Numerical Matrix Analysis

Notes #25
Arnoldi Iteration

Peter Blomgren
⟨blomgren.peter@gmail.com⟩

Department of Mathematics and Statistics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

http://terminus.sdsu.edu/

Spring 2020

Outline

Introduction, Definitions

The Arnoldi Iteration generates a sequence of matrices from which the eigenvalues λ of $A\vec{x} = \lambda\vec{x}$ for $A \neq A^*$ can easily be computed.

It is a Gram-Schmidt-style orthogonalization iteration for transforming the matrix into Hessenberg form.

	$A \rightsquigarrow QR$	$A \rightsquigarrow QHQ^*$
Orthogonal "Structuring"	Householder	Householder
"Structured" Orthogonalization	Gram-Schmidt	Arnoldi

Definition (Krylov Sequences and Subspaces)

Given a vector $\vec{x}$ and a matrix A, the associated **Krylov vector sequence** is $\{\vec{x}, A\vec{x}, \ldots, A^{k-1}\vec{x}, \ldots\}$, and the corresponding **Krylov subspaces** $K(A, \vec{b}; k) = \operatorname{span}\left(\vec{x}, A\vec{x}, \ldots, A^{k-1}\vec{x}\right)$.

Householder vs. Gram-Schmidt/Arnoldi

In the context of the QR factorization we have the Householder reflections which "triangularize" A by a sequence of orthogonal operations; and the Gram-Schmidt process which Orthogonalize A by a sequence of triangular operations.

The Householder reflection strategy performs better in finite precision.

The Gram-Schmidt process has the advantage that it can be stopped part-way; yielding an incomplete QR-factorization of the first n columns of A.

We have viewed the process of computing the Hessenberg (upper triangular + 1 sub-diagonal) form $A = QHQ^*$ as a two-sided application of Householder reflections. The Arnoldi iteration is the "interruptible" Gram-Schmidt "equivalent."

Hessenberg Reduction, at step n

Problem Structure and Notation

We consider the problem of computing $A \rightsquigarrow QHQ^*$ for $A \in \mathbb{R}^{m \times m}$. All norms $\|\cdot\|$ are 2-norms $\|\cdot\|_2$. n is the iteration number, and we operate in the regime where $n \ll m$.

The complete Hessenberg reduction can be written in the form $A = QHQ^*$, or $AQ = QH$, where $A, Q, M \in \mathbb{R}^{m \times m}$. However, seen at the n^{th} iterative step, we consider the partially computed matrix $Q_n \in \mathbb{R}^{m \times n}$, *i.e.* the first n columns of Q:

$$Q_n = \left[\begin{array}{c|c|c|c} & & & \\ \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \\ & & & \end{array}\right].$$

Hessenberg Reduction, at step n

We also consider the upper $((n+1) \times n)$ block of H, also a Hessenberg matrix

$$\tilde{H}_n = \begin{bmatrix} h_{1,1} & & \cdots & & h_{1,n} \\ h_{2,1} & h_{2,2} & & & \vdots \\ & \ddots & \ddots & & \\ & & & h_{n,n-1} & h_{n,n} \\ & & & & h_{n+1,n} \end{bmatrix},$$

so that

$$AQ_n = Q_{n+1}\tilde{H}_n$$

is well defined, since

$$[m \times n] = [m \times m] \otimes [m \times n] == [m \times (n+1)] \otimes [(n+1) \times n] = [m \times n].$$

Hessenberg Reduction, at step n

The n^{th} column of the relation $AQ_n = Q_{n+1}\tilde{H}_n$ is

$$A\vec{q}_n = \sum_{k=1}^{n+1} h_{k,n}\vec{q}_k,$$

which means that $\vec{q}_{n+1}$ satisfies an $(n+1)$-term recurrence relation involving itself, and previous Krylov basis vectors:

$$\vec{q}_{n+1} = \frac{1}{h_{n+1,h}}\left[A\vec{q}_n - \sum_{k=1}^{n} h_{k,n}\vec{q}_k\right].$$

The Arnoldi iteration implements this recurrence relation.

Algorithm

Algorithm (Arnoldi Iteration)

$\vec{b} \leftarrow \texttt{random}(\mathbb{R}^{m\times 1})$,
$\vec{q}_1 \leftarrow \vec{b}/\|\vec{b}\|$
for $n \in \{1, 2, \ldots\}$ **do**
  $\vec{v} \leftarrow A\vec{q}_n$
  for $j \in \{1, \ldots, n\}$ **do**
    $h_{j,n} \leftarrow \vec{q}_j^* \vec{v}$
    $\vec{v} \leftarrow \vec{v} - h_{j,n}\vec{q}_j$
  end for
  $h_{n+1,n} \leftarrow \|\vec{v}\|$
  $\vec{q}_{n+1} \leftarrow \vec{v}/h_{n+1,n}$
end for

Under the rug: $h_{n+1,n} = 0$

Arnoldi: Simplicity and Power, All-in-One i/ii

The power of the Arnoldi process is the various interpretations that can be made of it (and its by-products), and the algorithms these suggest.

Arnoldi: Generated Krylov Subspaces

From construction, the generated vectors $\{\vec{q}_k\}$ form orthonormal bases for the Krylov subspaces generated by A and $\vec{b}$:

$$\mathcal{K}_n = \text{span}\left(\vec{b}, A\vec{b}, \ldots, A^{n-1}\vec{b}\right) = \text{span}\left(\vec{q}_1, \vec{q}_2, \ldots, \vec{q}_n\right) \subseteq \mathbb{C}^m.$$

Arnoldi: Simplicity and Power, All-in-One i/ii

The $(m \times n)$ Krylov matrix $\mathcal{K}_n$ must have a reduced QR-factorization:

$$\mathcal{K}_n = \left[\begin{array}{c|c|c|c} & & & \\ \vec{b} & A\vec{b} & \cdots & A^{n-1}\vec{b} \\ & & & \end{array} \right] = Q_n R_n,$$

where Q_n is the matrix previously defined ($\vec{q}_k$ are its columns).

Neither $\mathcal{K}_n$, nor R_n can be stably formed during the Arnoldi process, but they hint at why the Arnoldi process leads to effective methods for determining certain (the dominant) eigenvalues of A.

Arnoldi: Simplicity and Power, All-in-One i/ii

The relationship between the Arnoldi iteration, and direct QR factorization of $\mathcal{K}_n$ is reminiscent of the Simultaneous-(Power)-Iteration vs. QR-algorithm approaches:

Straight-forward, unstable	*Subtle, stable*
Simultaneous Iteration $\mathcal{K}_n = Q_n R_n$	QR-algorithm Arnoldi

Arnoldi: Simplicity and Power, All-in-One ii/ii

We can also view the Arnoldi process as a computation of projections onto successive Krylov subspaces.

Note: $(n \times (n+1))$ Identity

The product

$$Q_n^* Q_{n+1} = I = \delta_{k,\ell}, \quad k = 1, \ldots, n, \quad \ell = 1, \ldots, (n+1).$$

Therefore $Q_n^* Q_{n+1} \tilde{H}_n$ is the $(n \times n)$ Hessenberg matrix obtained by removing the last row of $\tilde{H}_n$:

$$H_n = \begin{bmatrix} h_{1,1} & & \cdots & h_{1,n} \\ h_{2,1} & h_{2,2} & & \vdots \\ & \ddots & \ddots & \\ & & h_{n,n-1} & h_{n,n} \end{bmatrix}, \quad \text{and } H_n = Q_n^* A Q_n.$$

Arnoldi: Simplicity and Power, All-in-One ii/ii 2 of 2

This matrix can be interpreted as the orthogonal projection of A onto $\mathcal{K}_n$, represented by the basis $\{\vec{q}_1, \ldots, \vec{q}_n\}$.

This type of projection shows up in a variety of contexts, and is sometimes referred to as the *Rayleigh-Ritz* procedure. It turns out that the diagonal elements is H_n are the Rayleigh coefficients of A with respect to the vectors $\vec{q}_j$.

This projection process is one of the ideas underlying the *Finite Element Method* for solution of PDEs, as well as *spectral methods*.

Since H_n is a projection of A, its eigenvalues $\lambda(H_n)$ are related to the eigenvalues of A; they are referred to as the *Arnoldi eigenvalue estimates*, or *Ritz values* (with respect to $\mathcal{K}_n$) of A.

Arnoldi: Summary (so far)

Theorem

The matrices Q_n generated by the Arnoldi iteration are reduced QR-factors of the Krylov matrix:

$$\mathcal{K}_n = Q_n R_n.$$

The Hessenberg matrices H_n are the corresponding projections

$$H_n = Q_n^* A Q_n,$$

and the successive iterates are related by

$$AQ_n = Q_{n+1} \tilde{H}_n.$$

Arnoldi Eigenvalue Computation Strategy

Strategy

1. Perform Arnoldi Iteration
2. Compute the Eigenvalues of H_n using *e.g.* the QR-algorithm at regular intervals.
   - $\lambda(H_n)$ are the *Arnoldi eigenvalue estimates*, or *Ritz values*.
   - a (growing) subset of $\lambda(H_n)$ typically converge quickly, and are eigenvalues of A.
   - [$\exists$ Movie]
   - Typically, the largest-modulus eigenvalues are located first; and usually (but not always) those are the ones we are interested in.

Some aspects of the convergence can be quantified...

Minimization Problem Over the Monic Polynomials

Let $\vec{x} \in \mathcal{K}_n(A, \vec{b})$, then

$$\vec{x} = \sum_{k=0}^{n-1} c_0 A^k \vec{b} = q(A)\vec{b},$$

where $q(A)$ is a (matrix) polynomial in A.

With

$$\mathcal{P}^n = \{ \text{All Monic Polynomials of Degree } n \},$$

where *monic* means $c_n = 1$, we can state the

Arnoldi (and Lanczos) Approximation Problem (AAP)

Find $p_n \in \mathcal{P}^n$ such that

$$\|p_n(A)\vec{b}\| \text{ is minimized.}$$

Solution to the Minimization Problem

The Arnoldi iteration solves the stated minimization problem:

Theorem

As long as the Arnoldi iteration does not break down (i.e. $\mathcal{K}_n$ is of full rank n), (AAP) has a unique solution p_n; the characteristic minimal polynomial of H_n. *[Proof in Trefethen & Bau, pp. 259–260.]*

Hence, the Ritz values generated by the Arnoldi iteration are the roots of the optimal polynomial.

Since the class of monic polynomials $\mathcal{P}^n$ is invariant with respect to translations $[z \to (z + \alpha)]$, so is the Arnoldi iteration:

Ref: See [MATH 524 (NOTES#8)] for discussions on Cayley–Hamilton Theorem, the Minimal and Characteristic Polynomials.

Invariance Theorem

Theorem

Let the Arnoldi iteration be applied to a matrix $A \in \mathbb{C}^{m \times m}$; each of the following invariance properties hold:

- [Translation-Invariance] *If A is changed to $(A + \sigma I)$ for some $\sigma \in \mathbb{C}$, and $\vec{b}$ is left unchanged, the Ritz values $\{\theta_j\}$ change to $\{\theta_j + \sigma\}$.*
- [Scale-Invariance] *If A is changed to σA for some $\sigma \in \mathbb{C}$, and $\vec{b}$ is left unchanged, the Ritz values $\{\theta_j\}$ change to $\{\sigma\theta_j\}$.*
- [Invariance under Unitary Similarity Transforms] *If A is changed to UAU^* for some unitary matrix U, and $\vec{b}$ is changed to $U\vec{b}$, the Ritz values do not change.*

In all three cases the Ritz vectors, the vectors $Q_n\vec{y}_j$ corresponding to the eigenvectors $\vec{y}_j$ of H_n, do not change.

Invariance Theorem: Discussion

We know (Schur Factorization) that every matrix is similar to an upper triangular matrix.

Therefore, the invariance property shows that we can understand the properties of the Arnoldi iteration by considering upper triangular matrices.

Note: NON-HERMITIAN MATRICES ↔ TRIANGULAR MATRICES, and HERMITIAN MATRICES ↔ DIAGONAL MATRICES. Hence in the non-Hermitian case off-diagonal element complete the description of the matrix action beyond "just" the eigenvalues.

Convergence

Easy Case: *A is diagonalizable, and has $n \ll m$ distinct eigenvalues.* In this case the Arnoldi iteration finds the n eigenvalues exactly, as long as the "seed vector" $\vec{b}$ contains components in directions associated with every eigenvalue; and $\|p_n(A)\vec{b}\| = 0$. [In infinite precision]

General Case: Here, the Ritz values are approximations to eigenvalues, and p_n is a *pseudo-minimal* polynomial; *i.e.* $\|p_n(A)\vec{b}\|$ is small.

The convergence process can be illustrated by a particular level-curve (*lemniscate*) of the polynomial.

Arnoldi Lemniscates

Definition (Lemniscate, Level Set)

A lemniscate is a curve, or collection of curves

$$\{z \in \mathbb{C} : |p(z)| = C\},$$

where $p(z)$ is a polynomial and C is a real constant.

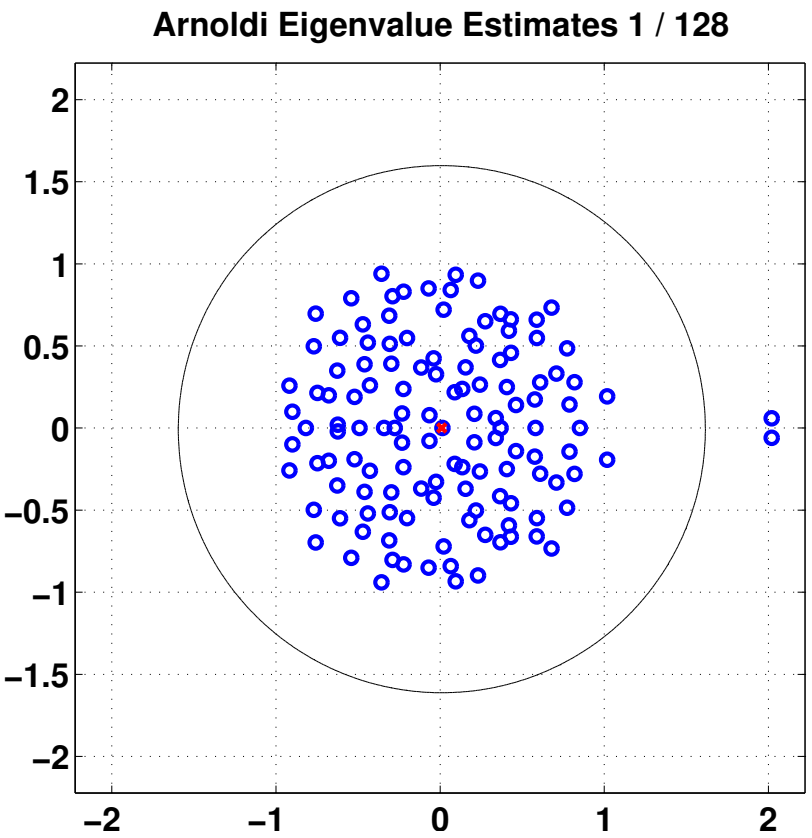
If $p(z)$ is the Arnoldi polynomial $p_n(\cdot)$ for an Arnoldi iteration, and

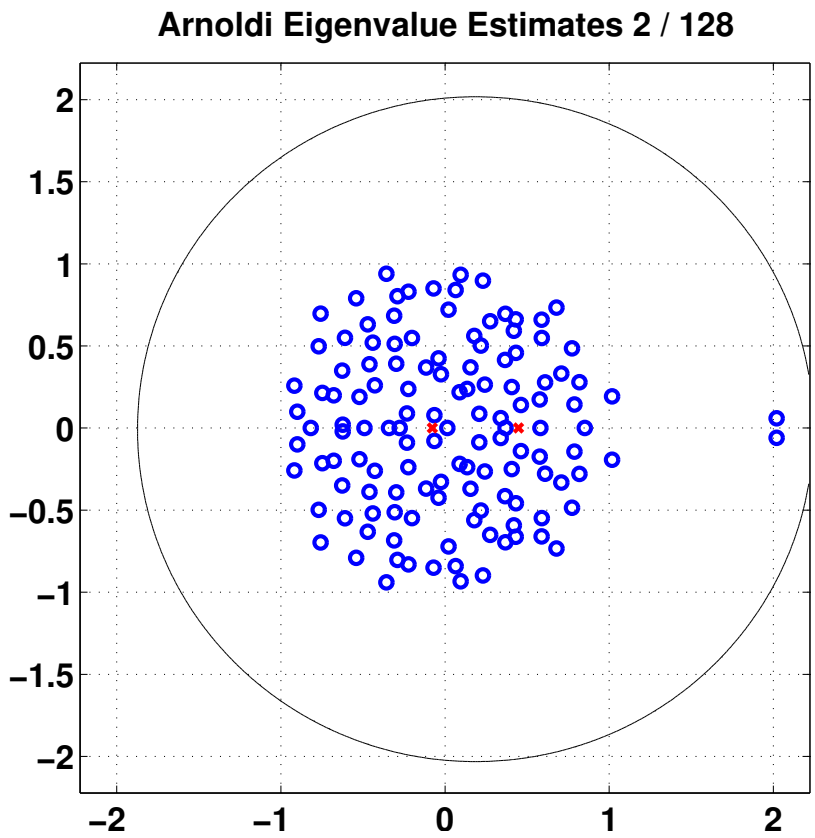
$$C_n = \frac{\|p_n(A)\vec{b}\|}{\|\vec{b}\|},$$

then the curves are called *Arnoldi Lemniscates*.

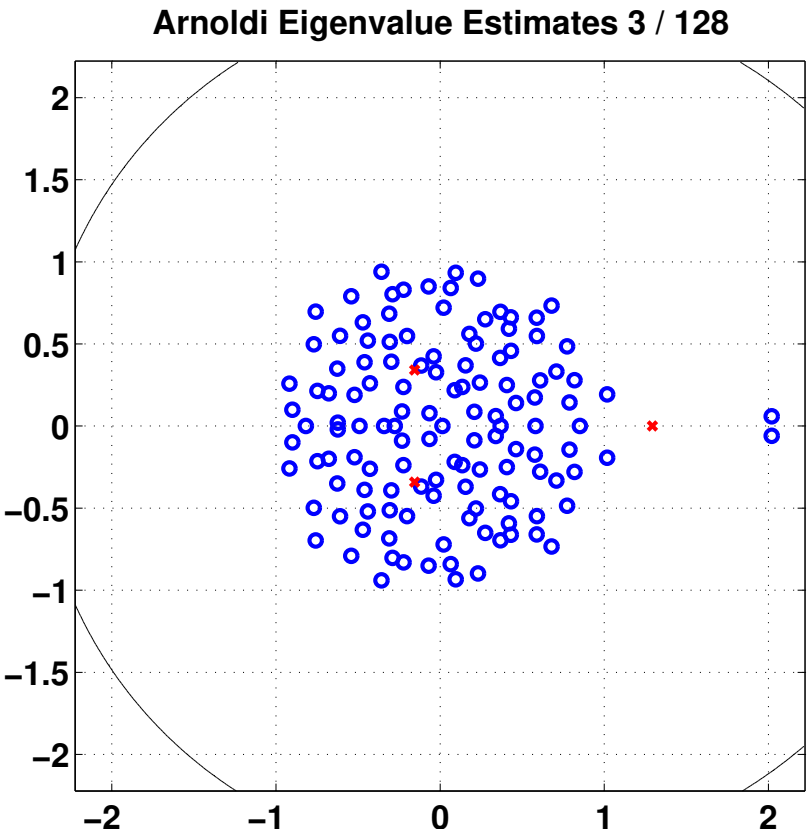
The components of the *Arnoldi Lemniscates* tend to surround the extreme eigenvalues of A and the shrink rapidly to a point (the eigenvalue).

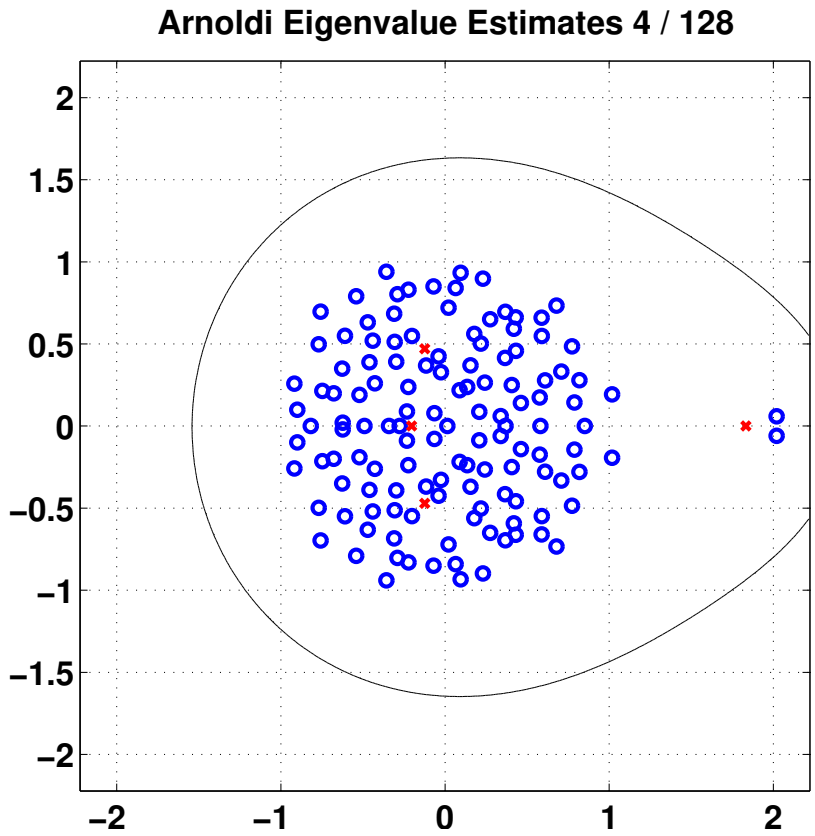
Arnoldi Lemniscates



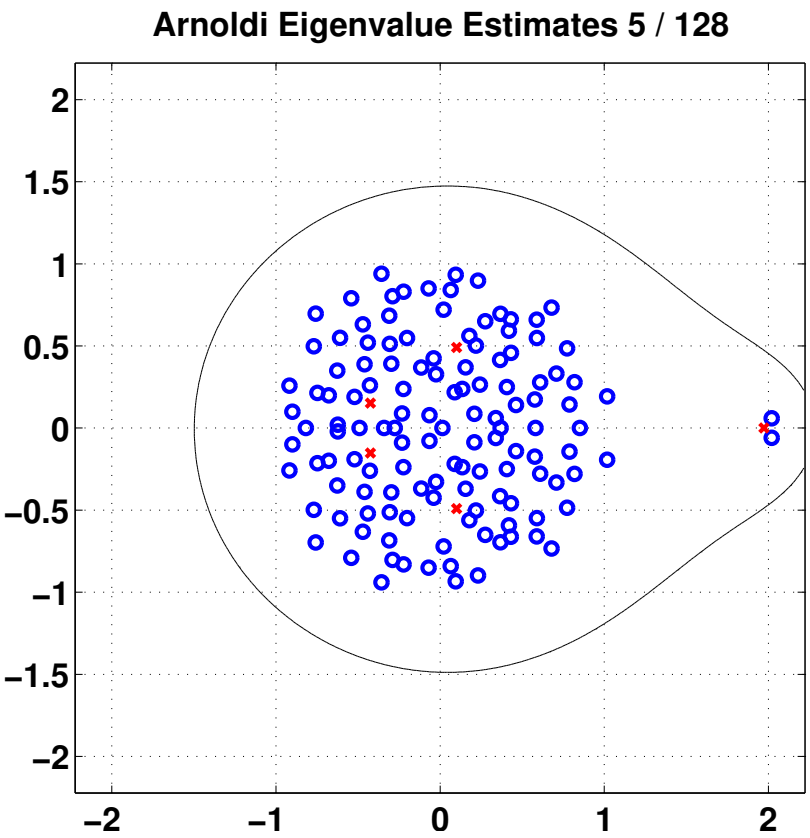


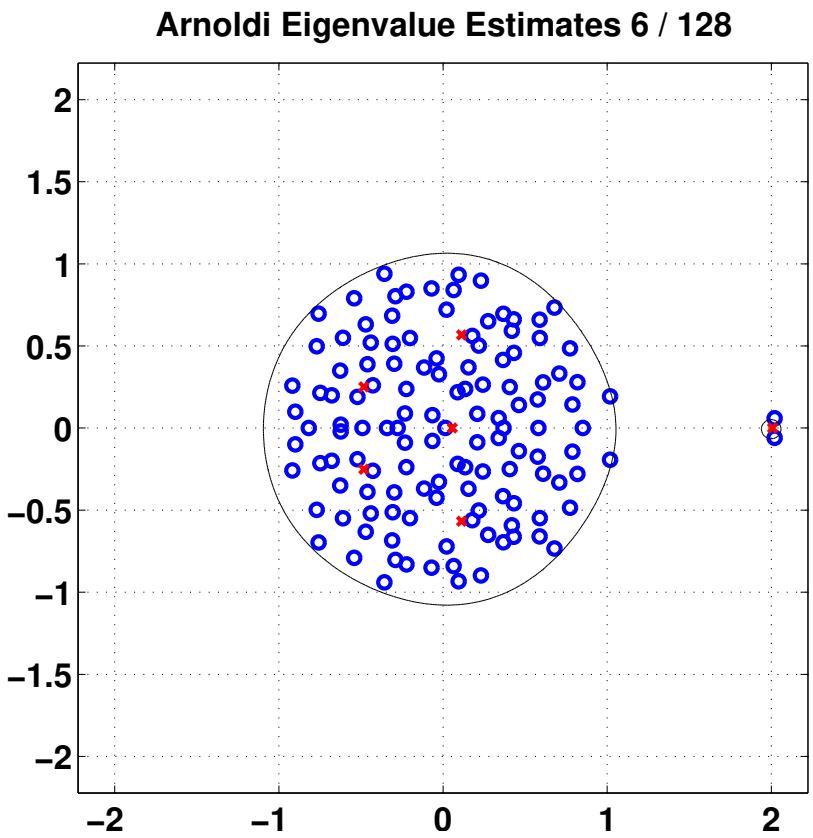
Arnoldi Lemniscates



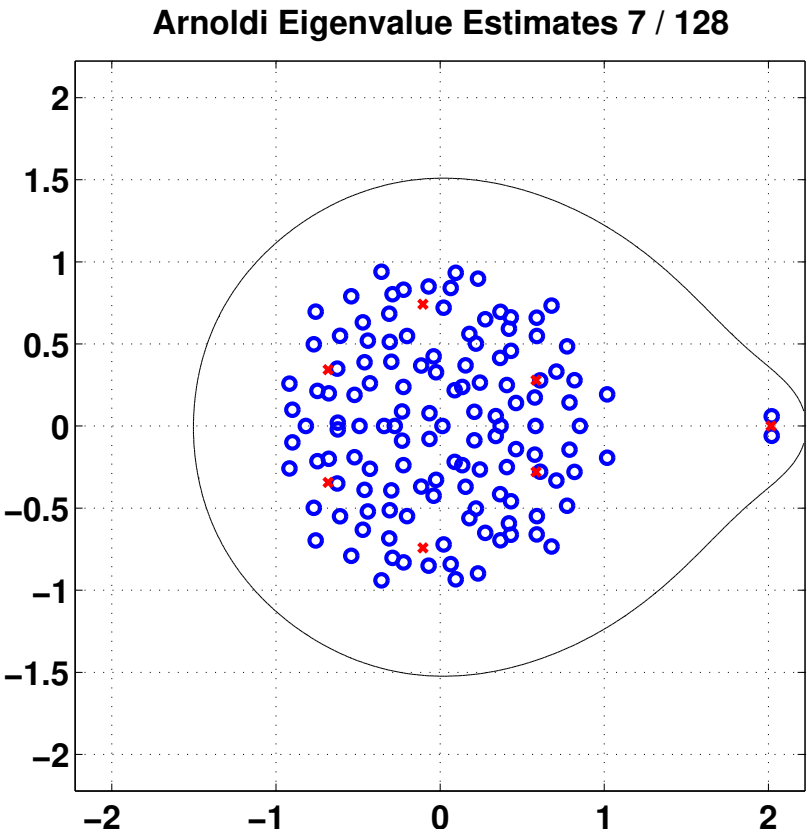


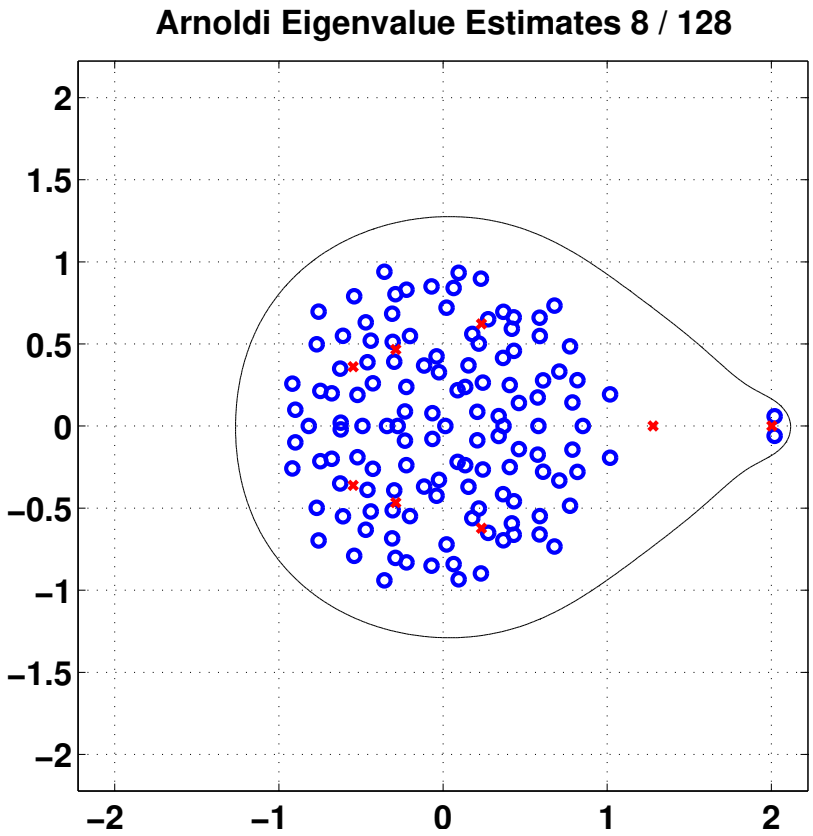
Arnoldi Lemniscates



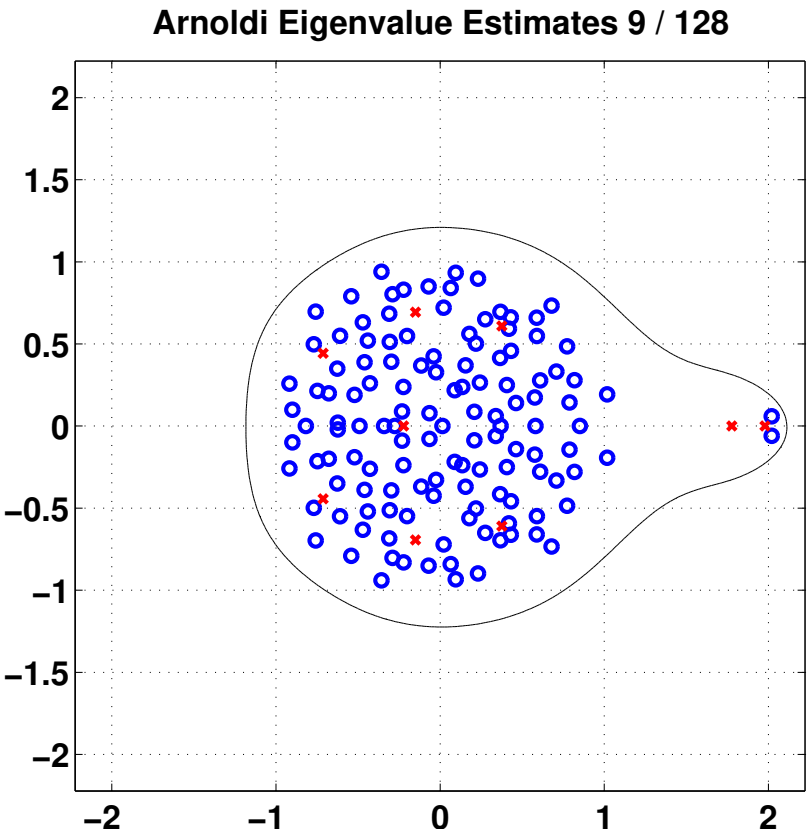


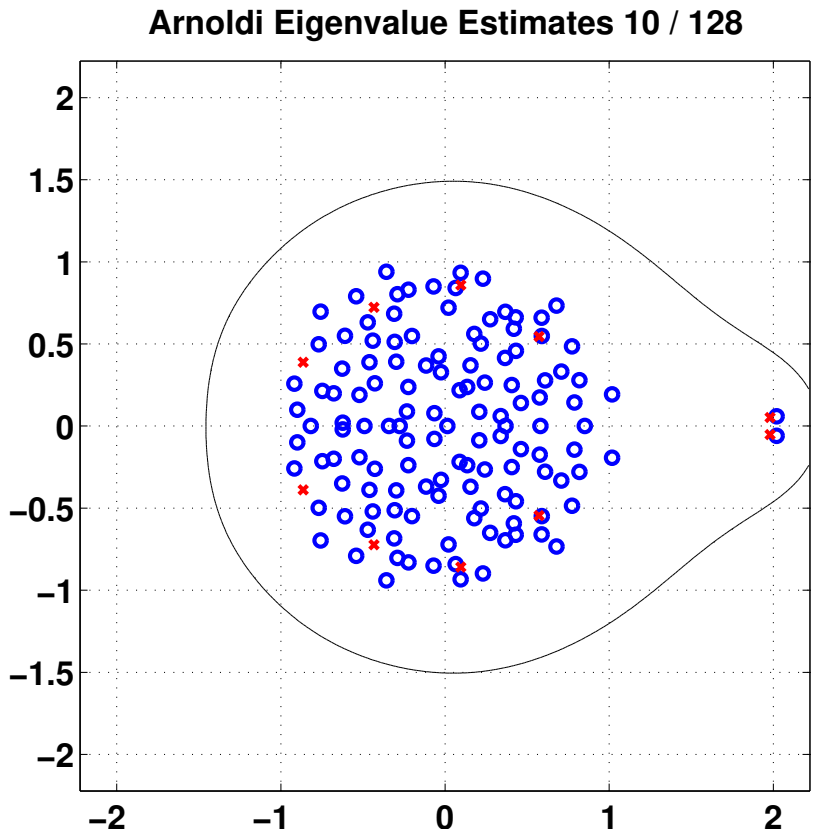
Arnoldi Lemniscates



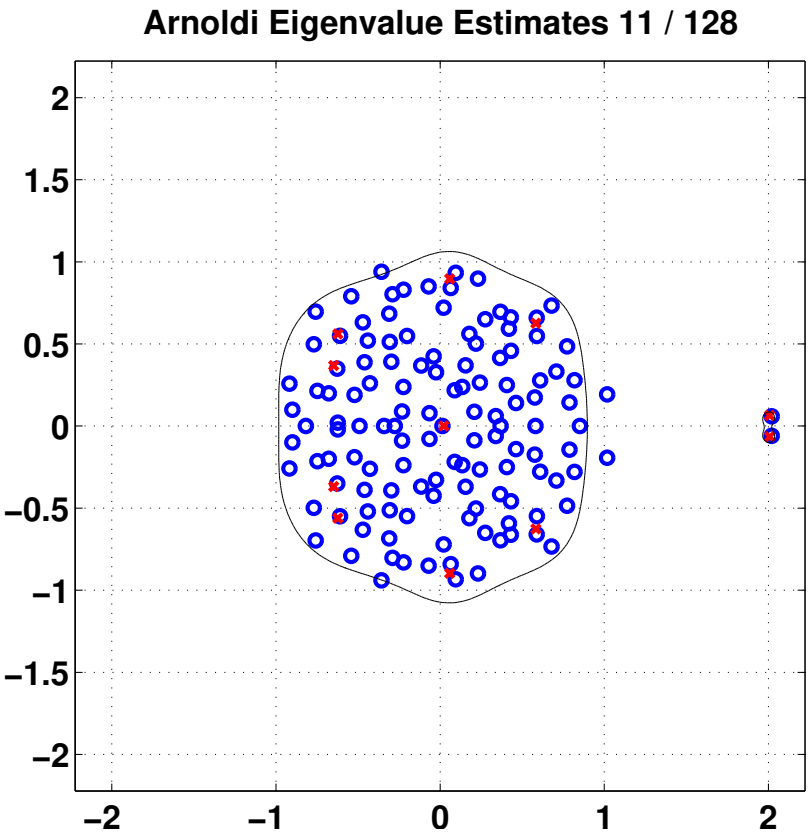


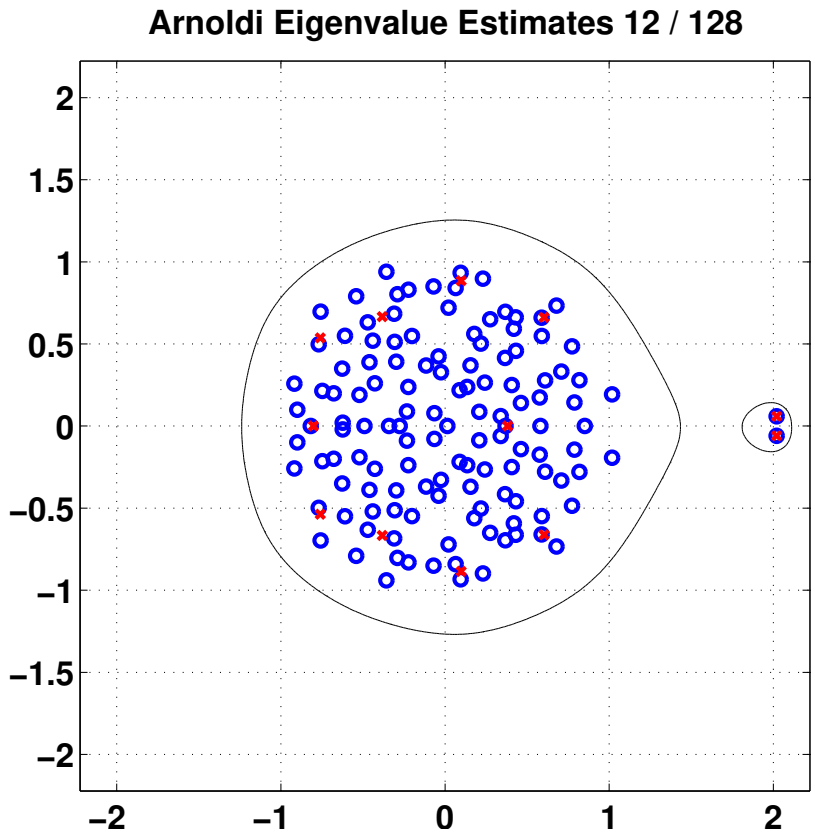
Arnoldi Lemniscates





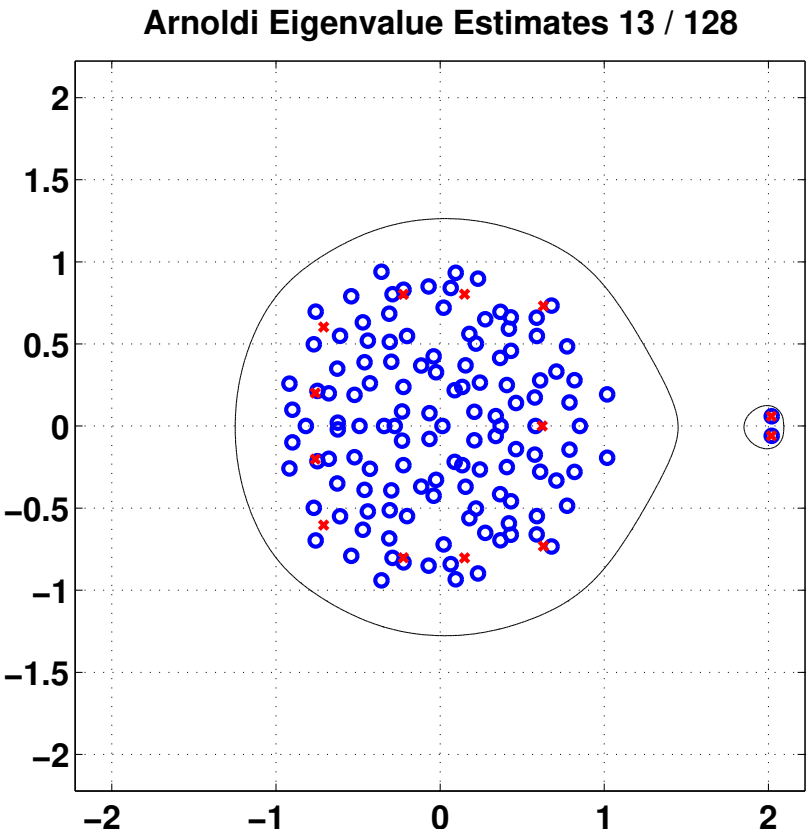
Arnoldi Lemniscates

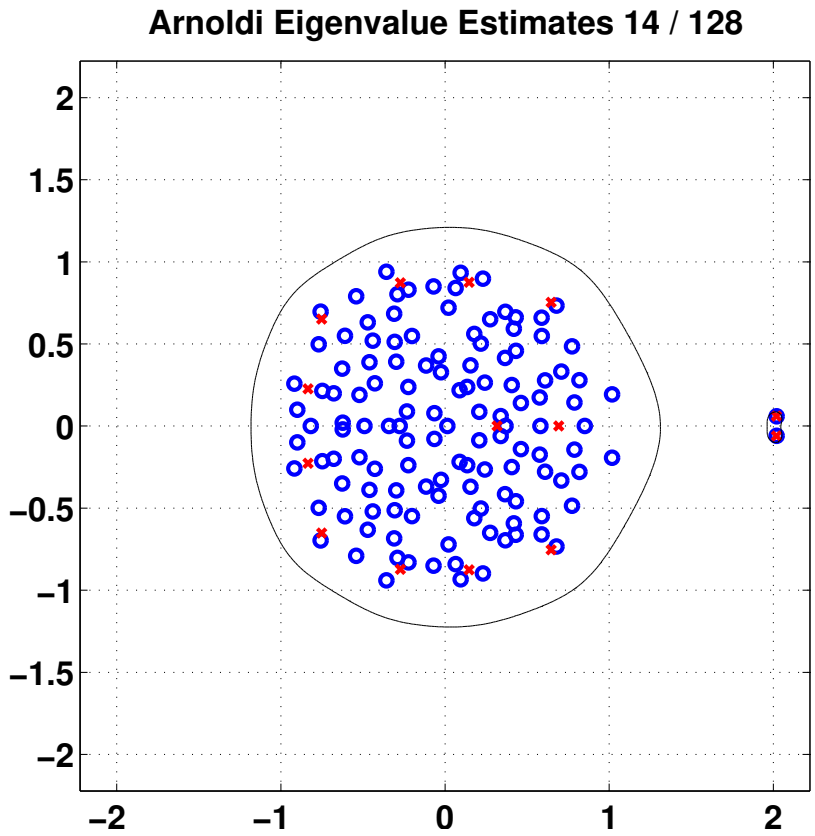




San Diego State University

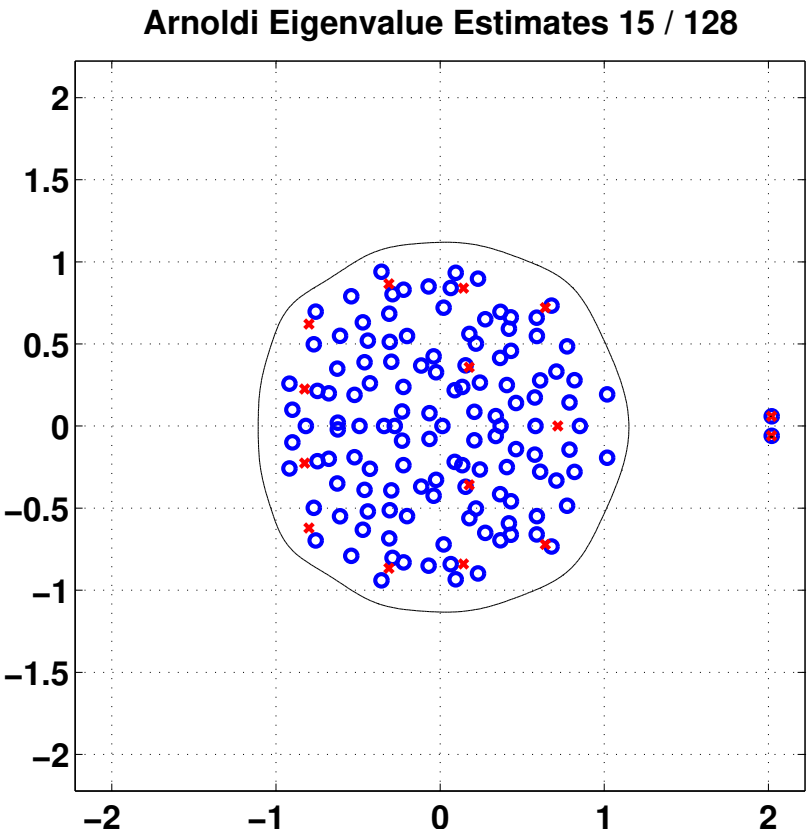
Arnoldi Lemniscates

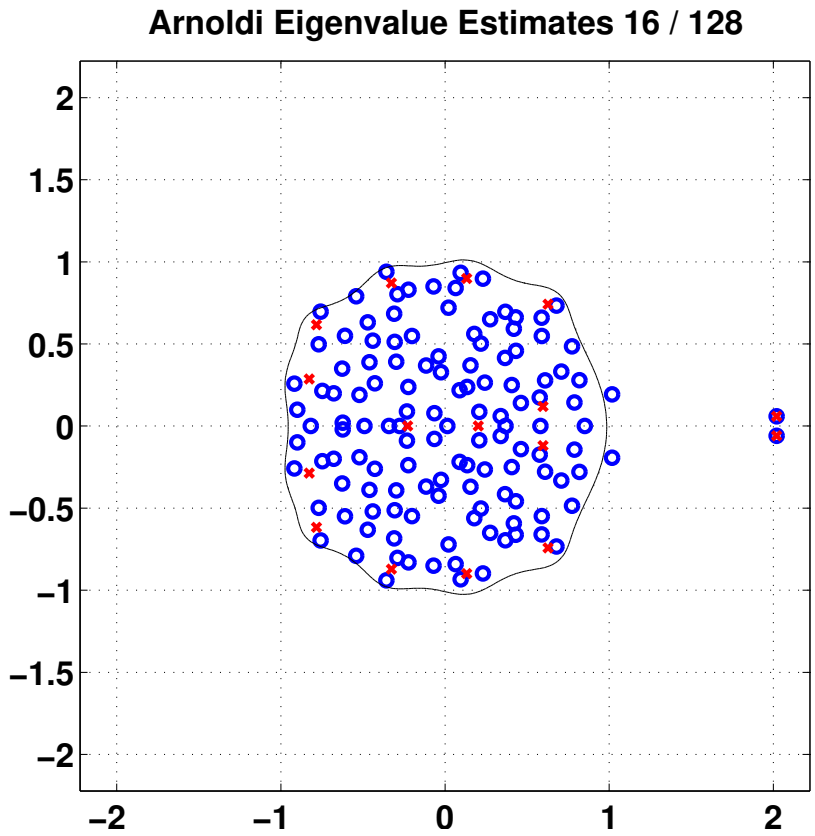




San Diego State University

Arnoldi Lemniscates





Rate of Convergence

The rate of convergence is not well understood in general.

For a single outlying eigenvalue, the convergence rate is typically geometric (in the first few iterations, then faster); *e.g.* Trefethen & Bau show an example where

$$|\lambda^{(n)} - \lambda| \sim \left(\frac{2}{3}\right)^n.$$

Convergence is highly dependent on the shape of the eigenvalue spectrum $\lambda(A)$.