Matrix Procrustes Problems

Nick Higham
Department of Mathematics
University of Manchester

higham@ma.man.ac.uk

Pythagoras Papadimitriou. *Parallel Solution of SVD-Related Problems, With Applications*. PhD thesis, University of Manchester, October 1993.

THE UNIVERSITY *of* MANCHESTER

Orthogonal Procrustes Problem (OPP)

Given $A, B \in \mathbb{R}^{m\times n}$,

$$\min\left\{\|A - BQ\|_F : Q^TQ = I_n\right\}.$$

Special case: $B = I$, *nearest orthonormal matrix:*

$$\min\left\{\|A - Q\|_F : Q^TQ = I_n\right\}, \qquad A \in \mathbb{R}^{m\times n}.$$

Here, the Frobenius norm

$$\|A\|_F = \Big(\sum_{i,j} a_{ij}^2\Big)^{1/2} = \mathrm{trace}(A^TA)^{1/2}.$$

Applications: OPP

- Factor analysis, statistics: are matrices A, B equivalent up to rotation?
- Satellite tracking.
- Rigid body movement in robotics.
- Structural and system identification.
- Vibration tests of large, complex structures (e.g., space station); Smith et al. (1993).

Applications: Nearest Orthogonal Matrix

- Real-time graphics. 4×4 orthogonal matrix describes orientation of hypercube rotating under user's control. A 5% deviation from orthogonality noticeable to eye.
- Aerospace computations: direction cosine matrix (DCM) $D \in \mathbb{R}^{3\times 3}$ satisfies the matrix ODE

  $$\frac{d}{dt}D(t) = SD(t), \quad S = -S^T, \quad D(0) \text{ orthogonal.}$$

  Solved by Euler's method. Approximate DCMs need to be re-orthogonalized periodically.
- Orthogonalizing a basis (Löwdin orthogonalization).
  Nearest orthogonal matrix is independent of basis ordering; result from Gram-Schmidt is not.

Norms

$A \in \mathbb{R}^{m \times n}$.

Frobenius norm:

$$\|A\|_F = \Big(\sum_{i,j} a_{ij}^2\Big)^{1/2} = \mathrm{trace}(A^T A)^{1/2}.$$

2-norm:

$$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \rho(A^T A)^{1/2}.$$

Unitarily invariant norm:

$$\|UAV\| = \|A\|$$

for all orthogonal U and V.

Useful Decompositions

$A \in \mathbb{R}^{m\times n}$, $m \geq n$.

Polar Decomposition

$$A = UH, \qquad U^TU = I_n, \quad H \text{ symmetric psd.}$$

Singular Value Decomposition (SVD)

$$A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T, \quad U^TU = I_m, \quad V^TV = I_n,$$

$$\Sigma = \mathrm{diag}(\sigma_i), \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

Connection

$$\begin{aligned} A &= \begin{bmatrix} \overset{n}{U_1} & \overset{m-n}{U_2} \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T = U_1 \Sigma V^T \\ &= U_1 V^T \cdot V \Sigma V^T \equiv UH. \end{aligned}$$

Solution of OPP

$$\min\{\,\|A - BQ\|_F : Q^TQ = I_n\}.$$

$$\|A - BQ\|_F^2 = \mathrm{trace}(A^TA - 2Q^TB^TA + B^TB),$$

so need to maximize $\mathrm{trace}(Q^TB^TA)$.

Now

$$\begin{aligned} \mathrm{trace}(Q^TC) &= \mathrm{trace}(Q^TU\Sigma V^T) \\ &= \mathrm{trace}(V^TQ^TU \cdot \Sigma) \\ &\le \sum_i \sigma_i, \end{aligned}$$

with equality for $Q = UV^T$ = orthogonal polar factor of B^TA!

Note: Orthogonal polar factor is not the solution for an arbitrary unitarily invariant norm (Mathias, 1993).

Nearest Orthogonal Matrix

Given $A \in \mathbb{R}^{m\times n}$ $(m \geq n)$,

$$\min\{\|A - Q\| : Q^TQ = I_n\}.$$

Let $A = UH$, polar decomposition.

- ▷ $Q = U$ is a solution for 2 and Frobenius norms.
- ▷ If $m = n$ then $Q = U$ is a solution for any unitarily invariant norm (Fan & Hoffman, 1955).

Lemma 1 $A = UH \in \mathbb{R}^{m\times n}$ $(m \geq n)$. *For any unitarily invariant norm,*

$$\frac{\|A^TA - I\|}{\|A\|_2 + 1} \leq \|A - U\| \leq \|A^TA - I\|.$$

Proof. Take norms in the equations

$$\begin{aligned} A^TA - I &= (A - U)^T(A + U), \\ (A - U)^TU &= (A^TA - I)(H + I)^{-1}. \end{aligned}$$

Rotation OPP

Wahba (1965): “A Least Squares Estimate of Satellite Attitude”.

Given $A, B \in \mathbb{R}^{m \times n}$,

$$\min\left\{\|A - BQ\|_F : Q^T Q = I_n,\ \det(Q) = 1\right\}.$$

As before, need to maximize $\operatorname{trace}(Q^T B^T A)$. Let $B^T A = U \Sigma V^T$ be an SVD. Then

$$\begin{aligned}
\operatorname{trace}(Q^T B^T A) &= \operatorname{trace}(Q^T U \Sigma V^T) \\
&= \operatorname{trace}(V^T Q^T U \cdot \Sigma) \\
&=: \operatorname{trace}(Z\Sigma).
\end{aligned}$$

Now

$$\det(Z) = \det(V^T U) \det(Q) = \det(V^T U) = \pm 1.$$

If $\det(V^T U) = 1$, max is attained for $Z = I$; otherwise for

$$Z = \operatorname{diag}(1, 1, \ldots, -1).$$

Permutation Procrustes Problem

Gower (1984): Given $A, B \in \mathbb{R}^{m \times n}$,

$$\min\{\|A - BQ\|_F : Q \text{ a permutation matrix}\}.$$

Do A and B represent the same (or similar) objects in a different order?

Again, want to maximize trace$(Q^T B^T A)$ =: trace(PC). Generalizing P to be *doubly stochastic*, we have

$$\text{maximize} \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} c_{ij},$$

subject to $p_{ij} \geq 0$ and

$$\sum_{j=1}^{n} p_{ij} = 1, \quad i = 1{:}\,n, \qquad \sum_{i=1}^{n} p_{ij} = 1, \quad j = 1{:}\,n.$$

Max must occur at a vertex, which corresponds to a permutation matrix.

This is a linear programming problem, in fact an assignment problem, Can be solved by the *Hungarian method.*

Two-Sided OPP

Schönemann (1966): Given $A, B \in \mathbb{R}^{m\times n}$,

$$\min\{\|A - PBQ\|_F : P^TP = I_m,\ Q^TQ = I_n\}.$$

$$\|A-PBQ\|_F^2 = \mathrm{trace}(A^TA - 2Q^TB^TP^TA + B^TB),$$

so need to maximize $\mathrm{trace}(Q^TB^TP^TA)$.

Answer: if $A = U_A\Sigma_AV_A$ and $B = U_B\Sigma_BV_B$, then

$$P = U_AU_B^T, \qquad Q = V_B^TV_A,$$

so that

$$\begin{aligned}\|A - PBQ\|_F &= \|U_A(\Sigma_A - \Sigma_B)V_A\|_F = \|\Sigma_A - \Sigma_B\|_F \\ &= \Big(\sum_{i=1}^{n}(\sigma_i(A) - \sigma_i(B))^2\Big)^{1/2}.\end{aligned}$$

Symmetric Procrustes Problem (SPP)

Larson (1966), Brock (1968): Given $A, B \in \mathbb{R}^{m\times n}$ $(m \geq n)$

$$\min\left\{ \|A - BX\|_F : X = X^T \right\}.$$

Various analogies with the least squares problem. Let S = set of minimizers.

- S is convex.
- $X \in S$ iff $X = X^T$ and
$$A^T AX + XA^T A = A^T B + B^T A.$$
- S has a unique element of minimal Frobenius norm.

Can solve SPP using SVD of B:

$$\begin{aligned}
\|A - BX\|_F^2 &= \left\| A - U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T X \right\|_F^2 \\
&= \left\| U^T A V - \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T X V \right\|_F^2 \\
&= \left\| C - \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} Y \right\|_F^2 \\
&= \|C_1 - \Sigma Y\|_F^2 + \|C_2\|_F^2, \quad C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}.
\end{aligned}$$

Minimum norm solution easily found to be given by, with $r = \operatorname{rank}(A)$,

$$y_{ij} = \begin{cases} 0, & i > r \text{ and } j > r, \\ \dfrac{\sigma_i c_{ij} + \sigma_j c_{ji}}{\sigma_i^2 + \sigma_j^2}, & \text{otherwise.} \end{cases}$$

Numerical Example

$\min \|A - BX\|_F$:

$$B = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

OPP:

$$X = \begin{bmatrix} 0.2919 & -0.6052 & -0.7406 \\ -0.4698 & 0.5838 & -0.6622 \\ 0.8331 & 0.5412 & -0.1139 \end{bmatrix}, \quad \min = 1.7323.$$

$\det(X) = 1.$

SPP:

$$X = \begin{bmatrix} 0.3793 & -0.6207 & 0.1034 \\ -0.6207 & 0.6552 & -0.0690 \\ 0.1034 & -0.0690 & -0.0345 \end{bmatrix}, \quad \min = 1.8099.$$

Unconstrained (least squares):

$$X = \begin{bmatrix} 0 & 1 & -2 \\ -1 & 2 & -2 \\ 0 & 1 & -1 \end{bmatrix}, \quad \min = 0.$$

Convex Set Case

Andersson and Elfving (1993) consider Procrustes problems of the form

$$\min_{X \in C} \|A - BX\|_F, \quad C = \text{convex set}.$$

Particular choices:

- $X \geq 0$ (nonnegative elements),
- $X = X^T$ and $X \geq 0$,
- X symmetric positive semidefinite.

A & E derive optimality conditions and use projected gradient methods to compute numerical solutions.

General Two-Sided Procrustes Problem

Given $A, B \in \mathbb{R}^{m\times n}$ and sets S and T,

$$\min\left\{\|A - XBY\|_F : X \in S,\ Y \in T\right\}.$$

Possible choices of S and T include Z that are

- orthogonal,
- a permutation,
- symmetric,
- symmetric positive semidefinite,
- diagonal,
- $Z \geq 0$ (nonnegative elements),
- arbitrary (reduces to one-sided problem).

Note, if no constraints on X and Y then,

$$\min = \begin{cases} 0, & \text{rank}(A) \leq \text{rank}(B), \\ \left(\sum_{\text{rank}(B)+1}^{\text{rank}(A)} \sigma_i(A)^2\right)^{1/2}, & \text{otherwise.} \end{cases}$$

Flip-Flop Algorithm

$$\min\{\|A - XBY\|_F : X \in S,\ Y \in T\}.$$

For most of these two-sided problems closed form solutions are not known. One way to try to solve them numerically is as follows.

Guess X_0
for $k = 1, 2, \ldots$ until converged
    $Y_k := \operatorname{argmin}\{\ \|A - X_{k-1}BY\|_F : Y \in T\ \}$.
    $X_k := \operatorname{argmin}\{\ \|A - XBY_k\|_F : X \in S\ \}$.
end

Generates a nonincreasing sequence $\|A - X_kBY_k\|_F$, hence convergent. But

- May not converge to global minimum.
- Convergence is linear, so can be very slow.

Numerical Solution of Procrustes Problems

Many Procrustes problems can be solved using the SVD or polar decomposition.

Can obtain SVD from polar decomposition and vice versa!

Polar decomposition known explicitly for $A \in \mathbb{R}^{2\times 2}$. Uhlig (1981):

$$\begin{aligned} U &= \gamma\big(A + |\det(A)|A^{-T}\big), \\ H &= \gamma\big(A^T A + |\det(A)|I\big), \end{aligned}$$

where

$$\gamma = \big|\det\big(A + |\det(A)|A^{-T}\big)\big|^{-1/2}.$$

Some Iterations

To compute U in $A = UH$, with $X_0 = A$:

- Newton:
$$X_{k+1} = \tfrac{1}{2}(X_k + X_k^{-T}).$$
Converges quadratically for any full rank A.

- Newton-Schulz:
$$X_{k+1} = \tfrac{1}{2}X_k(3I - X_k^T X_k),$$
converges for $\|X_0^T X_0 - I\|_2 < 1$.

- Halley's method (cubically convergent):
$$X_{k+1} = X_k(3I + X_k^T X_k)(I + 3X_k^T X_k)^{-1}.$$

- Quartically convergent method:
$$\begin{aligned} X_{k+1} &= 4X_k(I + X_k^T X_k)(I + 6X_k^T X_k + (X_k^T X_k)^2)^{-1} \\ &= (4 \text{ steps of Newton on } X_k)^{-T}. \end{aligned}$$

How to Derive Iterations

▷ Apply Newton's method to $X^T X = I$.

▷ Apply quadrature to the formula

$$U = \frac{2}{\pi} A \int_0^\infty (t^2 I + A^T A)^{-1} dt.$$

(Change variable then use Gauss-Chebyshev quadrature.)

▷ Since $H = (A^T A)^{1/2}$, apply iterations for matrix square root.

Open Problems

- ▷ Find analytic solutions to one or two-sided Procrustes problems.
- ▷ Handle sparsity constraints. E.g., $X = X^T$ with given sparsity pattern.
- ▷ Handle equality constraints: $(A - BQ)_{ij} = 0$, $i \in I$, $j \in J$.
- ▷ Develop better numerical methods for the solution.
- ▷ Updating: cheaply compute new solution when A undergoes a low rank change.
- ▷ Solve Procrustes problems in other norms. (E.g., Watson (1993) for Schatten p-norms.)
- ▷ Develop perturbation theory for Procrustes problems. (E.g., Higham (1988), Söderkvist (1992)).

References

[1] Lars-Erik Andersson and Tommy Elfving. Two constrained Procrustes problems. Report LiTH-MAT-R-1993-39, Department of Mathematics, Linköping University, Sweden, December 1993.

[2] C. A. Beattie and S. W. Smith. Optimal matrix approximants in structural identification. *J. Optimization Theory and Applications*, 74(1):23–56, 1992.

[3] Ky Fan and A. J. Hoffman. Some metric inequalities in the space of matrices. *Proc. Amer. Math. Soc.*, 6:111–116, 1955.

[4] J. C. Gower. Multivariate analysis: Ordination, multidimensional scaling and allied topics. In E. H. Lloyd, editor, *Statistics*, volume VI of *Handbook of Applicable Mathematics*, pages 727–781. John Wiley, Chichester, 1984.

[5] Bert F. Green. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 17(4):429–440, 1952.

[6] Richard J. Hanson and Michael J. Norris. Analysis of measurements based on the singular value decomposition. *SIAM J. Sci. Stat. Comput.*, 2(3):363–373, 1981.

[7] Nicholas J. Higham. Computing the polar decomposition—with applications. *SIAM J. Sci. Stat. Comput.*, 7(4):1160–1174, October 1986.

[8] Nicholas J. Higham. The symmetric Procrustes problem. *BIT*, 28:133–143, 1988.

[9] Nicholas J. Higham. Matrix nearness problems and applications. In M. J. C. Gover and S. Barnett, editors, *Applications of Matrix Theory*, pages 1–27. Oxford University Press, 1989.

[10] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985. ISBN 0-521-30586-1.

[11] Charles Kenney and Alan J. Laub. On scaling Newton's method for polar decomposition and the matrix sign function. *SIAM J. Matrix Anal. Appl.*, 13(3):688–706, 1992.

[12] Roy Mathias. Perturbation bounds for the polar decomposition. *SIAM J. Matrix Anal. Appl.*, 14(2):588–597, 1993.

[13] Pythagoras Papadimitriou. *Parallel Solution of SVD-Related Problems, With Applications*. PhD thesis, University of Manchester, England, October 1993.

[14] Peter H. Schönemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

[15] Suzanne Weaver Smith, John R. Baker, Mohamed Kaouk, and David C. Zimmerman. Mode shape expansion for visualization and model correlation. In L. Meirovitch, editor, *Dynamics and Control of Large Structures*, pages 385–396. Proceedings of the Ninth VPI&SU Symposium held in Blacksburg, Virginia, 1993.

[16] Inge Söderkvist. Perturbation analysis of the orthogonal Procrustes problem. *BIT*. To appear.

[17] Inge Söderkvist and Per-Åke Wedin. A practical condition number for a configuration of points defining a rigid body movement. Report UMINF-91.29, Institute of Information Processing, University of Umeå, Sweden, December 1991.

[18] Frank Uhlig. Explicit polar decomposition and a near-characteristic polynomial: The 2×2 case. *Linear Algebra and Appl.*, 38:239–249, 1981.

[19] G. Wahba. Problem 65-1: A least squares estimate of satellite attitude. *SIAM Review*, 7:409, 1965. Solutions in vol. 8, 1966, 384–386.

[20] G. A. Watson. Solving generalizations of orthogonal Procrustes problems. *WSSIA A (World Scientific)*, 2:413–426, 1993.