Chapter 7

Stochastic Inverse Eigenvalue Problem

- Overview
- Relation to nonnegative matrices
- Basic formulation
- ASVD flow
- Convergence
- Numerical Experiment

Overview

- Inverse Eigenvalue Problem (IEP):
  - ◇ Reconstruction of matrices from prescribed spectral data.
  - ◇ Spectral data may involve complete or partial information of eigenvalues or eigenvectors.
  - ◇ Often necessary to restrict the construction to special classes of matrices.
- Fundamental questions:
  - ◇ Solvability: Determine a necessary or a sufficient condition under which an IEP has a solution.
  - ◇ Computability: Develop a scheme through which, knowing a priori that the given spectral data are feasible, a matrix can be constructed numerically.

An Example: Parametrized IEP

- Given
  - ◇ Symmetric matrices $A_0, A_1, \ldots, A_n \in R^{n\times n}$;
  - ◇ Real numbers $\lambda_1^* \geq \ldots \geq \lambda_n^*$,
- Find
  - ◇ Values of $c := (c_1, \ldots, c_n)^T \in R^n$
  - ◇ Eigenvalues of the matrix

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

    are precisely $\lambda_1^*, \ldots, \lambda_n^*$.
- Not always does the PIEP has a solution.
- Iterative and continuous methods exist (Friedland et al. '87, Chu et al., '90).

Inverse Stochastic Spectrum Problem

- Construct a stochastic matrix with prescribed spectrum.
  - ◇ Stochastic structure.
  - ◇ No strings of symmetry.
  - ◇ Eigenvalues can appear in complex conjugate pairs.
- A hard problem (Karpelevič '51, Minc '88).
  - ◇ The set Θ_n of points in the complex plane that are eigenvalues of stochastic $n \times n$ matrices is completely characterized.
  - ◇ The Karpelevič theorem characterizes only one complex value a time and does not provide further insights into when two or more points in Θ_n are eigenvalues of the *same* stochastic matrix.

Karpelevič's Theorem

- A number λ is an eigenvalue for a stochastic matrix if and only if it belongs to a region Θ_n such as the one shown below for $n = 4$.

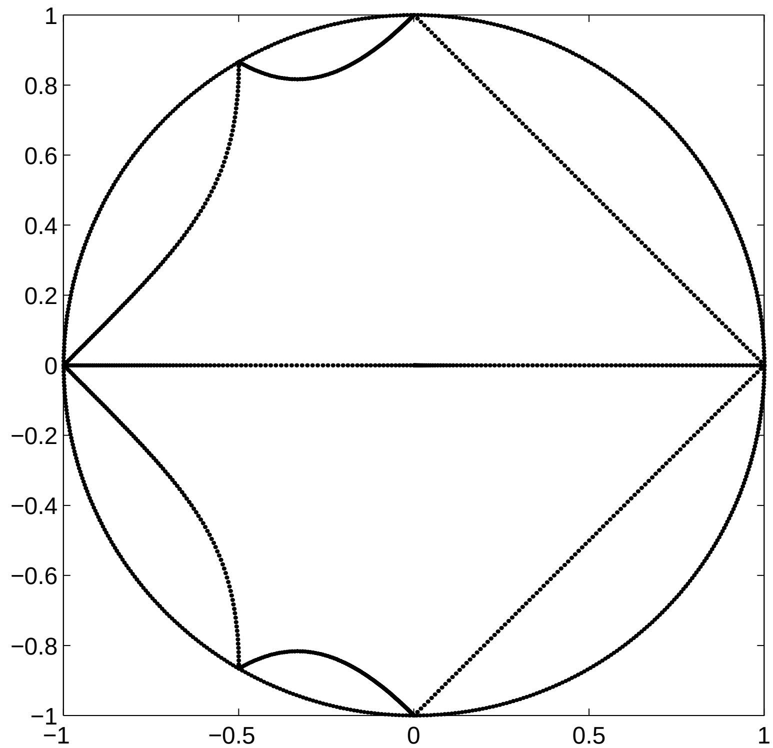

Figure 1: Θ_4 by the Karpelevič Theorem.

- ◇ Region is symmetric about the real axis.
- ◇ The points on the unit circles are given by $e^{2\pi a/b}$ where a and b range over all integers such that $0 \leq a < b \leq n$.
- ◇ The boundary of Θ_n consists of curvilinear arcs connecting these points in circular order. These arcs are characterized by specific parametric equations (Minc, '88).

Relation to Nonnegative Matrices

- A complex nonzero number α is an eigenvalue of a nonnegative matrix with a positive maximal eigenvalue r if and only if α/r is an eigenvalue of a stochastic matrix.
- If A is a nonnegative matrix with positive maximal eigenvalue r and a positive maximal eigenvector x, then $D^{-1}r^{-1}AD$ is a stochastic matrix where $D :=$ diag$\{x_1, \ldots, x_n\}$.
  - ◇ The IEP for nonnegative matrices (NIEP) has received considerable interest in the literature (Berman et al., 94).
  - ◇ Some necessary and a few sufficient conditions for the NIEP are available.
- A continuous method for the NIEP of *symmetric* matrices has been studied (Chu, '91).

Basic Formulation

- Notation:

$$\mathcal{M}(\Lambda) := \{P\Lambda P^{-1} | P \in R^{n\times n} \text{ is nonsingular}\}$$
$$\pi(R^n_+) := \{B \circ B | B \in R^{n\times n}\}$$

  - ◇ Λ = real-valued matrix carrying the spectrum information.
  - ◇ $\circ$ = Hadamard product.

- Idea:
  - ◇ Find the intersection of $\mathcal{M}(\Lambda)$ and $\pi(R^n+)$.
  - ◇ The intersection, if exists, results in a nonnegative matrix isospectral to Λ.
  - ◇ Reduce the nonnegative matrix, if its maximal eigenvector is positive, to a stochastic matrix by diagonal similarity transformation.

Reformulation

$$
\begin{array}{ll}
\text{Minimize} & F(P,R) := \frac{1}{2}||PJP^{-1} - R \circ R||^2 \\
\text{Subject to} & P \in Gl(n),\ R \in gl(n)
\end{array}
$$

- P and R are used as coordinates to maneuver elements in $\mathcal{M}(\Lambda)$ and $\pi(R_+^n)$ to reduce the objective value.
- Feasible domains are open sets.
- A minimum may not exist.

Gradient of F

- Inner product in the product topology:

$$\langle (X_1, Y_1), (X_2, Y_2) \rangle := \langle X_1, X_2 \rangle + \langle Y_1, Y_2 \rangle.$$

- With respect to the product topology:

$$\nabla F(P, R) =$$
$$\big((\Delta(P, R) M(P)^T - M(P)^T \Delta(P, R)) P^{-T},$$
$$-2\Delta(P, R) \circ R \big) .$$

  ◇ Abbreviation:

$$\begin{aligned} M(P) &:= PJP^{-1} \\ \Delta(P, R) &:= M(P) - R \circ R. \end{aligned}$$

Steepest Descent Flow

- Steepest descent flow:

$$\begin{aligned} \frac{dP}{dt} &= [M(P)^T, \Delta(P,R)]P^{-T} \\ \frac{dR}{dt} &= 2\Delta(P,R) \circ R. \end{aligned}$$

- Advantages:
  - ◇ No longer need the projection of $\nabla F(P,R)$ as does in the symmetric case.
  - ◇ The zero structure in the original matrix $R(0)$ is preserved throughout the integration — may be used to explore the possibility of constructing a Markov chain with prescribed linkages and spectrum.
- Disadvantage:
  - ◇ The solution flow $P(t)$ is susceptible to becoming unbounded — a possible frailty.
  - ◇ The involvement of P^{-1} is somewhat worrisome.

ASVD flow

- An analytic singular value decomposition of the path of matrices $P(t)$ is an analytic path of factorizations
$$P(t) = X(t)S(t)Y(t)^T$$
where $X(t)$ and $Y(t)$ are orthogonal and $S(t)$ is diagonal.
- An ASVD exists if $P(t)$ is analytic (Bunse-Gerstner et al., '91).
- The $P(t)$ defined by the differential system is analytic follows from the Cauchy-Kovalevskaya theorem since the coefficients of the vector field are analytic.

New Coordinate System

- The two matrices P and R are used, respectively, as *coordinates* to describe the isospectral matrices and nonnegative matrices.
  - ◇ May have used more dimensions of variables than necessary — does no harm.
  - ◇ When flows $P(t)$ and $R(t)$ are introduced, in a sense a flow in $\mathcal{M}(\Lambda)$ and a flow in $\pi(R_+^n)$ are also introduced.
- The motion of the coordinate P is further described by three other variables X, S, and Y according to the ASVD.
- To produce the steepest descent flow, a coordinate system $(X(t), S(t), Y(t), R(t))$ is eventually imposed on matrices in $\mathcal{M}(\Lambda) \times \pi(R_+^n)$.

Calculating the ASVD

- Differentiate $P(t) = X(t)S(t)Y(t)^T$: (Wright '92):

$$\begin{aligned} \dot{P} &= \dot{X}SY^T + X\dot{S}Y^T + XS\dot{Y}^T \\ X^T\dot{P}Y &= \underbrace{X^T\dot{X}}_{Z}S + \dot{S} + S\underbrace{\dot{Y}^TY}_{W} \end{aligned}$$

  - ◇ Z, W are skew-symmetric matrices.

- Define $Q := X^T\dot{P}Y$.
  - ◇ Q is known since $\dot{P}$ is already specified.
  - ◇ The inverse of $P(t)$ is calculated from

$$P^{-1} = YS^{-1}X^T.$$

  - ◇ The diagonal entries of $S = \text{diag}\{s_1, \ldots, s_n\}$ provide us with information about the proximity of $P(t)$ to singularity.

- Flow for $S(t)$:

$$\frac{dS}{dt} = \mathrm{diag}(Q).$$

- Obtain $W(t)$ and $Z(t)$:

$$\begin{aligned} q_{jk} &= z_{jk}s_k + s_j w_{jk}, \\ -q_{kj} &= z_{jk}s_j + s_k w_{jk}. \end{aligned}$$

  ◇ If $s_k^2 \neq s_j^2$, then

$$z_{jk} = \frac{s_k q_{jk} + s_j q_{kj}}{s_k^2 - s_j^2},$$
$$w_{jk} = \frac{s_j q_{jk} + s_k q_{kj}}{s_j^2 - s_k^2}$$

  for all $j > k$.

- Flow for $X(t)$ and $Y(t)$:

$$\begin{aligned} \frac{dX}{dt} &= XZ. \\ \frac{dY}{dt} &= YW. \end{aligned}$$

- The flow is now ready to be integrated by any IVP solvers.

Convergence

- The approach fails only when:
  - ◇ $P(t)$ becomes singular in finite time — requires a restart.
  - ◇ $F(P(t), R(t))$ converges to a nonzero constant — a LS local solution is found.
- Gradient flows enjoy global convergence:
  - ◇ $G(t) := F(P(t), R(t))$ enjoys the property:

$$\frac{dG}{dt} = -\|\nabla F(P(t), R(t))\|^2 \leq 0$$

    along any solution curve $(P(t), R(t))$.
  - ◇ Suppose $P(t)$ remains nonsingular. Then $G(t)$ converges.

Numerical Experiment

- Integrator: MATLAB ODE SUITE
  - ◇ **ode113** = ABM, PECE, non-stiff system.
  - ◇ **ode15s** = Klopfenstein-Shampine, quasi-constant step size, stiff system.
- Stopping criteria:
  - ◇ ABSERR = RELERR = 10^{-12}.
  - ◇ $\|\Delta(P, R)\| \leq 10^{-9} \Rightarrow$ a stochastic matrix has been found.
  - ◇ Relative improvement of $\Delta(P, R)$ between two consecutive output points $\leq 10^{-9} \Rightarrow$ a LS solution is found.

Example 1

- Spectrum:

$$\{1.0000, -0.2403, 0.1186 \pm 0.1805i, -0.1018\}$$

- Initial values:

$$P_0 = \begin{bmatrix} 0.2002 & 0.4213 & 0.9229 & 0.7243 & 0.4548 \\ 0.6964 & 0.0752 & 0.9361 & 0.2235 & 0.0981 \\ 0.7538 & 0.3620 & 0.2157 & 0.5272 & 0.2637 \\ 0.4366 & 0.3220 & 0.8688 & 0.1729 & 0.8697 \\ 0.8897 & 0.1436 & 0.7097 & 0.5343 & 0.7837 \end{bmatrix}$$

$$R_0 = .8328\mathbf{1}$$

- Limit point:

$$B = \begin{bmatrix} 0.1679 & 0.0522 & 0.4721 & 0.0000 & 0.3078 \\ 0.1436 & 0.1779 & 0.4186 & 0.1901 & 0.0698 \\ 0.0000 & 0.1377 & 0.5291 & 0.3034 & 0.0299 \\ 0.0560 & 0.4690 & 0.2404 & 0.0038 & 0.2309 \\ 0.1931 & 0.1011 & 0.5339 & 0.1553 & 0.0165 \end{bmatrix}.$$

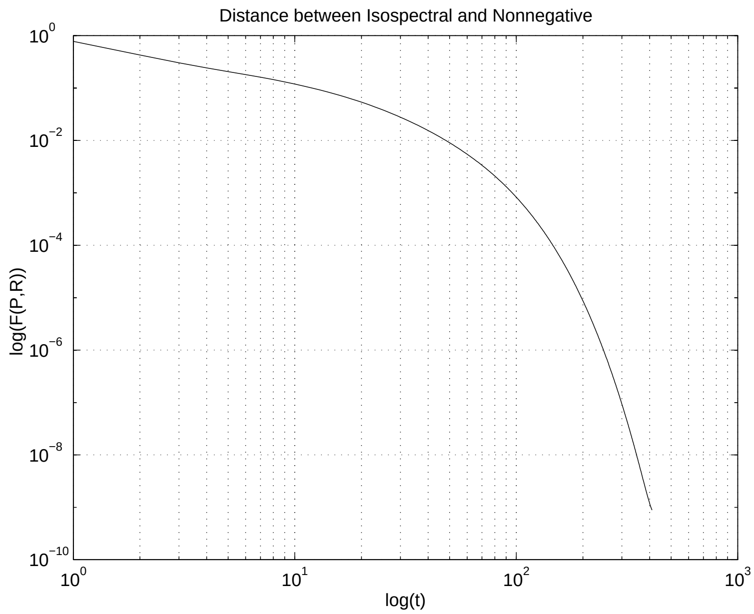

Figure 2: A log-log plot of $F(P(t), R(t))$ versus t for Example 1.

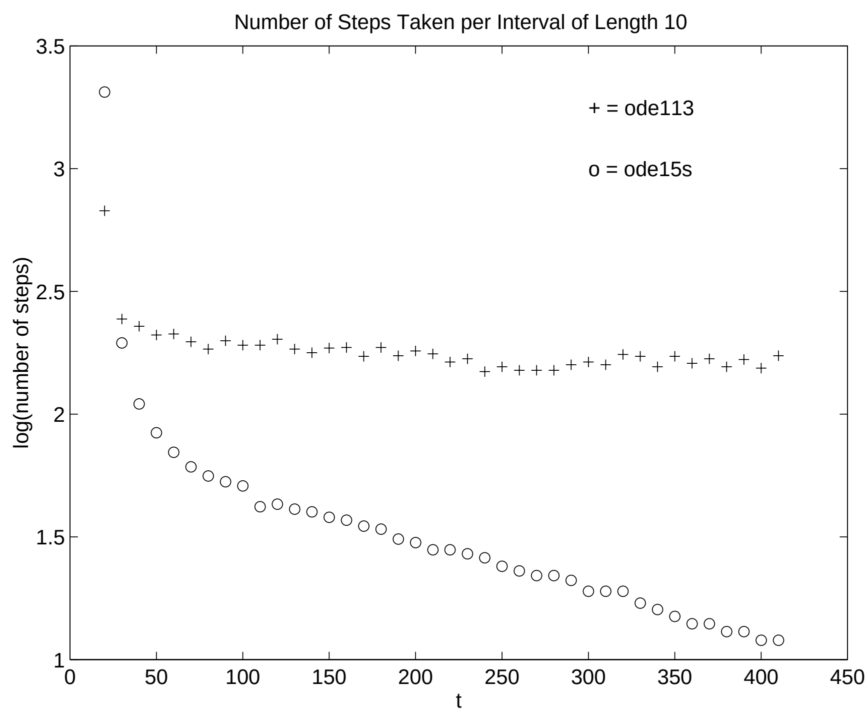

Figure 3: A comparison of steps taken by **ode113** and **ode15s** for Example 1.

- Both solvers work reasonably.
  - ◇ **ode15s** advances with larger step sizes at the cost of solving implicit algebraic equations.
  - ◇ Jacobians are calculated by finite difference. Function calls could be reduced by fewer output points.
- Different initial values lead to different stochastic matrices.

Example 2

- Spectrum:

$$\{1.0000, -0.2608, 0.5046, 0.6438, -0.4483\}$$

- Looking for a Markov chain with ring linkage, i.e., each state is linked at most to its two immediate neighbors.

- Initial values:

$$P_0 = \begin{bmatrix} 0.1825 & 0.7922 & 0.2567 & 0.9260 & 0.9063 \\ 0.1967 & 0.5737 & 0.7206 & 0.5153 & 0.0186 \\ 0.5281 & 0.2994 & 0.9550 & 0.6994 & 0.1383 \\ 0.7948 & 0.6379 & 0.5787 & 0.1005 & 0.9024 \\ 0.5094 & 0.8956 & 0.3954 & 0.6125 & 0.4410 \end{bmatrix}$$

$$R_0 = 0.9210 \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

- Limit point:

$$D = \begin{bmatrix} 0.0000 & 0.3094 & 0 & 0 & 0.6906 \\ 0.0040 & 0.5063 & 0.4896 & 0 & 0 \\ 0 & 0.0000 & 0.5134 & 0.4866 & 0 \\ 0 & 0 & 0.7733 & 0.2246 & 0.0021 \\ 0.4149 & 0 & 0 & 0.3900 & 0.1951 \end{bmatrix}$$

Example 3

- Spectrum

$$\{1.0000, -0.2403, 0.3090 \pm 0.5000i, -0.1018\}$$

- Initial values: same as Example 1 (or modify R_0).
- Slow convergence:

$$E = \begin{bmatrix} 0.3818 & 0.0000 & 0.4568 & 0.0000 & 0.1614 \\ 0.5082 & 0.3314 & 0.0871 & 0.0049 & 0.0684 \\ 0.0000 & 0.0000 & 0.5288 & 0.4712 & 0.0000 \\ 0.0266 & 0.7634 & 0.0292 & 0.0310 & 0.1498 \\ 0.5416 & 0.0524 & 0.3835 & 0.0196 & 0.0029 \end{bmatrix}$$

$$F = \begin{bmatrix} 0.3237 & 0 & 0.4684 & 0 & 0.2079 \\ 0.4742 & 0.3184 & 0.1303 & 0.0007 & 0.0764 \\ 0 & 0.0000 & 0.5231 & 0.4769 & 0 \\ 0.0066 & 0.7536 & 0.0372 & 0.0958 & 0.1068 \\ 0.5441 & 0.0429 & 0.3959 & 0.0022 & 0.0149 \end{bmatrix}$$

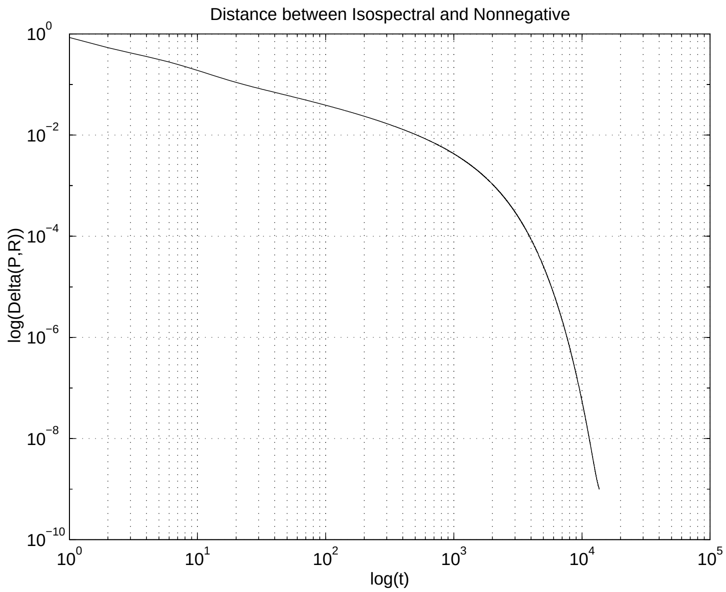

Figure 4: A log-log plot of $F(P(t), R(t))$ versus t for Example 3.

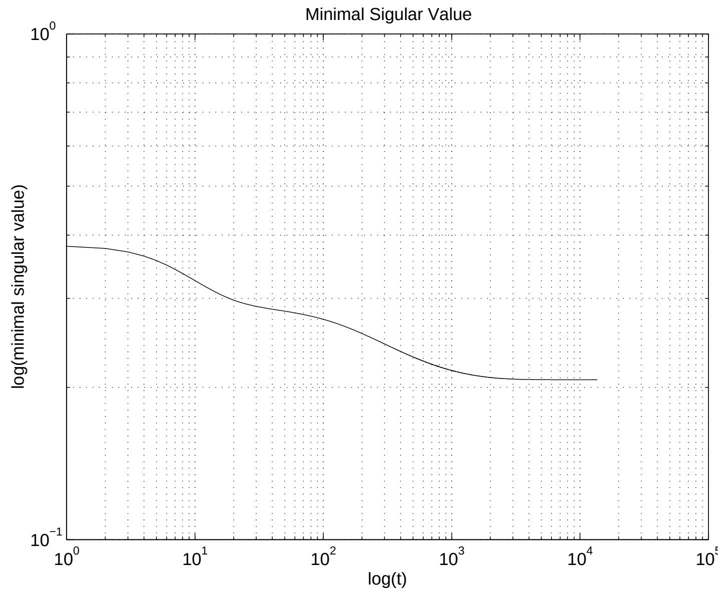

Figure 5: History of the smallest singular value for Example 3.

Conclusion

- The theory of solvability on the StIEP or the NIEP is yet to be developed.
- An ODE approach capable of solving the StIEP or the NIEP numerically, if the prescribed spectrum is feasible, is proposed.
- The method is easy to implement by existing ODE solvers.
- The method can also be used to approximate least squares solutions or linearly structured matrices.