Invariant theory and scaling algorithms for maximum likelihood estimation

Kathlén Kohn
KTH Stockholm

joint with

Carlos Améndola
TU Munich

Philipp Reichenbach
TU Berlin

Anna Seigal
University of Oxford

May 20, 2020
Given: statistical model

Sample data \( S \)

Task: find maximum likelihood estimate (MLE) = closest distance of orbit to origin = point in model that best fits \( S \).

Statistics

Invariant theory

Global picture
Given: statistical model sample data $S_Y$

Task: find maximum likelihood estimate (MLE) = closest distance of orbit to origin = point in model that best fits $S_Y$
Given: statistical model
sample data $S_Y$

Task: find maximum likelihood estimate (MLE)
= point in model that best fits $S_Y$
Given: statistical model sample data \( S_Y \)

Task: find \textbf{maximum likelihood estimate (MLE)}

\[ = \text{point in model that best fits } S_Y \]
Given: statistical model
sample data $S_Y$

Task: find **maximum likelihood estimate (MLE)**
= point in model that best fits $S_Y$

Given: orbit $G \cdot v = \{g \cdot v \mid g \in G\}$

Task: compute **capacity**
= closest distance of orbit to origin
The **orbit** of a vector $v$ in a vector space $V$ under an action by a group $G$ is

$$G.v = \{g \cdot v \mid g \in G\} \subset V.$$
Invariant theory
Stability notions

The **orbit** of a vector $v$ in a vector space $V$ under an action by a group $G$ is

$$G.v = \{g \cdot v \mid g \in G\} \subset V.$$ 

- $v$ is **unstable** iff $0 \in \overline{G.v}$ (i.e. $v$ can be scaled to 0 in the limit)

The **null cone** of the action by $G$ is the set of unstable vectors $v$.
Invariant theory

Stability notions

The orbit of a vector $v$ in a vector space $V$ under an action by a group $G$ is

$$G.v = \{g \cdot v \mid g \in G\} \subset V.$$ 

- $v$ is unstable iff $0 \in \overline{G.v}$ (i.e. $v$ can be scaled to 0 in the limit)
- $v$ semistable iff $0 \notin \overline{G.v}$

The null cone of the action by $G$ is the set of unstable vectors $v$. 
Invariant theory

Stability notions

The **orbit** of a vector \( v \) in a vector space \( V \) under an action by a group \( G \) is

\[
G \cdot v = \{ g \cdot v \mid g \in G \} \subset V.
\]

- \( v \) is **unstable** iff \( 0 \in G \cdot v \) (i.e. \( v \) can be scaled to 0 in the limit)
- \( v \) **semistable** iff \( 0 \notin G \cdot v \)
- \( v \) **polystable** iff \( v \neq 0 \) and its orbit \( G \cdot v \) is closed

The **null cone** of the action by \( G \) is the set of unstable vectors \( v \).
Invariant theory
Stability notions

The **orbit** of a vector \( v \) in a vector space \( V \) under an action by a group \( G \) is

\[
G \cdot v = \{ g \cdot v \mid g \in G \} \subset V.
\]

- \( v \) is **unstable** iff \( 0 \in \overline{G \cdot v} \) (i.e. \( v \) can be scaled to 0 in the limit)
- \( v \) **semistable** iff \( 0 \notin \overline{G \cdot v} \)
- \( v \) **polystable** iff \( v \neq 0 \) and its orbit \( G \cdot v \) is closed
- \( v \) is **stable** iff \( v \) is polystable and its stabilizer is finite

The **null cone** of the action by \( G \) is the set of unstable vectors \( v \).
Invariant theory
Null cone membership testing

Classical and often hard question: Describe null cone
(essentially equivalent to finding generators for the ring of polynomial invariants)

Modern approach: Provide a test to determine if a vector $v$ lies in null cone

$$\text{capacity of } v = \text{cap}_G(v) := \inf_{g \in G} \|g \cdot v\|^2.$$ 

Observation: $\text{cap}_G(v) = 0$ iff $v$ lies in null cone

Hence: Testing null cone membership is a minimization problem.

$\Rightarrow$ algorithms: [series of 3 papers in 2017 – 2019 by Bürgisser, Franks, Garg, Oliveira, Walter, Wigderson]
Invariant theory
Null cone membership testing

Classical and often hard question: Describe null cone
(essentially equivalent to finding generators for the ring of polynomial invariants)

Modern approach: Provide a test to determine if a vector $v$ lies in null cone

The **capacity** of $v$ is

$$\text{cap}_G(v) := \inf_{g \in G} \|g \cdot v\|_2^2.$$ 

Observation: $\text{cap}_G(v) = 0$ iff $v$ lies in null cone
Invariant theory
Null cone membership testing

Classical and often hard question: Describe null cone
(essentially equivalent to finding generators for the ring of polynomial invariants)

Modern approach: Provide a test to determine if a vector $v$ lies in null cone

The \textbf{capacity} of $v$ is

$$\text{cap}_G(v) := \inf_{g \in G} \|g \cdot v\|_2^2.$$ 

Observation: $\text{cap}_G(v) = 0$ iff $v$ lies in null cone

Hence: Testing null cone membership is a minimization problem.

$\iff$ algorithms: [series of 3 papers in 2017 – 2019 by Bürgisser, Franks, Garg, Oliveira, Walter, Wigderson]
Maximum likelihood estimation

Given:

- $M$: a statistical model = a set of probability distributions
- $Y = (Y_1, \ldots, Y_n)$: $n$ samples of observed data

Goal:

find a distribution in the model $M$ that best fits the empirical data $Y$

Approach:

maximize the likelihood function $L_Y(\rho) := \rho(Y_1) \cdots \rho(Y_n)$, where $\rho \in M$.

A maximum likelihood estimate (MLE) is a distribution in the model $M$ that maximizes the likelihood $L_Y$. 
Maximum likelihood estimation

Given:
- $\mathcal{M}$: a statistical model $= \text{a set of probability distributions}$
- $Y = (Y_1, \ldots, Y_n)$: $n$ samples of observed data

Goal: find a distribution in the model $\mathcal{M}$ that best fits the empirical data $Y$
Maximum likelihood estimation

**Given:**
- $\mathcal{M}$: a statistical **model** = a set of probability distributions
- $Y = (Y_1, \ldots, Y_n)$: $n$ samples of observed **data**

**Goal:** find a distribution in the model $\mathcal{M}$ that best fits the empirical data $Y$

**Approach:** maximize the **likelihood function**

$$L_Y(\rho) := \rho(Y_1) \cdots \rho(Y_n), \quad \text{where } \rho \in \mathcal{M}.$$  

A maximum likelihood estimate (MLE) is a distribution in the model $\mathcal{M}$ that maximizes the likelihood $L_Y$. 
Discrete statistical models

A probability distribution on \( m \) states is determined by its probability mass function \( \rho \), where \( \rho_j \) is the probability that the \( j \)-th state occurs.

\( \rho \) is a point in the probability simplex

\[
\Delta_{m-1} = \left\{ q \in \mathbb{R}^m \mid q_j \geq 0 \text{ and } \sum q_j = 1 \right\}.
\]

A discrete statistical model \( \mathcal{M} \) is a subset of the simplex \( \Delta_{m-1} \).
Given data is a **vector of counts** \( Y \in \mathbb{Z}_{\geq 0}^m \), where \( Y_j \) is the number of times the \( j \)-th state occurs.

The **empirical distribution** is \( S_Y = \frac{1}{n} Y \in \Delta_{m-1} \), where \( n = Y_1 + \ldots + Y_m \).
Discrete statistical models

maximum likelihood estimation

Given data is a **vector of counts** $Y \in \mathbb{Z}_{\geq 0}^m$, where $Y_j$ is the number of times the $j$-th state occurs.

The **empirical distribution** is $S_Y = \frac{1}{n} Y \in \Delta_{m-1}$, where $n = Y_1 + \ldots + Y_m$.

The **likelihood function** takes the form $L_Y(\rho) = \rho_{Y_1} \cdots \rho_{Y_m}$, where $\rho \in \mathcal{M}$.

An **MLE** is a point in model $\mathcal{M}$ that maximizes the likelihood $L_Y$ of observing $Y$. 
Log-linear models

= set of distributions whose logarithms lie in a fixed linear space.

Let $A \in \mathbb{Z}^{d \times m}$, and define

$$\mathcal{M}_A = \{ \rho \in \Delta_{m-1} \mid \log \rho \in \text{rowspan}(A) \}.$$ 

We assume that $1 := (1, \ldots, 1) \in \text{rowspan}(A)$ (i.e., uniform distribution in $\mathcal{M}_A$).
Log-linear models

= set of distributions whose logarithms lie in a fixed linear space.

Let $A \in \mathbb{Z}^{d \times m}$, and define

$$\mathcal{M}_A = \{ \rho \in \Delta_{m-1} \mid \log \rho \in \text{rowspan}(A) \}.$$ 

We assume that $1 := (1, \ldots, 1) \in \text{rowspan}(A)$ (i.e., uniform distribution in $\mathcal{M}_A$).

Matrix $A = [a_1 \mid a_2 \mid \ldots \mid a_m]$ also defines an action by the torus

- $\mathbb{G}T_d = \text{group of complex, diagonal, invertible } d \times d \text{ matrices}$
- on $\mathbb{C}^m$:
- $g \in \mathbb{G}T_d$ acts on $x \in \mathbb{C}^m$ by left multiplication with

$$\begin{bmatrix}
g^{a_1} \\
\vdots \\
g^{a_m}
ge^{a_{1j}} \ldots e^{a_{dj}}
\end{bmatrix}, \quad \text{where } g^{a_j} = g_1^{a_{1j}} \ldots g_d^{a_{dj}}.$$
Log-linear models

= set of distributions whose logarithms lie in a fixed linear space.

Let $A \in \mathbb{Z}^{d \times m}$, and define

$$\mathcal{M}_A = \{ \rho \in \Delta_{m-1} \mid \log \rho \in \text{rowspan}(A) \}.$$ 

We assume that $1 := (1, \ldots, 1) \in \text{rowspan}(A)$ (i.e., uniform distribution in $\mathcal{M}_A$).

Matrix $A = [a_1 \mid a_2 \mid \ldots \mid a_m]$ also defines an action by the torus

- $\text{GT}_d =$ group of complex, diagonal, invertible $d \times d$ matrices 
- on $\mathbb{C}^m$:
- $g \in \text{GT}_d$ acts on $x \in \mathbb{C}^m$ by left multiplication with

$$
\begin{bmatrix}
g^{a_1} \\
\vdots \\
g^{a_m}
\end{bmatrix}, \quad \text{where } g^{a_j} = g_1^{a_{1j}} \ldots g_d^{a_{dj}}.
$$

$\mathcal{M}_A$ is the orbit of the uniform distribution in $\Delta_{m-1} \cap \mathbb{R}^m_{>0}$. 
Example

\[ M_A = \{ \rho \in \Delta_{m-1} \mid \log \rho \in \text{rowspan}(A) \} \]. \quad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}

\[ g \in G_{T_2} \text{ acts on } x \in \mathbb{C}^3 \text{ by } \begin{bmatrix} g^{a_1} & g^{a_2} & g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 & g_1g_2 & g_2^2 \end{bmatrix} \].
Example

\[ \mathcal{M}_A = \{ \rho \in \Delta_{m-1} \mid \log \rho \in \text{rowspan}(A) \} \, . \quad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \]

\( g \in \text{GT}_2 \) acts on \( x \in \mathbb{C}^3 \) by

\[
\begin{bmatrix} g^{a_1} \\ g^{a_2} \\ g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 \\ g_1 g_2 \\ g_2^2 \end{bmatrix}.
\]

\[ \mathcal{M}_A = (\text{GT}_2 \cdot \frac{1}{3} \mathbb{I}) \cap \Delta_2 \cap \mathbb{R}_{>0}^3 \]

\[ = \left\{ \frac{1}{3} (g_1^2, g_1 g_2, g_2^2) \mid g_1, g_2 > 0, \ g_1^2 + g_1 g_2 + g_2^2 = 3 \right\} \]
Example

\[ M_A = \{ \rho \in \Delta_{m-1} \mid \log \rho \in \text{rowspan}(A) \} . \]

\[ A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \]

\[ g \in \text{GT}_2 \text{ acts on } x \in \mathbb{C}^3 \text{ by } \begin{bmatrix} g^{a_1} \\ g^{a_2} \\ g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 \\ g_1 g_2 \\ g_2^2 \end{bmatrix}. \]

\[ M_A = (\text{GT}_2 \cdot \frac{1}{3} \mathbb{I}) \cap \Delta_2 \cap \mathbb{R}_>^3 \]

\[ = \left\{ \frac{1}{3} (g_1^2, g_1 g_2, g_2^2) \mid g_1, g_2 > 0, \ g_1^2 + g_1 g_2 + g_2^2 = 3 \right\} \]

\[ = \{ \rho \in \mathbb{R}_>^3 \mid \rho_2^2 = \rho_1 \rho_3, \ \rho_1 + \rho_2 + \rho_3 = 1 \} \]
Example

\[ \mathcal{M}_A = \{ \rho \in \Delta_{m-1} \mid \log \rho \in \text{rowspan}(A) \} \quad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \]

\( g \in \text{GT}_2 \) acts on \( x \in \mathbb{C}^3 \) by

\[
\begin{bmatrix} \ g^{a_1} \\ \ g^{a_2} \\ \ g^{a_3} \end{bmatrix} \ = \begin{bmatrix} \ g_1^2 \\ \ g_1g_2 \\ \ g_2^2 \end{bmatrix}.
\]

\[ \mathcal{M}_A = (\text{GT}_2 \cdot \frac{1}{3} \mathbb{1}) \cap \Delta_2 \cap \mathbb{R}^3_{>0} \]

\[ = \left\{ \frac{1}{3} \left( g_1^2, g_1g_2, g_2^2 \right) \mid g_1, g_2 > 0, \ g_1^2 + g_1g_2 + g_2^2 = 3 \right\} \]

\[ = \left\{ \rho \in \mathbb{R}^3_{>0} \mid \rho_2^2 = \rho_1 \rho_3, \ \rho_1 + \rho_2 + \rho_3 = 1 \right\} \]

other examples: independence model, graphical models, hierarchical models, ...
Maximum likelihood estimation
for log-linear models

An MLE in $\mathcal{M}_A$ given data $Y$ is a point $\hat{\rho}$ in the model such that

$$A\hat{\rho} = AS_Y,$$
where $S_Y = \frac{1}{n} Y$.

The MLE is unique if it exists!
Maximum likelihood estimation

for log-linear models

An MLE in $\mathcal{M}_A$ given data $Y$ is a point $\hat{\rho}$ in the model such that

$$A\hat{\rho} = AS_Y,$$

where $S_Y = \frac{1}{n} Y$.

The MLE is unique if it exists!

Model $\mathcal{M}_A$ is not closed: MLE may not exist if $S_Y$ has zeroes. True maximizer could be on boundary of model.
Maximum likelihood estimation
for log-linear models

An MLE in $\mathcal{M}_A$ given data $Y$ is a point $\hat{\rho}$ in the model such that

$$A\hat{\rho} = AS_Y,$$

where $S_Y = \frac{1}{n}Y$.

The MLE is unique if it exists!

Model $\mathcal{M}_A$ is not closed: MLE may not exist if $S_Y$ has zeroes. True maximizer could be on boundary of model.

Polyhedral condition for MLE existence:
For $A = [a_1 \mid a_2 \mid \ldots \mid a_m] \in \mathbb{Z}^{d \times m}$, we define

$$P(A) = \text{conv} \{a_1, a_2, \ldots, a_m\} \subset \mathbb{R}^d.$$

Theorem (Eriksson, Fienberg, Rinaldo, Sullivant '06)
MLE given $Y$ exists in $\mathcal{M}_A$ iff $AS_Y$ is in relative interior of $P(A)$. 
Stability for torus actions

The action of the torus $\mathbb{G}_T^d$ given by the matrix $A \in \mathbb{Z}^{d \times m}$ is in fact well-defined on projective space $\mathbb{P}^{m-1}$. 

A linearization is a consistent action on $\mathbb{C}^{m}$, given by a character $b \in \mathbb{Z}^{d}$:

$g \in \mathbb{G}_T^d$ acts on $x \in \mathbb{C}^{m}$ by

$$
\begin{pmatrix}
g a_1 - b \\
g a_2 - b \\
\vdots \\
g a_m - b
\end{pmatrix}
$$

Polyhedral conditions for stability:

Define sub-polytopes of $P(A) = \text{conv}\{a_1, a_2, \ldots, a_m\}$ that depend on $x \in \mathbb{C}^{m}$:

$P_x(A) = \text{conv}\{a_j | j \in \text{supp}(x)\}$.

Theorem (standard, proof via Hilbert-Mumford criterion)

Consider the action of $\mathbb{G}_T^d$ given by matrix $A \in \mathbb{Z}^{d \times m}$ with linearization $b \in \mathbb{Z}^{d}$.

(a) $x$ unstable $\iff b \not\in P_x(A)$ can be scaled to 0 in the limit

(b) $x$ semistable $\iff b \in P_x(A)$ cannot be scaled to 0 in the limit

(c) $x$ polystable $\iff b \in \text{relint} P_x(A)$ closed orbit

(d) $x$ stable $\iff b \in \text{int} P_x(A)$ finite stabilizer
Stability for torus actions

The action of the torus $G T_d$ given by the matrix $A \in \mathbb{Z}^{d \times m}$ is in fact well-defined on projective space $\mathbb{P}^{m-1}$.

A linearization is a consistent action on $\mathbb{C}^m$, given by a character $b \in \mathbb{Z}^d$:

$$g \in G T_d \text{ acts on } x \in \mathbb{C}^m \text{ by } \begin{bmatrix} g^{a_1-b} & \cdots & g^{a_m-b} \end{bmatrix}.$$
Stability for torus actions

The action of the torus $\mathbb{G}T_d$ given by the matrix $A \in \mathbb{Z}^{d \times m}$ is in fact well-defined on projective space $\mathbb{P}^{m-1}$.

A **linearization** is a consistent action on $\mathbb{C}^m$, given by a character $b \in \mathbb{Z}^d$:

$$g \in \mathbb{G}T_d \text{ acts on } x \in \mathbb{C}^m \text{ by } \begin{bmatrix} g^{a_1-b} & & \cdots & & \cdots & & g^{a_m-b} \end{bmatrix}.$$

**Polyhedral conditions for stability:**

Define sub-polytopes of $P(A) = \text{conv}\{a_1, a_2, \ldots, a_m\}$ that depend on $x \in \mathbb{C}^m$:

$$P_x(A) = \text{conv}\{a_j \mid j \in \text{supp}(x)\}.$$  

**Theorem** (standard, proof via Hilbert-Mumford criterion)

Consider the action of $\mathbb{G}T_d$ given by matrix $A \in \mathbb{Z}^{d \times m}$ with linearization $b \in \mathbb{Z}^d$.

(a) $x$ unstable $\iff$ $b \notin P_x(A)$ can be scaled to 0 in the limit
(b) $x$ semistable $\iff$ $b \in P_x(A)$ cannot be scaled to 0 in the limit
(c) $x$ polystable $\iff$ $b \in \text{relint } P_x(A)$ closed orbit
(d) $x$ stable $\iff$ $b \in \text{int } P_x(A)$ finite stabilizer

$X - XVII$
Stability for torus actions

The action of the torus $\Gamma^d$ given by the matrix $A \in \mathbb{Z}^{d \times m}$ is in fact well-defined on projective space $\mathbb{P}^{m-1}$.

A linearization is a consistent action on $\mathbb{C}^m$, given by a character $b \in \mathbb{Z}^d$:

$$g \in \Gamma^d \text{ acts on } x \in \mathbb{C}^m \text{ by } \begin{bmatrix} g^{a_1-b} & & \\ & \ddots & \\ & & g^{a_m-b} \end{bmatrix}.$$  

polyhedral conditions for stability:

Define sub-polytopes of $P(A) = \text{conv}\{a_1, a_2, \ldots, a_m\}$ that depend on $x \in \mathbb{C}^m$:

$$P_x(A) = \text{conv}\{a_j \mid j \in \text{supp}(x)\}.$$  

**Theorem** (standard, proof via Hilbert-Mumford criterion)

Consider the action of $\Gamma^d$ given by matrix $A \in \mathbb{Z}^{d \times m}$ with linearization $b \in \mathbb{Z}^d$.

(a) $x$ unstable $\iff b \not\in P_x(A)$ can be scaled to 0 in the limit

(b) $x$ semistable $\iff b \in P_x(A)$ cannot be scaled to 0 in the limit

(c) $x$ polystable $\iff b \in \text{relint } P_x(A)$ closed orbit

(d) $x$ stable $\iff b \in \text{int } P_x(A)$ finite stabilizer
Combining both worlds

Theorem (Amédola, Kohn, Reichenbach, Seigal)
Consider a vector of counts $Y \in \mathbb{Z}^m$ with $n = \sum Y_j$, matrix $A \in \mathbb{Z}^{d \times m}$, and $b = AY \in \mathbb{Z}^d$. The MLE given $Y$ in $\mathcal{M}_A$ exists iff $1 \in \mathbb{C}^m$ is polystable under the action of $\text{GT}_d$ given by matrix $nA$ with linearization $b$.
Combining both worlds

**Theorem** (Amédola, Kohn, Reichenbach, Seigal)

Consider a vector of counts $Y \in \mathbb{Z}^m$ with $n = \sum Y_j$, matrix $A \in \mathbb{Z}^{d \times m}$, and $b = AY \in \mathbb{Z}^d$. The MLE given $Y$ in $\mathcal{M}_A$ exists iff $1 \in \mathbb{C}^m$ is polystable under the action of $G \mathcal{T}_d$ given by matrix $nA$ with linearization $b$.

How are the two optimal points related?

**Theorem** (cont’d)

If $x \in \mathbb{C}^m$ is a point of minimal norm in the orbit $G \mathcal{T}_d \cdot 1$, then the MLE is

$$\frac{x^{(2)}}{\|x\|^2},$$

where $x^{(2)}$ is the vector with $j$-th entry $|x_j|^2$. 
Algorithmic consequences

- Algorithms for finding MLE, e.g. iterative proportional scaling (IPS)
- Scaling algorithms to compute capacity
Algorithmic consequences

- Algorithms for finding MLE, e.g. iterative proportional scaling (IPS)
  - Maximize likelihood $\Leftrightarrow$ Minimize KL divergence

- Scaling algorithms to compute capacity
  - Minimize $\ell_2$-norm
Algorithmic consequences

- Algorithms for finding MLE, e.g. iterative proportional scaling (IPS)
- Maximize likelihood $\Leftrightarrow$ Minimize $KL$ divergence
- Model lives in $\Delta_{m-1} \cap \mathbb{R}^m_{>0}$
- Orbit lives in $\mathbb{C}^m$

$\Leftrightarrow$ Scaling algorithms to compute capacity
- Minimize $\ell_2$-norm
Algorithmic consequences

algorithms for finding MLE, e.g. iterative proportional scaling (IPS)

maximize likelihood ⇔ minimize KL divergence

model lives in $\Delta_{m-1} \cap \mathbb{R}^m_{>0}$

trivial linearization $b = 0$
(defines model and steps of IPS)

scaling algorithms to compute capacity

minimize $\ell_2$-norm

orbit lives in $\mathbb{C}^m$

linearization $b = AY$
The density function of an $m$-dimensional Gaussian with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is

$$
\rho_\Sigma(y) = \frac{1}{\sqrt{\det(2\pi \Sigma)}} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right), \quad \text{where } y \in \mathbb{R}^m.
$$

The **concentration matrix** $\Psi = \Sigma^{-1}$ is positive definite.
Gaussian statistical models

The density function of an $m$-dimensional Gaussian with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is

$$\rho_{\Sigma}(y) = \frac{1}{\sqrt{\det(2\pi \Sigma)}} \exp \left( -\frac{1}{2} y^T \Sigma^{-1} y \right), \quad \text{where } y \in \mathbb{R}^m.$$ 

The concentration matrix $\Psi = \Sigma^{-1}$ is positive definite.

A Gaussian model $\mathcal{M}$ is a set of concentration matrices, i.e. a subset of the cone of $m \times m$ positive definite matrices.
Gaussian statistical models

The density function of an $m$-dimensional Gaussian with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is

$$\rho_\Sigma(y) = \frac{1}{\sqrt{\det(2\pi \Sigma)}} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right), \quad \text{where } y \in \mathbb{R}^m.$$ 

The **concentration matrix** $\Psi = \Sigma^{-1}$ is positive definite.

A **Gaussian model** $\mathcal{M}$ is a set of concentration matrices, i.e. a subset of the cone of $m \times m$ positive definite matrices. Given data $Y = (Y_1, \ldots, Y_n)$, the likelihood is

$$L_Y(\Psi) = \rho_{\Psi^{-1}}(Y_1) \cdots \rho_{\Psi^{-1}}(Y_n), \quad \text{where } \Psi \in \mathcal{M}.$$
Gaussian statistical models

The density function of an $m$-dimensional Gaussian with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is

$$\rho_\Sigma(y) = \frac{1}{\sqrt{\det(2\pi \Sigma)}} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right), \text{ where } y \in \mathbb{R}^m.$$  

The concentration matrix $\Psi = \Sigma^{-1}$ is positive definite.

A Gaussian model $\mathcal{M}$ is a set of concentration matrices, i.e. a subset of the cone of $m \times m$ positive definite matrices. Given data $Y = (Y_1, \ldots, Y_n)$, the likelihood is

$$L_Y(\Psi) = \rho_{\Psi^{-1}}(Y_1) \cdots \rho_{\Psi^{-1}}(Y_n), \text{ where } \Psi \in \mathcal{M}.$$  

likelihood $L_Y$ can be unbounded from above  
MLE might not exist  
MLE might not be unique
Gaussian group model

The **Gaussian group model** of a group $G$ with a representation $G \xrightarrow{\varphi} \text{GL}_m$ on $\mathbb{R}^m$ is

$$
\mathcal{M}_G := \{\Psi_g = \varphi(g)^T \varphi(g) \mid g \in G\}.
$$

(depend only on image of $G$ in $\text{GL}_m$, hence may assume $G \subseteq \text{GL}_m$)
The **Gaussian group model** of a group $G$ with a representation $G \xrightarrow{\varphi} \text{GL}_m$ on $\mathbb{R}^m$ is

$$
\mathcal{M}_G := \{ \psi_g = \varphi(g)^T \varphi(g) \mid g \in G \}.
$$

(depends only on image of $G$ in $\text{GL}_m$, hence may assume $G \subseteq \text{GL}_m$)

We want to find an MLE, i.e. a maximizer of

$$
L_Y(\psi_g)
$$
Gaussian group model

The **Gaussian group model** of a group $G$ with a representation $G \xrightarrow{\varphi} \text{GL}_m$ on $\mathbb{R}^m$ is

$$\mathcal{M}_G := \left\{ \Psi_g = \varphi(g)^T \varphi(g) \mid g \in G \right\}.$$  

(depending only on image of $G$ in $\text{GL}_m$, hence may assume $G \subseteq \text{GL}_m$)

We want to find an MLE, i.e. a maximizer of

$$\log L_Y(\Psi_g) = \frac{1}{2} \left( n \log \det \Psi_g - \| g \cdot Y \|_2^2 \right) - \frac{nm}{2} \log(2\pi)$$  

for $g \in G$. 

[Diagram showing MLE and $S_Y$]
Combining both worlds

\[
\sup_{g \in G} \ell_Y(\Psi_g) = - \inf_{\tau \in \mathbb{R}_{>0}} \left\{ \tau \left( \inf_{h \in G \cap SL_m} \|h \cdot Y\|^2 \right) - nm \log \tau \right\}.
\]
Combining both worlds

Invariant theory classically over \( \mathbb{C} \) – can also define Gaussian (group) models over \( \mathbb{C} \)

**Proposition** (Améndola, Kohn, Reichenbach, Seigal)

For \( Y = (Y_1, \ldots, Y_n) \) with \( Y_i \in \mathbb{C}^m \) and a group \( G \subset \text{GL}_m(\mathbb{C}) \) closed under non-zero scalar multiples (i.e., \( g \in G, \lambda \in \mathbb{C}, \lambda \neq 0 \Rightarrow \lambda g \in G \)),

\[
\sup_{g \in G} \ell_Y(\Psi_g) = - \inf_{\tau \in \mathbb{R}_{>0}} \left( \tau \left( \inf_{h \in G \cap \text{SL}_m} \| h \cdot Y \|_2^2 \right) - nm \log \tau \right).
\]
Combining both worlds
Invariant theory classically over $\mathbb{C}$ – can also define Gaussian (group) models over $\mathbb{C}$

**Proposition** (Améndola, Kohn, Reichenbach, Seigal)
For $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{C}^m$ and a group $G \subset GL_m(\mathbb{C})$ closed under non-zero scalar multiples (i.e., $g \in G, \lambda \in \mathbb{C}, \lambda \neq 0 \Rightarrow \lambda g \in G$),

$$\sup_{g \in G} \ell_Y(\Psi_g) = -\inf_{\tau \in \mathbb{R}_{>0}} \left( \tau \left( \inf_{h \in G \cap SL_m} \|h \cdot Y\|_2^2 \right) - nm \log \tau \right).$$

If $h \cdot Y$ is a point of minimal norm in the $G \cap SL_m$-orbit of $Y$, then an MLE for the Gaussian group model $M_G$ is

$$\tau h^* h,$$ where $\tau$ is the unique value minimizing $\tau \|h \cdot Y\|_2^2 - nm \log \tau$. 
Combining both worlds
Invariant theory classically over $\mathbb{C}$ – can also define Gaussian (group) models over $\mathbb{C}$

**Proposition** (Améndola, Kohn, Reichenbach, Seigal)

For $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{C}^m$ and a group $G \subset GL_m(\mathbb{C})$ closed under non-zero scalar multiples (i.e., $g \in G, \lambda \in \mathbb{C}, \lambda \neq 0 \Rightarrow \lambda g \in G$),

$$
\sup_{g \in G} \ell_Y(\Psi_g) = - \inf_{\tau \in \mathbb{R}_{>0}} \left( \tau \left( \inf_{h \in G \cap SL_m} \|h \cdot Y\|_2^2 \right) - nm \log \tau \right).
$$

If $h \cdot Y$ is a point of minimal norm in the $G \cap SL_m$-orbit of $Y$, then an MLE for the Gaussian group model $\mathcal{M}_G$ is

$$
\tau h^* h, \text{ where } \tau \text{ is the unique value minimizing } \tau \|h \cdot Y\|_2^2 - nm \log \tau.
$$

**Theorem** (Améndola, Kohn, Reichenbach, Seigal)

Let $Y$ and $G$ as above. If $G$ is linearly reductive, ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap SL_m(\mathbb{C})$ as follows:

(a) $Y$ unstable $\iff \ell_Y$ not bounded from above
(b) $Y$ semistable $\iff \ell_Y$ bounded from above
(c) $Y$ polystable $\iff$ MLE exists
(d) $Y$ stable $\iff$ finitely many MLEs exist $\iff$ unique MLE
Combining both worlds

Real examples

Theorem

( Amendola, Kohn, Reichenbach, Seigal)

Let $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \text{GL}_m(\mathbb{R})$ be a linearly reductive group which is closed under non-zero scalar multiples.

ML estimation for $M_G$ relates to the action by $G \cap \text{SL}_m(\mathbb{R})$ as follows:

(a) $Y$ unstable $\iff$ $\ell Y$ not bounded from above

(b) $Y$ semistable $\iff$ $\ell Y$ bounded from above

(c) $Y$ polystable $\iff$ MLE exists

(d) $Y$ stable $\implies$ finitely many MLEs exist $\iff$ unique MLE

Examples: full Gaussian model, independence model, matrix normal model

Theorem

( Amendola, Kohn, Reichenbach, Seigal)

Let $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \text{GL}_m(\mathbb{R})$ be a group which is closed under non-zero scalar multiples, but not necessarily linearly reductive.

ML estimation for $M_G$ relates to the action by $G \cap \text{SL}_m(\mathbb{R})$ as follows:

(a) $Y$ unstable $\iff$ $\ell Y$ not bounded from above

(b) $Y$ semistable $\iff$ $\ell Y$ bounded from above

(c) $Y$ polystable $\implies$ MLE exists

Example: Gaussian graphical models
Combining both worlds

Real examples

**Theorem** (Améndola, Kohn, Reichenbach, Seigal)

Let $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \text{GL}_m(\mathbb{R})$ be a linearly reductive group which is closed under non-zero scalar multiples.

ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \text{SL}_m(\mathbb{R})$ as follows:

(a) $Y$ unstable $\iff \ell_Y$ not bounded from above
(b) $Y$ semistable $\iff \ell_Y$ bounded from above
(c) $Y$ polystable $\iff$ MLE exists
(d) $Y$ stable $\Rightarrow$ finitely many MLEs exist $\iff$ unique MLE

Example: Gaussian graphical models

XVI - XVII
Combining both worlds

Real examples

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let \( Y = (Y_1, \ldots, Y_n) \) with \( Y_i \in \mathbb{R}^m \), and let \( G \subset \text{GL}_m(\mathbb{R}) \) be a linearly reductive group which is closed under non-zero scalar multiples.

ML estimation for \( M_G \) relates to the action by \( G \cap \text{SL}_m(\mathbb{R}) \) as follows:

(a) \( Y \) unstable \iff \( \ell_Y \) not bounded from above
(b) \( Y \) semistable \iff \( \ell_Y \) bounded from above
(c) \( Y \) polystable \iff MLE exists
(d) \( Y \) stable \implies finitely many MLEs exist \iff unique MLE

Examples: full Gaussian model, independence model, matrix normal model
Combining both worlds

Real examples

**Theorem** (Améndola, Kohn, Reichenbach, Seigal)

Let \( Y = (Y_1, \ldots, Y_n) \) with \( Y_i \in \mathbb{R}^m \), and let \( G \subset \text{GL}_m(\mathbb{R}) \) be a linearly reductive group which is closed under non-zero scalar multiples.

ML estimation for \( \mathcal{M}_G \) relates to the action by \( G \cap \text{SL}_m(\mathbb{R}) \) as follows:

(a) \( Y \) unstable \( \iff \ell_Y \) not bounded from above
(b) \( Y \) semistable \( \iff \ell_Y \) bounded from above
(c) \( Y \) polystable \( \iff \) MLE exists
(d) \( Y \) stable \( \Rightarrow \) finitely many MLEs exist \( \iff \) unique MLE

**Examples:** full Gaussian model, independence model, matrix normal model

**Theorem** (Améndola, Kohn, Reichenbach, Seigal)

Let \( Y = (Y_1, \ldots, Y_n) \) with \( Y_i \in \mathbb{R}^m \), and let \( G \subset \text{GL}_m(\mathbb{R}) \) be a group which is closed under non-zero scalar multiples, but not necessarily linearly reductive.

ML estimation for \( \mathcal{M}_G \) relates to the action by \( G \cap \text{SL}_m^\pm(\mathbb{R}) \) as follows:

(a) \( Y \) unstable \( \iff \ell_Y \) not bounded from above
(b) \( Y \) semistable \( \iff \ell_Y \) bounded from above
(c) \( Y \) polystable \( \Rightarrow \) MLE exists
Combining both worlds

Real examples

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let \( Y = (Y_1, \ldots, Y_n) \) with \( Y_i \in \mathbb{R}^m \), and let \( G \subset \text{GL}_m(\mathbb{R}) \) be a linearly reductive group which is closed under non-zero scalar multiples.

ML estimation for \( M_G \) relates to the action by \( G \cap \text{SL}_m(\mathbb{R}) \) as follows:

(a) \( Y \) unstable \( \Leftrightarrow \) \( \ell_Y \) not bounded from above
(b) \( Y \) semistable \( \Leftrightarrow \) \( \ell_Y \) bounded from above
(c) \( Y \) polystable \( \Leftrightarrow \) MLE exists
(d) \( Y \) stable \( \Rightarrow \) finitely many MLEs exist \( \Leftrightarrow \) unique MLE

Examples: full Gaussian model, independence model, matrix normal model

Theorem (Améndola, Kohn, Reichenbach, Seigal)
Let \( Y = (Y_1, \ldots, Y_n) \) with \( Y_i \in \mathbb{R}^m \), and let \( G \subset \text{GL}_m(\mathbb{R}) \) be a group which is closed under non-zero scalar multiples, but not necessarily linearly reductive.

ML estimation for \( M_G \) relates to the action by \( G \cap \text{SL}_m^\pm(\mathbb{R}) \) as follows:

(a) \( Y \) unstable \( \Leftrightarrow \) \( \ell_Y \) not bounded from above
(b) \( Y \) semistable \( \Leftrightarrow \) \( \ell_Y \) bounded from above
(c) \( Y \) polystable \( \Rightarrow \) MLE exists

Example: Gaussian graphical models

XVI - XVII
Summary

Invariant theory
describe null cone
algorithmic null cone
membership testing
historical progression

Statistics
algorithms to find MLE
convergence analysis