 3, thus, by Buhl's bound in Proposition 1.3, we have $\mathrm{mlt}(G) \leq 4$. Using Corollary 3.3, we can obtain the improved bound $\mathrm{mlt}(K_{3,3}) \leq 3$. Since $K_{3,3}$ is not a forest, we deduce that $\mathrm{mlt}(K_{3,3}) = 3$.

In rigidity theory, there are many operations that take an independent set and produce a new independent set on a larger number of vertices. We review two of these here, vertex addition and edge splitting. We begin with *vertex addition*, also called *0-extensions*, and elaborate on some of the implications with respect to the maximum likelihood threshold. Again, we state the theorem in terms of the rank of *G*, using the equivalence established in Theorem 2.5. A variation of Proposition 3.5 with rank replaced by Ben-David's Gaussian rank is proved independently in [2].

**Proposition 3.5 (Vertex addition, [28], Lemma 11.1.1).** *Let $G = (V, E)$ be a graph such that* rank$(G) \leq n$ *and* $\#V = m$. *Let $G'$ be a new graph obtained from $G$ by adding the vertex $v'$ and at most $n - 1$ edges connecting $v'$ to other vertices in $G$. Then* rank$(G') \leq n$, *and, in particular,* mlt$(G') \leq n$.

**Definition 3.6.** *Let $G$ be a graph and fix an integer $r$. The $r$-core of $G$ denoted $r$-core$(G)$ is the graph obtained from $G$ by successively deleting vertices of degree $< r$. A graph is said to have empty $r$-core, if $r$-core$(G)$ has no vertices.*

Using Proposition 3.5 inductively, we immediately deduce the following corollary.

**Corollary 3.7.** *Let $G$ be a graph with empty $n$-core. Then* mlt$(G) \leq n$.

Corollary 3.7 was also proven independently in [2] using bounds on the Gaussian rank of a graph. While not as powerful as the splitting result from the next section, Corollary 3.7 already implies a number of nice consequences in some simple cases.

**Corollary 3.8.** *Let $Gr_{k_1,k_2}$ denote the $k_1 \times k_2$ grid graph with $k_1, k_2 \geq 2$. Then* mlt$(Gr_{k_1,k_2}) = 3$.

**Proof.** First of all, mlt$(Gr_{k_1,k_2}) \geq 3$, since $Gr_{k_1,k_2}$ contains a cycle. On the other hand, rank$(Gr_{k_1,k_2}) \leq 3$, since $Gr_{k_1,k_2}$ has empty 3-core. This can be seen by removing the corner vertices, which successively leaves a new vertex of degree 2. Hence,

$$3 \leq \text{mlt}(Gr_{k_1,k_2}) \leq \text{rank}(Gr_{k_1,k_2}) \leq 3$$

completes the proof.                                                                                    □

Another well-known graph operation preserving rank is *edge-splitting*.

**Theorem 3.9 (Edge splitting, [28], Theorem 11.1.7).** *Let $G = (V, E)$ be a graph such that* rank$(G) \leq n$, *and let $e = \{v_1, v_2\} \in E$. Let $G'$ be a graph obtained from $G$ by removing $e$ and then adding a new vertex $v'$ such that $v'$ is attached to the vertices $v_1$ and $v_2$ and at most $n - 2$ other vertices in $V$. Then* rank$(G') \leq n$, *and, in particular* mlt$(G') \leq n$.

**Example 3.10.** Consider the lattice graph $L_{(2,4)}$ pictured in Figure 1. The graph $L_{(2,4)}$ has tree-width 4 and contains the complete graph on 4 vertices, therefore $4 \leq \text{mlt}(L_{(2,4)}) \leq 5$. Denote the graph pictured in Figure 2 by $G$. The graph $G$ has an empty $4 - core$, thus mlt$(G) \leq 4$. We can obtain $L_{(2,4)}$ by removing the edge $\{2, 5\}$ and adding the vertex 8 and the edges $\{1, 8\}$, $\{2, 8\}$, $\{5, 8\}$, and $\{7, 8\}$, thus $L_{(2,4)}$ can be obtained from $G$ through an edge splitting and we have mlt$(L_{(2,4)}) = 4$.

Via a more advanced application of results of combinatorial rigidity theory, we can deduce the following bound on the rank of any planar graph.

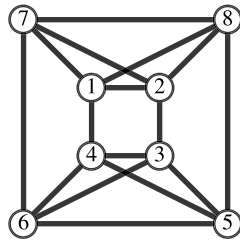**Corollary 3.11.** *Let $G$ be a planar graph. Then* mlt$(G) \leq 4$.

**Figure 1.** The lattice graph $L_{(2,4)}$.

This proof is essentially due to Gluck [10] and depends on Dehn's [5] strengthening of Cauchy's theorem.

**Proof of Corollary 3.11.** Every planar graph is a subgraph of a maximal planar graph, that is, a planar graph where it is not possible to add any further edges and maintain planarity. By Proposition 3.1, it suffices to prove the bound for such maximal subgraphs. The theorem is clearly true if $\#V \le 3$, so assume that $\#V \ge 4$.

Now, every maximal planar graph with $\#V \ge 4$ is 3-connected. Indeed, if a graph were not 3-connected, 2 vertices could be removed from $G$ leaving a disconnected graph, then an edge could be added from one of the components to another without disrupting the planarity property. Thus, if a planar graph is not 3-connected, it is not maximal.

Every 3-connected planar graph is the edge graph of a simplicial convex polytope via Steinitz theorem (see [31], Chapter 3). Dehn's theorem [5] implies that the framework of any 3-dimensional simplicial convex polytope is infinitesimally rigid in three dimensions, and hence the associated graph is generically rigid in three dimensions. Since a maximal planar graph with $m$ vertices has exactly $3m - 6$ edges, Dehn's theorem combined with Laman's criterion implies that any maximal planar graph is isostatic in 3 dimensions. Hence, the set of edges of a maximal planar graph is an independent set in $\mathcal{A}(3)$, which implies that rank$(G) \le 4$. $\square$

Graphical models associated to planar graphs are especially important in spatial statistics where they are used for applications including in weather forecasting and image analysis. Corollary 3.11 shows that only four data points are needed to estimate parameters in such models.
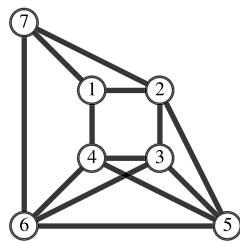


**Figure 2.** Graph with mlt$(G) \le 4$.

# 4. Splitting theorem

In this section, we prove a theorem that allows us to relate the rank of a graph to the rank of smaller subgraphs, at the expense of needing to calculate the birank of some associated bipartite graphs. The birank is the bipartite analogue of rank of a graph, and is naturally related to the theory of *bipartite rigidity* introduced in [16]. This splitting theorem allows us to give a number of simple computations of rank$(G)$ and hence gives us a simple way to compute bounds on mlt$(G)$.

For a bipartite graph $G = (V_1, V_2, E)$ with a fixed bipartition of the vertices $V = V_1 \sqcup V_2$ where $\#V_1 = m_1$, $\#V_2 = m_2$ and two integers $r_1, r_2$, define the following linear space for generic points $X \in \mathbb{C}^{m_1 \times r_1}$ and $Y \in \mathbb{C}^{r_2 \times m_2}$:

$$L_{r_1, r_2}^{(X,Y)} := \left\{ X \cdot A + B \cdot Y : A \in \mathbb{C}^{r_1 \times m_2}, B \in \mathbb{C}^{m_1 \times r_2} \right\}.$$

Let

$$\phi_E : \mathbb{C}^{m_1 \times m_2} \to \mathbb{C}^E, \qquad \phi_E(\Sigma) = (\sigma_{ij})_{ij \in E}$$

be the coordinate projection that extracts the entries corresponding to edges of $G$. Notice that while $\phi_G$ from (1) extracts entries corresponding to the diagonal, $\phi_E$ does not.

**Definition 4.1.** *Let $G = (V_1, V_2, E)$ be bipartite graph. Define the* bipartite rank *of $G$, denoted* birank$(G)$, *to be the set of all pairs of integers $(r_1, r_2)$ such that $\phi_E(L_{r_1, r_2}^{(X,Y)}) = \mathbb{C}^E$ for generic $X \in \mathbb{C}^{m_1 \times r_1}$ and $Y \in \mathbb{C}^{r_2 \times m_2}$.*

The case where $r_1 = r_2 = r$, the linear space $L_{r,r}^{(X,Y)}$ is the tangent space of the set of $m_1 \times m_2$ matrices of rank $r$ at the point $XY$. Hence, bipartite rank in this case tells us about independent sets in the algebraic matroid of the ideal of $(r+1)$-minors of a generic matrix. This matroid was studied in the context of matrix completion problems in [17,18,24].

The following proposition describes a method for constructing a new bipartite graph $G'$ from $G$ such that if $(r_1, r_2) \in$ birank$(G)$ then $(r_1, r_2) \in$ birank$(G')$.

**Proposition 4.2.** *Let $G = (V_1, V_2, E)$ be a bipartite graph with fixed partition $V = V_1 \sqcup V_2$ such that $\#V_1 = m_1$ and $\#V_2 = m_2$. Let $(r_1, r_2) \in$ birank$(G)$ and let $G'$ be a new bipartite graph obtained from $G$ by adding the vertex $v'$ to $V_1$ and at most $r_2$ edges connecting $v'$ to other vertices in $V_2$. Then $(r_1, r_2) \in$ birank$(G')$.*

This result is essentially Lemma 3.7 of [16].

**Proof of Proposition 4.2.** Let $X' \in \mathbb{C}^{(m_1 + 1) \times r_1}$ and $Y \in \mathbb{C}^{r_2 \times m_2}$ be generic. Write

$$X' = \begin{bmatrix} X \\ x \end{bmatrix},$$

where $X \in \mathbb{C}^{m_1 \times r_1}$ and $x \in \mathbb{C}^{r_1}$. Since $X'$ is generic, $X$ and $x$ are both generic as well.

Let $E'$ be the edge set of $G'$. Let $w' \in \mathbb{C}^{E'}$, which we will write as $w' = \binom{w}{u}$ where $w \in \mathbb{C}^E$ and $u \in \mathbb{C}^{E'-E}$. Since $(r_1, r_2) \in \mathrm{birank}(G)$, by the definition of bipartite rank, the image $\phi_E(L_{r_1,r_2}^{(X,Y)}) = \mathbb{C}^E$, and thus there exists $A \in \mathbb{C}^{r_1 \times m_2}$, $B \in \mathbb{C}^{m_1 \times r_2}$ such that

$$\phi_E(X \cdot A + B \cdot Y) = w.$$

Now, note that if $b \in (\mathbb{C}^{r_2})^*$, then

$$\begin{bmatrix} X \\ x \end{bmatrix} \cdot A + \begin{bmatrix} B \\ b \end{bmatrix} \cdot Y = \begin{bmatrix} X \cdot A + B \cdot Y \\ x \cdot A + b \cdot Y \end{bmatrix}.$$

Thus, to show surjectivity of $\phi_{E'}$, we need to find a $b \in \mathbb{C}^{r_2}$ such that

$$\phi_{E'-E}(x \cdot A + b \cdot Y) = u. \tag{3}$$

However, the equation (3) results a linear system with generic coefficients and $r_2$ unknowns (the entries of $b$). Therefore, a solution always exists when $\#(E' - E) \leq r_2$ if $r_2 \leq m_2$, or $\#(E' - E) \leq m_2$ if $r_2 > m_2$. $\qquad\square$

For a bipartite graph $G$ with fixed bipartition of the vertices $V_1$, $V_2$ and two integers $r_1, r_2$, let $\mathrm{core}_{r_1,r_2}(G)$ be the graph obtained from $G$ by repeatedly removing vertices of $G$ whenever $j \in V_1$ has degree less than or equal to $r_2$ or $j \in V_2$ has degree less than or equal to $r_1$. Note that the $\mathrm{core}_{r_1,r_2}(G)$ is uniquely determined, despite the fact that we have choices in the order we choose to remove vertices. A graph is said to have empty $(r_1, r_2)$-core if $\mathrm{core}_{r_1,r_2}(G)$ has no vertices. Clearly if $G$ has empty $(r_1, r_2)$-core, it will have $(r_1, r_2) \in \mathrm{birank}(G)$, in analogy to the relationship between ordinary rank of a graph and core.

***Example 4.3.*** A bipartite graph $G$ has empty $(1, 1)$-core if and only if $G$ has no cycles.

The notion of bipartite rank can help us understand the rank of an arbitrary (not necessarily bipartite) graph $G$. For a graph $G = (V, E)$ and disjoint subsets $V_1$, $V_2 \subseteq V$, let $G(V_1, V_2)$ be the bipartite graph consisting of all edges $ij \in E(G)$ such that $i \in V_1$ and $j \in V_2$. Let $G_{V_1}$ denote the induced subgraph of vertex set $V_1$.

**Theorem 4.4 (Splitting theorem).** *Let $G$ be a graph, $r_1, \ldots, r_k$ integers and $V_1, \ldots, V_k$ a partition of the vertices of $G$, such that*:

(1) *for all $i$, $\mathrm{rank}(G_{V_i}) \leq r_i$ and*
(2) *for all $i \neq j$, $(r_i, r_j) \in \mathrm{birank}(G(V_i, V_j))$.*

*Then $\mathrm{rank}(G) \leq r_1 + \cdots + r_k$, and, in particular, $\mathrm{mlt}(G) \leq r_1 + \cdots + r_k$.*

**Proof.** Let $m_i = \#V_i$ for all $1 \leq i \leq k$. Let $m = \sum_{i=1}^k m_i$ and $n = \sum_{i=1}^k r_i$. Recall that $\mathrm{Sym}(m, n)$ can be parameterized over $\mathbb{C}$ as

$$\mathrm{Sym}(m, n) = \left\{ P^T P : P \in \mathbb{C}^{n \times m} \right\}.$$

Using this parameterization, we see that the tangent space of $\mathrm{Sym}(m, n)$ at the point $X = P^T P$ is

$$T_X\big(\mathrm{Sym}(m, n)\big) = \big\{P^T A + A^T P : A \in \mathbb{C}^{n \times m}\big\}.$$

To show that $\dim \phi_G(\mathrm{Sym}(m, n)) = \#V + \#E$, we will show that the differential of $\phi_G$ at $X$

$$(D\phi_G)_X : T_X\big(\mathrm{Sym}(m, n)\big) \to \mathbb{C}^{V+E},$$

$$(D\phi_G)_X(\Sigma) = \phi_G(\Sigma)$$

is surjective for a particular $X \in \mathrm{Sym}(m, n)$. This will imply $(D\phi_G)_X$ is surjective for generic $X \in \mathrm{Sym}(m, n)$, and consequently $\phi_G$ restricted to $\mathrm{Sym}(m, n)$ is dominant.

Let $X = P^T P$ where $P$ is a block diagonal matrix of the form

$$P = \begin{pmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_k \end{pmatrix}$$

such that $P_i \in \mathbb{C}^{r_i \times m_i}$ is generic for $1 \le i \le k$. For every $A \in \mathbb{C}^{n \times m}$, write $A$ as the block matrix

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & A_{22} & \cdots & A_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{kk} \end{pmatrix},$$

where $A_{ij} \in \mathbb{C}^{r_i \times m_j}$. Then $P^T A + A^T P \in T_X(\mathrm{Sym}(m, n))$ is a symmetric block matrix where the $ij$th block is $P_i^T A_{ij} + A_{ji}^T P_j$, i.e.,

$$P^T A + A^T P = \begin{pmatrix} P_1^T A_{11} + A_{11}^T P_1 & \cdots & P_1^T A_{1k} + A_{k1}^T P_k \\ \vdots & \ddots & \vdots \\ P_k^T A_{k1} + A_{1k}^T P_1 & \cdots & P_k^T A_{kk} + A_{kk}^T P_k \end{pmatrix}.$$

To prove surjectivity of $(D\phi_G)_X$, let $w \in \mathbb{C}^{V+E}$, which can be written in the block form $w = (w_{11}, w_{12}, \ldots, w_{kk})$ where $w_{ii} \in \mathbb{C}^{V_i + E(G_i)}$ for all $1 \le i \le k$ and $w_{ij} \in \mathbb{C}^{E(G(V_i, V_j))}$ for all $1 \le i < j \le k$. Since $\mathrm{rank}(G_{V_i}) \le r_i$ and $P_i$ is generic, the image of the linear space $\{P_i^T A_{ii} + A_{ii}^T P_i : A_{ii} \in \mathbb{C}^{r_i \times m_i}\}$ under the map $\phi_{G_{V_i}}$ is $\mathbb{C}^{V_i + E(G_i)}$, which means there exists a $A_{ii}' \in \mathbb{C}^{r_i \times m_i}$ such that $\phi_{G_{V_i}}(P_i^T A_{ii}' + A_{ii}'^T P_i) = w_{ii}$. Furthermore, since $(r_i, r_j) \in \mathrm{birank}(G(V_i, V_j))$ for $i \neq j$, there exists $A_{ij}' \in \mathbb{C}^{r_i \times m_j}$ and $A_{ji}' \in \mathbb{C}^{r_j \times m_i}$ such that $\phi_{E(G(V_i, V_j))}(P_i^T A_{ij}' + A_{ji}'^T P_i) = w_{ij}$. Let $A' \in \mathbb{C}^{n \times m}$ with $ij$th block $A_{ij}'$. Then $(D\phi_G)_X(P^T A' + A'^T P) = \phi_G(P^T A' + A'^T P) = w$, and we have shown surjectivity of the differential $(D\phi_G)_X$. $\qquad \square$

Theorem 4.4 and repeated application of Proposition 4.2 gives us the following corollary.

**Corollary 4.5.** *Let G be a graph, $r_1, \ldots, r_k$ integers and $V_1, \ldots, V_k$ a partition of the vertices of G, such that*:

   (1) *for all i*, $\text{rank}(G_{V_i}) \leq r_i$ *and*
   (2) *for all $i \neq j$, $G(V_i, V_j)$ has an empty $(r_i, r_j)$ core.*

*Then* $\text{mlt}(G) \leq r_1 + \cdots + r_k$.

The special case where all the $r_i$ are equal to one is easy to understand.

**Corollary 4.6.** *Let G be a graph and $V_1, \ldots, V_k$ be a partition of the vertices of G such that*:

   (1) *for all i, $V_i$ is an independent set of G and*
   (2) *for all $i \neq j$, $G(V_i, V_j)$ has no cycles.*

*Then* $\text{mlt}(G) \leq k$.

Of course, a partition of the vertices of the graph into independent sets is a proper coloring of the graph, so we seek proper graph colorings where the induced subgraph on pairs of colors has no cycles. Such a coloring is called an acyclic coloring of a graph, and the smallest number of colors such that a graph has an acyclic coloring with that many colors is the acyclic coloring number of the graph [12]. We conclude with some examples illustrating the use of the splitting theorem and its corollaries.

***Example 4.7.*** Consider the lattice graph $L_{(2,4)}$ from Example 3.10. The partition of the vertices $V_1 = \{1, 5\}$, $V_2 = \{2, 6\}$, $V_3 = \{3, 7\}$, and $V_4 = \{4, 8\}$ has each $V_i$ an independent set in $L_{(2,4)}$ and each bipartite graph $L_{(2,4)}(V_i, V_j)$ without cycles. This implies that $\text{rank}(G) \leq 4$.

***Example 4.8.*** Consider the octahedral graph $O_6$, pictured in Figure 3. The partition of vertices $V_1 = \{1, 4, 5\}$, $V_2 = \{2, 3, 6\}$ yields a splitting that produces the bound $\text{rank}(O_6) \leq 4$. Indeed, since $(O_6)_{V_1}$ and $(O_6)_{V_2}$ are both trees, they have $\text{rank}((O_6)_{V_1}) = \text{rank}((O_6)_{V_2}) = 2$, and the bipartite graph $O_6(V_1, V_2)$ has empty $(2, 2)$-core so $(2, 2) \in \text{birank}(O_6(V_1, V_2))$. On the other hand $O_6$ has 12 edges, which by Theorem 3.2 implies that $\text{rank}(O_6) \geq 4$, so the splitting proves that $\text{rank}(O_6) = 4$.
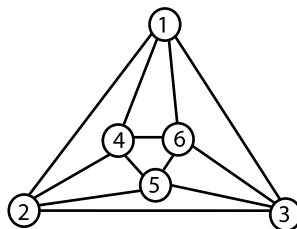


**Figure 3.** The octahedron graph $O_6$.

***Example 4.9.*** Consider the grid graphs $Gr_{k_1,k_2}$. Identify the vertices naturally with $[k_1] \times [k_2]$. Partition the vertices into three parts $V_0, V_1, V_2$ where

$$V_i = \big\{ (j_1, j_2) : j_1 + j_2 \equiv i \mod 3 \big\}.$$

Clearly each $V_i$ is an independent set and each graph $G(V_i, V_j)$ has no cycles, so by Corollary 4.6, $\mathrm{rank}(Gr_{k_1,k_2}) \leq 3$.

The three preceding examples illustrating the Splitting theorem can already be handled using the standard techniques from rigidity theory from Section 3. Furthermore, we should discuss the fact that the acyclic coloring number can grow very quickly. Let $d(G)$ be the maximal degree of the graph $G$ and $A(G)$ denote the acylic coloring number. In fact, Alon, McDiarmid and Reed [1] showed that $A(G) = O(d(G)^{4/3})$ and there exist graphs for which

$$A(G) = \Omega\left( \frac{d(G)^{4/3}}{(\log d(G))^{1/3}} \right).$$

On one hand, based on the results from the previous sections, $\mathrm{rank}(G) \leq d(G) + 1$, since any graph $G$ has empty $(d(G)+1)$-core. On the other hand, there are graphs where $A(G) < d(G)+1$.

***Example 4.10.*** Consider the graph $TGr_{k_1,k_2}$, the $k_1 \times k_2$ torus grid graph. This graph has $k_1 k_2$ vertices, each of degree 4, and hence $2k_1 k_2$ edges in total. The core argument implies that $\mathrm{rank}(G) \leq 5$ whereas from the edge count we see that $\mathrm{rank}(G) \geq 4$.

Suppose that $k_1$ is divisible by 4 and $k_2$ is divisible by 3. Consider the $4 \times 3$ blocks of colors:

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \\ 4 & 1 & 2 \end{pmatrix}$$

and consider the resulting coloring of $TGr_{k_1,k_2}$ obtained by repeating this block. This coloring shows that $A(TGr_{k_1,k_2}) \leq 4$ since each of the induced colorings on coloring classes $(i, i+1)$ mod 4 will consist of paths descending from the northeast to the southwest, that do not cross left-right boundaries from one $B$ to the next $B$. Coloring classes $(i, i+2)$ mod 4 only involve edges that cross between adjacent left-right blocks, so also do not produce cycles.

## 5. Weak maximum likelihood threshold

A weaker notion of maximum likelihood threshold was also introduced in [3] and further studied in [25], which asks not for maximum likelihood estimates to exist for almost all $\Sigma_0 \in \mathrm{Sym}(m, n) \cap \mathbb{S}^m_{\geq 0}$ but just for an open set of $\mathrm{Sym}(m, n) \cap \mathbb{S}^m_{\geq 0}$. This leads us to the notion of weak maximum likelihood threshold.

***Definition 5.1.*** *For each graph $G$, the* weak maximum likelihood threshold, *$\mathrm{wmlt}(G)$, is the smallest $n$ such that there exists a $\Sigma_0 \in \mathrm{Sym}(m, n) \cap \mathbb{S}^m_{\geq 0}$ and a $\Sigma \in \mathbb{S}^m_{>0}$ such that $\phi_G(\Sigma_0) = \phi_G(\Sigma)$.*

Note that because the positive definite cone $\mathbb{S}^m_{>0}$ is open, the existence of a single matrix $\Sigma_0$ with this property guarantees an open set of such matrices of positive measure in $\mathrm{Sym}(m, n) \cap \mathbb{S}^m_{\geq 0}$. Hence, we could also say that if $\mathrm{wmlt}(G) \geq n$, then maximum likelihood estimates for the Gaussian graphical model associated to $G$ exist with positive probability with $n$ data points. Evaluating this probability would depend on having a specific distribution to draw the data from, for example, Buhl [3] calculated this for data drawn from an $\mathcal{N}(0, I_m)$ distribution for the cycle graph.

Clearly we have $\mathrm{wmlt}(G) \leq \mathrm{mlt}(G)$. The two numbers can be equal, but often they are different. Analogous to the splitting theorem for $\mathrm{rank}(G)$, there is also a straightforward splitting lemma for $\mathrm{wmlt}(G)$.

**Lemma 5.2 (Splitting lemma).** *Let $G$ be a graph, $r_1, \ldots, r_k$ integers and $V_1, \ldots, V_k$ a partition of the vertices of $G$, such that for all $i$, $\mathrm{wmlt}(G_{V_i}) \leq r_i$. Then $\mathrm{wmlt}(G) \leq r_1 + \cdots + r_k$.*

**Proof.** For $i = 1, \ldots, k$, let $\Sigma_0^i \in \mathrm{Sym}(m, r_i)$ and $\Sigma^i \in \mathbb{S}^{\#V_i}_>$ such that

$$\phi_{G_{V_i}}\left(\Sigma_0^i\right) = \phi_{G_{V_i}}\left(\Sigma^i\right).$$

Then the block diagonal matrices

$$\Sigma_0 = \mathrm{diag}\left(\Sigma_0^1, \ldots, \Sigma_0^k\right) \quad \text{and} \quad \Sigma = \mathrm{diag}\left(\Sigma^1, \ldots, \Sigma^k\right)$$

satisfy $\phi_G(\Sigma_0) = \phi_G(\Sigma)$, $\Sigma \in \mathbb{S}^m_{>0}$, and $\Sigma_0 \in \mathrm{Sym}(m, r_1 + \cdots + r_k) \cap \mathbb{S}^m_{\geq 0}$. $\qquad\square$

The special case where all $r_i = 1$ yields the following corollary where $\chi(G)$ denotes the chromatic number of $G$.

**Corollary 5.3.** *Let $G$ be a graph. Then $\mathrm{wmlt}(G) \leq \chi(G)$.*

So for example, every bipartite graph $G$ that has an edge satisfies $\mathrm{wmlt}(G) = 2$. On the other hand, for the grid graphs $\mathrm{mlt}(Gr_{k_1, k_2}) = 3$, so $\mathrm{wmlt}(G)$ is typically smaller than $\mathrm{mlt}(G)$.

At this point we know very little about the weak maximum likelihood threshold, even for the graphs with $\mathrm{wmlt}(G) = 2$. Buhl showed that if $C_k$ is a cycle of length $k \geq 4$, then $\mathrm{wmlt}(C_k) = 2$, while $\mathrm{wmlt}(C_3) = 3$. A corollary to this result is the following necessary condition for a graph to have $\mathrm{wmlt}(G) = 2$.

**Corollary 5.4.** *Let $G = ([m], E)$ be a graph with $\mathrm{wmlt}(G) = 2$. Then $G$ is triangle free and there exists a cyclic order $w = w_1 w_2 \cdots w_m$ of the vertices of $G$ such that for any subset $V \subset [m]$ such $G_V$ is a cycle, the induced cyclic ordering $w_V$ is not a cycle ordering induced by the natural cyclic ordering from $G_V$.*

**Proof.** Let $\Sigma_0 \in \mathrm{Sym}(m, 2) \cap \mathbb{S}^m_{\geq 0}$. Then $\Sigma_0 = P^T P$ where $P = (\mathbf{p}_1, \ldots, \mathbf{p}_m)$ and each $\mathbf{p}_i \in \mathbb{R}^2$. Scaling the $\mathbf{p}_i$ by nonzero constants $\lambda_i$ does not change whether or not there exists a $\Sigma$ (since we could also scale the resulting $\Sigma$) so we can assume that all the $\mathbf{p}_i$ are in the upper half-plane. Buhl showed that for the cycle graph $C_k$ with edges $(i, i+1)$, there exists a $\Sigma \in \mathbb{S}^k_{>0}$ with

$\phi_{C_k}(\Sigma_0) = \phi_{C_k}(\Sigma)$ if and only if the vectors $\mathbf{p}_i$ are not in cyclic order when considered by their angles in the upper half plane.

Let $\Sigma_0 \in \mathrm{Sym}(m, n) \cap \mathbb{S}^m_{\geq 0}$ and $\Sigma \in \mathbb{S}^m_{>0}$ such that $\phi_G(\Sigma_0) = \phi_G(\Sigma)$. Then if $V$ is any subset of $[m]$ and $(\Sigma_0)_V$ is the submatrix of $\Sigma_0$ obtained by deleting all rows and columns not indexed by vertices in $V$, then $(\Sigma_0)_V \in \mathrm{Sym}(\#V, n) \cap \mathbb{S}^{\#V}_{\geq 0}$ and $\Sigma_V \in \mathbb{S}^{\#V}_{>0}$, with $\phi_{G_V}((\Sigma_0)_V) = \phi_{G_V}(\Sigma_V)$. Hence, taking $n = 2$, by Buhl's result the vectors $\mathbf{p}_i$ must not appear in cyclic order for any cycle. A necessary condition for finding such a set of vectors is the existence of a permutation with the prescribed property. □

If such an ordering $w$ of the vertices of a triangle free graph $G$ exists, then $G$ is said to satisfy *Buhl's cycle condition*. So a graph with $\mathrm{wmlt}(G) = 2$ satisfies Buhl's cycle condition, but we do not know if the converse of this statement is true. Also we know of no example of a triangle-free graph that does not satisfy Buhl's cycle condition. Note that every triangle free graph $G$ with $\chi(G) \leq 3$ satisfies Buhl's cycle condition, by choosing a 3-coloring and listing the vertices in blocks according to their color.

**Example 5.5.** Consider the Grötsch graph $G_{11}$, pictured in Figure 4, the smallest triangle free graph with $\chi(G_{11}) = 4$. This graph has a cyclic ordering of its vertices satisfying Buhl's cycle condition, namely $achjbdefikg$. On the other hand, the best upper bound on $\mathrm{wmlt}(G_{11})$ using Theorem 5.2 comes from the splitting $V_1 = \{a, b, d, g, j, k\}$, $V_2 = \{c, e, f, h, i\}$, which yields $\mathrm{wmlt}(G_{11}) \leq 3$. Is $\mathrm{wmlt}(G_{11}) = 2$?

# 6. Score matching threshold

An alternative estimator to the maximum likelihood estimator for Gaussian graphical models is the *score matching estimator* (SME) [13]. Unlike the MLE, the SME does not need to be computed iteratively, but instead is the solution to a set of linear equations.
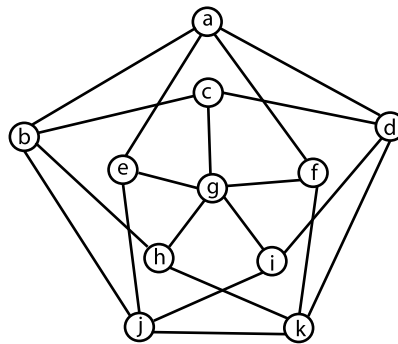


**Figure 4.** The Grötsch graph: the smallest triangle free graph with $\chi(G) = 4$.

The SME is an estimate of the concentration matrix $K = \Sigma^{-1}$. Let $G = (V, E)$ be a graph with $|V| = m$. Let $L_G$ be the following linear subspace of $\mathbb{S}_m$

$$L_G := \{K \in \mathbb{S}_m : K_{ij} = 0 \text{ if } ij \notin E(G) \text{ and } i \neq j\},$$

and let $\Pi_G$ be the orthogonal projection from $\mathbb{S}_m$ onto $L_G$. The estimating equations for the SME are

$$\frac{1}{2} \cdot \Pi_G\big(K \Sigma_0^T + \Sigma_0 K^T\big) = I_m, \tag{4}$$

where $\Sigma_0$ is the sample covariance matrix and $I_m$ is the $m \times m$ identity matrix.

Let $\Sigma_0 = P^T P$ where $P = (\mathbf{p}_1, \ldots, \mathbf{p}_m)$ and each $\mathbf{p}_i \in \mathbb{R}^n$. In [8], the authors give several equivalent conditions that are necessary and sufficient for the SME to exist, that is, for the equation (4) to have a unique solution. We will use the following proposition.

**Proposition 6.1 ([8]).** *Given a graph $G$, the SME exists if and only if $K = 0$ is the only element of $L_G$ such that $K P^T = 0$.*

While we do not know yet how large the difference between mlt($G$) and rank($G$) can be, we can show that given a graph $G$ the minimal observations $n$ needed to ensure that the SME exists almost surely is exactly equal to the rank of $G$.

**Definition 6.2.** *Let $G$ be a graph. The graph $G$ is $n$-estimable if the score matching estimator of $\Sigma$ exists with probability one. The score matching threshold of $G$, denoted smt($G$) is the minimal $n$ such that $G$ is $n$-estimable.*

**Theorem 6.3.** *Let $G$ be a graph. Then*

$$\text{smt}(G) = \text{rank}(G).$$

**Proof.** The system $K P^T = 0$ is a linear system in the entries of $K$ with coefficients in the entries of $P$. The coefficient matrix $C$ of the system $K P^T = 0$ is a $mn \times (\#V + \#E)$ matrix where the columns are indexed by the vertices and edges of $G$; the system $K P^T = 0$ has a unique solution if and only if the rank of $C$ is $\#V + \#E$.

Let $M$ be the matrix obtained from the Jacobian $J(g, P)$ from Section 2 by scaling the columns indexed by $ii$ by $\frac{1}{2}$. The coefficient matrix $C$ is the submatrix of $M$ obtained by selecting the columns indexed by the vertices and edges of $G$. Thus, the matrix $C$ has rank $\#V + \#E$ for generic $\mathbf{p}_1, \ldots, \mathbf{p}_m$ if and only if $\{\sigma_{ii} : i \in [m]\} \cup \{\sigma_{ij} : ij \in E(G)\}$ is an independent set of the symmetric minor matroid $S(m, n)$. The statement then follows by Proposition 2.3. $\square$

We can now apply all the results in the previous sections on the rank of a graph to the score matching threshold. For example:

**Corollary 6.4.** *Let $G = (V, E)$ be a graph. The smt($G$) $\leq 3$ if and only if for all subgraphs $G' = (V', E')$ of $G$*

$$\#E' \leq 2\big(\#V'\big) - 3.$$

**Corollary 6.5.** *Let G be a graph with empty n-core. Then* $\mathrm{smt}(G) \leq n$.

**Corollary 6.6.** *Let G be a graph and* $V_1, \ldots, V_k$ *be a partition of the vertices of G such that*:

(1) *for all i*, $V_i$ *is an independent set of G and*
(2) *for all* $i \neq j$, $G(V_i, V_j)$ *has no cycles.*

*Then* $\mathrm{smt}(G) \leq k$.

Lauritzen stated the following conjecture about the score matching threshold in his lecture at the 2014 Prague Stochastics meeting.

**Conjecture 6.7.** *The graph G is n-estimable if and only if*

$$\#V + \#E \leq nm - \binom{n}{2}.$$

The translation to rigidity theory immediately provides counterexamples.

***Example 6.8 (Counterexample to Conjecture 6.7).*** Let $G = (V, E)$ be the graph depicted in Figure 5. Let $n = 3$. Then

$$\#V + \#E = 12 = nm - \binom{n}{2}.$$

Thus, $G$ is conjectured to be 3-estimable. However, by Corollary 6.4, this cannot be the case since the complete graph $K_4$ is a subgraph of $G$.

Even with the stronger condition that the inequality $\#V' + \#E' \leq n\#V' - \binom{n}{2}$ for every induced subgraph $G' = (V', E')$ of $G$, there are known counterexamples of graphs satisfying all of these inequalities but not being rigid. The simplest such graph is the double banana graph, which satisfies all these inequalities for $n = 4$ but is not a rigid graph in $\mathcal{A}(3)$. So the Gaussian graphical model associated to the double banana graph is not 4-estimable.

# 7. Conclusion

The maximum likelihood threshold of a graph is an important measure of the complexity of the Gaussian graphical model associated to the graph. It measures how much data is needed to cal-
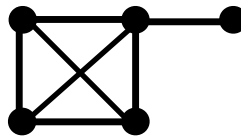


**Figure 5.** Graph satisfies conditions of Conjecture 6.7 for $n = 3$, but is not 3-estimable.

culate maximum likelihood estimates for the parameters of the model. We showed that a result of Uhler implies that the maximum likelihood threshold is closely related to combinatorial rigidity theory, and then imported a number of results from combinatorial rigidity theory to get new bounds on the maximum likelihood threshold. These new bounds significantly improve bounds that exist in the literature, and in some cases imply effective ways to check whether the MLE exists almost surely. Note that a probabilistic algorithm to check the rank of a graph can be employed by checking the rank of the appropriate Jacobian matrix at randomly sampled points.

We conclude here with a few remaining questions. First, as discussed in the Introduction, does there exists a graph $G$ such that $\text{mlt}(G)$ is strictly less than $\text{rank}(G)$? Second, just as we are interested in the possible discrepancy between $\text{mlt}(G)$ and $\text{rank}(G)$ of a given graph, we are also intrigued by the discrepancy between the $\text{mlt}(G)$ and the Gaussian rank of a graph as defined in [2].

In [2], Ben-David defines the Gaussian rank of a graph to be the smallest $n$ such that for every $\Sigma_0 \in \text{Sym}(m, n) \cap \mathbb{S}_{\geq 0}^m$ *in general position* there exists a $\Sigma \in \mathbb{S}_{>0}^m$ such that $\phi_G(\Sigma_0) = \phi_G(\Sigma)$. General position means that in any representation $\Sigma_0 = P^T P$ where $P$ is an $n \times m$ matrix, all $n \times n$ minors of $P$ are nonzero. Such an $n$ is an upper bound on the $\text{mlt}(G)$. The difference between the $\text{mlt}(G)$ and the Gaussian rank is that the $\text{mlt}(G)$ requires the condition in Problem 1.1 to hold everywhere except for *some* set of measure zero, while Gaussian rank prescribes a specific set of measure zero off of which the condition must hold. The Gaussian rank has the advantage that the general position condition can be explicitly checked in examples, but $\text{mlt}(G)$ is often smaller. In [2], the author gives an example of a planar graph whose Gaussian rank is 6, however, from Corollary 3.11, we know that the $\text{mlt}(G)$ of any planar graph is at most 4, using the bound coming from the rank of $G$. While both rank and Gaussian rank are both upper bounds for the maximum likelihood threshold, we do not know if it is always the case the Gaussian rank is greater than or equal to the rank of $G$.

Third, we are curious whether it possible to directly pin down the precise connection between rigidity theory and the maximum likelihood threshold. We provide a conjecture relating the maximum likelihood threshold to a stronger form of rigidity. Let $G = (V, E)$ be a graph with $\#V = m$. A *framework* in $\mathbb{R}^n$ with respect to $G$, denoted $(G, P)$, is an $n \times m$ matrix $P$ such that the $i$th column of $P$, denoted $\mathbf{p}_i$, is an embedding of the $i$th vertex of $G$ into $\mathbb{R}^n$.

**Definition 7.1.** *Two frameworks $(G, P)$ and $(G, Q)$ are* edge-equivalent *if*

$$\|\mathbf{p}_i - \mathbf{p}_j\|_2^2 = \|\mathbf{q}_i - \mathbf{q}_j\|_2^2 \qquad \forall ij \in E(G).$$

**Definition 7.2.** *Let $G = (V, E)$ be a graph with $\#V = m$. A framework $(G, P)$ in $\mathbb{R}^n$ is $n$-dependently rigid if for every edge equivalent framework $(G, Q)$ in $\mathbb{R}^m$ the set of point $\{\mathbf{q}_1, \ldots, \mathbf{q}_m\}$ is affinely dependent.*

We will say that $G$ is *generically $n$-dependently rigid* if every generic framework $(G, P)$ in $\mathbb{R}^n$ is $n$-dependently rigid.

**Conjecture 7.3.** *The maximum likelihood threshold for a graph $G$ is greater than $n$ if and only if $G$ is generically $n$-dependently rigid.*

We hope that once a precise connection between combinatorial rigidity theory and maximum likelihood estimation is established, new results on the mlt($G$), guaranteed to be sharp, could be obtained.

In addition to studying the maximum likelihood threshold, in this paper, we also looked at two related graph invariants, the weak maximum likelihood threshold and the score matching threshold. Little is understood about the weak maximum likelihood threshold, however, here we were able to show wmlt($G$) is bounded above by the chromatic number of $G$, and we were able to give a necessary condition on $G$ for wmlt($G$) = 2. As for the score matching threshold, we showed a direct connection between the smt($G$) and independent sets in the generic rigidity matroid. While we saw that for $n = 3$, conditions for independence in $\mathcal{A}(2)$ are efficient to check and some sufficient conditions for independence in $\mathcal{A}(n-1)$ are known for $n > 3$, it should be noted that it is still an open problem to characterize all independent sets in $\mathcal{A}(3)$. We hope that this connection though inspires more work on understanding the rigidity matroid for statistical applications.

# Acknowledgments

# References

[1] Alon, N., McDiarmid, C. and Reed, B. (1991). Acyclic coloring of graphs. *Random Structures Algorithms* **2** 277–288. MR1109695

[2] Ben-David, E. (2015). Sharp lower and upper bounds for the Gaussian rank of a graph. *J. Multivariate Anal.* **139** 207–218. MR3349487

[3] Buhl, S.L. (1993). On the existence of maximum likelihood estimators for graphical Gaussian models. *Scand. J. Stat.* **20** 263–270. MR1241392

[4] Connelly, R. and Whiteley, W.J. (2010). Global rigidity: The effect of coning. *Discrete Comput. Geom.* **43** 717–735. MR2610468

[5] Dehn, M. (1916). Über die Starrheit konvexer Polyeder. *Math. Ann.* **77** 466–473. MR1511873

[6] Dempster, A.P. (1972). Covariance selection. *Biometrics* **28** 157–175.

[7] Dobra, A., Hans, C., Jones, B., Nevins, J.R., Yao, G. and West, M. (2004). Sparse graphical models for exploring gene expression data. *J. Multivariate Anal.* **90** 196–212. MR2064941

[8] Forbes, P.G.M. and Lauritzen, S. (2015). Linear estimating equations for exponential families with application to Gaussian linear concentration models. *Linear Algebra Appl.* **473** 261–283. MR3338335

[9] Friedman, J., Hastie, T. and Tibsharani, T. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

[10] Gluck, H. (1975). Almost all simply connected closed surfaces are rigid. In *Geometric Topology* (*Proc. Conf.*, *Park City*, *Utah*, 1974). *Lecture Notes in Math.* **438** 225–239. Berlin: Springer. MR0400239

[11] Graver, J., Servatius, B. and Servatius, H. (1993). *Combinatorial Rigidity*. *Graduate Studies in Mathematics* **2**. Providence, RI: Amer. Math. Soc. MR1251062

[12] Grünbaum, B. (1973). Acyclic colorings of planar graphs. *Israel J. Math*. **14** 390–408. MR0317982

[13] Hyvärinen, A. (2005). Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res*. **6** 695–709. MR2249836

[14] Jackson, B., Jordán, T. and Tanigawa, S. (2014). Combinatorial conditions for the unique completability of low-rank matrices. *SIAM J. Discrete Math*. **28** 1797–1819. MR3268605

[15] Jacobs, D.J. and Hendrickson, B. (1997). An algorithm for two-dimensional rigidity percolation: The pebble game. *J. Comput. Phys*. **137** 346–365. MR1481894

[16] Kalai, G., Nevo, E. and Novik, I. (2016). Bipartite rigidity. *Trans. Amer. Math. Soc*. **368** 5515–5545. MR3458389

[17] Király, F., Rosen, Z. and Theran, L. (2013). Algebraic matroids with graph symmetry. Preprint. Available at arXiv:1312.3777.

[18] Király, F.J., Theran, L. and Tomioka, R. (2015). The algebraic combinatorial approach for low-rank matrix completion. *J. Mach. Learn. Res*. **16** 1391–1436. MR3417786

[19] Krumsiek, J., Suhre, K., Illig, T., Adamski, J. and Theis, F.J. (2011). Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst. Biol*. **5**.

[20] Laman, G. (1970). On graphs and rigidity of plane skeletal structures. *J. Engrg. Math*. **4** 331–340. MR0269535

[21] Lauritzen, S.L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. New York: Oxford Univ. Press. MR1419991

[22] Oxley, J. (2011). *Matroid Theory*, 2nd ed. *Oxford Graduate Texts in Mathematics* **21**. Oxford: Oxford Univ. Press. MR2849819

[23] Schäfer, J. and Strimmer, K. (2005). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* **21** 754–764.

[24] Singer, A. and Cucuringu, M. (2009/2010). Uniqueness of low-rank matrix completion by rigidity theory. *SIAM J. Matrix Anal. Appl*. **31** 1621–1641. MR2595541

[25] Uhler, C. (2012). Geometry of maximum likelihood estimation in Gaussian graphical models. *Ann. Statist*. **40** 238–261. MR3014306

[26] Welsh, D.J.A. (1976). *Matroid Theory*. London: Academic Press. MR0427112

[27] Whiteley, W. (1983). Cones, infinity and 1-story buildings. *Topol. Struct*. **8** 53–70. MR0721956

[28] Whiteley, W. (1996). Some matroids from discrete applied geometry. In *Matroid Theory* (*Seattle*, *WA*, 1995). *Contemp. Math*. **197** 171–311. Providence, RI: Amer. Math. Soc. MR1411692

[29] Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley. MR1112133

[30] Wu, X., Ye, Y. and Subramanian, K.R. (2003). Interactive analysis of gene interactions using graphical Gaussian model. *Proceedings of the ACM SIGKDD Workshop on Data Mining in Bioinformatics* **3** 63–69.

[31] Ziegler, G.M. (1995). *Lectures on Polytopes*. *Graduate Texts in Mathematics* **152**. New York: Springer. MR1311028