,

COMPUTING MAXIMUM LIKELIHOOD ESTIMATES FOR GAUSSIAN GRAPHICAL MODELS WITH MACAULAY2

CARLOS AMÉNDOLA, LUIS DAVID GARCÍA PUENTE, ROSER HOMS, OLGA KUZNETSOVA, AND HARSHIT J MOTWANI

Abstract. We introduce the package `GraphicalModelsMLE` for computing the maximum likelihood estimator (MLE) of a Gaussian graphical model in the computer algebra system `Macaulay2`. The package allows to compute for the class of loopless mixed graphs. Additional functionality allows to explore the underlying algebraic structure of the model, such as its ML degree and the ideal of score equations.

1. Introduction

The purpose of this package is to extend the functionality of `Macaulay2` [6] related to algebraic statistics, specifically allowing computations of maximum likelihood estimates of Gaussian graphical models. While `GraphicalModels` is an existing package that already provides useful information such as conditional independence ideals and vanishing ideals for such models, the fundamental statistical inference task of computing maximum likelihood estimates is missing. This package aims to fill this void and also extend the functionality of `GraphicalModels` to handle more general types of graphs.

The algebraic framework of `Macaulay2` permits us to use both commutative algebra and numerical algebraic geometry to obtain a guaranteed global optimal solution by computing all critical points of the log-likelihood function. This is a different insight from the classical statistical approach of the `R` package `ggm` [7], and more in line with the recent numerical algebraic geometry approach from the package `LinearCovarianceModels.jl` in *Julia* [11]. The package `GraphicalModelsMLE` is a great complement to these two, handling some Gaussian graphical models not covered by them (`LinearcovarianceModels.jl` version 0.2 and `ggm` version 2.5).

Given a sample of n i.i.d. random vectors $X^{(1)},\dots,X^{(n)}$ that follow an m-dimensional multivariate Gaussian distribution $\mathcal{N}(\mu,\Sigma)$, the *maximum likelihood estimate* (MLE) for the covariance matrix Σ is the matrix that best explains the observed data. More precisely, to compute the MLE one solves the optimization problem of maximizing the log-likelihood function of the Gaussian model:

$$\begin{aligned} \max_{\Sigma\in\mathbb{R}^{m\times m}} \quad & \ell(\Sigma) = -\log\det\Sigma - \operatorname{tr}(S\Sigma^{-1}) \\ \text{subject to:} \quad & \Sigma \succ 0, \end{aligned} \tag{1.1}$$

where S is the sample covariance matrix and $\Sigma \succ 0$ means that Σ is a positive definite matrix [12, Proposition 7.1.9].

In `GraphicalModelsMLE` we focus on computing the MLE for the covariance matrix of Gaussian models that arise from graphical models and, in particular, those that arise from *loopless mixed graphs* (LMG).

Our algebraic approach allows to study the main algebraic features of the MLE problem: the ideal of score equations and the ML-degree of the model, see Section 4 and Section 5 respectively.

2. Graphical models of loopless mixed graphs

A *mixed graph* $G = (V, E)$ is a graph with undirected edges $i - j$, directed edges $i \to j$ and bidirected edges $i \leftrightarrow j$. A *directed cycle* is a cycle on directed edges or a cycle formed by directed edges after identifying the vertices that are connected by undirected or bidirected edges. A *loopless mixed graph* (LMG) is a mixed graph without loops or directed cycles. We allow double edges of the types directed-undirected and directed-bidirected.

Following [13], we assume that the nodes of G are partitioned as $V = U \cup W$, such that:

- if $i - j$ in G then $i, j \in U$
- if $i \leftrightarrow j$ in G then $i, j \in W$
- there is no directed edge $i \to j$ in G such that $i \in W$ and $j \in U$

Our definition differs from the one in [10] in that we do not allow multiple edges of the same type, which is due to the set up of the `Graphs` package. We also prohibit directed cycles, which ensures there is an ordering on the vertices such that that all vertices in U come before vertices in W, and whenever $i \to j$ we have $i < j$.

A Gaussian graphical model imposes constraints on the covariance matrix of a Gaussian distribution. More precisely, a loopless mixed graph $G = (V, E)$ gives rise to the space of covariance matrices $\Sigma \in \mathbb{R}^{V \times V}$ of the form

$$\Sigma \quad = \quad (I - \Lambda)^{-T} \begin{bmatrix} K^{-1} & \mathbf{0} \\ \mathbf{0} & \Psi \end{bmatrix} (I - \Lambda)^{-1}, \tag{2.1}$$

where

(i) $\Lambda = [\lambda_{ij}] \in \mathbb{R}^{V \times V}$ is such that $\lambda_{ij} = 0$ whenever $i \to j \notin E$
(ii) $K = [k_{ij}] \in \mathbb{R}^{U \times U}$ is symmetric positive definite such that $k_{ij} = 0$ whenever $i - j \notin E$
(iii) $\Psi = [\psi_{ij}] \in \mathbb{R}^{W \times W}$ is symmetric positive definite such that $\psi_{ij} = 0$ whenever $i \leftrightarrow j \notin E$

3. Maximum likelihood estimator

Given n i.i.d. random vectors $X^{(1)}, \ldots, X^{(n)} \sim \mathcal{N}(\mu, \Sigma)$ and the parameter space $\Theta = \mathbb{R}^m \times \Theta_2 \subseteq \mathbb{R}^m \times PD_m$, the estimator for the covariance matrix is determined by maximizing

$$\ell(\Sigma) = -\frac{n}{2} \log\det\Sigma - \frac{n}{2} \operatorname{tr}\left(S\Sigma^{-1}\right) \tag{3.1}$$

over $\Sigma \in \Theta_2$ [12, Proposition 7.1.9]. The function `solverMLE` allows to compute this optimum when Θ_2 is induced by (2.1). It does so by calculating the critical points of the log-likelihood function and selecting the points corresponding to the maximum value in the cone of positive definite matrices. The default output is the maximum value of $\ell(\Sigma)$, the list of maximum likelihood estimates for the covariance matrix and the maximum likelihood degree of the model.

For undirected graphs, the MLE for the covariance matrix is known to be the unique positive definite critical point of the likelihood. In particular, it is a positive definite matrix completion to the partial sample covariance matrix. See [14, Theorem 2.1] or [4, Theorem 2.1.14] for more details.

Example 3.2. We compute the MLE for the covariance matrix of the graphical model associated to the undirected 4-cycle. We take as input data the sample covariance matrix S defined below.

```
i1 : loadPackage "GraphicalModelsMLE";
i2 : G=graph{{1,2},{2,3},{3,4},{4,1}};
i3 : S=matrix {{.105409, -.0745495, -.0186132, .0621907},
              {-.0745495, .0783734,-.00844503,-.0872842},
              {-.0186132, -.00844503, .128307, .0230245},
              {.0621907, -.0872842, .0230245,.109849}};
i4 : solverMLE(G,S,SampleData=>false)
o4 = (6.62005, | .105409   -.0745495  .0124099   .0621907  |, 5)
               | -.0745495 .0783734   -.00844503 -.0439427 |
               | .0124099  -.00844503 .128307    .0230245  |
               | .0621907  -.0439427  .0230245   .109849   |
```

Note that all entries in the MLE for the covariance matrix coincide with the entries in the sample covariance matrix except for those corresponding to non-edges of the graph. See [8, Example 12.16] for more on a positive definite matrix completion problem associated to the 4-cycle.

For more general types of graphs, uniqueness of the positive definite critical points is no longer guaranteed. In the mixed graph below, the optimization problem has a global maximum but there are also local maxima, see Example 4.3.

Example 3.3. We compute the MLE for the covariance matrix of the graphical model associated to the loopless mixed graph with undirected edge $1-2$, directed edges $1 \to 3, 2 \to 4$ and bidirected edge $3 \leftrightarrow 4$.

```
i2 : G = mixedGraph(graph{{1,2}},digraph{{1,3},{2,4}},bigraph{{3,4}});
i3 : S=matrix {{34183/50000, 716539/10000000, 204869/250000, 12213/25000},
              {716539/10000000, 112191/500000, 309413/1000000, 1803/4000},
              {204869/250000, 309413/1000000, 3849/3125,15172/15625},
              {12213/25000, 1803/4000, 15172/15625, 4487/4000}};
i4 : solverMLE(G,S,SampleData=>false)
o4 = (9.36624, {| .68366    .0716539 1.00282   .234375   |}, 5)
                | .0716539 .224382  .105105   .733937   |
                | 1.00282  .105105  1.76955   -.0700599 |
                | .234375  .733937  -.0700599 2.97432   |
```

4. Ideal of score equations

The critical points of the log-likelihood function $\ell(\Sigma)$ are the solutions to the system of equations obtained by taking partial derivatives of ℓ with respect to all variables in the entries of Σ from our construction in (2.1):

$$-\frac{\partial}{\partial(\cdot)}\det\Sigma - \det\Sigma\frac{\partial}{\partial(\cdot)}\operatorname{tr}(\mathrm{S}\Sigma^{-1})\,. \tag{4.1}$$

Such equations are called *score equations*. From an algebraic point of view, the ideal generated by the score equations of the model is already of interest on its own, see [12, Chapter 7].

Note that the log-likelihood function depends on the sample covariance matrix, therefore our implementation of `scoreEquations` requires as input both sample data and the graphical model encoded in a `gaussianRing`. See Section 6 for more details on Gaussian rings.

Example 4.2. We compute the ideal of score equations associated to the 4-cycle after creating the graph G as in Example 3.2. We now consider as input data the sample data encoded in the columns of matrix U below.

```
i5 : U=matrix{{3,5,9,5},{1,6,1,5},{2,9,6,6},{2,5,0,4}};
i6 : J=scoreEquations(gaussianRing G,U);
o6 : Ideal of QQ[k   , k   , k   , k   , k   , k   , k   , k   ]
                  1,1   2,2   3,3   4,4   1,2   1,4   2,3   3,4
i7 : dim J
o7 = 0
```

The ideal of score equations J is generated by 14 non-homogeneous polynomials in $\mathbb{Q}[k_{1,1}, k_{1,2}, k_{1,4}, k_{2,2}, k_{2,3}, k_{3,3}, k_{3,4}, k_{4,4}]$: 4 linear polynomials and 10 quadratic polynomials such as $1312002k_{3,4}^2 - 387081k_{1,2} + 109860k_{1,4} + 1972025k_{2,3} - 898518k_{3,4} - 291556$.

Example 4.3. We want to obtain all local maxima of the log-likelihood function associated to the graphical model in Example 3.3. The score equations generate an ideal in $\mathbb{Q}[k_{1,1}, k_{2,2}, k_{1,2}, l_{1,3}, l_{2,4}, p_{3,3}, p_{4,4}, p_{3,4}]$ and we display their solutions in the `Macaulay2` session below. We retrieve the covariance matrix Σ with rational entries in variables $k_{1,1}, k_{2,2}, k_{1,2}, l_{1,3}, l_{2,4}, p_{3,3}, p_{4,4}, p_{3,4}$ using the optional output `CovarianceMatrix` in `scoreEquations`.

```
i5 : R = gaussianRing G;
i6 : (J,Sigma)=scoreEquations(R,S,SampleData=>false,CovarianceMatrix=>true);
i7 : dim J, degree J
o7 = (0, 5)
i8 : sols=zeroDimSolve(J);netList sols
     +-------------------------------------------------------------------------+
o8 = |{1.51337, 4.61101, -.483277, 1.46684, 3.27093, .298576, .573665, -.41385} |
     +-------------------------------------------------------------------------+
     |{1.51337, 4.61101, -.483277, 1.39884+.440525*ii, 2.45466-.923165*ii,      |
     |.144129+.120574*ii,.0696297-.184692*ii,-.19668+.0553853*ii}               |
     +-------------------------------------------------------------------------+
     |{1.51337, 4.61101, -.483277, 1.39884-.440525*ii, 2.45466+.923165*ii,      |
     |.144129-.120574*ii, .0696297+.184692*ii, -.19668-.0553853*ii}             |
     +-------------------------------------------------------------------------+
     |{1.51337, 4.61101, -.483277, .684147, .979681, .430388, .453924, .381688} |
     +-------------------------------------------------------------------------+
     |{1.51337, 4.61101, -.483277, .988484, 1.64649, .279607, .245722, .0952865}|
     +-------------------------------------------------------------------------+
```

How many of the 3 real critical points correspond to positive definite matrices that are local maxima of the log-likelihood function? We first check that they correspond to positive definite matrices by substituting the three real solutions in the covariance matrix Σ.

```
i9 : checkPD(apply(sols,i->sub(Sigma,matrix{coordinates(i)})))
| .68366   .0716539 1.00282   .234375   |, | .68366   .0716539 .467724  .070198 |,
| .0716539 .224382  .105105   .733937   |  | .0716539 .224382  .0490218 .219823 |
| 1.00282  .105105  1.76955   -.0700599 |  | .467724  .0490218 .75038   .429714 |
| .234375  .733937  -.0700599 2.97432   |  | .070198  .219823  .429714  .66928  |
                          | .68366   .0716539 .675787  .117978 |
                          | .0716539 .224382  .0708287 .369443 |
                          | .675787  .0708287 .947611  .211905 |
                          | .117978  .369443  .211905  .854009 |
```

The MLE for the covariance obtained in Example 3.3 corresponds to the first positive definite matrix in the list above. The eigenvalues of the Hessian matrix computed below tell us which kind of critical point we have for each of the 3 real solutions.

```
-- compute Jacobian matrix (i.e. score equations)
i10 : scoreEq=-1/det Sigma*jacobianMatrixOfRationalFunction(det Sigma)-
jacobianMatrixOfRationalFunction(trace(S*(inverse Sigma)));
-- compute Hessian matrix
i11 : Hessian=matrix for f in flatten entries scoreEq list
flatten entries jacobianMatrixOfRationalFunction(f);
-- compute eigenvalues of the Hessian matrix evaluated at real points in sols
i12 : apply({sols_0,sols_3,sols_4},i->eigenvalues sub(Hessian,matrix{coordinates(i)}))
{{-.516478   }, {-.516478 }, {-.516478 }}
 {-.271913   }  {-.271913 }  {-.271913 }
 {-.0464172  }  {-.0464172}  {-.0464172}
 {-9869730000}  {-414.15  }  {-59.7135 }
 {-128887    }  {-28.6352 }  {1.52598  }
 {-58261.1   }  {-6.1936  }  {-2.80504 }
 {-773.513   }  {-1.64689 }  {-11.5533 }
```

There are two local maxima and a saddle point. This shows that the log-likelihood function of this model is not a concave function, see [3].

Example 4.4. Next we compute the ideal of score equations associated to a mixed graph that has a multi-edge: directed edges $1 \to 3, 1 \to 2, 2 \to 4, 3 \to 4$ and undirected edge $1 - 2$.

```
i2 : G = mixedGraph(digraph{{1,3},{1,2},{2,4},{3,4}},graph{{1,2}});
i3 : R = gaussianRing G;
i4 : U = random(RR^4,RR^4);
i5 : J=scoreEquations(R,U);
i6 : dim J
o6 = 1
```

Note that in this case the ideal of score equations is no longer zero-dimensional.

5. Maximum likelihood degree

The *maximum likelihood degree* (ML-degree) of a model is defined as the number of complex solutions of the score equations for generic sample data, see [12, Definition 7.1.4]. For a more algebraic flavour of the notion of ML-degree, see [9, Definition 5.4].

Note that the ML-degree is only well-defined when the ideal of score equations is zero dimensional. A typical way where this fails is where the model becomes non-idenfiable. See e.g. [1] for some sufficient conditions to avoid non-identifiability and preservation of dimension of the model in terms of the number of parameters.

It is important to observe that for generic data the solutions to score equations are all distinct, see [2, Remark 2.1, Lemma 2.2]. Computing the algebraic degree of the zero-dimensional score equations ideal via the `degree` function in M2 is equivalent to computing the number of complex solutions - without multiplicity - to the score equations.

In our implementation of the `MLdegree` function in `Macaulay2` a random sample matrix is used as sample data. Therefore, the ML-degree of the graphical model we provide is correct with probability 1.

Example 5.1. The ML-degree of the 4-cycle can be directly computed as follows:

```
i2 : G=graph{{1,2},{2,3},{3,4},{4,1}};
i3 : MLdegree(gaussianRing G)
o3 = 5
```

In the case of ideals of score equations with positive dimension, `MLdegree` will still compute the degree of the ideal but this no longer matches the number of solutions to the score equations.

Example 5.2. Continuing with Example 4.4, where the ideal of score equations is 1-dimensional, `MLdegree` does not provide a meaningful answer.

```
i2: G=mixedGraph(digraph{{1,3},{1,2},{2,4},{3,4}},graph{{1,2}});
i3: MLdegree(gaussianRing G)
error: the ideal of score equations has dimension 1 > 0,
so ML degree is not well-defined. The degree of this ideal is 2.
```

6. Updates in related packages

`GraphicalModelsMLE` relies on the new package `StatGraphs` 0.1 and the updated packages `Graphs` 0.3.3 and `GraphicalModels` 2.0 (see [5] for version 1.0).

We created a dedicated package StatGraphs for graph theoretic functions relevant to algebraic statistics. It contains the functions `isCyclic`, `isSimple`, `isLoopless` and `partitionLMG` to deal with loopless mixed graph. The package `Graphs` keeps graph-related functionality of general use.

The function `partitionLMG` computes the partition $V = U \cup W$ of vertices of a loopless mixed graph described in Section 2. Vertices in the input graph need to be ordered such that (1) all vertices in U come before vertices in W and (2) if there is a directed edge $i \to j$, then $i < j$.

Example 6.1. The vertices of the loopless mixed graph in Example 3.3 are partitioned into $U = \{1, 2\}$ and $W = \{3, 4\}$.

```
i1 : loadPackage "StatGraphs";
i2 : G = mixedGraph(digraph {{1,3},{2,4}},bigraph{{3,4}},graph{{1,2}});
i3 : partitionLMG G
o3 = ({1, 2}, {3, 4})
o3 : Sequence
```

The central object in the implementation of our MLE algorithm is `gaussianRing` from the package `GraphicalModels`.

Example 6.2. We compute the `gaussianRing` associated to the graph in Example 6.1 and display the variables of the ring as entries of matrices:

```
i4 : loadPackage "GraphicalModels";
i5 : R=gaussianRing G;
i6 : undirectedEdgesMatrix R
o6 = | k_(1,1) k_(1,2) |
     | k_(1,2) k_(2,2) |
i7 : directedEdgesMatrix R
o7 = | 0 0 l_(1,3) 0       |
     | 0 0 0       l_(2,4) |
     | 0 0 0       0       |
     | 0 0 0       0       |
```

```
i8 : bidirectedEdgesMatrix R
o8 = | p_(3,3) p_(3,4) |
     | p_(3,4) p_(4,4) |
i9 : covarianceMatrix R
o9 = | s_(1,1) s_(1,2) s_(1,3) s_(1,4) |
     | s_(1,2) s_(2,2) s_(2,3) s_(2,4) |
     | s_(1,3) s_(2,3) s_(3,3) s_(3,4) |
     | s_(1,4) s_(2,4) s_(3,4) s_(4,4) |
```

In version 2.0 of `GraphicalModels`, we updated the functionalities of the method `gaussianRing` – and its related methods – in order to accept loopless mixed graphs with undirected, directed and bidirected edges.

Note that mixed graphs that include undirected edges are required to have an ordering compatible with `partitionLMG`. For mixed graphs with only directed and bidirected edges this is no longer necessary, as in version 1.0 of `GraphicalModels`.

Acknowledgements. We thank Michael Stillman and Daniel Grayson for their support and help with the inner workings of `Macaulay2` during the 2020 online Warwick Macaulay2 Workshop. We are also grateful to Elina Robeva and David Swinarski for creating the first draft of `GraphicalModelsMLE`, and to Piotr Zwiernik for useful discussions.

References

[1] Carlos Améndola, Philipp Dettling, Mathias Drton, Federica Onori, and Jun Wu, *Structure learning for cyclic linear causal models*, Uncertainty of Artificial Intelligence (UAI) (2020).

[2] Carlos Améndola, Lukas Gustafsson, Kathlén Kohn, Orlando Marigliano, and Anna Seigal, *The maximum likelihood degree of linear spaces of symmetric matrices*, 2020.

[3] Mathias Drton, *Computing all roots of the likelihood equations of seemingly unrelated regressions*, Journal of Symbolic Computation **41** (2006), no. 2, 245 –254. Computational Algebraic Statistics.

[4] Mathias Drton, Bernd Sturmfels, and Seth Sullivant, *Lectures on algebraic statistics*, Vol. 39, Springer Science & Business Media, 2008.

[5] Luis García-Puente, Sonja Petrović, and Seth Sullivant, *Graphical models*, Journal of Software for Algebra and Geometry **5** (2013), no. 1, 1–7.

[6] Daniel R Grayson and Michael E Stillman, *Macaulay2, a software system for research in algebraic geometry*, Available at `http://www.math.uiuc.edu/Macaulay2/`, 2002.

[7] Giovanni M Marchetti et al., *Independencies induced from a graphical Markov model after marginalization and conditioning: the R package ggm*, Journal of Statistical Software **15** (2006), no. 6, 1–15.

[8] Mateusz Michałek and Bernd Sturmfels, *Invitation to nonlinear algebra*, Graduate Studies in Mathematics, American Mathematical Society, 2021.

[9] Mateusz Michałek, Bernd Sturmfels, Caroline Uhler, and Piotr Zwiernik, *Exponential varieties*, Proceedings of the London Mathematical Society **112** (2016), 27–56.

[10] Kayvan Sadeghi, Steffen Lauritzen, et al., *Markov properties for mixed graphs*, Bernoulli **20** (2014), no. 2, 676–696.

[11] Bernd Sturmfels, Sascha Timme, and Piotr Zwiernik, *Estimating linear covariance models with numerical nonlinear algebra*, Algebraic Statistics **11** (2020), no. 1, 31–52.

[12] Seth Sullivant, *Algebraic statistics*, Vol. 194, American Mathematical Soc., 2018.

[13] Seth Sullivant, Kelli Talaska, Jan Draisma, et al., *Trek separation for Gaussian graphical models*, The Annals of Statistics **38** (2010), no. 3, 1665–1685.

[14] Caroline Uhler, *Geometry of maximum likelihood estimation in Gaussian graphical models*, Ann. Statist. **40** (201202), no. 1, 238–261.

Email address: `carlos.amendola@tum.de`
Email address: `lgarcia@shsu.edu`
Email address: `roser.homs@tum.de`
Email address: `kuznetsova.olga@gmail.com`
Email address: `harshitjitendra.motwani@ugent.be`