Optimizing over Two Embeddings of the Grassmannian

Hannah Friedman

Example: Euclidean Distance

Let $\mathbf{u}$ be a point in $\mathbb{R}^2$. What is the closest point to $\mathbf{u}$ on the unit circle?

Function

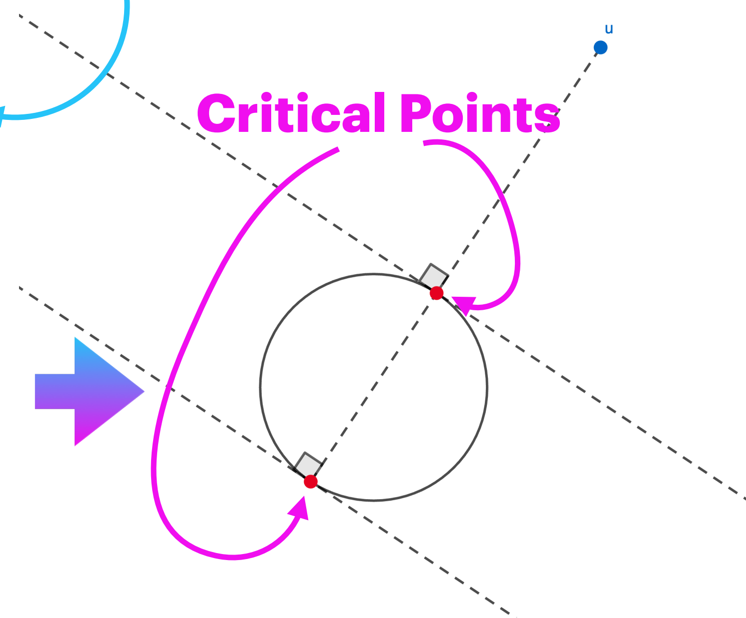
Variety

$$\begin{aligned} &\text{minimize} \quad ||\mathbf{x}-\mathbf{u}||^2 \\ &\text{subject to} \quad ||\mathbf{x}||^2 = 1 \end{aligned}$$

Lagrange Multiplier

$$\begin{aligned} 2(x_1-u_1)+2\lambda x_1 &= 0 \\ 2(x_2-u_2)+2\lambda x_2 &= 0 \\ x_1^2+x_2^2-1 &= 0 \end{aligned}$$

$$\mathbf{x} = \pm \frac{1}{||\mathbf{u}||}\mathbf{u}$$



Lagrange Multipliers

Any optimal solution $\mathbf{x}$ to the following optimization problem

maximize $f(\mathbf{x})$

subject to $G(\mathbf{x}) = \begin{pmatrix} g_1(\mathbf{x}) & \cdots & g_k(\mathbf{x}) \end{pmatrix}^T = 0$

must satisfy $\nabla \mathscr{L}(\mathbf{x}) = 0$ where $\mathscr{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum \lambda_i g_i(\mathbf{x})$.

$$\nabla \mathscr{L}(\mathbf{x}, \lambda) = \begin{cases} \nabla f(\mathbf{x}) - \sum \lambda_k \nabla g_k(\mathbf{x}) & = 0 \\ g_1(\mathbf{x}) = \cdots = g_k(\mathbf{x}) & = 0 \end{cases} \longleftrightarrow \begin{cases} \operatorname{rank} \left(\operatorname{Jac}(G(\mathbf{x})) \mid \nabla f(\mathbf{x}) \right) = \operatorname{rank} \operatorname{Jac}(G(\mathbf{x})) \\ G(\mathbf{x}) = 0 \end{cases}$$

where $\operatorname{Jac}(G(\mathbf{x})) = \begin{pmatrix} \nabla g_1(\mathbf{x}) & \cdots & \nabla g_k(\mathbf{x}) \end{pmatrix}$.

Algebraic Degree of an Optimization Problem

Optimization Problem

optimize $\quad f(\mathbf{x})$

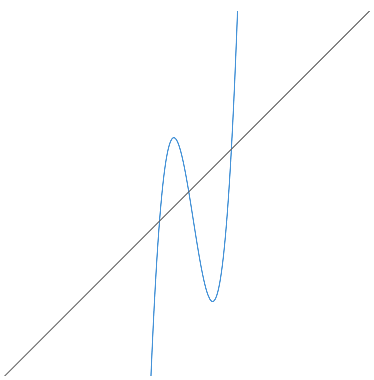
subject to $G(\mathbf{x}) = 0$

Critical Points

$$\operatorname{rank}\big(\operatorname{Jac}(G(\mathbf{x})) \mid \nabla f(\mathbf{x})\big) = \operatorname{rank}\ \operatorname{Jac}(G(\mathbf{x}))$$
$$G(\mathbf{x}) = 0$$

often 0-dimensional

Definition. The *algebraic degree* of an optimization problem is the number of critical points.



When the variety is not zero dimensional, its degree can still give an idea of the complexity of the problem.

The algebraic degree of a problem is a proxy for the difficulty of correctly solving the problem.

Degrees of Optimization Problems

optimize $\quad f(\mathbf{x})$

subject to $G(\mathbf{x}) = 0$

Fixed $f(x)$:

$f_{\mathbf{u}}(\mathbf{x}) = \|\|\mathbf{x} - \mathbf{u}\|\|^2$	Euclidean Distance Degree	Sum of Polar Degrees
$f_{\mathbf{u}}(\mathbf{x}) = \sum u_i \log(x_i)$	Maximum Likelihood Degree	Euler Characteristic
$f_{\mathbf{u}}(\mathbf{x}) = \sum u_i x_i$	Linear Optimization Degree	First Polar Degree
$\vdots$	$\vdots$	$\vdots$

The Grassmannian

The Grassmannian $\mathrm{Gr}(k, n)$ is the space of k-subspaces of n-space.

What's the best way to work with $\mathrm{Gr}(k, n)$?

- Equivalence classes of $n \times k$ matrices with the same column span
- Plücker coordinates
- Orthogonal projection matrices

The Two Lives of the Grassmannian

Plücker Coordinates	**Orthogonal Projection Matrices**
Pure Math	Applied Math
Projective Variety	Affine Variety
Algebraic Combinatorics	Numerics and Statistics
Particle Physics	Data Science
$\vdots$	$\vdots$

For more on different embeddings of the Grassmannian, check out my recent paper!
"The Two Lives of the Grassmannian," arXiv:2401.03684

Plücker Coordinates

L : k-dimensional subspace of $\mathbb{R}^n$ A : $n \times k$ matrix whose columns span L

The Plücker coordinates for L are $x_I = \det(A_I)$ for $I \subseteq [n], |I| = k$, where A_I is the $k \times k$ submatrix of A formed by taking the rows indexed by I.

Example (k = 2, n = 5).

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{42} \end{pmatrix}$$

$$x_{ij} = \det \begin{pmatrix} a_{i1} a_{i2} \\ a_{j1} a_{j2} \end{pmatrix}$$
$$= a_{i1}a_{j2} - a_{j1}a_{i2}$$

$$x_{12}x_{34} - x_{13}x_{24} + x_{14}x_{23} = 0$$
$$x_{12}x_{35} - x_{13}x_{25} + x_{15}x_{23} = 0$$
$$x_{12}x_{45} - x_{14}x_{25} + x_{15}x_{24} = 0$$
$$x_{13}x_{45} - x_{14}x_{35} + x_{15}x_{34} = 0$$
$$x_{23}x_{45} - x_{24}x_{35} + x_{25}x_{34} = 0$$

Orthogonal Projection Matrices

L : k-dimensional subspace of $\mathbb{R}^n$ A : $n \times k$ matrix whose columns span L

The $n \times n$ matrix $P = A(A^T A)^{-1} A^T$ is the unique projection matrix onto L.

The matrix P satisfies

$$P^T = P, P^2 = P \text{ and } \mathrm{trace}(P) = k\,.$$

If a symmetric matrix P satisfies the above equations, it is the projection matrix onto some k-dimensional subspace, so the *projection Grassmannian is*

$$\mathrm{pGr}(k,n) = \mathscr{V}(\langle P^T - P, P^2 - P, \mathrm{trace}(P) - k\rangle).$$

Theorem. (Devriendt, F., Reinke, Sturmfels 2024)

$$\mathscr{I}(\mathrm{pGr}(k,n)) = \langle P^T - P, P^2 - P, \mathrm{trace}(P) - k\rangle.$$

Moving Between the Two Lives

Projection matrix P

Plücker coordinates $\mathbf{x}$

Take maximal minors of the first k rows of P

$$\longrightarrow$$

Lemma. (Bloch, Karp 2023)

$$\longleftarrow$$

$$p_{ij} = \frac{\sum_{K \in \binom{[n]}{k-1}} x_{iK} x_{jK}}{\sum_{I \in \binom{[n]}{k}} x_I^2}$$

Two Maximum Likelihood Problems on the Grassmannian

maximize $\displaystyle\sum_{I \subseteq \binom{[n]}{k}} u_I \log\left(x_I\right)$

subject to $\mathbf{x} \in \mathrm{Gr}(k,n)$

Maximum Likelihood Degrees:

	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
$k=2$	4	22	156	1368	14400

maximize $\displaystyle\sum_{I \subseteq \binom{[n]}{k}} u_I \log\left(q_I\right)$

subject to $\mathbf{q} \in \mathrm{sGr}(k,n)$

$$\mathrm{sGr}(k,n) = \left\{ (x_I^2)_{I \in \binom{[n]}{k}} \mid (x_I)_{I \in \binom{[n]}{k}} \in \mathrm{Gr}(k,n) \right\}$$

Maximum Likelihood Degrees:

	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
$k=2$	3	12	60	360	2520

The lower maximum likelihood degrees indicate that the model on the right is a natural probability model: it is an example of a determinantal point process!

Probability Distributions on the Grassmannian

Let P be a real, symmetric matrix with eigenvalues in $[0,1]$. A *determinantal point process* with *kernel* P is a random variable Z on $2^{[n]}$ such that

$$\mathbb{P}[I \subseteq Z] = \det(P_I)$$

where P_I is the $k \times k$ principal submatrix of P obtained by selecting the k rows and columns indexed by I.

Example (n = 3).

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{23} & p_{33} \end{pmatrix}$$

$$\mathbb{P}[\{2\} \subseteq Z] = p_{22}$$

$$\mathbb{P}[\{1,3\} \subseteq Z] = \det\begin{pmatrix} p_{11} & p_{13} \\ p_{13} & p_{33} \end{pmatrix} = p_{11}p_{33} - p_{13}^2$$

Probability Distributions on the Grassmannian

Relationship between Plücker embedding
and projection matrix embedding

If $P \in \mathrm{pGr}(k, n)$, then $\mathbb{P}[I = Z] = \begin{cases} \det(P_I) = \dfrac{x_I^2}{\sum_{J \in \binom{[n]}{[k]}} x_J^2} = q_I & \text{if } |I| = k \\ 0 & \text{else} \end{cases}$

$\sum_{I \subseteq \binom{[n]}{k}} u_I \log(q_I)$ **is the log likelihood function for a determinantal point process with kernel in $\mathrm{pGr}(k, n)$!**

Degrees of Optimization Problems

$$\text{optimize} \quad f(\mathbf{x})$$

$$\text{subject to } G(\mathbf{x}) = 0$$

Fixed $f(x)$:

$f_{\mathbf{u}}(\mathbf{x}) = \|\|\mathbf{x} - \mathbf{u}\|\|^2$	Euclidean Distance Degree	Sum of Polar Degrees
$f_{\mathbf{u}}(\mathbf{x}) = \sum u_i \log(x_i)$	Maximum Likelihood Degree	Euler Characteristic
$f_{\mathbf{u}}(\mathbf{x}) = \sum u_i x_i$	Linear Optimization Degree	First Polar Degree
$\vdots$	$\vdots$	$\vdots$

Eigenvalue Problem

Let M be real, symmetric, positive definite $n \times n$ matrix with eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_n$.

Goal: compute an $n \times k$ matrix $X = [\mathbf{x}_1 \cdots \mathbf{x}_k]$ such that $M\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for $i = 1,\ldots,k$.

$$\begin{aligned} &\text{maximize} \quad \operatorname{trace}(X^T M X) \\ &\text{subject to} \quad X^T X = \mathrm{Id}_k. \end{aligned}$$

Critical Points of the Eigenvalue Problem

$$\begin{aligned} &\text{maximize} \quad \operatorname{trace}(X^T M X) \\ &\text{subject to} \quad X^T X = \mathrm{Id}_k \leftrightarrow \quad \langle \mathbf{x}_i, \mathbf{x}_j \rangle = \delta_{ij} \end{aligned}$$

$$\begin{cases} \operatorname{rank}\big(\operatorname{Jac}(G(\mathbf{x})) \mid \nabla f(\mathbf{x})\big) = \operatorname{rank} \operatorname{Jac}(G(\mathbf{x})) \\ G(\mathbf{x}) = 0 \end{cases}$$

The critical points of this problem are matrices X satisfying $X^T X = \mathrm{Id}_k$ and $\nabla \operatorname{trace}(X^T M X) = 2 \begin{pmatrix} M\mathbf{x}_1 \\ \vdots \\ M\mathbf{x}_k \end{pmatrix}$ is in column span of

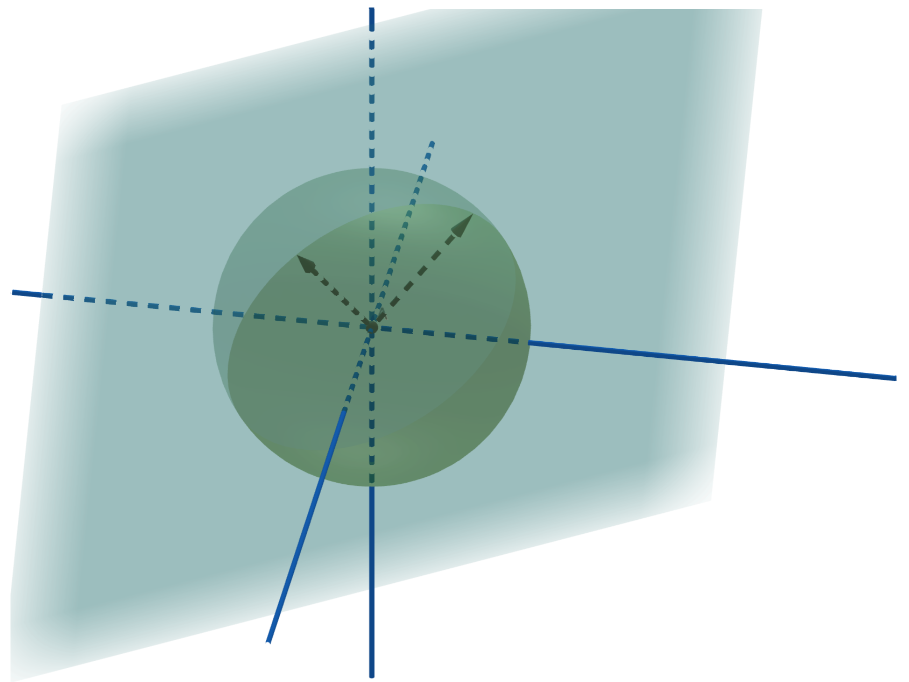
$$\operatorname{Jac} \begin{pmatrix} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle - 1 \\ \vdots \\ \langle \mathbf{x}_k, \mathbf{x}_k \rangle - 1 \\ \langle \mathbf{x}_1, \mathbf{x}_2 \rangle \\ \vdots \\ \langle \mathbf{x}_{k-1}, \mathbf{x}_k \rangle \end{pmatrix} = \begin{pmatrix} 2\mathbf{x}_1 & & & & & \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_k & & & & \cdots & \\ & 2\mathbf{x}_2 & & & & \mathbf{x}_1 & & & & \mathbf{x}_3 & \cdots & \mathbf{x}_k & & \\ & & 2\mathbf{x}_3 & & & & \mathbf{x}_1 & & & \mathbf{x}_2 & & & \cdots & \\ & & & \ddots & & & & \ddots & & & \ddots & & & \mathbf{x}_k \\ & & & & 2\mathbf{x}_k & & & & \mathbf{x}_1 & & & \mathbf{x}_2 & \cdots & \mathbf{x}_{k-1} \end{pmatrix}.$$

Critical Points of the Eigenvalue Problem

Let $\Theta \in O(k)$, where $O(k)$ is the group of orthogonal $k \times k$ matrices. Then

$$\mathrm{trace}(\Theta^T X^T M X \Theta) = \mathrm{trace}(X^T M X \Theta \Theta^T) = \mathrm{trace}(X^T M X)$$

$$X^T X = \mathrm{Id}_k \implies \Theta^T X^T X \Theta = \Theta^T \Theta = \mathrm{Id}_k$$



Critical Points of the Eigenvalue Problem

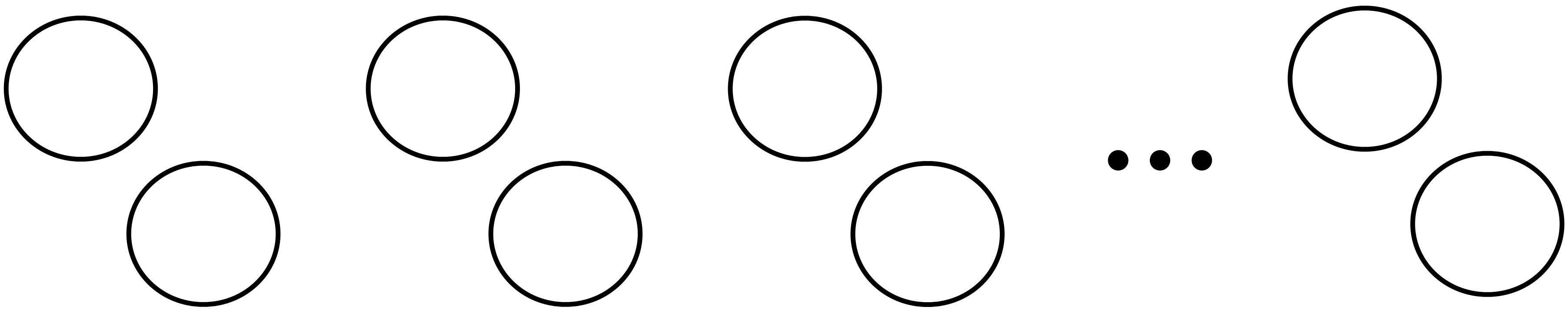
Theorem. (F., Hoşten 2024+) Let M be a generic real symmetric $n \times n$ matrix and let X be an $n \times k$ variable matrix. The algebraic set of complex critical points of the eigenvalue optimization problem is

$$\bigsqcup_{\{i_1,\ldots,i_k\}\in\binom{[n]}{k}} \{[q_{i_1}\, q_{i_2}\, \cdots\, q_{i_k}]\Theta \,:\, \Theta \in O(k)\}$$

where $q_1, \ldots, q_n$ is an orthonormal eigenbasis of M. This algebraic set is a disjoint union of $\binom{n}{k}$ irreducible varieties isomorphic to $O(k)$, and hence its dimension is equal to $\dim(O(k))$ and its degree is equal to $\deg(O(k)) \cdot \binom{n}{k}$.

Critical Points of the Eigenvalue Problem

Example. The variety $O(2)$ is the disjoint union of two circles, so the critical points of the eigenvalue problem for $k = 2$ are



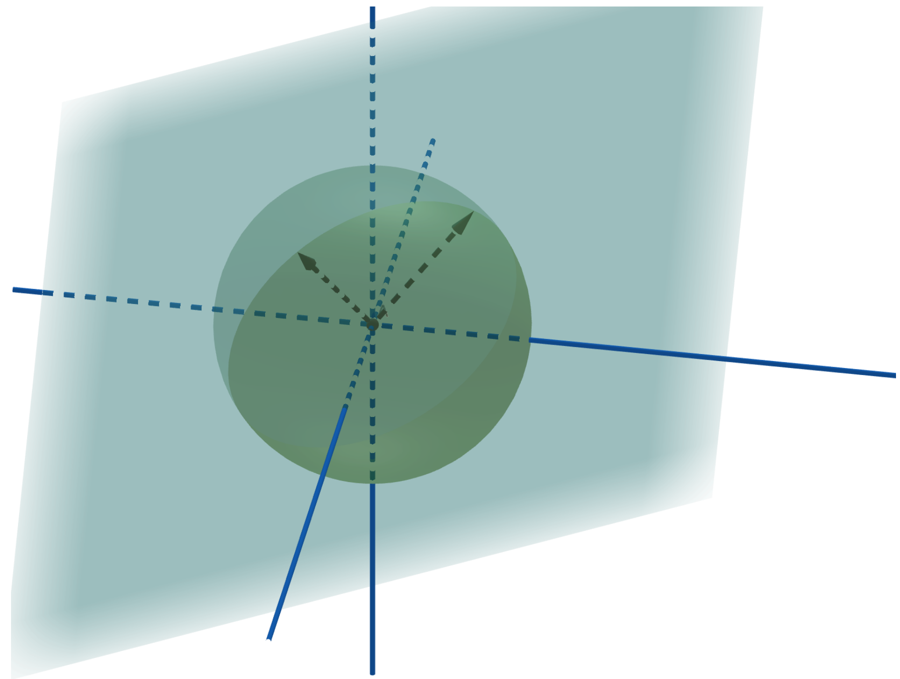
$\{1,2\}$ $\{1,3\}$ $\{1,4\}$ $\{n-1,n\}$

This variety has degree $4\binom{n}{2}$.

Rethinking Our Formulation

Optimizing over specific sets of basis vectors for our space

Optimizing over coordinate-free representations of our space

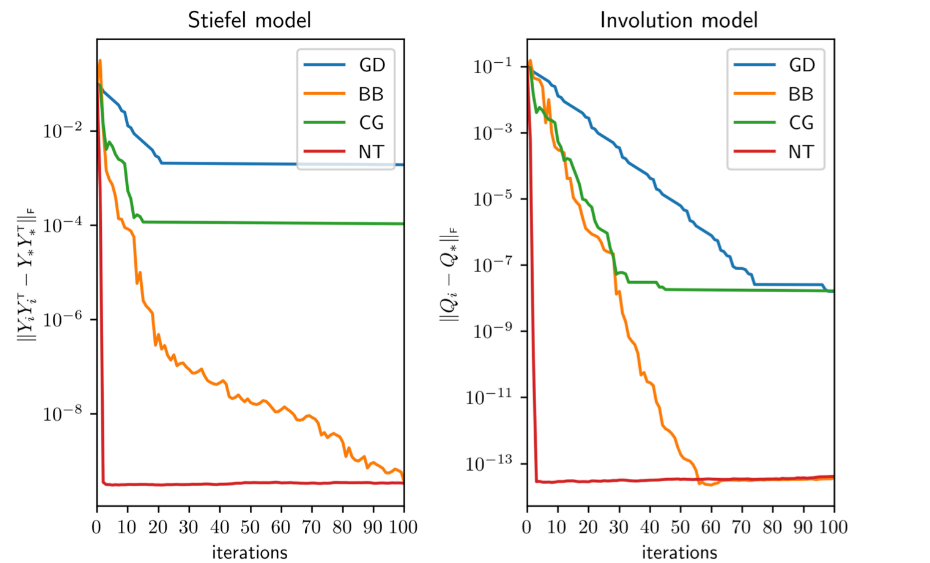


Rethinking Our Formulation

Optimizing over specific sets of basis vectors for our space

Optimizing over coordinate-free representations of our space

FIGURE 1. Convergence behavior of algorithms in the Stiefel and involution models.



"Simpler Grassmannian Optimization"
by Lai, Lim, and Ye
arXiv:2009.13502

Orthogonal Projection Matrices

L : k-dimensional subspace of $\mathbb{R}^n$ $\quad$ $A : n \times k$ matrix whose columns span L

The $n \times n$ matrix $P = A(A^TA)^{-1}A^T$ is the unique projection matrix onto L.

The matrix P satisfies

$$P^T = P,\ P^2 = P \text{ and } \mathrm{trace}(P) = k\,.$$

If a symmetric matrix P satisfies the above equations, it is the projection matrix onto some k-dimensional subspace, so the *projection Grassmannian is*

$$\mathrm{pGr}(k,n) = V(\langle P^T - P, P^2 - P, \mathrm{trace}(P) - k\rangle).$$

Eigenvalue Problem on the Projection Grassmannian

Since the columns of X are *orthonormal*, we have

$$P = X(X^TX)^{-1}X^T = X\mathrm{Id}_kX^T = XX^T.$$

maximize $\operatorname{trace}(XMX^T) = \operatorname{trace}(MXX^T)$

subject to $X^TX = \mathrm{Id}_k$.

maximize $\operatorname{trace}(MP)$

subject to $P \in \mathrm{pGr}(k, n)$

Eigenvalue Problem on the Projection Grassmannian

Theorem. (F., Hoşten 2024+) Let M be a generic real symmetric $n \times n$ matrix and let X be an $n \times k$ variable matrix. The algebraic set of complex critical points of the eigenvalue optimization problem over $\mathrm{pGr}(k,n)$ is

$$\left\{ [q_{i_1}\, q_{i_2}\, \cdots\, q_{i_k}][q_{i_1}\, q_{i_2}\, \cdots\, q_{i_k}]^T \mid \{i_1, \ldots, i_k\} \in \binom{[n]}{k} \right\}$$

where $q_1, \ldots, q_n$ is an orthonormal eigenbasis of M. Therefore the degree of the eigenvalue optimization problem over $\mathrm{pGr}(k,n)$ is $\binom{n}{k}$.

Eigenvalue Problem on the Projection Grassmannian

If the columns of X form an orthonormal basis for the subspace, then we have

$$P = X(X^TX)^{-1}X^T = X\mathrm{Id}_kX^T = XX^T.$$

maximize $\operatorname{trace}(MXX^T)$

subject to $X^TX = \mathrm{Id}_k$.

maximize $\operatorname{trace}(MP)$

subject to $P \in \mathrm{pGr}(k, n)$

Eigenspaces can be found by *linear* optimization over the Grassmanian!

"Eigendegree" = LO Degree

Example (n = 3).

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{12} & m_{22} & m_{23} \\ m_{13} & m_{23} & m_{33} \end{pmatrix} \qquad P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{23} & p_{33} \end{pmatrix}$$

$$\begin{aligned} \operatorname{trace}(MP) &= m_{11}p_{11} + 2m_{12}p_{12} + 2m_{13}p_{13} + m_{22}p_{22} + 2m_{23}p_{23} + m_{33}p_{33} \\ &= u_{11}p_{11} + u_{12}p_{12} + u_{13}p_{13} + u_{22}p_{22} + u_{23}p_{23} + u_{33}p_{33} \end{aligned}$$

$$\begin{array}{ll} u_{11} = m_{11} & u_{12} = 2m_{12} \\ u_{22} = m_{22} & u_{13} = 2m_{13} \\ u_{33} = m_{33} & u_{23} = 2m_{23} \end{array}$$

"Eigendegree" = LO Degree

maximize $\operatorname{trace}(MP)$

subject to $P \in \mathrm{pGr}(k,n)$

$\longleftrightarrow$

maximize $\sum u_{ij} p_{ij}$

subject to $P \in \mathrm{pGr}(k,n)$

We can write the objective function as a generic linear form over pGr(k,n):

$$\operatorname{trace}(MP) = \sum_{i \le j} u_{ij} p_{ij}$$

where

$$u_{ij} = \begin{cases} 2m_{ij} & \text{if } i \neq j \\ m_{ij} & \text{if } i = j \end{cases}$$

Eigenvalue Problem on the Projection Grassmannian

degree of the eigenvalue problem over $\mathrm{pGr}(k, n)$ = **linear optimization degree of** $\mathrm{pGr}(k, n)$ = $\binom{n}{k}$

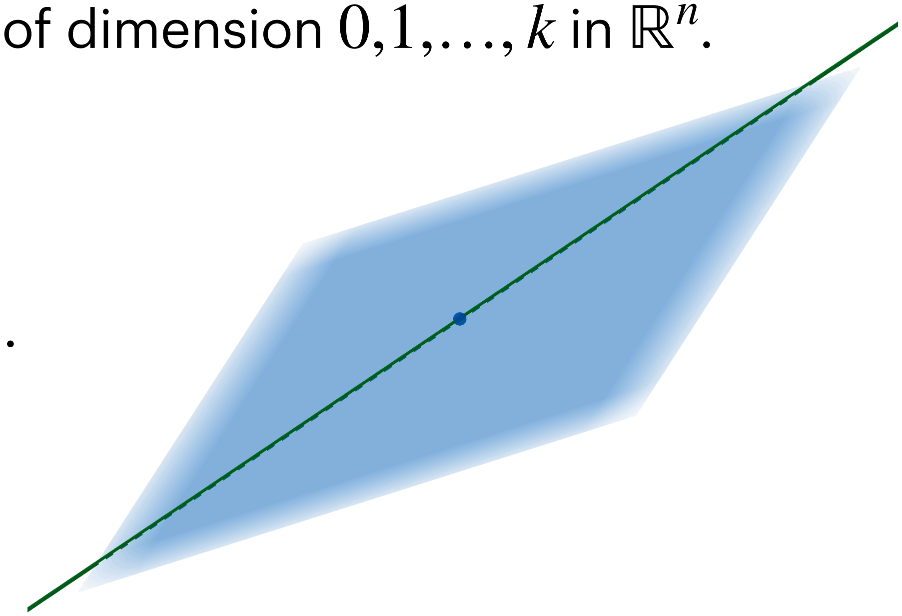
$\mathrm{trace}(MP)$

$f_u(x) = \sum u_{ij} p_{ij}$

Beyond the Grassmannian

Definition. The complete flag variety, denoted $\mathrm{Fl}(0,1,\ldots,k;n)$, is the space of nested subspaces of dimension $0,1,\ldots,k$ in $\mathbb{R}^n$.

Example. A point in $\mathrm{Fl}(0,1,2;3)$.



Proposition. (Ye, Wong, Lim 2022)

$$\mathrm{pFl}(0,1,\ldots,k;n) = \{(P_1,\ldots,P_k) \mid P_{i+1}P_i = P_i, P_i^2 = P_i, \mathrm{trace}(P_i) = i\}$$

Extending to Flag Varieties

degree of the eigenvalue problem over $\mathrm{pGr}(k, n)$ = **linear optimization degree of** $\mathrm{pGr}(k, n)$ = $\binom{n}{k}$

degree of the heterogeneous quadrics minimization problem over $\mathrm{pFl}(1, \ldots, k; n)$ = **linear optimization degree of** $\mathrm{pFl}(1, \ldots, k; n)$ = $???$

Heterogeneous Quadrics Minimization Problem

Given k generic real symmetric $n \times n$ matrices $M_1, \ldots, M_k$,

$$\text{maximize } \sum_{i=1}^{k} \mathbf{x}_i^T M_i \mathbf{x}_i \qquad\qquad \text{maximize } \sum_{i=1}^{k} \operatorname{trace}(M_i P_i)$$

$$\text{subject to } X^T X = \mathrm{Id}_k \qquad\qquad \text{subject to } (P_1, \ldots, P_k) \in \mathrm{pFl}(0,1,\ldots,k;n)$$

Compute!

Degrees of the problem

$$\text{maximize} \sum_{i=1}^{k} \mathbf{x}_i^T M_i \mathbf{x}_i$$

$$\text{subject to } X^T X = \mathrm{Id}_k$$

	n = 3	n = 4	n = 5	n = 6	n = 7	n
k = 2	40	112	240	440	728	$4\sum_{j=1}^{n-1} 2j^2$
k = 3		960	5536	21,440	???	???????

These numbers were produced with `HomotopyContinuation.jl`.

Degrees of Optimization Problems

optimize $\quad f(\mathbf{x})$

subject to $G(\mathbf{x}) = 0$

Fixed $f(x)$:

✓ $f_{\mathbf{u}}(\mathbf{x}) = \vert\vert\mathbf{x} - \mathbf{u}\vert\vert^2$	Euclidean Distance Degree	Sum of Polar Degrees
✓ $f_{\mathbf{u}}(\mathbf{x}) = \sum u_i \log(x_i)$	Maximum Likelihood Degree	Euler Characteristic
✓ $f_{\mathbf{u}}(\mathbf{x}) = \sum u_i x_i$	Linear Optimization Degree	First Polar Degree
$\vdots$	$\vdots$	$\vdots$

References

Maximum Likelihood Degree

"The Maximum Likelihood Degree"

by Catanese, Hoşten, Khetan, Sturmfels

Euclidean Distance Degree

"The Euclidean Distance Degree of an Algebraic Variety"

by Draisma, Horobeț, Ottaviani, Sturmfels, Thomas

Linear Optimization Degree

"Linear Optimization on Varieties and Chern-Mather Classes"

by Maxim, Rodriguez, Wang, Wu

The Grassmannian and Flags

"The Two Lives of the Grassmannian"

by Devriendt, F. Reinke, Sturmfels

"Simpler Grassmannian Optimization"

by Lai, Lim, and Ye

"Gradient Flows, Adjoint Orbits, and the Topology of Totally Nonnegative Flag Varieties"

by Bloch and Karp

"Optimization on Flag Manifolds"

by Ye, Wong, and Lim

Thank you!