A Fast Nonsymmetric Eigensolver for Structured Matrices

David Bindel

Ming Gu

David Garmire

James Demmel

Shivkumar Chandrasekaran

Outline

- QR iteration
- Rank symmetry and nearly rank-symmetric matrices
- Examples of nearly rank-symmetric matrices
- A fast QR variant
- Stability and balancing
- Implementation and performance

QR iteration

Goal: Convert A to a quasi-upper-triangular matrix T by a sequence of orthogonal similarities.

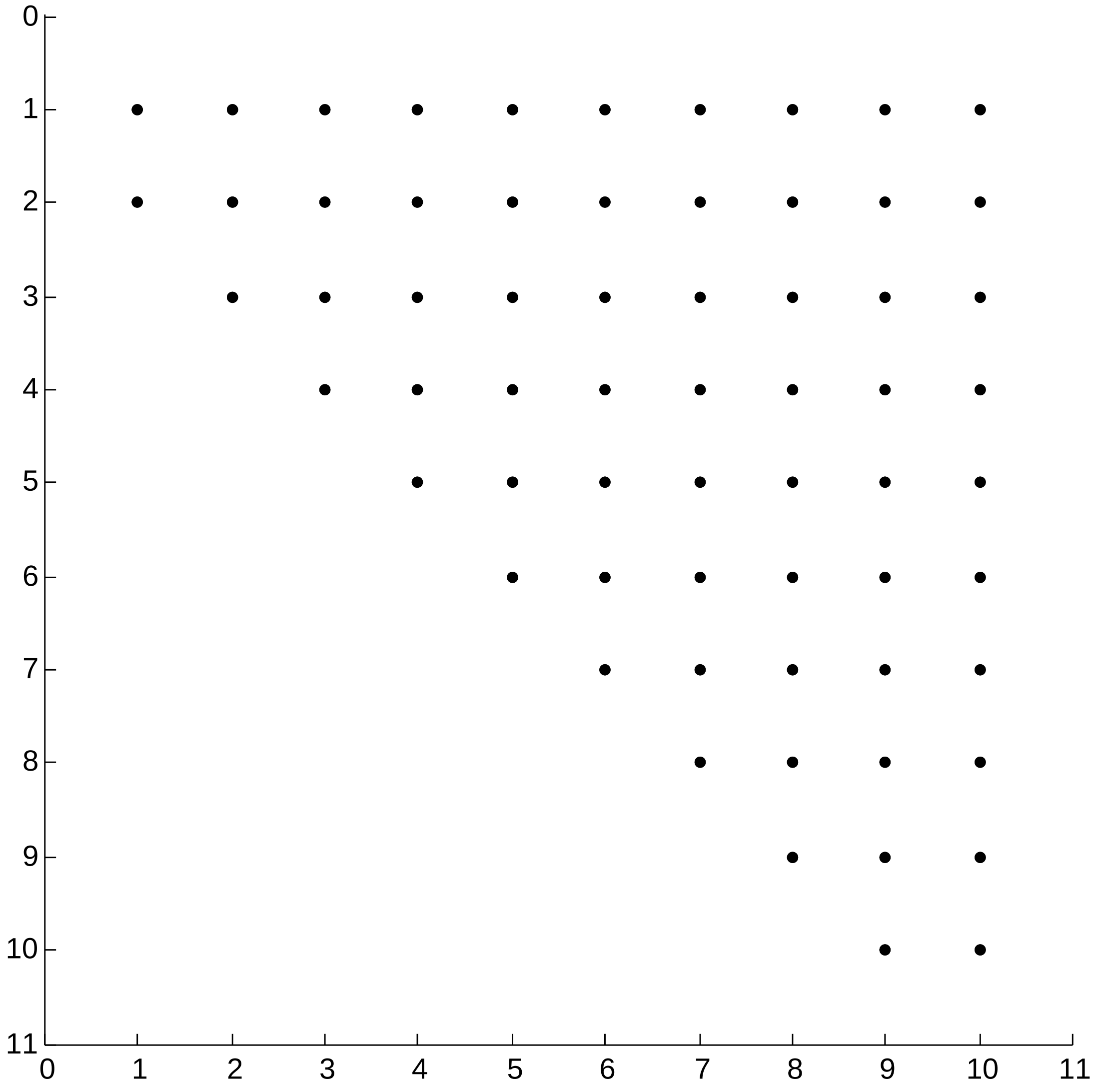
- Choose a shift σ_i
- $A_i - \sigma_i I = Q_i R_i$
- $A_{i+1} = R_i Q_i + \sigma_i I = Q_i^T A_i Q_i$

Cost per iteration: $O(n^3)$. Total cost is $O(n^4)$.

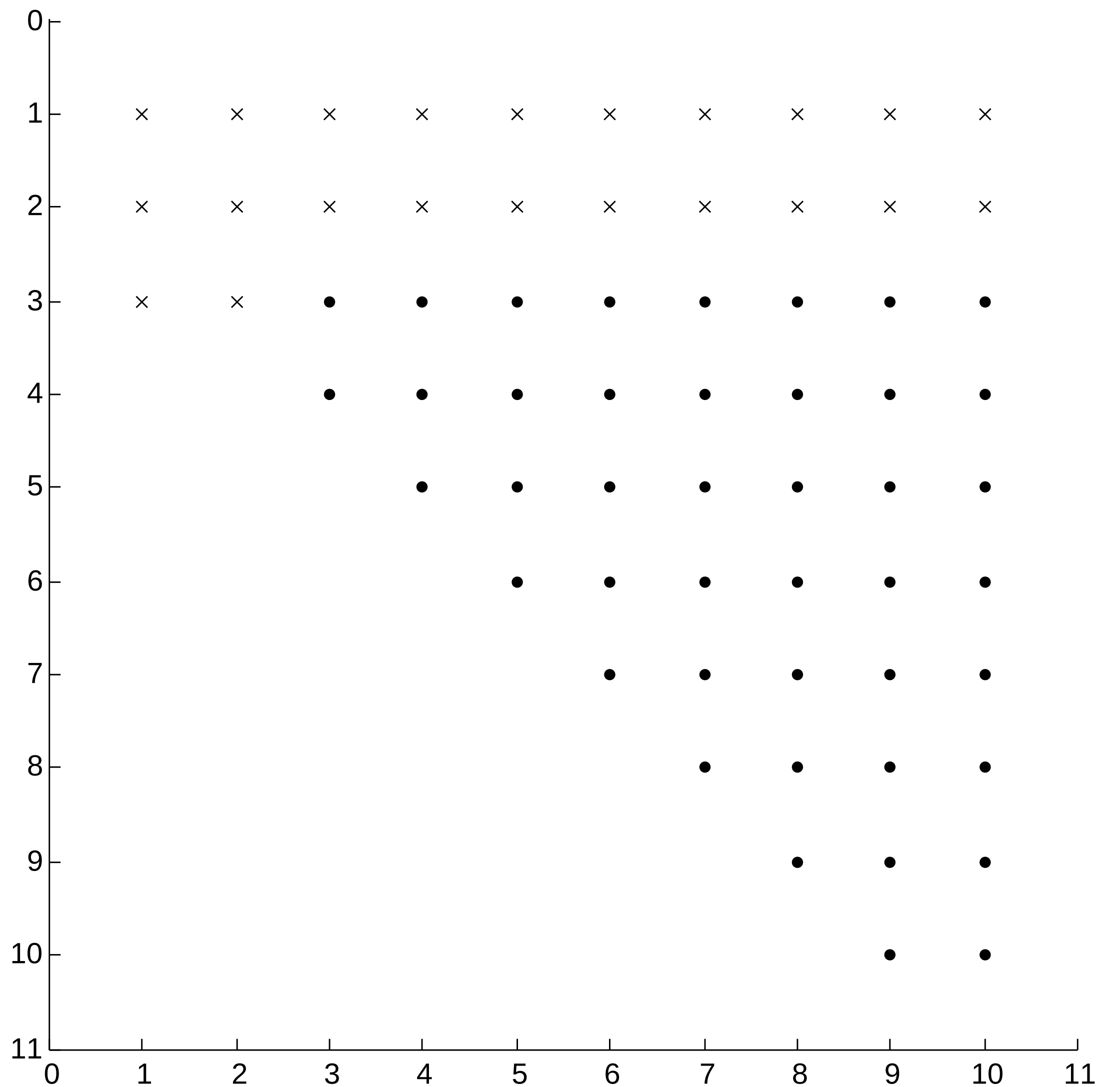
Hessenberg QR iteration

- Find an upper Hessenberg matrix $H = Q^T A Q$
- QR iteration preserves the Hessenberg structure
- Q is determined up to signs by first column
- Implicitly apply Q by *bulge-chasing*
- Cost per iteration is now $O(n^2)$.

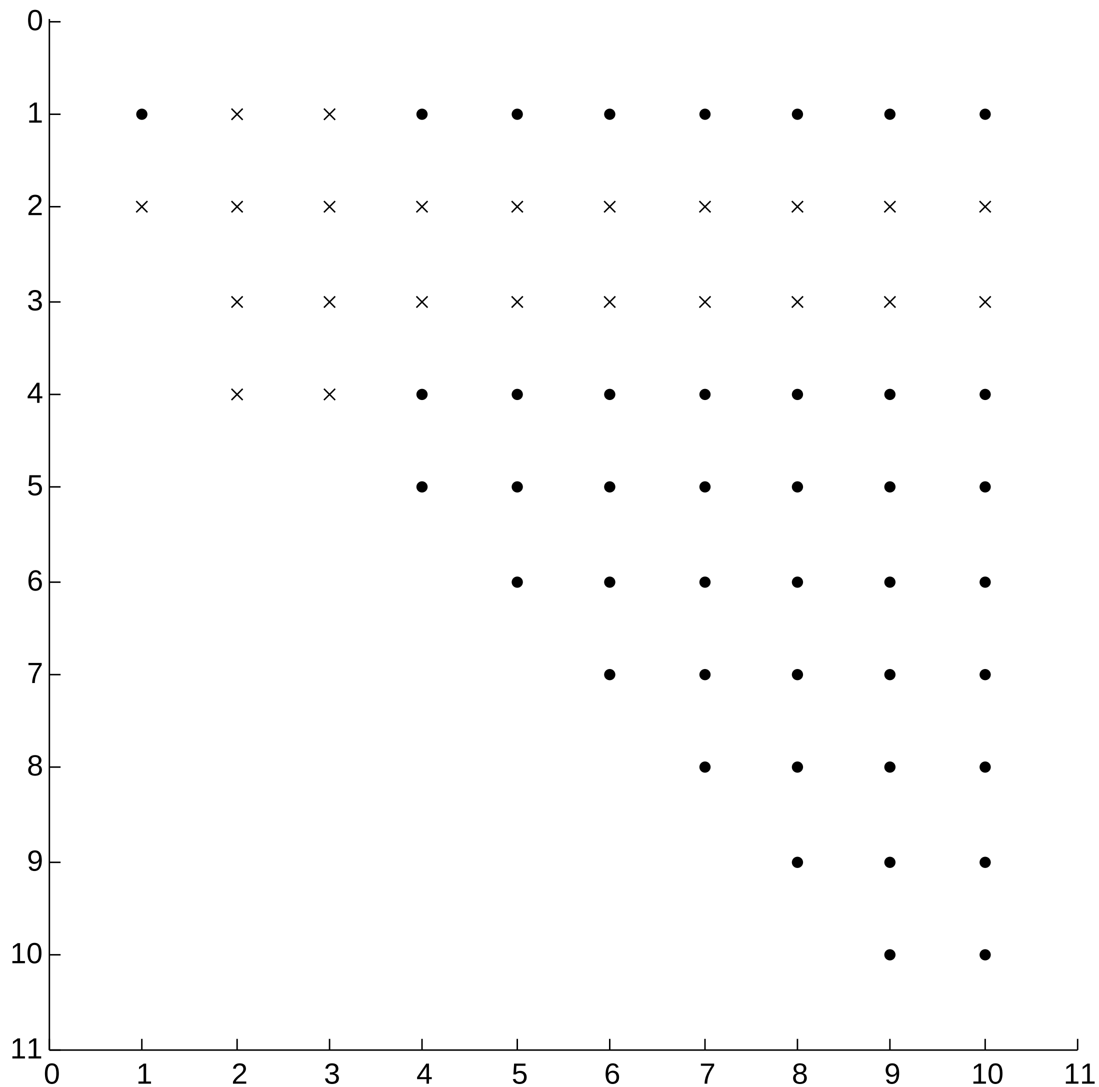
Hessenberg QR iteration



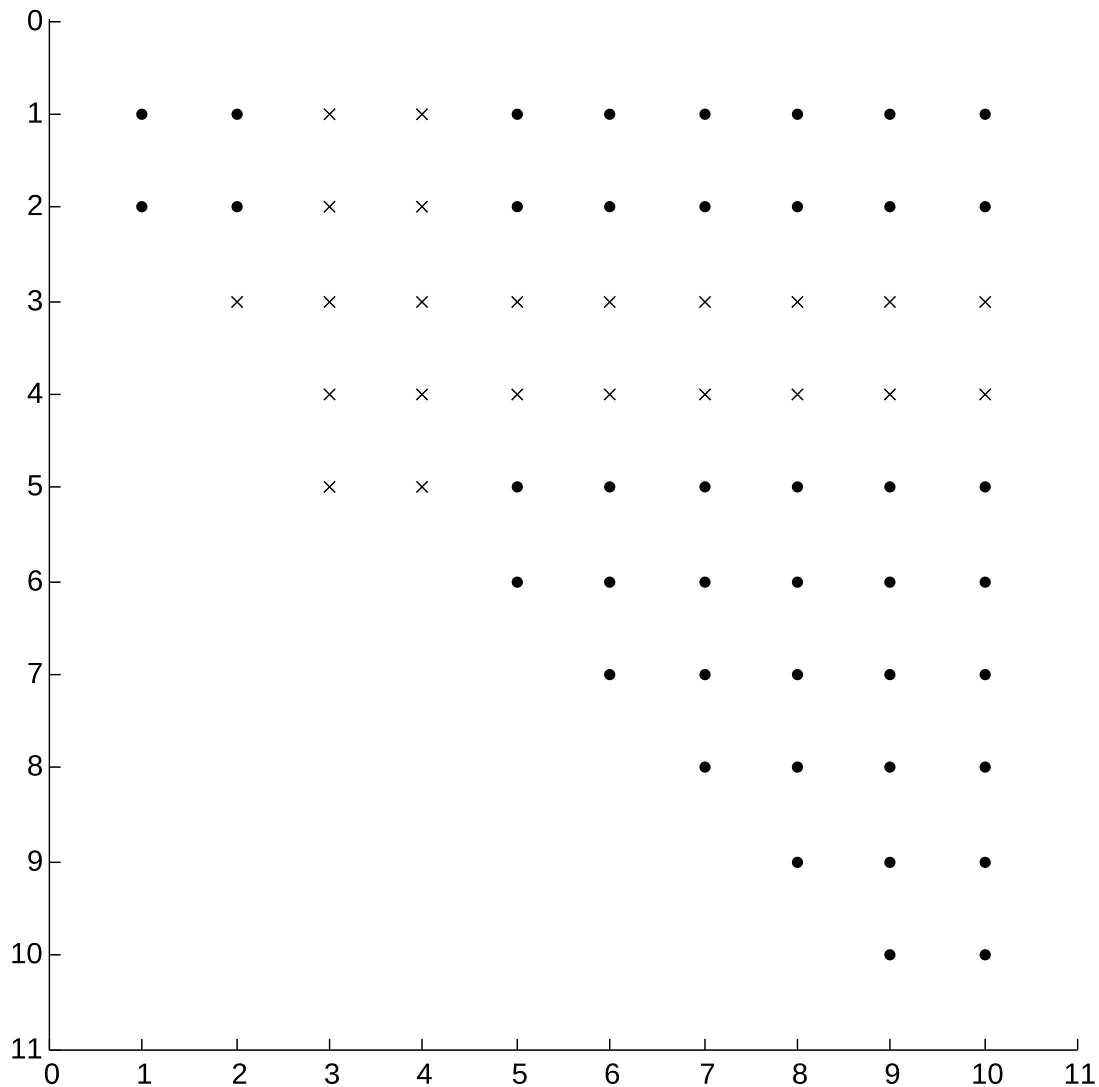
Hessenberg QR iteration



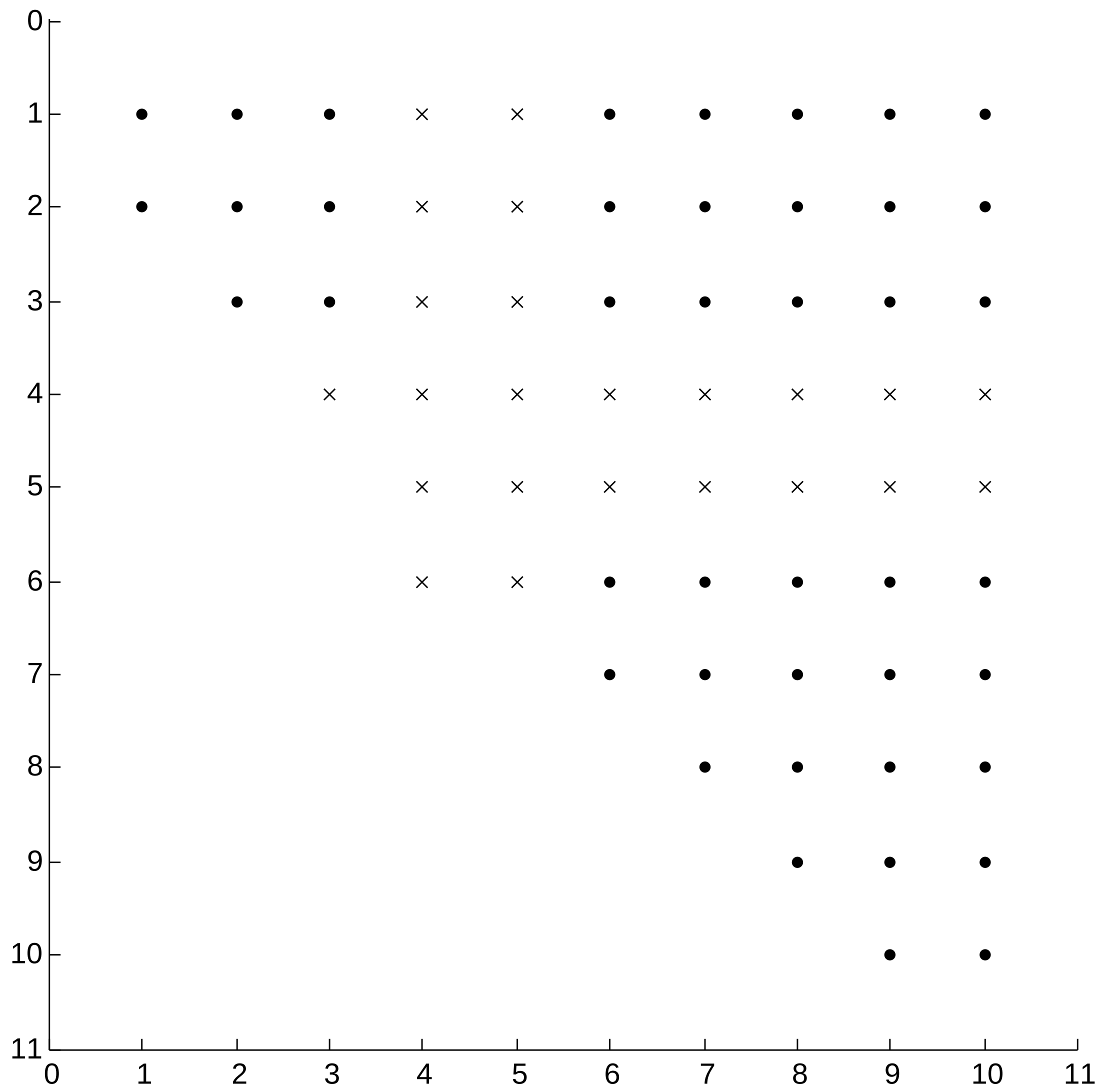
Hessenberg QR iteration



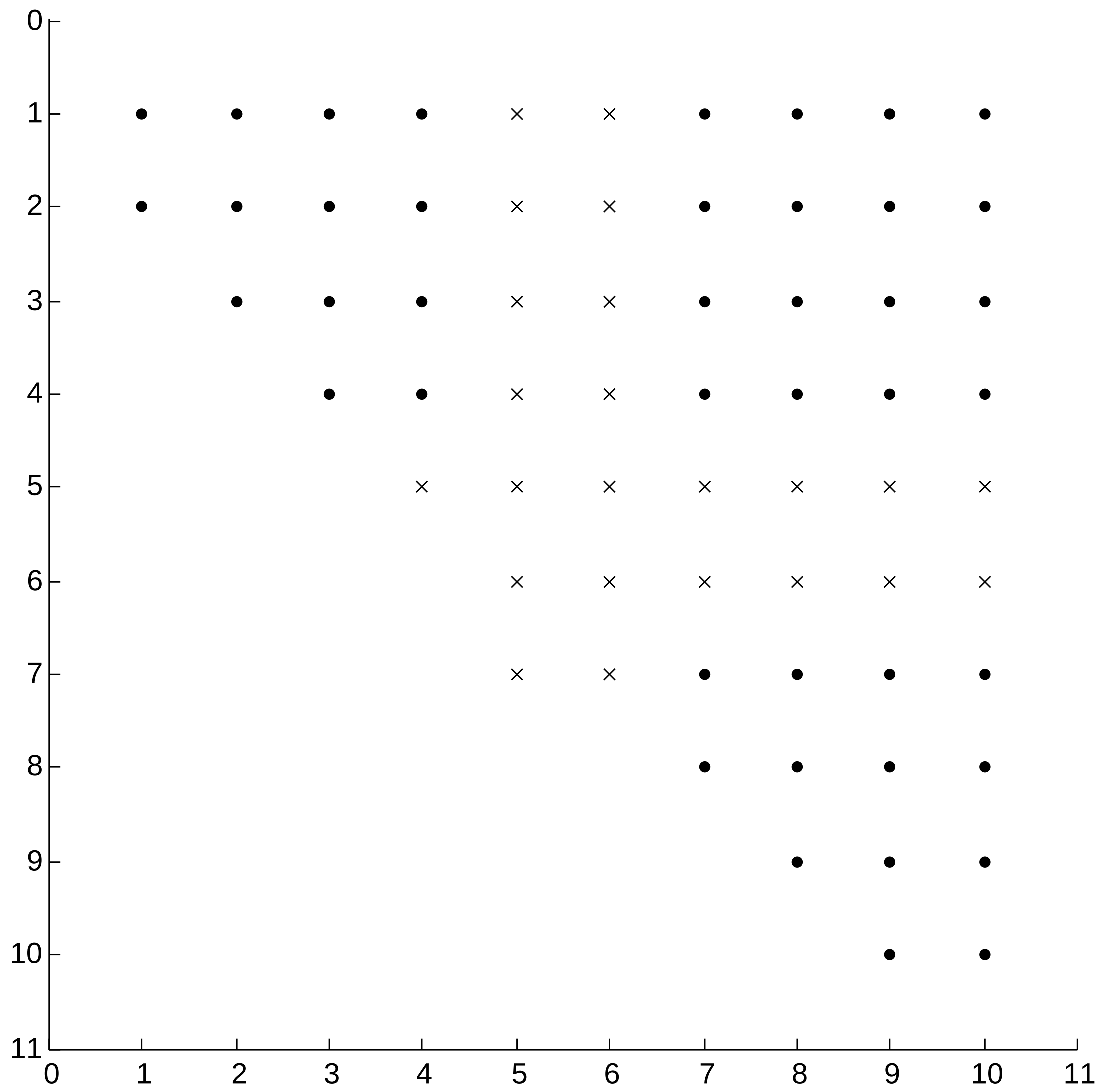
Hessenberg QR iteration



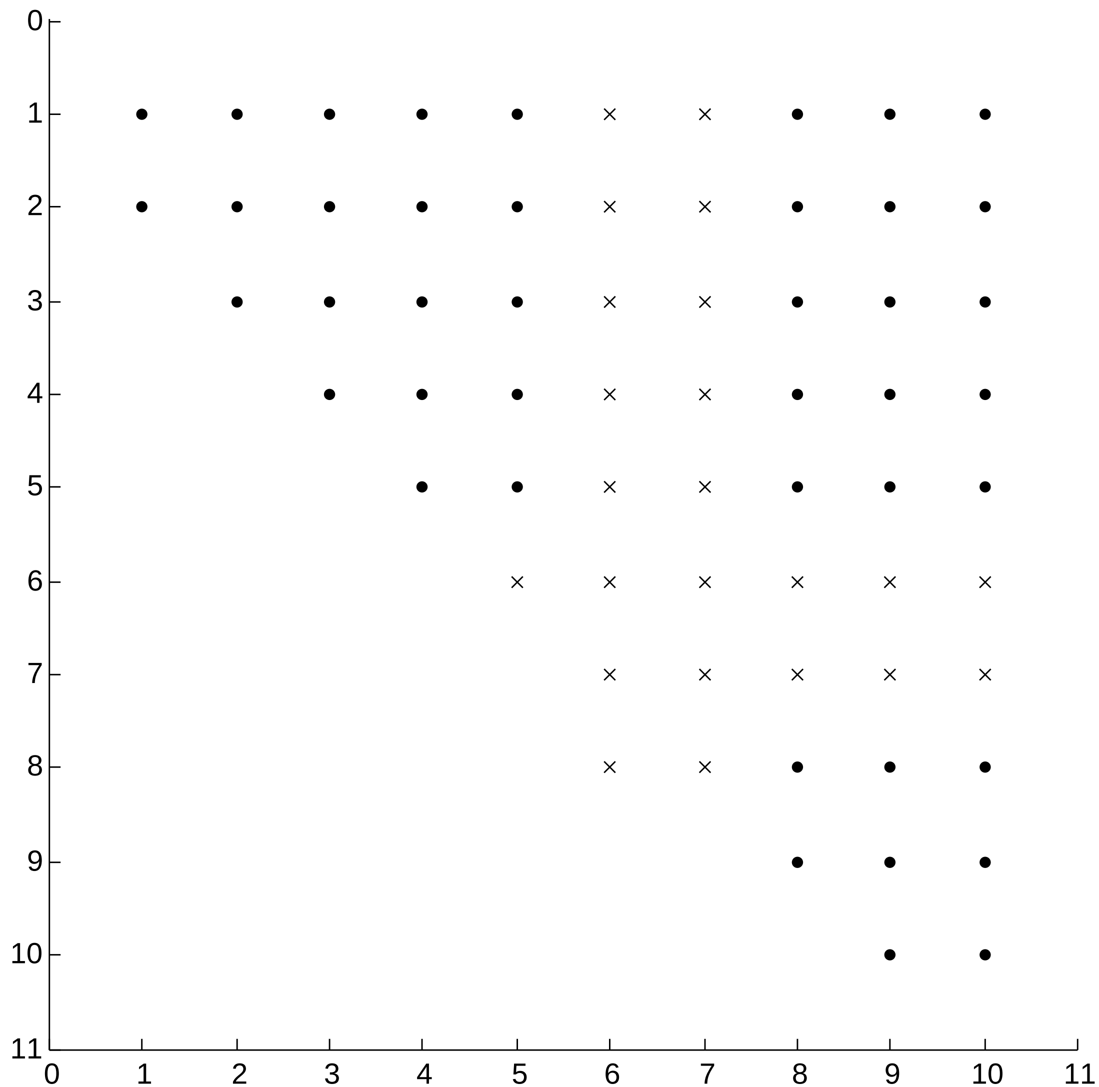
Hessenberg QR iteration



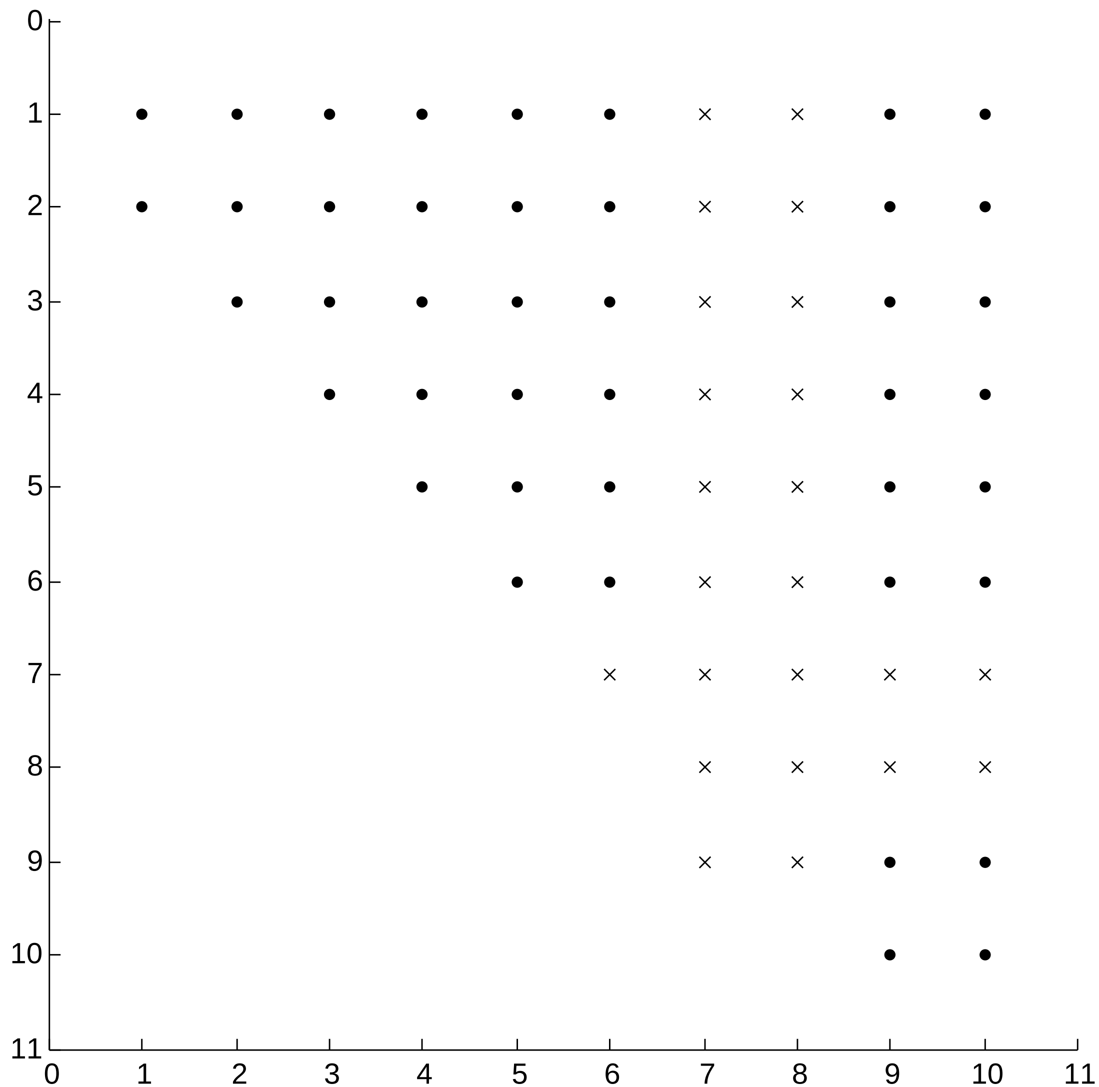
Hessenberg QR iteration



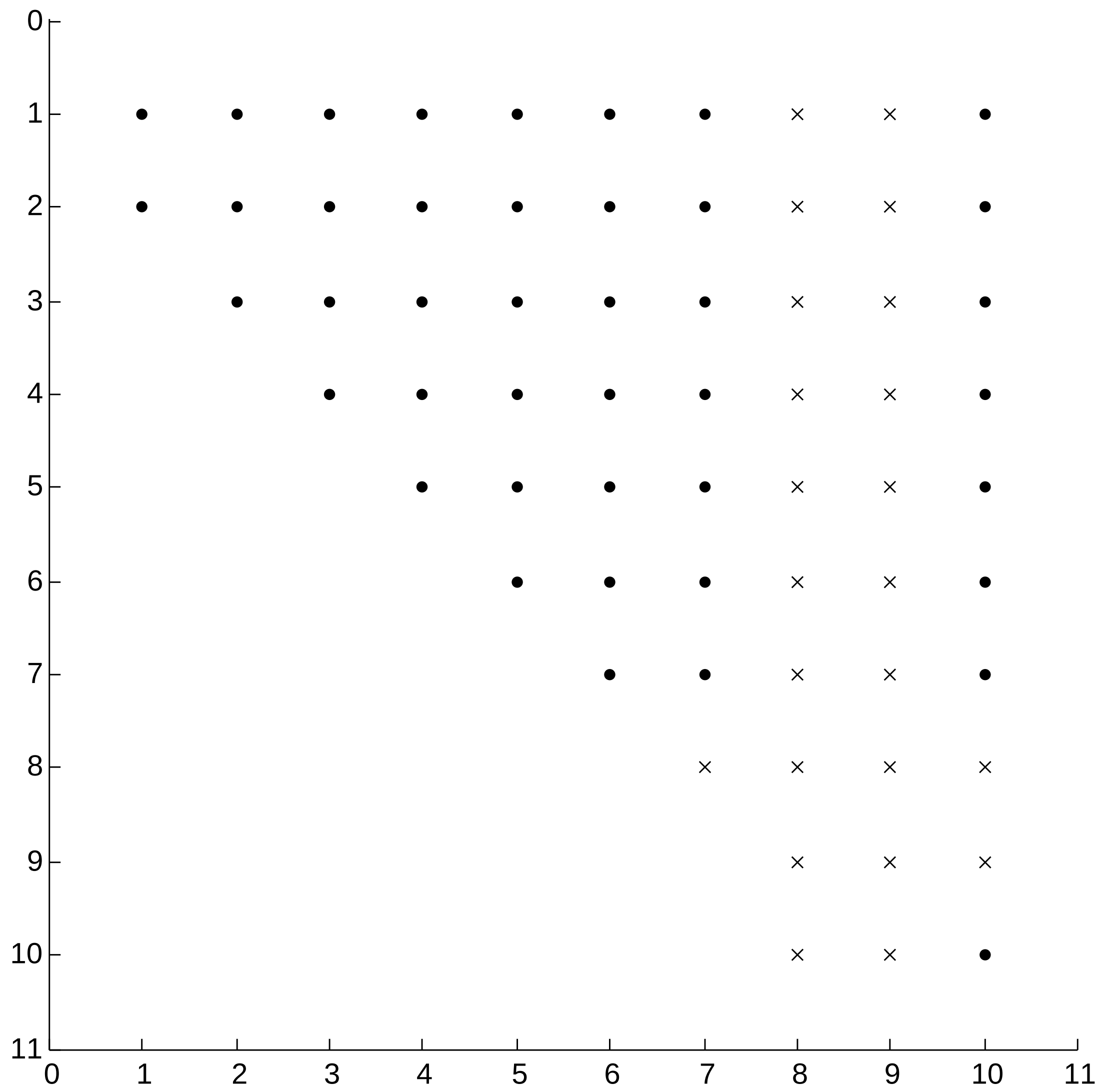
Hessenberg QR iteration



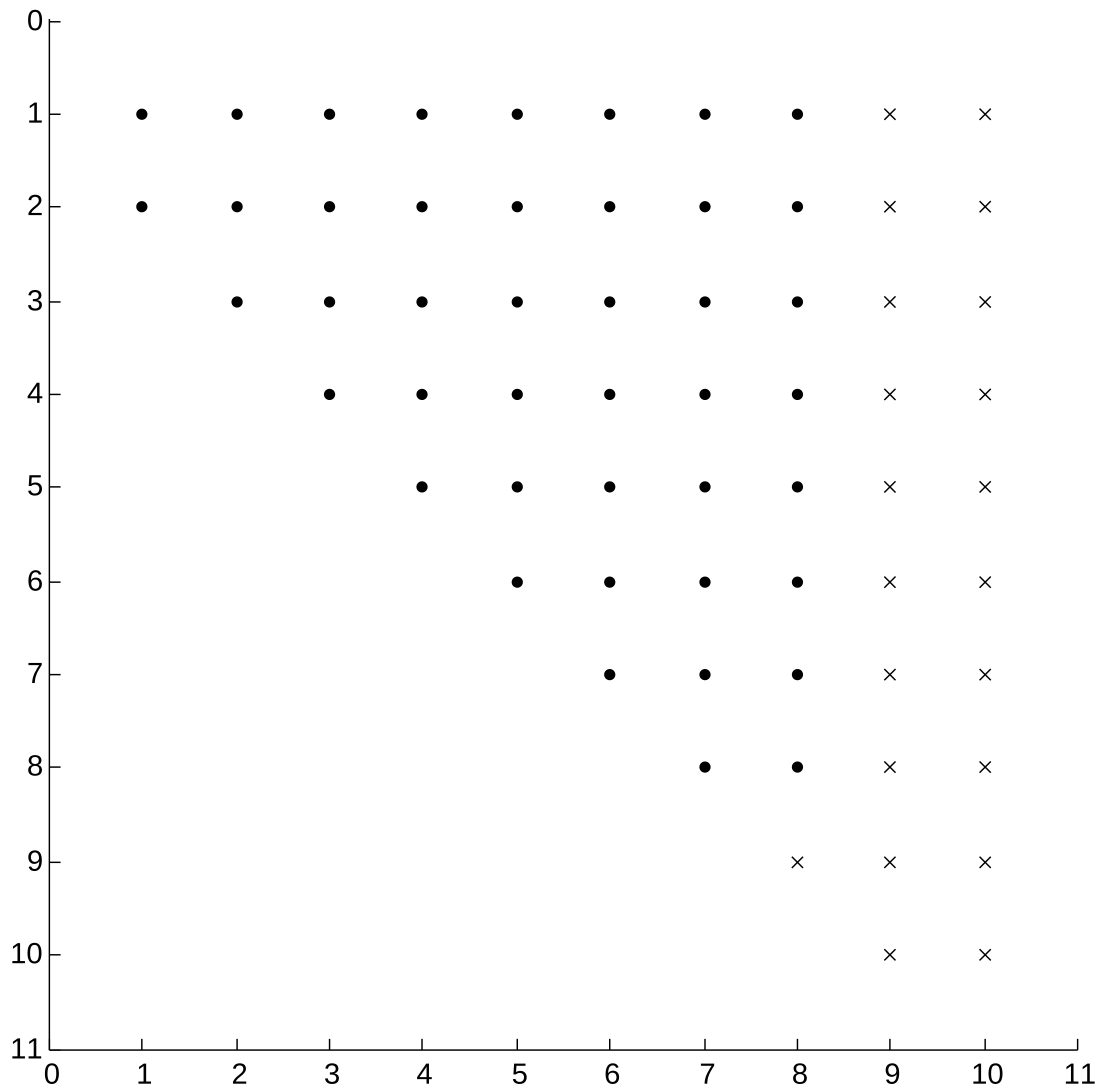
Hessenberg QR iteration



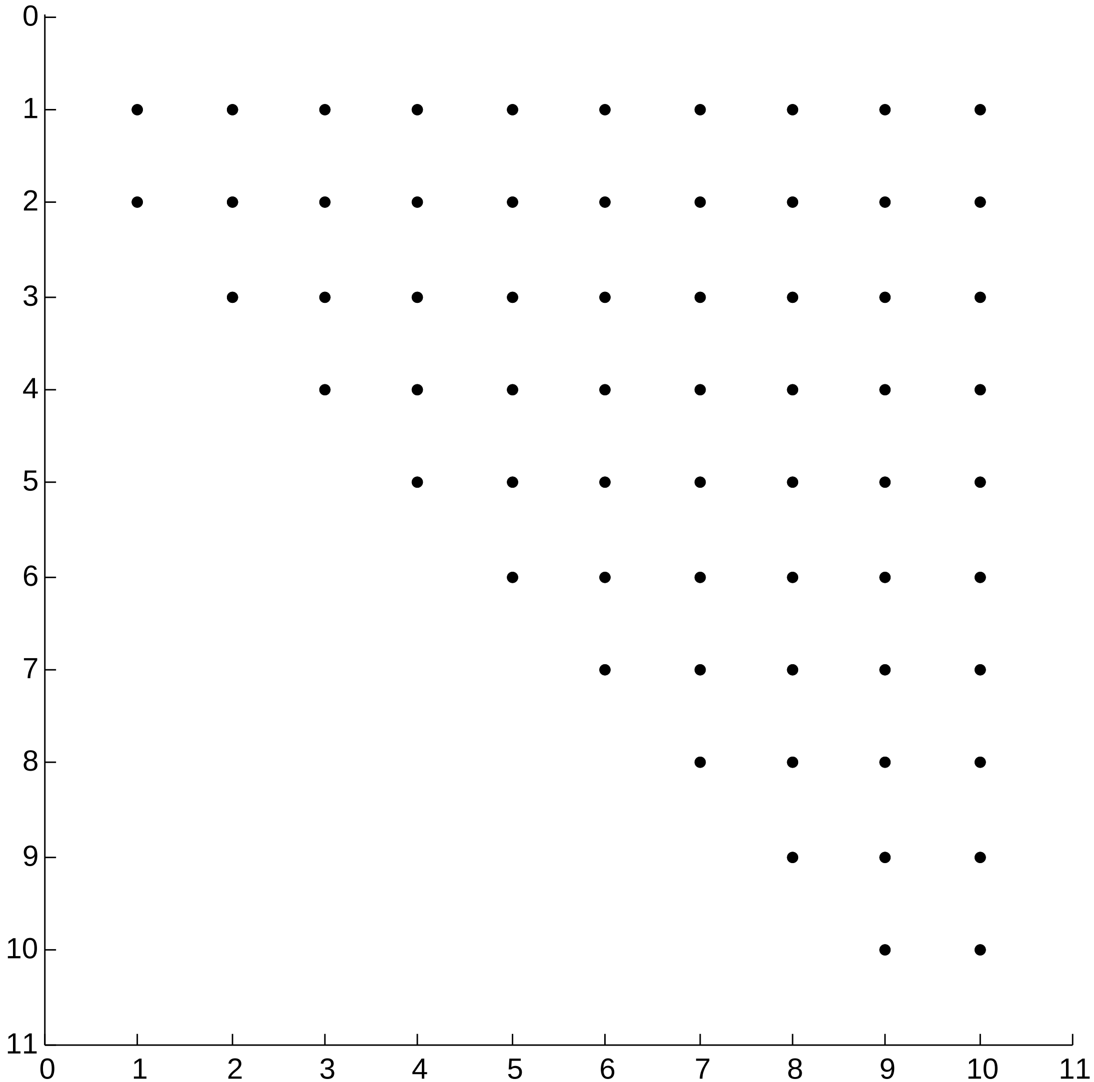
Hessenberg QR iteration



Hessenberg QR iteration



Hessenberg QR iteration



Even faster QR

Existing fast QR iterations:

- QR iteration preserves symmetry, orthogonality
- Symmetric tridiagonals: $O(n)$ parameters
- Unitary Hessenberg matrices: $O(n)$ parameters
- Use to do one QR iteration in $O(n)$ time

Even faster QR

New work:

- QR preserves symmetric or orthogonal + rank k
- Describe Hessenberg(symmetric or orthogonal + low rank k) with $O(nk)$ parameters.
- Use to do one QR iteration in $O(nk^2)$ time.

So who cares about matrices with this structure?

Example: Polynomial root finding

A standard algorithm: QR iteration on a companion matrix

- Robust software exists
- It's normwise backward stable
- It's used in Matlab
- But it takes $O(n^3)$ time and $O(n^2)$ storage

Use structure in QR iterates to get $O(n^2)$ time, $O(n)$ space.

Companion matrices

Given

$$p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0$$

Define a companion matrix C with $p(\lambda) = \det(C - \lambda I)$:

$$C = \begin{bmatrix} -a_{n-1} & -a_{n-2} & -a_{n-3} & \ldots & -a_1 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

For matrix polynomials, replace a_i by A_i and 1 by I.
C is a permutation plus a rank one matrix.

Example: Damped oscillators

Differential equations for a damped linear oscillator:

$$Mu'' + Du' + Ku = 0$$

Scaled time-harmonic problem:

$$(\lambda^2 I + \lambda C + K)x = 0.$$

Let L be a Cholesky factor of K. Then

$$\begin{bmatrix} 0 & L \\ -L^T & -C \end{bmatrix} \begin{bmatrix} Lx \\ \lambda x \end{bmatrix} = \begin{bmatrix} Lx \\ \lambda x \end{bmatrix} \lambda$$

If damping comes from localized boundary conditions or effects in a small part of a domain, C can have low rank.

Rank symmetry

Call A *rank-symmetric* if rank$(A_{12}) =$ rank(A_{12}) for any 2-by-2 blocking of A with square A_{11} and A_{12}. Examples:

- Symmetric matrices
- Skew-symmetric matrices
- Orthogonal matrices (by the CS decomposition)
- J-orthogonal matrices (by hyperbolic CS decomposition)

For symmetric, skew-symmetric, and orthogonal matrices, something stronger: the singular values of A_{12} and A_{21} are the same.

Rank symmetric + low rank

For A rank symmetric and L with rank k, both square block 2-by-2:

$$|\text{rank}(A_{12} + L_{12}) - \text{rank}(A_{21} + L_{21})| \leq 2k$$

Proof:

rank$(L_{ij}) \leq k$, so $|\text{rank}(A_{ij} + L_{ij}) - \text{rank}(A_{ij})| \leq k$.
Note rank$(A_{ij}) = $ rank(A_{ji}) and apply triangle inequality.

Structure in Hessenberg form

Suppose $H = A + L$, where A is rank symmetric, L is rank k, and H is Hessenberg. Then

$$\text{rank}(H_{12}) \leq 2k + \text{rank}(H_{21}) \leq 2k + 1$$

If H is a Schur form, $H_{21} = 0$ and $\text{rank}(H_{12}) \leq 2k$.

If A is symmetric, skew-symmetric, or orthogonal, the structure is preserved under QR iteration. All the Hessenberg QR iterates will have low rank superdiagonal blocks.

SSS structure

Write Hessenberg matrices with low off-diagonal rank as narrow band matrix + SSS matrix.

$$H = B + \begin{bmatrix} 0 & U_1V_2^T & U_1W_2V_3^T & U_1W_2W_3V_4^T \\ 0 & 0 & U_2V_3^T & U_2W_3V_4^T \\ 0 & 0 & 0 & U_3V_4^T \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Write $H_{ij} = U_iW_{i+1}\ldots W_{j-1}V_j$ for $i < j$ where $U_i \in \mathbb{R}^{m_b \times 2k+1}$, $V_i \in \mathbb{R}^{m_b \times 2k+1}$, and $W_i \in \mathbb{R}^{2k+1 \times 2k+1}$. Make the block size $m_b \geq 2k + 1$.

Total storage cost: $O(nk)$.

SSS structure transformed

$$\hat{H} = \begin{bmatrix} B_{11} & 0 & 0 & 0 \\ \hat{B}_{21} & \hat{B}_{22} & 0 & 0 \\ 0 & \hat{B}_{32} & B_{33} & 0 \\ 0 & 0 & B_{43} & B_{44} \end{bmatrix} + \begin{bmatrix} 0 & U_1 V_2^T & U_1 W_2 \hat{V}_3^T & U_1^T W_2 W_3 V_4^T \\ 0 & 0 & \hat{U}_2 \hat{V}_3^T & \hat{U}_2^T W_3 V_4^T \\ 0 & 0 & 0 & U_3 V_4^T \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Transform second row and column by reflection.
Only affects the band, U_2, and V_2.

What about transformations that cross block edges?
Merge blocks so it never happens – then split them.

Result: A bulge-chasing pass takes $O(nk^2)$ time vs $O(n^2)$

Splitting blocks

Define C_j and G_j by the recurrences

$$\begin{aligned} C_1 &= U_1 \text{ and } C_j = \begin{bmatrix} C_{j-1}W_j \\ U_j \end{bmatrix} \text{ for } j > 1 \\ G_n &= V_n \text{ and } G_j = \begin{bmatrix} V_v \\ G_{j+1}W_j^T \end{bmatrix} \text{ for } j < n. \end{aligned}$$

C_j and G_j are row and column bases for the submatrix left of block column j and above block row j.

All C_j orthonormal: left proper form.
All G_j orthonormal: right proper form.

Splitting blocks

$$\begin{bmatrix} H_{1,i} & \dots & H_{1,n} \\ \vdots & & \vdots \\ H_{i,i} & \dots & H_{i,n} \end{bmatrix} = \begin{bmatrix} C_{i-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} V_i^T & W_i \\ H_{ii} & U_i \end{bmatrix} \begin{bmatrix} 0 & G_{i+1} \\ I & 0 \end{bmatrix}^T .$$

$$\begin{bmatrix} V_j^T & W_j \\ B_{jj} & U_j \end{bmatrix} \rightarrow \begin{bmatrix} V_j^{\alpha T} & W_j^{\alpha} V_j^{\beta T} & W_j^{\alpha} W_j^{\beta} \\ B_{jj}^{\alpha\alpha} & U_j^{\alpha} V_j^{\beta T} & U_j^{\alpha} W_j^{\beta} \\ B_{jj}^{\beta\alpha} & B_{jj}^{\beta\beta} & U_j^{\beta} \end{bmatrix}$$

Do pivoted QR or SVD decomposition to split U, V, W.

Choose $\begin{bmatrix} W_j^{\alpha} \\ U_j^{\alpha} \end{bmatrix}$ with orthonormal columns.

Issue: Balancing

- Would like matrix to be *balanced* – scaled by diagonal similarity to reduce norm, improve accuracy
- General diagonal similarities destroy the structure!
- Rescaling a few rows and columns corresponds to making an additional low rank modification.
- For companion matrices, optimal balancing in infinity norm respects the structure.

Balancing

If C is a companion matrix, so is $|C|$. The Perron vector x of $|C|$ therefore has the form

$$x_i = c\lambda^i$$

Let

$$D = \text{diag}(x_i^{-1}) = \text{diag}(\lambda^{-i}).$$

Then DCD^{-1} is optimally balanced in the ∞-norm. Corresponds to scaling the polynomial indeterminate by λ.

Issue: Normalization error

Write Hessenberg $H = A + L + E$ where

- $A_{12} = U_1 \Sigma V_2^T$ and $A_{21} = U_2 \Sigma V_1^T$
- rank$(L) = k$
- E is a perturbation

Then

$$H_{12} = (\widehat{H_{21}} - \widehat{L_{21}} + L_{12}) + (E_{12} - \widehat{E_{21}})$$

where

$$\begin{aligned} \widehat{H_{21}} &= (U_1 U_2^T) H_{21} (V_1 V_2^T) \\ \widehat{L_{21}} &= (U_1 U_2^T) L_{21} (V_1 V_2^T) \\ \widehat{E_{21}} &= (U_1 U_2^T) E_{21} (V_1 V_2^T) \end{aligned}$$

Normalization error

Can error grow during splitting?

- Only if we move away from (symmetric, skew, orthogonal) + low rank structure.
- Some growth with QR-based split (not with SVD).
- We can monitor error from splitting.
- Can we quickly perturb back to the right structure?
- Still thinking about this one.

Implementation

Modified DLAHQR (basic double-shift QR) to use the new data structure.

- Convert to right proper form.
- Choose shifts (logic from LAPACK).
- Bulge chase top to bottom / convert to left proper form.
- Check for convergence / deflate (logic from LAPACK).

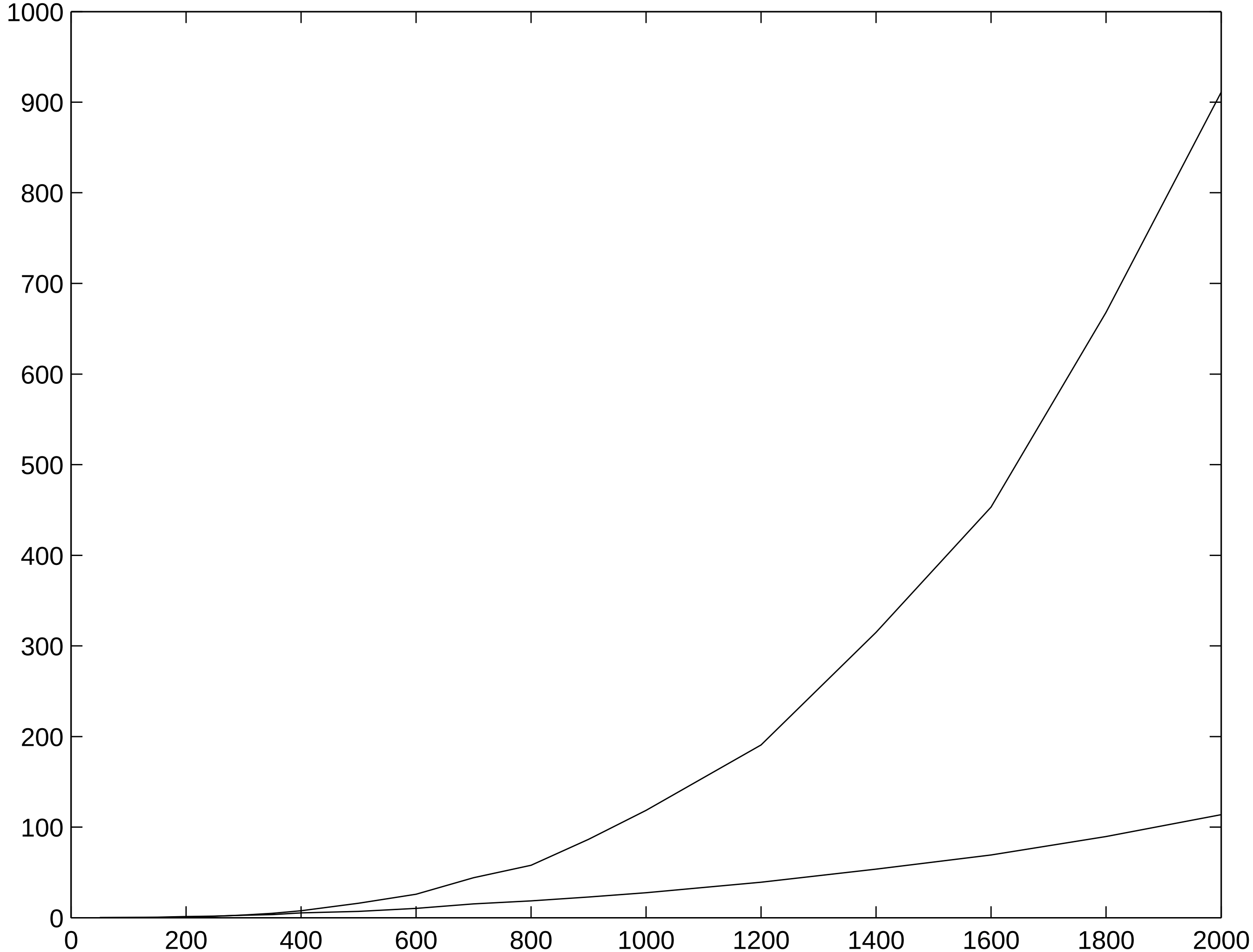
Most of the code accesses data in the band part of the structure. This text is *identical* to the standard code – just change `LDH`.

Performance

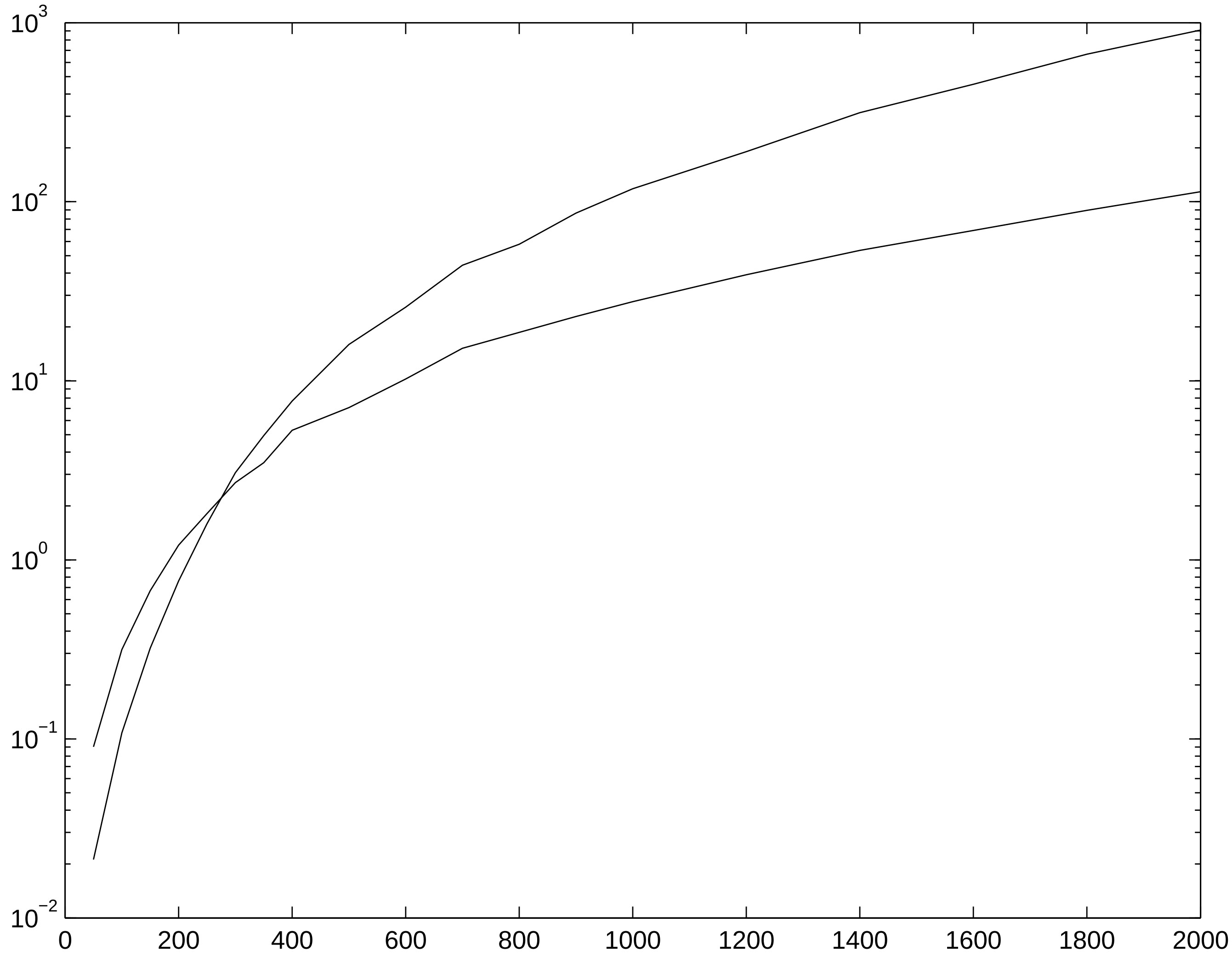
Update with most recent version

- Compared against LAPACK DLAHQR (not DHSEQR).
- Compiled using g77 without optimizations.
- LAPACK with standard optimizations, ATLAS BLAS
- Pentium 3 laptop at 700 MHz.
- Polynomials of degree 50-2000.
- For degree 10000, current code takes 41 minutes.

Performance



Performance



Conclusions

- Hessenberg({ symmetric, skew, unitary } + low rank) has low off-diagonal rank.
- Applies to (matrix) polynomial root finding, some damped vibration calculations.
- Can use low-rank structure for fast bulge-chasing.
- Can use-use most existing QR lore and code.
- Have no direct fast method for Hessenberg reduction (though know how to do fast Arnoldi variant).
- Can balance companion matrices in ∞ norm, but general balancing destroys needed structure.
- Backward stable – except maybe for splitting step?