The Unbalanced Procrustes Problem and Algebraic Optimization

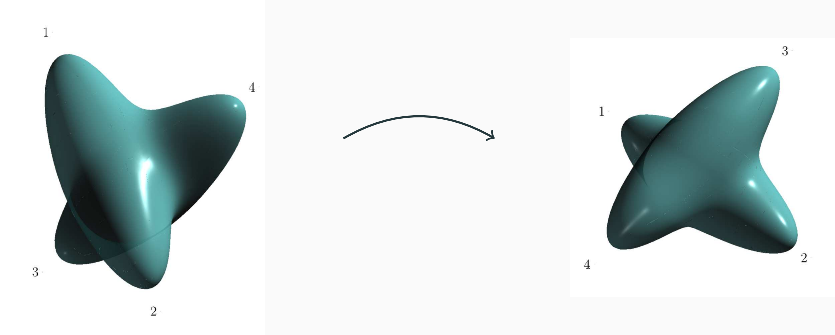
Thomas Yahl
tyahl@wisc.edu
University of Wisconsin – Madison

Applied Algebra Seminar Spring 2024

Joint work with Jose Israel Rodriguez

The Balanced Procrustes Problem

The *Balanced Procrustes Problem* concerns finding an optimal rotation which aligns two sets of data.



- This problem arises in aviation, satellite positioning, and image reconstruction.
- The case that the data sets live in $\mathbb{R}^3$ is known as Wahba's Problem (1965).

The Balanced Procrustes Problem

- We represent the data sets as $m \times n$ matrices A and B.
- Similarity of data sets is measured by the Frobenius norm $|| \cdot ||_F$.

The *Balanced Procrustes Problem* is the optimization problem

$$\operatorname*{argmin}_{X \in O(n)} ||AX - B||_F^2.$$

- We will view this problem and solutions from the lens of algebraic optimization.

Algebraic Optimization

Algebraic optimization looks to describe the variety of complex critical points of an optimization problem, its dimension, degree, symmetries, etc.

- Euclidean distance minimization
- Maximal likelihood estimation
- Machine learning
- Robotic mechanisms

The Balanced Procrustes Problem

Using Lagrange multipliers, one finds defining equations for the locus of critical points of the balanced Procrustes problem ($m = 3$, $n = 2$):

$$
\begin{aligned}
o1 = \{ & -x_{1,1}l_1 - x_{1,2}l_2 + (a_{1,1}^2 + a_{2,1}^2 + a_{3,1}^2)x_{1,1} + (a_{1,1}a_{1,2} + a_{2,1}a_{2,2} + a_{3,1}a_{3,2})x_{2,1} + (a_{1,1}a_{1,3} + a_{2,1}a_{2,3} + a_{3,1}a_{3,3})x_{3,1} - a_{1,1}b_{1,1} - a_{2,1}b_{2,1} - a_{3,1}b_{3,1}, \\
& -x_{1,1}l_2 - x_{1,2}l_3 + (a_{1,1}^2 + a_{2,1}^2 + a_{3,1}^2)x_{1,2} + (a_{1,1}a_{1,2} + a_{2,1}a_{2,2} + a_{3,1}a_{3,2})x_{2,2} + (a_{1,1}a_{1,3} + a_{2,1}a_{2,3} + a_{3,1}a_{3,3})x_{3,2} - a_{1,1}b_{1,2} - a_{2,1}b_{2,2} - a_{3,1}b_{3,2}, \\
& -x_{2,1}l_1 - x_{2,2}l_2 + (a_{1,1}a_{1,2} + a_{2,1}a_{2,2} + a_{3,1}a_{3,2})x_{1,1} + (a_{1,2}^2 + a_{2,2}^2 + a_{3,2}^2)x_{2,1} + (a_{1,2}a_{1,3} + a_{2,2}a_{2,3} + a_{3,2}a_{3,3})x_{3,1} - a_{1,2}b_{1,1} - a_{2,2}b_{2,1} - a_{3,2}b_{3,1}, \\
& -x_{2,1}l_2 - x_{2,2}l_3 + (a_{1,1}a_{1,2} + a_{2,1}a_{2,2} + a_{3,1}a_{3,2})x_{1,2} + (a_{1,2}^2 + a_{2,2}^2 + a_{3,2}^2)x_{2,2} + (a_{1,2}a_{1,3} + a_{2,2}a_{2,3} + a_{3,2}a_{3,3})x_{3,2} - a_{1,2}b_{1,2} - a_{2,2}b_{2,2} - a_{3,2}b_{3,2}, \\
& -x_{3,1}l_1 - x_{3,2}l_2 + (a_{1,1}a_{1,3} + a_{2,1}a_{2,3} + a_{3,1}a_{3,3})x_{1,1} + (a_{1,2}a_{1,3} + a_{2,2}a_{2,3} + a_{3,2}a_{3,3})x_{2,1} + (a_{1,3}^2 + a_{2,3}^2 + a_{3,3}^2)x_{3,1} - a_{1,3}b_{1,1} - a_{2,3}b_{2,1} - a_{3,3}b_{3,1}, \\
& -x_{3,1}l_2 - x_{3,2}l_3 + (a_{1,1}a_{1,3} + a_{2,1}a_{2,3} + a_{3,1}a_{3,3})x_{1,2} + (a_{1,2}a_{1,3} + a_{2,2}a_{2,3} + a_{3,2}a_{3,3})x_{2,2} + (a_{1,3}^2 + a_{2,3}^2 + a_{3,3}^2)x_{3,2} - a_{1,3}b_{1,2} - a_{2,3}b_{2,2} - a_{3,3}b_{3,2} \}
\end{aligned}
$$

- This is incredibly uninsightful.. Let's do better.

The Balanced Procrustes Problem

A critical point is a point $X \in O(n)$ where the differential of the function $||AX - B||_F^2$ annihilates the tangent space $T_X O(n)$.

- The differential of $||AX - B||_F^2$ is given by $A^T(AX - B)$.
- The annihilator of the tangent space $\text{Ann}\, T_X O(n)$ is parameterized by

$$\text{Ann}\, T_X O(n) = \{XS \in \mathbb{R}^{n \times n} : S \in \text{Sym}(n)\}.$$

- The variety of critical points is defined by the equations

$$X^T X - I = 0, \;\; A^T(AX - B) - XS = 0.$$

The Balanced Procrustes Problem

The variety of (complex) critical points of the balanced Procrustes problem is defined by the equations

$$X^T X - I = 0, \quad A^T(AX - B)X^T - X(AX - B)^T A = 0.$$

- The critical points lie in the *(algebraic) orthogonal group*

$$O_{\mathbb{C}}(n) = \{X \in \mathbb{C}^{n\times n} : X^T X - I = 0\}.$$

- The *optimization degree* $\mathrm{PDeg}(n)$ is the number of (complex) critical points for general $A, B \in \mathbb{C}^{m\times n}$.

The Balanced Procrustes Problem

The balanced Procrustes problem is equivalent to a Euclidean distance minimization problem.

- There is a constant C depending on A and B such that

$$||AX - B||_F^2 = ||X - A^T B||_F^2 + C.$$

- The optimization degree of the balanced Procrustes problem is equal to the Euclidean distance degree of $O_{\mathbb{C}}(n)$:

$$\text{PDeg}(n) = \text{EDDeg}(O_{\mathbb{C}}(n)).$$

The Balanced Procrustes Problem

Use the structure of $O_{\mathbb{C}}(n)$ to compute the Euclidean distance degree!

- $O_{\mathbb{C}}(n)$ is invariant under the left and right action of $O(n)$.
- The set of real diagonal matrices in $O_{\mathbb{C}}(n)$ is invariant under signed permutations.

$$\begin{pmatrix} \pm 1 & 0 & \dots & 0 \\ 0 & \pm 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pm 1 \end{pmatrix} \in \mathrm{Diag}(O_{\mathbb{C}}(n)).$$

- Rekha Thomas et al. [1] show the Euclidean distance degree can be computed via the algebraic singular value decomposition.

The Balanced Procrustes Problem

By [1], there are equalities:

$$\mathrm{PDeg}(n) = \mathrm{EDDeg}(O_{\mathbb{C}}(n)) = \mathrm{EDDeg}(\mathrm{Diag}(O(n))) = 2^n.$$

- If A and B are general (real), and $A^T B = USV^T$ is a SVD, then the critical points have the form

$$UDV^T \quad \text{for} \quad D \in \mathrm{Diag}(O(n)).$$

- If S has non-negative entries, then

$$UV^T \quad \text{is an } \underline{\text{optimal}} \text{ solution.}$$

- If $A^T B$ has full rank, then this optimal solution is unique.

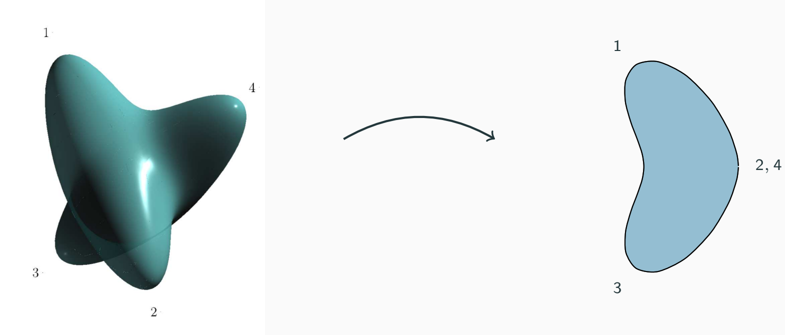
The Balanced Procrustes Problem

- The critical points of the balanced Procrustes problem are completely understood by the (algebraic) SVD.
- Our understanding of the critical points aided in computing the optimal solution(s).

We would like to be able to produce a similar analysis for a related problem, the *Unbalanced* Procrustes Problem.

The Unbalanced Procrustes Problem

The *Unbalanced Procrustes Problem* looks to find the orthogonal projection for which a higher-dimensional data set best models a lower-dimensional data set.



- Appears in facial recognition, data mining, and canonical correlation analysis.
- No known formulas for the optimal solutions!

The Unbalanced Procrustes Problem

- The data set A is represented by a $m \times \ell$ matrix and the data set B is represented by a $m \times n$ matrix with $\ell > n$.
- The *Stiefel Manifold* is $\mathrm{St}(\ell, n) = \{X \in \mathbb{R}^{\ell \times n} : X^T X = I\}$.

The *Unalanced Procrustes Problem* is the optimization problem

$$\underset{X \in \mathrm{St}(\ell,n)}{\operatorname{argmin}} \; ||AX - B||_F^2.$$

- We again study the variety of (complex) critical points.

The Unbalanced Procrustes Problem

- We again have defining equations:

$$X^T X - I = 0, \quad A^T(AX - B)X^T - X(AX - B)^T A = 0.$$

- The solutions lie on the *(algebraic) Stiefel manifold*

$$\mathrm{St}_{\mathbb{C}}(\ell, n) = \{X \in \mathbb{C}^{\ell \times n} : X^T X - I = 0\}.$$

- The *optimization degree* $\mathrm{PDeg}(\ell, n)$ is the number of isolated (complex) critical points for general $(A, B) \in \mathbb{C}^{m \times \ell} \times \mathbb{C}^{m \times n}$.

The Unbalanced Procrustes Problem

Our analysis of the balanced Procrustes problem relied on the reduction to a Euclidean distance minimization problem.

- The unbalanced Procrustes problem cannot be reduced to a Euclidean distance minimization problem!
- We can't use Euclidean distance degree results such as in [1]!
- We don't know that $\mathrm{PDeg}(\ell, n)$ is even finite!

We have to get our hands dirty instead of relying on known results.

The Unbalanced Procrustes Problem

Why are there finitely many critical points?

- We reparameterize $C = A^T A \in \mathrm{Sym}(\ell)$ and $D = A^T B \in \mathbb{C}^{\ell \times n}$,

$$(CX - D)X^T - X(CX - D)^T = 0, \; X \in \mathrm{St}(\ell, n).$$

- There is an incidence correspondence:

$$\begin{array}{c} \{(C, D, X) : (CX - D)X^T - X(CX - D)^T = 0\} \\ \downarrow \pi \\ \mathrm{Sym}(\ell) \times \mathbb{C}^{\ell \times n} \end{array}$$

- The top variety is isomorphic to $\mathrm{Sym}(\ell) \times \mathrm{St}(\ell, n) \times \mathrm{Sym}(n)$.
- There are finitely many critical points for general parameters $(A, B) \in \mathbb{C}^{m \times \ell} \times \mathbb{C}^{m \times n}$.

The Unbalanced Procrustes Problem

The degrees $\mathrm{PDeg}(\ell, n)$ for various ℓ and n are given below.

$n \backslash \ell$	1	2	3	4	5	6	7	8
1	2	4	6	8	10	12	14	16
2		4	16	36	64	100	144	196
3			8	64	248	640	1320	2368
4				16	256	1808	7552	22288
5					32	1024	13654	-
6						64	4096	-

Computed numerically using `HomotopyContinuation.jl`

- $\mathrm{EDDeg}(\mathrm{St}(\ell, n)) = 2^n$ is a lower bound.

The (Diagonal) Unbalanced Procrustes Problem

We consider the relaxation $A \in \mathrm{Diag}(\mathbb{C}^{m\times\ell})$ and $B \in \mathrm{Diag}(\mathbb{C}^{m\times n})$.

- The number of critical points for general $A \in \mathrm{Diag}(\mathbb{C}^{m\times\ell})$ and $B \in \mathrm{Diag}(\mathbb{C}^{m\times n})$ gives a lower bound for $\mathrm{PDeg}(\ell, n)$.
- Experimentally, this family is general in the parameter space.

The solutions in this case are more structured.

- For any critical point $X \in \mathrm{St}(\ell, n)$, the bottom $\ell - n$ rows are orthogonal.
- If $\ell > 2n$, then X has a row of zeros.

The (Diagonal) Unbalanced Procrustes Problem

We can partition the critical points by the number of rows of zeros to enumerate.

- If a critical point $X \in \mathrm{St}(\ell, n)$ has k rows of zeros, then it is a solution of the smaller unbalanced Procrustes problem with $\ell - k$ rows.
- We can permute the bottom $\ell - n$ rows freely.
- We need to use inclusion-exclusion to not overcount.

The (Diagonal) Unbalanced Procrustes Problem

Let $d_{\ell,n}$ denote the number of critical points for general $A \in \mathrm{Diag}(\mathbb{C}^{m\times\ell})$ and $B \in \mathrm{Diag}(\mathbb{C}^{m\times n})$.

- There is a recursive formula

$$d_{\ell,n} = \sum_{k=1}^{\ell-n} (-1)^{k+1} \binom{\ell-n}{k} d_{\ell-k,n}.$$

- For fixed n, the number of critical points $d_{\ell,n}$ is a polynomial in ℓ of degree n.

$$\begin{aligned} d_{\ell,1} &= 2\ell \\ d_{\ell,2} &= 4(\ell-1)^2 \\ d_{\ell,3} &= 8(\ell-2)^3 + \frac{16}{3}(\ell-2)(\ell-3)(\ell-4) \end{aligned}$$

Future Work

- We'd like to show the diagonal case is general!
- For fixed n, describe the polynomials $d_{\ell,n}/\operatorname{PDeg}(\ell, n)$.
- Find some analogue of SVD to describe the critical points.
- Find the optimal solutions!
- Alternatively: Use our understanding of the critical points for other iterative methods of solving.

Reference

D. Drusvyatskiy, H.-L. Lee, G. Ottaviani, and R. R. Thomas.
The Euclidean distance degree of orthogonally invariant matrix varieties.
Israel J. Math., 221(1):291–316, 2017.

Thank You!

Thank you for attending!