Chapter 6

Inverse Problem in the Schur-Horn Theorem

- Overview
- Schur-Horn theorem
- Lift and projection
- A projected gradient method
- Convergence
- Numerical experiment

Overview

- Given vectors $a = [a_i], \lambda = [\lambda_i] \in R^n$, a majorizes λ if and only if
  - ◇ Arranged in increasing order:

$$a_{j_1} \leq \ldots \leq a_{j_n},$$
$$\lambda_{m_1} \leq \ldots \leq \lambda_{m_n};$$

  - ◇ For all $k = 1, 2, \ldots, n$

$$\sum_{i=1}^{k} a_{j_i} \geq \sum_{i=1}^{k} \lambda_{m_i};$$

  - ◇ Equality holds for $k = n$.
- Majororization theory has important applications (Marshall el al., 79, Arnold '87).
- Would like to construct a Hermitian matrix with specified diagonal entries and eigenvalues.
  - ◇ Can this be done?
  - ◇ How to do it?

- Two methods are proposed:
  - ◇ Lift and projection method
    - ▷ Iterative approach
    - ▷ Linear convergence
    - ▷ Connects to the Wielandt-Hoffman theorem.
  - ◇ Projected gradient method
    - ▷ Continuous approach
    - ▷ Easy to implement
    - ▷ Offers a new proof of existence.

Schur-Horn Theorem

- The Theorem: Hermitian matrix H with eigenvalues λ and diagonal entries a if and only if a majorizes λ.
  - ◇ The known proof is not constructive.
- An inverse eigenvalue problem (SHIEP): Construct a Hermitian matrix with given eigenvalues and diagonal entries.
  - ◇ Known as the harder part of the Schur-Horn Theorem.
  - ◇ Far more variable in the SHIEP than constraints $\Rightarrow$ Solution is far from unique.

Notion

- Notation:

$$\begin{aligned}
\mathrm{diag}(M) &= \text{Diagonal matrix from matrix } M \\
\mathrm{diag}(v) &= \text{Diagonal matrix from vector } v \\
\mathcal{T}(a) &:= \{T \in R^{n\times n} | \mathrm{diag}(T) = \mathrm{diag}(a)\} \\
\mathcal{M}(\Lambda) &:= \{Q^T \Lambda Q | Q \in \mathcal{O}(n)\} \\
\Lambda &:= \mathrm{diag}(\lambda) \\
\mathcal{O}(n) &= \text{Orthogonal matrices in } R^{n\times n}.
\end{aligned}$$

- Idea:

$$\min_{T\in\mathcal{T}(a), Z\in\mathcal{M}(\Lambda)} \|T - Z\|$$

  ◇ Find the shortest distance between $\mathcal{T}(a)$ and $\mathcal{M}(\Lambda)$.

  ◇ Schur-Horn Theorem $\Rightarrow \mathcal{T}(a) \cap \mathcal{M}(\Lambda) \neq \emptyset$.

  ◇ SHIEP $\equiv$ Find the intersection.

SHIEP versus PIEP

- PIEP:
  - ◇ Given symmetric matrices $A_0, A_1, \ldots, A_n \in R^{n\times n}$ and $\lambda \in R^n$,
  - ◇ Find values of $c := (c_1, \ldots, c_n)^T \in R^n$ such that eigenvalues of

$$A(c) := A_0 + c_1 A_1 + \ldots + c_n A_n$$

    are precisely λ.
- SHIEP in terms of PIEP?
  - ◇ Needs to specify A_i a priori so that a SHIEP solution may be written in the form of a PIEP.
  - ◇ Not easy because off-diagonal elements are free and too many.
  - ◇ Numerical techniques proposed for PIEP are not directly applicable for SHIEP unless A_i are properly selected.

Structure SHIEP

- Totally $2n-1$ given data elements — a and λ.
- Sensible to restrict the structure of the matrix, say a Jacobi matrix?
  - ◇ Interesting, but
  - ◇ An example: No real numbers b_1, b_2 such that

$$\begin{bmatrix} 1 & b_1 & 0 \\ b_1 & 2 & b_2 \\ 0 & b_2 & 3 \end{bmatrix}$$

has eigenvalues $\{-5, -4, 15\}$.

Lift and Projection

- Alternate between $\mathcal{T}$ and $\mathcal{M}(\Lambda)$ in the following way:
  - ◇ A lift: From $T^{(k)} \in \mathcal{T}$, find $Z^{(k)} \in \mathcal{M}(\Lambda)$ such that
$$\|T^{(k)} - Z^{(k)}\| = \operatorname{dist}(T^{(k)}, \mathcal{M}(\Lambda)).$$
  - ◇ A projection: Find $T^{(k+1)} \in \mathcal{T}$ such that
$$\|T^{(k+1)} - Z^{(k)}\| = \operatorname{dist}(\mathcal{T}, Z^{(k)}).$$

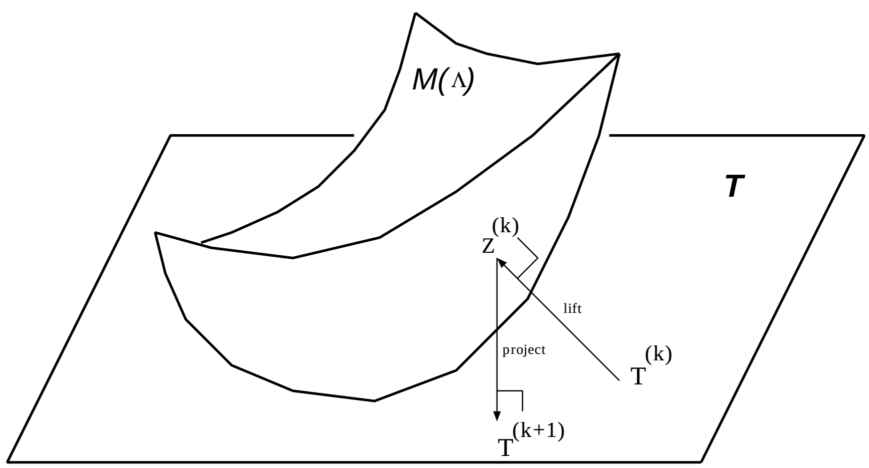

Figure 1: Geometric sketch of lifting and projection.

Calculation

- Projection is easy.
  - ◇ If $T = [t_{ij}] = P(Z = [z_{ij}])$ onto $\mathcal{T}$, then

$$t_{ij} := \begin{cases} z_{ij}, & \text{if } i \neq j \\ a_i, & \text{if } i = j. \end{cases}$$

- Lifting is by Wielandt-Hoffman theorem.
  - ◇ Assume Λ and $T \in \mathcal{T}$ have simple spectrum.
    - ▷ Multiple eigenvalues needs only a slight modification.
  - ◇ Spectral decomposition $T = Q^T D Q$.
  - ◇ π = permutation so that $\lambda_{\pi_1}, \ldots, \lambda_{\pi_n}$ and D are in the same algebraic ordering.
  - ◇ Then the lift of T onto $\mathcal{M}(\Lambda)$ is

$$Z := Q^T \text{diag}(\lambda_{\pi_1}, \ldots, \lambda_{\pi_n}) Q$$

- Both lifting or projection minimize the distance between a point and a set:

  $$\|T^{(k+1)}-Z^{(k+1)}\|^2 \leq \|T^{(k+1)}-Z^{(k)}\|^2 \leq \|T^{(k)}-Z^{(k)}\|^2.$$

- The lift and projection is a descent method.
- The method is essentially the same as Von Neumann's alternating projection method for convex sets (Cheney '59, Deutsch '83, Boyle et al. '89).
  - ◇ $\mathcal{M}(\Lambda)$ is not convex.
  - ◇ A stationary point is not necessarily in the intersection $\mathcal{T} \cap \mathcal{M}(\Lambda)$.
  - ◇ The proximity map is defined by applying the Wielandt-Hoffman theorem.
  - ◇ Linear convergence.

Gradient Flow

- Solve the problem:

$$\min_{Q\in\mathcal{O}(n)} F(Q) := \frac{1}{2}\|\mathrm{diag}(Q^T\Lambda Q) - \mathrm{diag}(a)\|^2.$$

- Schur-Horn theorem $\Rightarrow$ Existence of a Q at which F vanishes.
- Fréchet derivative of F:

$$\begin{aligned} F'(Q)U &= 2\langle \mathrm{diag}(Q^T\Lambda Q) - \mathrm{diag}(a), \mathrm{diag}(Q^T\Lambda U)\rangle \\ &= 2\langle \mathrm{diag}(Q^T\Lambda Q) - \mathrm{diag}(a), Q^T\Lambda U\rangle \\ &= 2\langle \Lambda Q(\mathrm{diag}(Q^T\Lambda Q) - \mathrm{diag}(a)), U\rangle. \end{aligned}$$

  - ◇ Diagonal matrix in the first entry of the inner product $\Rightarrow$ The second equality.
  - ◇ Adjoint property $\Rightarrow$ The third equality.

- Gradient ∇F can be interpreted as:

$$\nabla F(Q) = 2\Lambda Q\beta(Q)$$

  - ◇ $\beta(Q) := \mathrm{diag}(Q^T\Lambda Q) - \mathrm{diag}(a)$.

- The projected gradient of $\nabla F(Q)$ onto $\mathcal{O}(n)$:

$$g(Q) = Q[Q^T \Lambda Q, \beta(Q)]$$

- The projected Hessian:

$$\langle g'(Q)QK, QK\rangle = \langle \mathrm{diag}[Q^T\Lambda Q, K] - [\beta(Q), K], [Q^T \Lambda Q, K]\rangle.$$

- The steepest descent flow on $\mathcal{O}(n)$:

$$\dot{Q} = -g(Q).$$

- An isospectral flow on $\mathcal{M}(\Lambda)$:

$$\dot{X} = [X, [\alpha(X), X]]$$

  - ⋄ $X := Q^T \Lambda Q$.
  - ⋄ $\alpha(X) := \beta(Q) = \mathrm{diag}(X) - \mathrm{diag}(a)$.
  - ⋄ Reducing the distance between $\mathrm{diag}(X)$ and $\mathrm{diag}(a)$.

- The SHIEP can be solved by integrating the differential equation.

Convergence

- First order necessary condition:

$$[\alpha(X), X] = 0.$$

- Second order necessary condition if $\beta(Q) = 0$:

$$\langle g'(Q)QK, QK\rangle = \|\mathrm{diag}[Q^T \Lambda Q, K]\|q^2 \geq 0$$

for all skew-symmetric matrices K.

- The strict inequality is not true in general.
  - ◇ Denote $\Omega := \mathrm{diag}[X, K] = \mathrm{diag}\{\omega_1, \ldots, \omega_n\}$.
  - ◇ Then

$$\omega_i = \sum_{s=1}^{i-1} x_{si}k_{si} - \sum_{t=i+1}^{n} x_{it}k_{it}.$$

  - ◇ The system $\omega_i = 0$ for $i = 1, \ldots, n$ contains only $n-1$ independent equations in the $\frac{n(n-1)}{2}$ unknowns k_{ij}.
  - ◇ Can find a non-trivial skew symmetric matrix K that makes $\Omega = 0$.

Asymptotically Stable Equilibrium

- If $\beta(Q) \neq 0$ at a stationary point Q, then there exists a skew-symmetric matrix K such that $\langle g'(Q)QK, QK\rangle < 0$.
- If $\beta(Q) \neq 0$ at a stationary point Q, there exists an unstable direction along which F is increasing.
- Converge to an unstable equilibrium point is numerically impossible.
- Only X's such that $\beta(Q) = 0$ are the possible *asymptotically stable* equilibrium points.

Proof of Unstable Manifold

- Assume $\beta(Q) = \text{diag}\{\beta_1 I_{n_1}, \ldots, \beta_k I_{n_k}\}$.
- $[Q^T \Lambda Q, \beta(Q)] = 0 \Rightarrow$

$$X = Q^T \Lambda Q = \text{diag}\{X_{11}, \ldots, X_{kk}\}.$$

  - ⋄ $X_{ii} = n_i \times n_i$ real symmetric matrix.
- Define $E := Q\beta(Q)Q^T$.
- $[\Lambda, E] = 0 \Rightarrow$
  - ⋄ $E = \text{diag}(e_1, \ldots, e_n)$.
  - ⋄ $\{e_1, \ldots, e_n =$ a permutation of elements of $\beta(Q)$.
- Q^T = Matrix of eigenvectors of $X \Rightarrow Q$ has the same block structure as X.

- Consider a skew-symmetric matrix $K = [K_{ij}]$ such that,
  - ◇ Partitioned in the same way as X
  - ◇ $K_{ii} = 0$ for all $i = 1, \ldots, k$.
- Observe
  - ◇ $\mathrm{diag}[Q^T \Lambda Q, K] = 0$.
  - ◇ The projected Hessian:

$$\begin{aligned} & \langle g'(Q)QK, QK\rangle \\ = & -\langle [\beta(Q), K], [Q^T\Lambda Q, K]\rangle \\ = & -\langle E\tilde{K} - \tilde{K}E, \Lambda\tilde{K} - \tilde{K}\Lambda\rangle \\ = & -2\sum_{i<j}(\lambda_i - \lambda_j)(e_i - e_j)\tilde{k}_{ij}^2 \end{aligned}$$

  - ◇ Easy to pick up values of $\tilde{k}_{ij}$ so that

$$\langle g'(Q)QK, QK\rangle < 0.$$

Numerical Experiment

- Initial value:
  - ◇ Cannot use Λ as the initial value.
  - ◇ $X_0 := Q^T \Lambda Q$ with Q a random orthogonal matrix.
- Integrator:
  - ◇ Subroutine ODE
  - ◇ $RELERR = ABSERR = 10^{-12}$.
  - ◇ Check output values at interval of 1.

Example 1

- Test data:

$$a = [4.3792\times10^{-1}, 1.0388\times10^{+0}, 1.5396\times10^{-2}, 1.8609\times10^{+0}, 1.4024\times10^{+0}]$$

$$\lambda = [-1.4169\times10^{+0}, -5.6698\times10^{-1}, 4.3890\times10^{-1}, 1.4162\times10^{+0}, 4.8842\times10^{+0}]$$

- Random orthogonal matrix:

$$\begin{bmatrix} -6.4009\times10^{-1} & -5.3594\times10^{-1} & -1.8454\times10^{-1} & -3.3375\times10^{-2} & -5.1757\times10^{-1} \\ 2.1804\times10^{-1} & -1.2359\times10^{-1} & -5.0336\times10^{-1} & -8.2193\times10^{-1} & 9.0802\times10^{-2} \\ -7.2099\times10^{-1} & 5.6072\times10^{-1} & 1.4302\times10^{-2} & -2.4876\times10^{-1} & 3.2199\times10^{-1} \\ 2.8417\times10^{-3} & -1.9828\times10^{-1} & 8.4401\times10^{-1} & -4.9375\times10^{-1} & -6.7297\times10^{-2} \\ -1.5134\times10^{-1} & -5.8632\times10^{-1} & 3.0406\times10^{-3} & 1.3284\times10^{-1} & 7.8464\times10^{-1} \end{bmatrix}$$

- Limit point: At $t \approx 11$, the gradient flow converges to:

$$\begin{bmatrix} 4.3792\times10^{-1} & 2.6691\times10^{-1} & -1.9178\times10^{-1} & -6.1356\times10^{-1} & -1.5920\times10^{+0} \\ 2.6691\times10^{-1} & 1.0388\times10^{+0} & -7.2845\times10^{-1} & -8.6726\times10^{-1} & -1.9618\times10^{+0} \\ -1.9178\times10^{-1} & -7.2845\times10^{-1} & 1.5396\times10^{-2} & -6.3601\times10^{-1} & 1.6256\times10^{-1} \\ -6.1356\times10^{-1} & -8.6726\times10^{-1} & -6.3601\times10^{-1} & 1.8609\times10^{+0} & 1.5032\times10^{+0} \\ -1.5920\times10^{+0} & -1.9618\times10^{+0} & 1.6256\times10^{-1} & 1.5032\times10^{+0} & 1.4024\times10^{+0} \end{bmatrix}$$

- Different random orthogonal matrix $\Rightarrow$ Different limit point.

Example 2

- Repeat the experiment with $2,000$ test data.
  - ◇ Entries in a and λ are from random symmetric matrices with distribution $\mathcal{N}(0,1)$.
  - ◇ Orthogonal matrices Q are from the QR decomposition of non-symmetric random matrices (Stewart ,80).
- Collect the length of integration required for reaching convergence in each case.
  - ◇ Inherent only to the individual problem data (and the stopping criterion).
  - ◇ Independent of the machine used.
- Histogram:
  - ◇ $\approx 77\%$ of the cases converge with the length of integration less than 7.
  - ◇ $\approx 93\%$ converge with length less than 17.
  - ◇ Maximal length of integration $= 296$.
  - ◇ All $2,000$ cases converge.

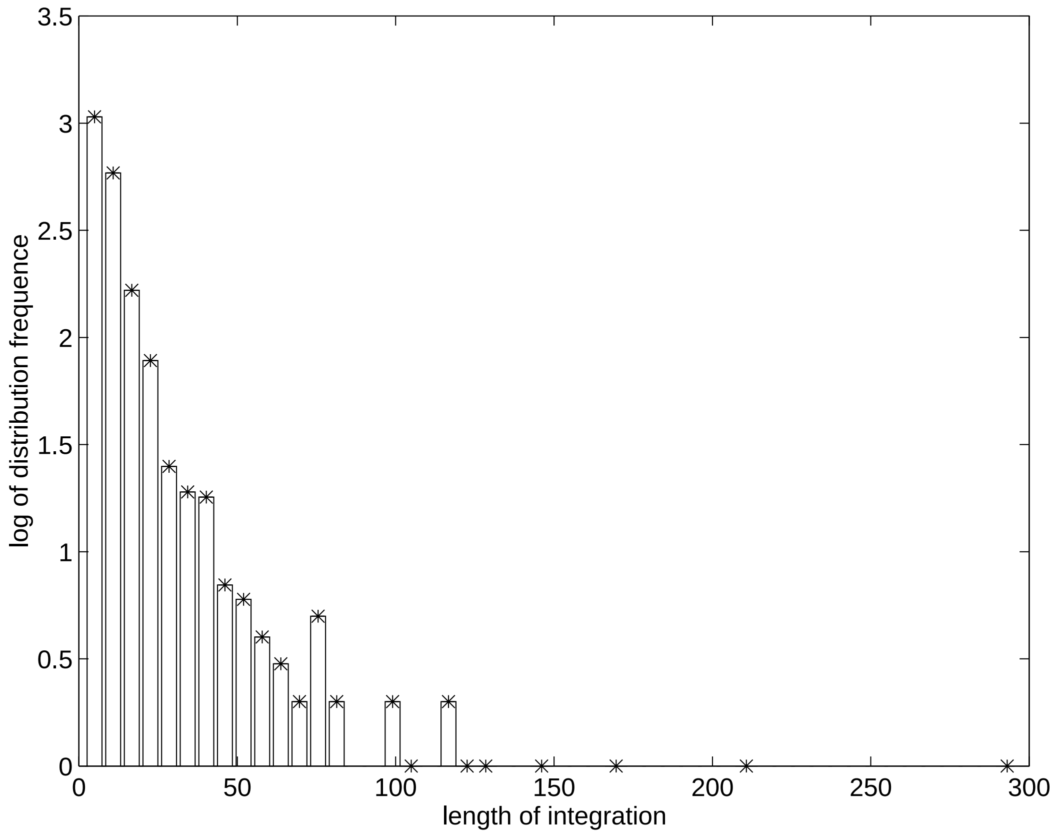

Figure 2: Histogram on the length of integration required for convergence.

Conclusion

- The lift-and-project method makes a connection with the Wielandt-Hoffman theorem.
- The gradient flow method can be integrated by any available ordinary differential equation solver.
- Numerical methods for general PIEP will not work.
- The gradient flow method always converges.
- A constructive proof of the Schur-Horn theorem.