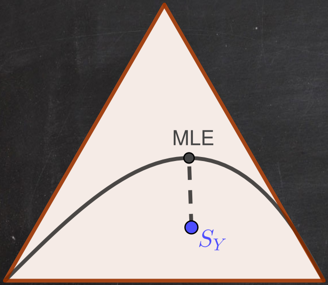
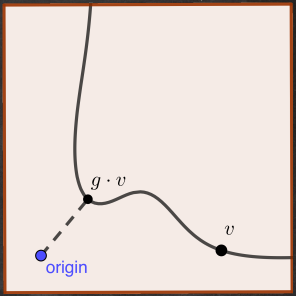
Invariant theory for maximum likelihood estimation

Statistics



Invariant theory



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**

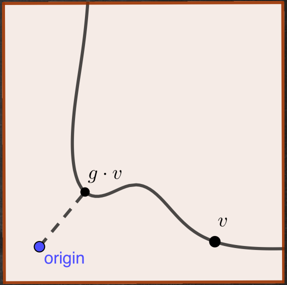
$=$ closest distance of orbit to origin

Invariant theory

Stability notions

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

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

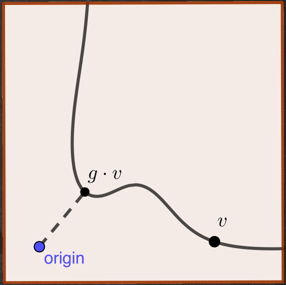


Invariant theory

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)

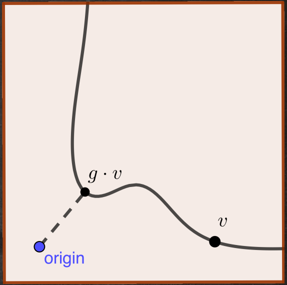
The **null cone** of the action by G is the set of unstable vectors v.

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 $0 \notin \overline{G.v}$

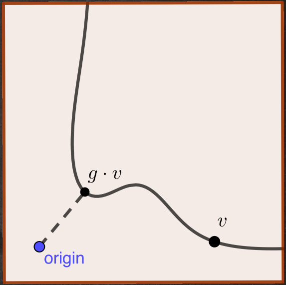
The **null cone** of the action by G is the set of unstable vectors v.

Invariant theory

Stability notions

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

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



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

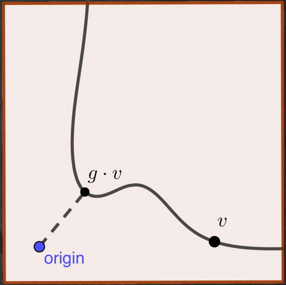
The **null cone** of the action by G is the set of unstable vectors v.

Invariant theory

Stability notions

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

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



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

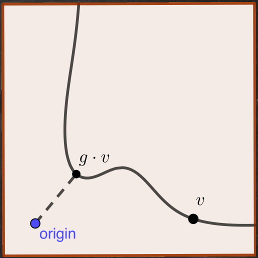
The **null cone** of the action by G is the set of unstable vectors v.

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



Invariant theory

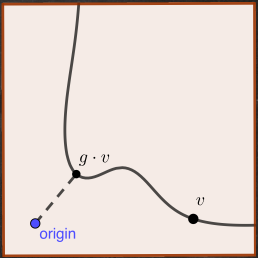
Null cone membership testing

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

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



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

Invariant theory

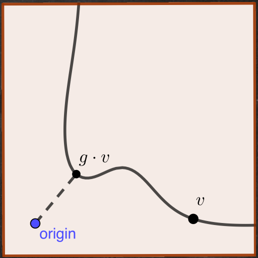
Null cone membership testing

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 **capacity** of v is

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



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

Hence: Testing null cone membership is a minimization problem.
$\rightsquigarrow$ algorithms: [series of 3 papers in 2017 – 2019 by
Bürgisser, Franks, Garg, Oliveira, Walter, Wigderson]

Moment map

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a subgroup.

Moment map

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a subgroup. Assume that

- G is Zariski closed and
- self-adjoint (i.e., $g \in G \Rightarrow g^* \in G$).

Moment map

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a subgroup. Assume that

- G is Zariski closed and
- self-adjoint (i.e., $g \in G \Rightarrow g^* \in G$).

For $v \in \mathbb{K}^m$, consider $\gamma_v : G \longrightarrow \mathbb{R}$, $g \longmapsto \|gv\|^2$.
Note: $\mathrm{cap}_G(v) = \inf_{g \in G} \gamma_v(g)$

Moment map

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a subgroup. Assume that

- G is Zariski closed and
- self-adjoint (i.e., $g \in G \Rightarrow g^* \in G$).

For $v \in \mathbb{K}^m$, consider $\gamma_v : G \longrightarrow \mathbb{R}$, $g \longmapsto \|gv\|^2$.
Note: $\mathrm{cap}_G(v) = \inf_{g \in G} \gamma_v(g)$

Its differential at the identity matrix I_m is $D_{I_m}\gamma_v : T_{I_m}G \longrightarrow \mathbb{R}$, $\dot{g} \longmapsto 2\mathrm{Re}[\mathrm{tr}(\dot{g}vv^*)]$.

Moment map

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a subgroup. Assume that

- G is Zariski closed and
- self-adjoint (i.e., $g \in G \Rightarrow g^* \in G$).

For $v \in \mathbb{K}^m$, consider $\gamma_v : G \longrightarrow \mathbb{R}$, $g \longmapsto \|gv\|^2$.
Note: $\mathrm{cap}_G(v) = \inf_{g \in G} \gamma_v(g)$

Its differential at the identity matrix I_m is $D_{I_m}\gamma_v : T_{I_m}G \longrightarrow \mathbb{R}$, $\dot{g} \longmapsto 2\mathrm{Re}[\mathrm{tr}(\dot{g}vv^*)]$.

The **moment map** assigns this differential to each vector v:

$$\mu : \mathbb{K}^m \longrightarrow \mathrm{Hom}_{\mathbb{R}}(T_{I_m}G, \mathbb{R}),\ v \longmapsto D_{I_m}\gamma_v.$$

Moment map

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a subgroup. Assume that

- G is Zariski closed and
- self-adjoint (i.e., $g \in G \Rightarrow g^* \in G$).

For $v \in \mathbb{K}^m$, consider $\gamma_v : G \longrightarrow \mathbb{R}$, $g \longmapsto \|gv\|^2$.
Note: $\mathrm{cap}_G(v) = \inf_{g \in G} \gamma_v(g)$

Its differential at the identity matrix I_m is $D_{I_m}\gamma_v : T_{I_m}G \longrightarrow \mathbb{R}$, $\dot{g} \longmapsto 2\mathrm{Re}[\mathrm{tr}(\dot{g}vv^*)]$.

The **moment map** assigns this differential to each vector v:

$$\mu : \mathbb{K}^m \longrightarrow \mathrm{Hom}_{\mathbb{R}}(T_{I_m}G, \mathbb{R}), \ v \longmapsto D_{I_m}\gamma_v .$$

Note:
$\mu(v) = 0 \Leftrightarrow I_m$ is a critical point of γ_v

Moment map

Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a subgroup. Assume that

- G is Zariski closed and
- self-adjoint (i.e., $g \in G \Rightarrow g^* \in G$).

For $v \in \mathbb{K}^m$, consider $\gamma_v : G \longrightarrow \mathbb{R}$, $g \longmapsto \|gv\|^2$.
Note: $\mathrm{cap}_G(v) = \inf_{g \in G} \gamma_v(g)$

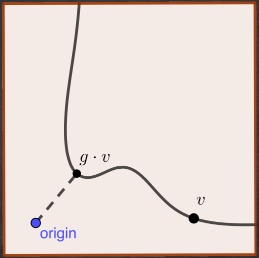
Its differential at the identity matrix I_m is $D_{I_m}\gamma_v : T_{I_m}G \longrightarrow \mathbb{R}$, $\dot{g} \longmapsto 2\mathrm{Re}[\mathrm{tr}(\dot{g}vv^*)]$.

The **moment map** assigns this differential to each vector v:

$$\mu : \mathbb{K}^m \longrightarrow \mathrm{Hom}_{\mathbb{R}}(T_{I_m}G, \mathbb{R}),\ v \longmapsto D_{I_m}\gamma_v.$$

Note:
$\mu(v) = 0 \Leftrightarrow I_m$ is a critical point of γ_v
$\Leftrightarrow v$ is a critical point of the norm minimization problem along its orbit.

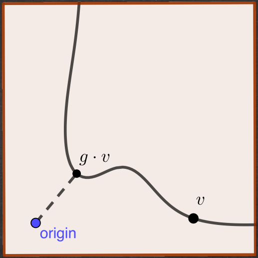
Kempf-Ness theorem



Theorem (Kempf, Ness '79 over $\mathbb{C}$ / Richardson, Slodowy '90 over $\mathbb{R}$)
Let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a Zariski closed, self-adjoint subgroup with moment map μ.
For $v \in \mathbb{K}^m$, we have:

(a) v is of minimal norm in its orbit $\Leftrightarrow \mu(v) = 0$.

Kempf-Ness theorem

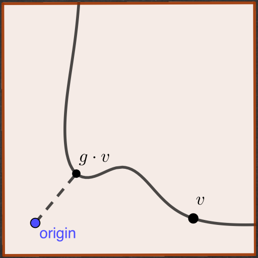


Theorem (Kempf, Ness '79 over $\mathbb{C}$ / Richardson, Slodowy '90 over $\mathbb{R}$)
Let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a Zariski closed, self-adjoint subgroup with moment map μ. For $v \in \mathbb{K}^m$, we have:

(a) v is of minimal norm in its orbit $\Leftrightarrow \mu(v) = 0$.

(b) If the orbit $G \cdot v$ is closed, there exists some $w \in G \cdot v$ with $\mu(w) = 0$.

Kempf-Ness theorem



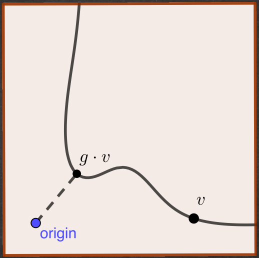
Theorem (Kempf, Ness '79 over $\mathbb{C}$ / Richardson, Slodowy '90 over $\mathbb{R}$)
Let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a Zariski closed, self-adjoint subgroup with moment map μ.
For $v \in \mathbb{K}^m$, we have:

(a) v is of minimal norm in its orbit $\Leftrightarrow \mu(v) = 0$.

(b) If the orbit $G \cdot v$ is closed, there exists some $w \in G \cdot v$ with $\mu(w) = 0$.

(c) If $\mu(v) = 0$, the orbit $G \cdot v$ is closed.

Kempf-Ness theorem



Theorem (Kempf, Ness '79 over $\mathbb{C}$ / Richardson, Slodowy '90 over $\mathbb{R}$)
Let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a Zariski closed, self-adjoint subgroup with moment map μ.
For $v \in \mathbb{K}^m$, we have:

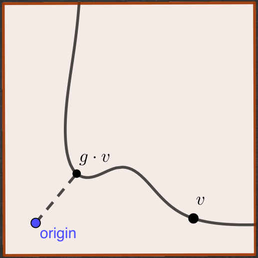
(a) v is of minimal norm in its orbit $\Leftrightarrow \mu(v) = 0$.

(b) If the orbit $G \cdot v$ is closed, there exists some $w \in G \cdot v$ with $\mu(w) = 0$.

(c) If $\mu(v) = 0$, the orbit $G \cdot v$ is closed.

(d) v is polystable $\Leftrightarrow \exists 0 \neq w \in G \cdot v : \mu(w) = 0$.

Kempf-Ness theorem



Theorem (Kempf, Ness '79 over $\mathbb{C}$ / Richardson, Slodowy '90 over $\mathbb{R}$)
Let $G \subset \mathrm{GL}_m(\mathbb{K})$ be a Zariski closed, self-adjoint subgroup with moment map μ.
For $v \in \mathbb{K}^m$, we have:

(a) v is of minimal norm in its orbit $\Leftrightarrow$ $\mu(v) = 0$.

(b) If the orbit $G \cdot v$ is closed, there exists some $w \in G \cdot v$ with $\mu(w) = 0$.

(c) If $\mu(v) = 0$, the orbit $G \cdot v$ is closed.

(d) v is polystable $\Leftrightarrow \exists 0 \neq w \in G \cdot v : \mu(w) = 0$.

(e) v is semistable $\Leftrightarrow \exists 0 \neq w \in \overline{G \cdot v} : \mu(w) = 0$.

Maximum likelihood estimation

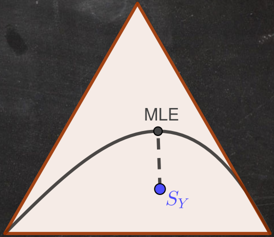
Given:

- $\mathscr{M}$: 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 $\mathscr{M}$ that best fits the empirical data Y

Approach: maximize the **likelihood function**

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



A **maximum likelihood estimate (MLE)** is a distribution in the model $\mathscr{M}$ that maximizes the likelihood L_Y.

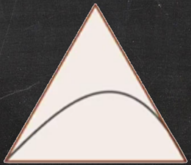
Discrete statistical models

A probability distribution on m states is determined by is **probability mass function** ρ, where ρ_j is the probability that the j-th state occurs.

ρ 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** $\mathscr{M}$ is a subset of the simplex Δ_{m-1}.



Discrete statistical models

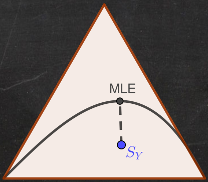
maximum likelihood estimation

Given data is a **vector of counts** $Y \in \mathbb{Z}^m_{\geq 0}$,
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 $\quad L_Y(\rho) = \rho_1^{Y_1} \cdots \rho_m^{Y_m}, \quad$ where $\rho \in \mathscr{M}$.

An **MLE** is a point in model $\mathscr{M}$ 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

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

We assume that $\mathbb{1} := (1, \dots, 1) \in \mathrm{rowspan}(A)$ (i.e., uniform distribution in $\mathscr{M}_A$).

Log-linear models

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

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

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

We assume that $\mathbb{1} := (1,\dots,1) \in \mathrm{rowspan}(A)$ (i.e., uniform distribution in $\mathscr{M}_A$).

Matrix $A = [a_1 \mid a_2 \mid \dots \mid a_m]$ also defines an **action by the torus** $(\mathbb{C}^\times)^d$ on $\mathbb{C}^m$:

$g \in (\mathbb{C}^\times)^d$ acts on $x \in \mathbb{C}^m$ by left multiplication with

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

Log-linear models

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

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

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

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

Matrix $A = [a_1 \mid a_2 \mid \ldots \mid a_m]$ also defines an **action by the torus** $(\mathbb{C}^\times)^d$ on $\mathbb{C}^m$:

$g \in (\mathbb{C}^\times)^d$ acts on $x \in \mathbb{C}^m$ by left multiplication with

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

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

Example

$\mathscr{M}_A = \{\rho \in \Delta_{m-1} \mid \log\rho \in \mathrm{rowspan}(A)\}$. $\quad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$

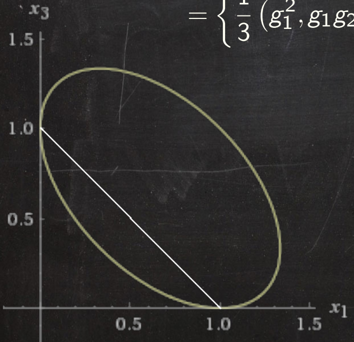
$g \in (\mathbb{C}^\times)^2$ acts on $x \in \mathbb{C}^3$ by $\begin{bmatrix} g^{a_1} & & \\ & g^{a_2} & \\ & & g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 & & \\ & g_1 g_2 & \\ & & g_2^2 \end{bmatrix}$.

Example

$$\mathscr{M}_A = \{\rho \in \Delta_{m-1} \mid \log\rho \in \mathrm{rowspan}(A)\}. \qquad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

$g \in (\mathbb{C}^\times)^2$ acts on $x \in \mathbb{C}^3$ by $\begin{bmatrix} g^{a_1} & & \\ & g^{a_2} & \\ & & g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 & & \\ & g_1 g_2 & \\ & & g_2^2 \end{bmatrix}$.

$$\begin{aligned} \mathscr{M}_A &= ((\mathbb{C}^\times)^2 \cdot \frac{1}{3}\mathbb{1}) \cap \Delta_2 \cap \mathbb{R}^3_{>0} \\ &= \left\{ \frac{1}{3}\left(g_1^2, g_1 g_2, g_2^2\right) \mid g_1, g_2 > 0,\ g_1^2 + g_1 g_2 + g_2^2 = 3 \right\} \end{aligned}$$



Example

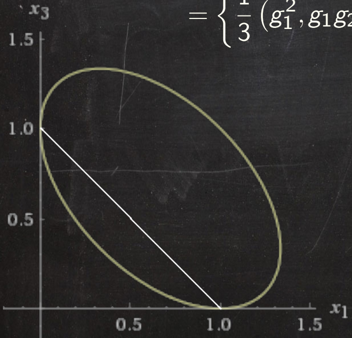
$$\mathscr{M}_A = \{\rho \in \Delta_{m-1} \mid \log \rho \in \mathrm{rowspan}(A)\}. \qquad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

$g \in (\mathbb{C}^\times)^2$ acts on $x \in \mathbb{C}^3$ by $\begin{bmatrix} g^{a_1} & & \\ & g^{a_2} & \\ & & g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 & & \\ & g_1 g_2 & \\ & & g_2^2 \end{bmatrix}.$

$$\mathscr{M}_A = ((\mathbb{C}^\times)^2 \cdot \frac{1}{3}\mathbb{1}) \cap \Delta_2 \cap \mathbb{R}^3_{>0}$$

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

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



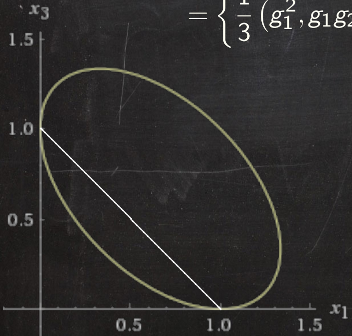
Example

$$\mathscr{M}_A = \{\rho \in \Delta_{m-1} \mid \log \rho \in \mathrm{rowspan}(A)\}. \qquad A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

$g \in (\mathbb{C}^\times)^2$ acts on $x \in \mathbb{C}^3$ by $\begin{bmatrix} g^{a_1} & & \\ & g^{a_2} & \\ & & g^{a_3} \end{bmatrix} = \begin{bmatrix} g_1^2 & & \\ & g_1 g_2 & \\ & & g_2^2 \end{bmatrix}.$

$$\mathscr{M}_A = ((\mathbb{C}^\times)^2 \cdot \frac{1}{3}\mathbb{1}) \cap \Delta_2 \cap \mathbb{R}^3_{>0}$$

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



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

other examples: independence model, graphical models, hierarchical models, . . .

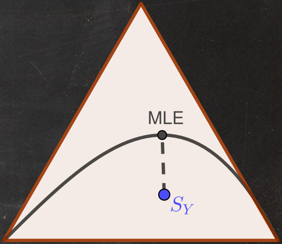
Maximum likelihood estimation

for log-linear models

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

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

The MLE is unique **if it exists**!



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

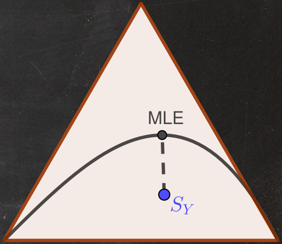
Maximum likelihood estimation

for log-linear models

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

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

The MLE is unique **if it exists**!



Model $\mathscr{M}_A$ 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

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

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

Stability for torus actions

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

Stability for torus actions

The action of the torus GT_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 $b \in \mathbb{Z}^d$:

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

Stability for torus actions

The action of the torus GT_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 $b \in \mathbb{Z}^d$:

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

polyhedral conditions for stability:

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

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

Theorem (standard, proof via Hilbert-Mumford criterion)

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

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

Stability for torus actions

The action of the torus GT_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 $b \in \mathbb{Z}^d$:

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

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

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

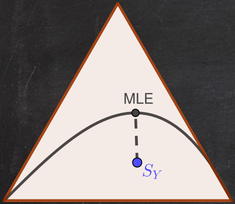
Theorem (standard, proof via Hilbert-Mumford criterion)
Consider the action of GT_d given by matrix $A \in \mathbb{Z}^{d\times m}$ with linearization $b \in \mathbb{Z}^d$.

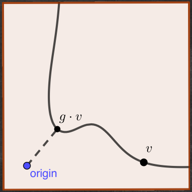
(a)	x unstable	$\Leftrightarrow$	$b \notin P_x(A)$	can be scaled to 0 in the limit
(b)	x semistable	$\Leftrightarrow$	$b \in P_x(A)$	cannot be scaled to 0 in the limit
(c)	x **polystable**	$\Leftrightarrow$	$b \in \mathrm{relint}\, P_x(A)$	closed orbit
(d)	x stable	$\Leftrightarrow$	$b \in \mathrm{int}\,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

Combining both worlds

Theorem
Let $A = [a_1|\dots|a_m] \in \mathbb{Z}^{d\times m}$ and $Y \in \mathbb{Z}^m$ be a vector of counts with $n = \sum Y_j$.

MLE given Y exists in $\mathscr{M}_A$ $\Leftrightarrow$ $\mathbb{1} \in \mathbb{C}^m$ is polystable under the action of $(\mathbb{C}^\times)^d$ given by the matrix $[na_1 - AY|\dots|na_m - AY]$



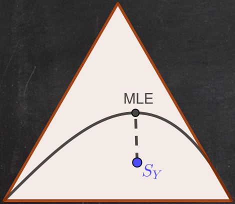


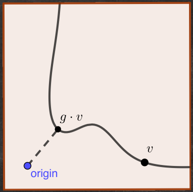
Combining both worlds

Theorem

Let $A = [a_1|\dots|a_m] \in \mathbb{Z}^{d\times m}$ and $Y \in \mathbb{Z}^m$ be a vector of counts with $n = \sum Y_j$.

MLE given Y exists in $\mathscr{M}_A$ $\Leftrightarrow$ $\mathbb{1} \in \mathbb{C}^m$ is polystable under the action of $(\mathbb{C}^\times)^d$ given by the matrix $[na_1 - AY|\dots|na_m - AY]$



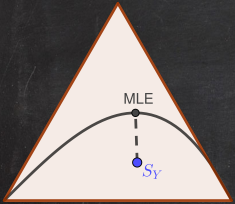


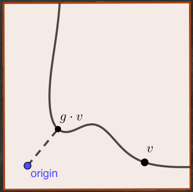
attains its maximum $\Leftrightarrow$ attains its minimum

Combining both worlds

Theorem
Let $A = [a_1 | \ldots | a_m] \in \mathbb{Z}^{d \times m}$ and $Y \in \mathbb{Z}^m$ be a vector of counts with $n = \sum Y_j$.

MLE given Y exists in $\mathscr{M}_A$ $\Leftrightarrow$ $\mathbb{1} \in \mathbb{C}^m$ is polystable under the action of $(\mathbb{C}^\times)^d$ given by the matrix $[na_1 - AY | \ldots | na_m - AY]$





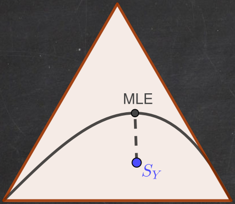
attains its maximum $\Leftrightarrow$ attains its minimum

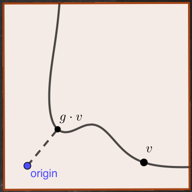
How are the two optimal points related?

Theorem (cont'd)
If $x \in \mathbb{C}^m$ is a point of minimal norm in the orbit $(\mathbb{C}^\times)^d \cdot \mathbb{1}$, then the MLE is

$$\frac{x^{(2)}}{\|x\|^2}, \quad \text{where } x^{(2)} \text{ is the vector with } j\text{-th entry } |x_j|^2.$$

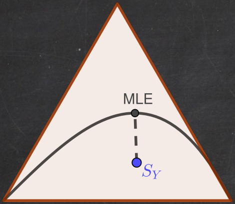
Algorithmic consequences

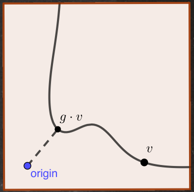




algorithms for finding MLE, e.g. iterative proportional scaling (IPS) $\leftrightarrow$ scaling algorithms to compute capacity

Algorithmic consequences



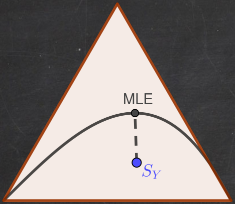


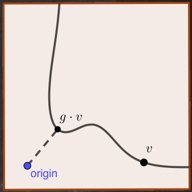
algorithms for finding MLE, e.g. iterative proportional scaling (IPS) $\leftrightarrow$ scaling algorithms to compute capacity

maximize likelihood $\Leftrightarrow$ minimize KL divergence

minimize ℓ_2-norm

Algorithmic consequences





algorithms for finding MLE, e.g. iterative proportional scaling (IPS) $\leftrightarrow$ scaling algorithms to compute capacity

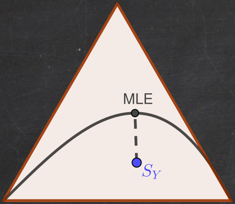
maximize likelihood $\Leftrightarrow$ minimize KL divergence

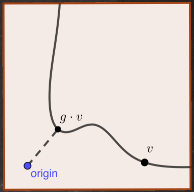
minimize ℓ_2-norm

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

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

Algorithmic consequences





algorithms for finding MLE, e.g. iterative proportional scaling (IPS) $\leftrightarrow$ scaling algorithms to compute capacity

maximize likelihood $\Leftrightarrow$ minimize KL divergence

minimize ℓ_2-norm

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

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

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

linearization $b = AY$

Gaussian statistical models

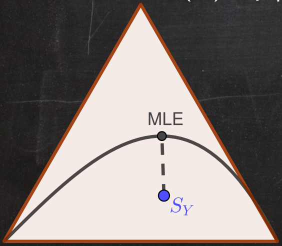
The density function of an m-dimensional Gaussian with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is

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

The **concentration matrix** $\Psi = \Sigma^{-1}$ is symmetric and positive definite.
A **Gaussian model** $\mathscr{M}$ is a set of concentration matrices, i.e. a subset of the cone of $m \times m$ symmetric positive definite matrices.

Given data $Y = (Y_1, \ldots, Y_n)$, the likelihood is

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



likelihood L_Y can be unbounded from above

MLE might not exist

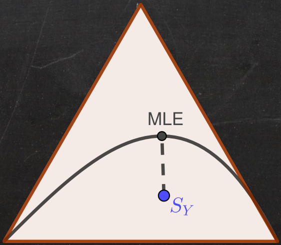
MLE might not be unique

Gaussian group model

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

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

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



Gaussian group model

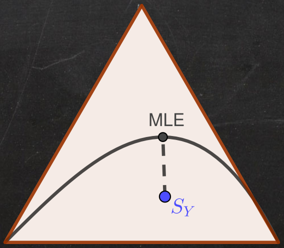
The **Gaussian group model** of a group G with a representation $G \xrightarrow{\varphi} \mathrm{GL}_m$ on $\mathbb{R}^m$ is

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

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

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

$$L_Y(\Psi_g)$$



Gaussian group model

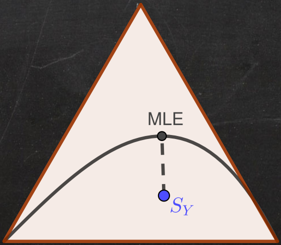
The **Gaussian group model** of a group G with a representation $G \xrightarrow{\varphi} \mathrm{GL}_m$ on $\mathbb{R}^m$ is

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

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

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

$$\log L_Y(\Psi_g) = \frac{1}{2} \underbrace{\left(n \log\det \Psi_g - \|g \cdot Y\|_2^2 \right)}_{\ell_Y(\Psi_g)} - \frac{nm}{2} \log(2\pi) \quad \text{for } g \in G.$$



Combining both worlds

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

Combining both worlds

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

For a group $G \subset \mathrm{GL}_m(\mathbb{C})$, define $\mathscr{M}_G := \{g^* g \mid g \in G\}$.

Proposition

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

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

Combining both worlds

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

For a group $G \subset \mathrm{GL}_m(\mathbb{C})$, define $\mathscr{M}_G := \{g^* g \mid g \in G\}$.

Proposition

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

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

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

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

Combining both worlds

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

For a group $G \subset \mathrm{GL}_m(\mathbb{C})$, define $\mathscr{M}_G := \{g^* g \mid g \in G\}$.

Proposition

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

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

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

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

All MLEs, if they exist, are of this form.

Combining both worlds

Theorem
Let Y and G as above.

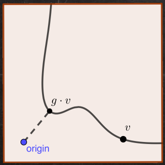
Combining both worlds

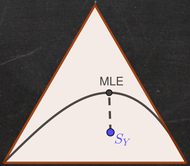
Theorem
Let Y and G as above.
If G is Zariski closed and self-adjoint (i.e., $g \in G \Rightarrow g^* \in G$),

Combining both worlds

Theorem
Let Y and G as above.
If G is Zariski closed and self-adjoint (i.e., $g \in G \Rightarrow g^* \in G$),
ML estimation for $\mathscr{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{C})$ as follows:

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





Combining both worlds

Real examples

Combining both worlds

Real examples

Theorem
Let $Y = (Y_1, \dots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a Zariski closed, self-adjoint group that is closed under non-zero scalar multiples.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{R})$ as follows:

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

Combining both worlds

Real examples

Theorem
Let $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a Zariski closed, self-adjoint group that is closed under non-zero scalar multiples.
ML estimation for $\mathscr{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{R})$ as follows:

(a)	Y unstable	$\Leftrightarrow$	ℓ_Y not bounded from above		
(b)	Y semistable	$\Leftrightarrow$	ℓ_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**

Combining both worlds

Real examples

Theorem

Let $Y = (Y_1, \dots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a Zariski closed, self-adjoint group that is closed under non-zero scalar multiples.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{R})$ as follows:

(a)	Y unstable	$\Leftrightarrow$	ℓ_Y not bounded from above		
(b)	Y semistable	$\Leftrightarrow$	ℓ_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

Let $Y = (Y_1, \dots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a group that is closed under non-zero scalar multiples, but not necessarily Zariski closed and self-adjoint.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m^{\pm}(\mathbb{R})$ as follows:

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

Combining both worlds

Real examples

Theorem

Let $Y = (Y_1, \dots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a Zariski closed, self-adjoint group that is closed under non-zero scalar multiples.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m(\mathbb{R})$ as follows:

(a) Y unstable $\Leftrightarrow$ ℓ_Y not bounded from above
(b) Y semistable $\Leftrightarrow$ ℓ_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

Let $Y = (Y_1, \dots, Y_n)$ with $Y_i \in \mathbb{R}^m$, and let $G \subset \mathrm{GL}_m(\mathbb{R})$ be a group that is closed under non-zero scalar multiples, but not necessarily Zariski closed and self-adjoint.
ML estimation for $\mathcal{M}_G$ relates to the action by $G \cap \mathrm{SL}_m^{\pm}(\mathbb{R})$ as follows:

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

Example: **Gaussian graphical models defined by transitive DAGs**

Combining both worlds

Real examples

Proposition

For $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$ and a group $G \subset \mathrm{GL}_m(\mathbb{R})$ closed under non-zero scalar multiples,

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

If $h \cdot Y$ is a point of minimal norm in the $G \cap \mathrm{SL}_m^{\pm}$-orbit of Y, then an MLE for the Gaussian group model $\mathscr{M}_G$ is

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

All MLEs, if they exist, are of this form.

Combining both worlds

Real examples

Proposition

For $Y = (Y_1, \ldots, Y_n)$ with $Y_i \in \mathbb{R}^m$ and a group $G \subset \mathrm{GL}_m(\mathbb{R})$ closed under non-zero scalar multiples,

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

If $h \cdot Y$ is a point of minimal norm in the $G \cap \mathrm{SL}_m^{\pm}$-orbit of Y, then an MLE for the Gaussian group model $\mathscr{M}_G$ is

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

All MLEs, if they exist, are of this form.

Remark

If G contains an orthogonal matrix of determinant -1, then we can work with SL_m instead of $\mathrm{SL}_m^{\pm}$.

Gaussian graphical models

Directed acyclic graphs

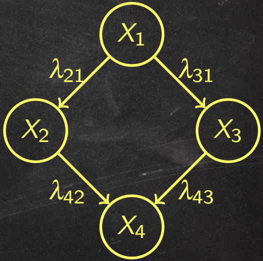
Important family of statistical models that represent interaction structures between several random variables.

- Consider a directed acyclic graph (DAG) $\mathscr{G}$ with m nodes.
- Each node j represents a random variable X_j (e.g., Gaussian).
- Each edge $j \to i$ encodes (conditional) dependence: X_j 'causes' X_i.
- The parents of i are $\mathrm{pa}(i) = \{j \mid j \to i\}$.

The model is defined by the recursive linear equation:

$$X_i = \sum_{j \in \mathrm{pa}(i)} \lambda_{ij} X_j + \varepsilon_i$$

where λ_{ij} is the edge coefficient and ε_i is Gaussian error.



It can be written as $\boldsymbol{X} = \boldsymbol{\Lambda X} + \boldsymbol{\varepsilon}$ where $\Lambda \in \mathbb{R}^{m \times m}$ satisfies $\lambda_{ij} = 0$ for $j \not\to i$ in $\mathscr{G}$ and $\varepsilon \sim N(0, \Omega)$ with Ω diagonal, positive definite.

Gaussian graphical models

coming from groups

From $X = \Lambda X + \varepsilon$, we rewrite

$$X = (I - \Lambda)^{-1}\varepsilon$$

so that $X \sim N(0, \Sigma)$ with

$$\Sigma = (I - \Lambda)^{-1}\Omega(I - \Lambda)^{-T} \quad \& \quad \Psi = (I - \Lambda)^{T}\Omega^{-1}(I - \Lambda).$$

The **Gaussian graphical model** $\mathscr{M}_{\vec{\mathscr{G}}}$ consists of concentration matrices Ψ of this form.

Gaussian graphical models

coming from groups

From $X = \Lambda X + \varepsilon$, we rewrite

$$X = (I - \Lambda)^{-1}\varepsilon$$

so that $X \sim N(0, \Sigma)$ with

$$\Sigma = (I-\Lambda)^{-1}\Omega(I-\Lambda)^{-T} \quad \& \quad \Psi = (I-\Lambda)^{T}\Omega^{-1}(I-\Lambda).$$

The **Gaussian graphical model** $\mathscr{M}_{\mathscr{G}}^{\rightarrow}$ consists of concentration matrices Ψ of this form. Consider the set

$$G(\mathscr{G}) = \{g \in \mathrm{GL}_m \mid g_{ij} = 0 \text{ for } i \neq j \text{ with } j \not\to i \text{ in } \mathscr{G}\}.$$

Gaussian graphical models

coming from groups

From $X = \Lambda X + \varepsilon$, we rewrite

$$X = (I - \Lambda)^{-1}\varepsilon$$

so that $X \sim N(0, \Sigma)$ with

$$\Sigma = (I - \Lambda)^{-1}\Omega(I - \Lambda)^{-T} \quad \& \quad \Psi = (I - \Lambda)^{T}\Omega^{-1}(I - \Lambda).$$

The **Gaussian graphical model** $\mathscr{M}_{\mathscr{G}}^{\rightarrow}$ consists of concentration matrices Ψ of this form. Consider the set

$$G(\mathscr{G}) = \{g \in \mathrm{GL}_m \mid g_{ij} = 0 \text{ for } i \neq j \text{ with } j \not\to i \text{ in } \mathscr{G}\}.$$

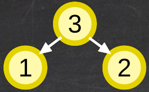
Proposition

The set of matrices $G(\mathscr{G})$ is a group if and only if $\mathscr{G}$ is a **transitive** directed acyclic graph (TDAG), i.e., $k \to j$ and $j \to i$ in $\mathscr{G}$ imply $k \to i$. In this case,

$$\mathscr{M}_{\mathscr{G}}^{\rightarrow} = \mathscr{M}_{G(\mathscr{G})}.$$

TDAG group models

Example

Let $\mathscr{G}$ be the TDAG 

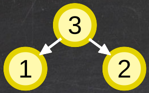
.

The corresponding group $G(\mathscr{G}) \subseteq \mathrm{GL}_3$ consists of invertible matrices g of the form

$$g = \begin{bmatrix} * & 0 & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}.$$

TDAG group models

Example

Let $\mathscr{G}$ be the TDAG 

.

The corresponding group $G(\mathscr{G}) \subseteq \mathrm{GL}_3$ consists of invertible matrices g of the form

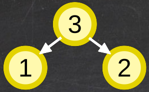
$$g = \begin{bmatrix} * & 0 & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}.$$

The Gaussian graphical model $\mathscr{M}_{\vec{\mathscr{G}}}$ is a 5-dimensional linear subspace of the cone of symmetric positive definite 3×3 matrices:

$$\mathscr{M}_{\vec{\mathscr{G}}} = \{g^T g \mid g \in G(\mathscr{G})\} = \{\Psi \in \mathrm{PD}_3 \mid \psi_{12} = \psi_{21} = 0\}.$$

TDAG group models

Example

Let $\mathscr{G}$ be the TDAG 

.

The corresponding group $G(\mathscr{G}) \subseteq \mathrm{GL}_3$ consists of invertible matrices g of the form

$$g = \begin{bmatrix} * & 0 & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}.$$

The Gaussian graphical model $\mathscr{M}_{\vec{\mathscr{G}}}$ is a 5-dimensional linear subspace of the cone of symmetric positive definite 3×3 matrices:

$$\mathscr{M}_{\vec{\mathscr{G}}} = \{g^T g \mid g \in G(\mathscr{G})\} = \{\Psi \in \mathrm{PD}_3 \mid \psi_{12} = \psi_{21} = 0\}.$$

Note that $G(\mathscr{G})$ is not self-adjoint!

MLE existence

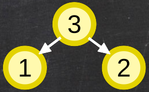
Theorem
Let $Y \in \mathbb{R}^{m \times n}$ be a tuple of n samples. If some row of Y corresponding to vertex i is in the linear span of the rows corresponding to the parents of i,

- then Y is unstable under the action by $G(\mathscr{G}) \cap \mathrm{SL}_m$,
  i.e. the likelihood is unbounded;
- otherwise, Y is polystable, i.e. an MLE exists.

MLE existence

Theorem
Let $Y \in \mathbb{R}^{m\times n}$ be a tuple of n samples. If some row of Y corresponding to vertex i is in the linear span of the rows corresponding to the parents of i,

- then Y is unstable under the action by $G(\mathscr{G}) \cap \mathrm{SL}_m$,
  i.e. the likelihood is unbounded;
- otherwise, Y is polystable, i.e. an MLE exists.

Example Let $n = 2$ in 

and consider three different pairs of samples:

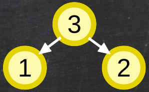
$$Y^1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Y^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 4 \end{pmatrix}, \quad Y^3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 3 & 2 \end{pmatrix}.$$

Using the theorem, we see that

MLE existence

Theorem
Let $Y \in \mathbb{R}^{m\times n}$ be a tuple of n samples. If some row of Y corresponding to vertex i is in the linear span of the rows corresponding to the parents of i,

- then Y is unstable under the action by $G(\mathscr{G}) \cap \mathrm{SL}_m$, i.e. the likelihood is unbounded;
- otherwise, Y is polystable, i.e. an MLE exists.

Example Let $n = 2$ in 

and consider three different pairs of samples:

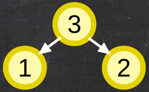
$$Y^1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Y^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 4 \end{pmatrix}, \quad Y^3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 3 & 2 \end{pmatrix}.$$

Using the theorem, we see that Y^1 and Y^2 are unstable and Y^3 is polystable.

MLE existence

Theorem
Let $Y \in \mathbb{R}^{m\times n}$ be a tuple of n samples. If some row of Y corresponding to vertex i is in the linear span of the rows corresponding to the parents of i,

- then Y is unstable under the action by $G(\mathscr{G}) \cap \mathrm{SL}_m$, i.e. the likelihood is unbounded;
- otherwise, Y is polystable, i.e. an MLE exists.

Example Let $n = 2$ in 

and consider three different pairs of samples:

$$Y^1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Y^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 4 \end{pmatrix}, \quad Y^3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 3 & 2 \end{pmatrix}.$$

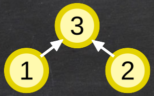
Using the theorem, we see that Y^1 and Y^2 are unstable and Y^3 is polystable.
The null cone has two components: $V(y_{11}y_{32} - y_{12}y_{31}) \cup V(y_{21}y_{32} - y_{22}y_{31})$.

Null cones of TDAGs

Corollary Let $\mathscr{G}$ be a TDAG with m nodes and n samples.
Each irreducible component of the Zariski closure of the null cone under the action of $G(\mathscr{G}) \cap \mathrm{SL}_m$ on $\mathbb{R}^{m \times n}$ is defined by the maximal minors of the submatrix whose rows are a childless node and its parents.

Null cones of TDAGs

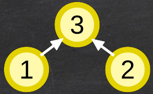
Corollary Let $\mathscr{G}$ be a TDAG with m nodes and n samples.
Each irreducible component of the Zariski closure of the null cone under the action of $G(\mathscr{G}) \cap \mathrm{SL}_m$ on $\mathbb{R}^{m\times n}$ is defined by the maximal minors of the submatrix whose rows are a childless node and its parents.

Example
Let $\mathscr{G}$ be the TDAG .

- The null cone is **not** Zariski closed for $n \geq 2$.
  Its Zariski closure is the variety of matrices of rank at most two.

Null cones of TDAGs

Corollary Let $\mathscr{G}$ be a TDAG with m nodes and n samples.
Each irreducible component of the Zariski closure of the null cone under the action of $G(\mathscr{G}) \cap \mathrm{SL}_m$ on $\mathbb{R}^{m\times n}$ is defined by the maximal minors of the submatrix whose rows are a childless node and its parents.

Example
Let $\mathscr{G}$ be the TDAG 

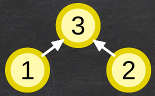
.

- The null cone is **not** Zariski closed for $n \geq 2$.
  Its Zariski closure is the variety of matrices of rank at most two.
- For $n = 2$, Y is not in the null cone but in its Zariski closure ($= \mathbb{R}^{3\times 2}$):

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

Null cones of TDAGs

Corollary Let $\mathscr{G}$ be a TDAG with m nodes and n samples.
Each irreducible component of the Zariski closure of the null cone under the action of $G(\mathscr{G}) \cap \mathrm{SL}_m$ on $\mathbb{R}^{m\times n}$ is defined by the maximal minors of the submatrix whose rows are a childless node and its parents.

Example
Let $\mathscr{G}$ be the TDAG 

.

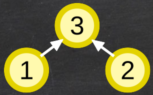
- The null cone is **not** Zariski closed for $n \geq 2$.
  Its Zariski closure is the variety of matrices of rank at most two.
- For $n = 2$, Y is not in the null cone but in its Zariski closure ($= \mathbb{R}^{3\times 2}$):

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

Hence, an MLE given Y exists.

Null cones of TDAGs

Corollary Let $\mathscr{G}$ be a TDAG with m nodes and n samples.
Each irreducible component of the Zariski closure of the null cone under the action of $G(\mathscr{G}) \cap \mathrm{SL}_m$ on $\mathbb{R}^{m\times n}$ is defined by the maximal minors of the submatrix whose rows are a childless node and its parents.

Example
Let $\mathscr{G}$ be the TDAG 

.

- The null cone is **not** Zariski closed for $n \geq 2$.
  Its Zariski closure is the variety of matrices of rank at most two.
- For $n = 2$, Y is not in the null cone but in its Zariski closure ($= \mathbb{R}^{3\times 2}$):

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

Hence, an MLE given Y exists. What is it? Is it unique? Homework!

Undirected Graphical Models

Which TDAGs have Zariski closed null cones?

Undirected Graphical Models

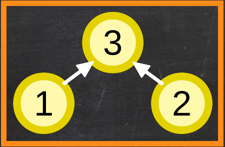
Which TDAGs have Zariski closed null cones?

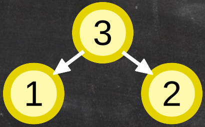
Corollary Let $\mathscr{G}$ be a TDAG with m nodes. The null cone under the action of $G(\mathscr{G}) \cap \mathrm{SL}_m$ on $\mathbb{R}^{m\times n}$ is Zariski closed for every n iff $\mathscr{G}$ has no unshielded colliders.

Undirected Graphical Models

Which TDAGs have Zariski closed null cones?

Corollary Let $\mathscr{G}$ be a TDAG with m nodes. The null cone under the action of $G(\mathscr{G}) \cap \mathrm{SL}_m$ on $\mathbb{R}^{m \times n}$ is Zariski closed for every n iff $\mathscr{G}$ has no unshielded colliders.



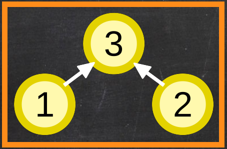


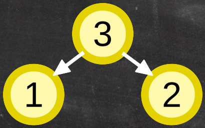
An **unshielded collider** of $\mathscr{G}$ is a subgraph $j \to i \leftarrow k$ with no edge between j and k.

Undirected Graphical Models

Which TDAGs have Zariski closed null cones?

Corollary Let $\mathscr{G}$ be a TDAG with m nodes. The null cone under the action of $G(\mathscr{G}) \cap \mathrm{SL}_m$ on $\mathbb{R}^{m \times n}$ is Zariski closed for every n iff $\mathscr{G}$ has no unshielded colliders.

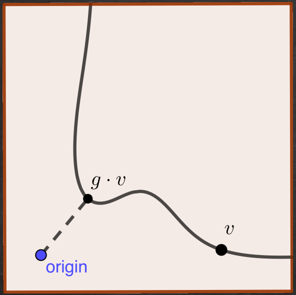


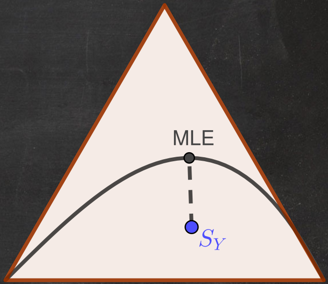


An **unshielded collider** of $\mathscr{G}$ is a subgraph $j \to i \leftarrow k$ with no edge between j and k.

This is a very interesting condition in statistics! $\mathscr{G}$ has no unshielded colliders if and only if it has the same graphical model as its underlying undirected graph.

Summary





	Invariant theory		**Statistics**
historical progression ↓	describe null cone	×	algorithms to find MLE
	algorithmic null cone membership testing		convergence analysis