Fixed-Point Iterations

Solve nonsingular linear system: $Ax = b$ (solution $\hat{x} = A^{-1}b$)
Solve an approximate, but simpler system: $Mx = b \rightarrow x_0 = M^{-1}b$

Improve the solution using the residual: $r_0 = b - Ax_0$ (iterative refinement)
Error, $e_0 = \hat{x} - x_0$, satisfies $Ae_0 = b - Ax_0 = r_0$
Don’t compute exact error, instead solve $Mz_0 = r_0$ and set $x_1 = x_0 + z_0$

Iterate:
\[
\begin{align*}
  r_k &= b - Ax_k = b - A(x_{k-1} + z_{k-1}) = r_{k-1} - Az_{k-1} \\
  z_k &= M^{-1}r_k \quad \text{(solve } Mz_k = r_k) \\
  x_{k+1} &= x_k + z_k
\end{align*}
\]

Methods: Jacobi iteration (diagonal), Gauss-Seidel (upper triangular), many others such as (S)SOR, ...
Fixed-Point Iterations

Convergence of such iterations?
Let $A = M - N$ (matrix splitting).
Then $Mz_k = b - Ax_k \iff Mz_k = b - Mx_k + Nx_k \iff Mx_{k+1} = Nx_k + b$
(fixed-point iteration $x_{k+1} = M^{-1}Nx_k + M^{-1}b$)

Note that the fixed-point is the solution (proof?)

Error:
$e_k = \hat{x} - x_k = M^{-1}N\hat{x} + M^{-1}b - M^{-1}Nx_{k-1} - M^{-1}b$
$= M^{-1}N(\hat{x} - x_{k-1}) = M^{-1}Ne_{k-1}$
$= (M^{-1}N)^k e_0$

Residual: $r_k = (NM^{-1})^k r_0$ and $M^{-1}r_k = (M^{-1}N)^k (M^{-1}r_0)$ (proof?)

To analyze convergence we need to introduce/review a number of concepts

Rate of Convergence

Let $\hat{x}$ be the solution of $Ax = b$, and we have iterates $x_0, x_1, x_2, \ldots$

\{x_k\} converges (q-)linearly to $\hat{x}$ if there are $N \geq 0$ and $c \in [0,1)$ such that for
$k \geq N: \|x_{k+1} - \hat{x}\| \leq c \|x_k - \hat{x}\|$,

\{x_k\} converges (q-)superlinearly to $\hat{x}$ if there are $N \geq 0$ and a sequence $\{c_k\}$
that converges to 0 such that for $k \geq N: \|x_{k+1} - \hat{x}\| \leq c_k \|x_k - \hat{x}\|$

\{x_k\} converges to $\hat{x}$ with (q-)order at least $p$ if there are $p > 1$, $c \geq 0$, and
$N \geq 0$ such that $k \geq N: \|x_{k+1} - \hat{x}\| \leq c \|x_k - \hat{x}\|^p$ (quadratic if $p = 2$, cubic
if $p = 3$, and so on)

\{x_k\} converges to $\hat{x}$ with j-step (q-)order at least $p$ if there are a fixed integer
$j \geq 1$, $p > 1$, $c \geq 0$, and $N \geq 0$, such that $k \geq N: \|x_{k+j} - \hat{x}\| \leq c \|x_k - \hat{x}\|^p$
Norms

A norm on a vector space $V$ is any function $f : V \to \mathbb{R}$ such that
1. $f(x) \geq 0$ and $f(x) = 0 \Leftrightarrow x = 0$,
2. $f(\alpha x) = |\alpha| f(x)$,
3. $f(x + y) \leq f(x) + f(y)$,
where $x \in V$ and $\alpha \in \mathbb{R}$.

Important vector spaces in this course: $\mathbb{R}^n$, $\mathbb{C}^n$, and $\mathbb{R}^{m\times n}$, $\mathbb{C}^{m\times n}$ (matrices).

Note that the set of all m-by-n matrices (real or complex) is a vector space.

Many matrix norms possess the submultiplicative or consistency property:
$f(AB) \leq f(A)f(B)$ for all $A \in \mathbb{C}^{m\times k}$ and $B \in \mathbb{C}^{k\times n}$ (or real matrices).

Note that strictly speaking this is a property of a family of norms, because in general ‘each’ $f$ is defined on a different vector space.

We can define a matrix norm using a vector norm (an induced matrix norm):

$$
\| A \|_\alpha = \max_{x=0} \frac{\| Ax \|_\alpha}{\| x \|_\alpha} = \max_{\| x \|_\alpha = 1} \| Ax \|_\alpha
$$

Induced norms are always consistent (satisfy consistency property).

Two norms $\| \cdot \|_\alpha$ and $\| \cdot \|_\beta$ are equivalent if there exist positive, real constants $a$ and $b$ such that

$$
\forall x : a \| x \|_\alpha \leq \| x \|_\beta \leq b \| x \|_\alpha
$$

The constants depend on the two norms but not on $x$.

All norms on a finite dimensional vector space are equivalent.
Norms

Some useful norms on $\mathbb{R}^n$, $\mathbb{C}^n$, $\mathbb{R}^m \times \mathbb{R}^n$, $\mathbb{C}^m \times \mathbb{C}^n$:

p-norms: $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$, especially $p = 1, 2, \infty$, where $\|x\|_\infty = \max_i |x_i|$. Induced matrix p-norms are:

$$\|A\|_1 = \max \sum_{i=1}^n |a_{ij}| \quad \text{(max absolute column sum)}$$

$$\|A\|_\infty = \max \sum_{j=1}^n |a_{ij}| \quad \text{(max absolute row sum)}$$

Matrix Frobenius norm:

$$\|A\|_F = \left(\sum_{i,j=1}^n |a_{ij}|^2\right)^{1/2} \quad \text{(similar to vector 2-norm for a matrix)}$$

All these norms are consistent (satisfy the submultiplicative property)

Eigenvalues and Eigenvectors

Let $Ax = \lambda x$ and $\bar{y}A = \lambda \bar{y}$ (for same $\lambda$).

We call the column vector $x$ a (right) eigenvector, the row vector $\bar{y}$ a left eigenvector, and $\lambda$ an eigenvalue, the triple together is called an eigentriple, and $(\lambda, x)$ and $(\lambda, \bar{y})$ a (right) eigenpair or left eigenpair.

The set of all eigenvalues of $A$, $\Lambda(A)$, is called the spectrum of $A$.

If the matrix $A$ is diagonalizable (has a complete set of eigenvectors) we have

$$A = VAV^{-1} \iff AV = VA,$$

where $V$ is a matrix with the right eigenvectors as columns and $A$ is a diagonal matrix with the eigenvalues as coefficients.

A similar decomposition can be given for the left eigenvectors.
Spectral Radius

The spectral radius $\rho(A)$ is defined as $\rho(A) = \max \{ |\lambda| : \lambda \in \sigma(A) \}$.

Theorem:
For all $A$ and $\varepsilon > 0$ a consistent norm $\| \cdot \|$ exists such that $\| A \| \leq \rho(A) + \varepsilon$.

So, if $\rho(A) < 1$, then a consistent norm $\| \cdot \|$ exists such that $\| A \| < 1$.

Take $\varepsilon = \frac{1}{2} (1 - \rho(A))$ and apply theorem above.

Define $A^* = \overline{A^T}$ (complex conjugate transpose).

If $A$ is Hermitian ($A = A^*$), then $\rho(A) = \| A \|$.

If $A$ is normal ($AA^* = A^*A$), then $\rho(A) = \| A \|$.

Fixed-Point Iterations

Under what conditions does $e_k \to 0$ and $x_k \to \hat{x}$ (convergence) for arbitrary $e_0$?

Theorem: $e_k = (M^{-1}N)^k e_0 \to 0$ for arbitrary $e_0$ iff $(M^{-1}N)^k \to 0$.

Proof:
Let $G = M^{-1}N$ and the matrix norm $\| \cdot \|$ be induced by a vector norm.

1. Assume $G^k \to 0$
Then $G^k \to 0 \Rightarrow \| G^k \| \to 0$ and $\| G^k e_0 \| \leq \| G^k \| \| e_0 \| \to 0$ for any $e_0$.

2. Assume $G^k e_0 \to 0$ for all $e_0$.
Consider the identity matrix $I = [\eta_1 \eta_2 \cdots \eta_n]$.
$G^k I = [G^k \eta_1 G^k \eta_2 \cdots G^k \eta_n] \to [00\cdots0]$; so $G^k \to 0$ since $G^k I = G^k$.
Alternatively, consider $\| [G^k \eta_1 G^k \eta_2 \cdots G^k \eta_n] \|$ (note that $G^k \eta_i \to 0$).
Norms

Note that we can generalize the result for the one-norm to all norms by using the equivalence of norms on finite dimensional vector spaces.

Similarly, the results are readily generalized for inconsistent matrix norms (with \( \| AB \| > \| A \| \| B \| \) possible), by using the equivalence of norms on finite dimensional spaces.

Fixed-Point Iterations

Theorem: \( G^k \to 0 \) iff \( \rho(G) < 1 \).

Proof:
1. \( G^k \to 0 \Rightarrow \rho(G) < 1 \).

For each eigenvalue \( \lambda \) of \( G \) there exists at least one eigenvector \( v \) s.t. \( Gv = \lambda v \).

Then \( \| G^k v \| = \| \lambda^k v \| = \| \lambda \|^k \| v \| \) and \( G^k v \to 0 \). So, \( | \lambda |^k \| v \| \to 0 \Rightarrow | \lambda | < 1 \).

Since this holds for each eigenvalue, \( \rho(G) < 1 \) must hold.

2. \( \rho(G) < 1 \Rightarrow G^k \to 0 \).

There exists a consistent norm \( \| \cdot \| \) s.t. \( \| G \| < 1 \).

Hence, \( \| G^k \| \leq \| G \|^k \to 0 \). Therefore \( \| G^k \| \to 0 \Rightarrow G^k \to 0 \).

So \( e_k \to 0 \) (\( x_k \to \hat{x} \)) for arbitrary \( e_0 \) iff \( \rho(M^{-1}N) < 1 \).
**Krylov Spaces**

Given $x_0$, set $r_0 = b - Ax_0$.

For $k = 0, 1, 2, \ldots$

\[
    z_k = M^{-1}r_k,
\]

\[
x_{k+1} = x_k + z_k,
\]

\[
r_{k+1} = b - Ax_{k+1} = r_k - Az_k.
\]

Note that $x_{k+1} - x_k = z_k = M^{-1}r_k$ and hence $x_{k+1} - x_0 = z_0 + z_1 + \cdots + z_k$.

This implies $x_{k+1} - x_0 = M^{-1}r_0 + (M^{-1}N)M^{-1}r_0 + \cdots + (M^{-1}N)^k M^{-1}r_0$.

So, correction $x_{k+1} - x_0$ is given by polynomial $S_k(t) = 1 + t + t^2 + \cdots + t^k$:

\[
x_{k+1} - x_0 = \sum_{j=0}^{k} (M^{-1}N)^j M^{-1}r_0 = S_k(M^{-1}N) \cdot M^{-1}r_0.
\]

Note also $c_k = (M^{-1}N)^k c_0$ and $r_k = (NM^{-1})^k r_0 = M(M^{-1}N)^k M^{-1}r_0$.

---

**Similarity Transformation**

Let $A$ have eigenpairs $(\lambda_i, v_i)$: $Av_i = \lambda_i v_i$

Define the similarity transformation: $BAB^{-1}$

The matrix $BAB^{-1}$ has the same eigenvalues, $\lambda_i$, as $A$ and eigenvectors $Bv_i$:

\[
    BAB^{-1}(Bv_i) = BA v_i = \lambda_i Bv_i
\]

Show that $(M^{-1}N)^k$ has the same eigenvalues as $(NM^{-1})^k$.
### Polynomials and Spaces

Main computational cost is in the multiplication by \( A \) and solving for \( M \).
So, we can try to generate better polynomials (faster convergence) at same cost.

Also correction \( x_{i+1} - x_0 \in \text{span} \left\{ M^{-1}r_0, (M^{-1}N)M^{-1}r_0, \ldots, (M^{-1}N)^iM^{-1}r_0 \right\} \)

We call a space \( K_m \left( B, y \right) \equiv \text{span} \left\{ y, By, B^2y, \ldots, B^{m-1}y \right\} \)
the Krylov (sub)space of dimension \( m \) associated with \( B \) and \( y \).
So, \( x_m - x_0 \in K_m \left( M^{-1}N, M^{-1}r_0 \right) \)

\[ e_m = \left( M^{-1}N \right)^m e_0 \in K_{m+1} \left( M^{-1}N, e_0 \right) \quad \text{and} \]
\[ M^{-1}r_m = \left( M^{-1}N \right)^m M^{-1}r_0 \in K_{m+1} \left( M^{-1}N, M^{-1}r_0 \right). \]

Therefore, alternatively we can compute better approximate solutions from the same space (faster convergence) at same cost.

### Krylov Spaces

So, a Krylov space is a space of polynomials in a matrix times a vector.

These spaces inherit the many important approximation properties that polynomials on the real line or in the complex plane possess.

For simplicity let the matrix \( B \) be diagonalizable, \( B = VAV^{-1} \).
Then \( B^2 = VAV^{-1}VAV^{-1} = VAV^{-1} \) and generally \( B^k = VAV^{-1} \).
So, the polynomial \( p_m \left( t \right) = \alpha_0 + \alpha_1 t + \cdots + \alpha_m t^m \) applied to \( B \) gives

\[ p_m \left( B \right) = V \left( \alpha_0 I + \alpha_1 A + \alpha_2 A^2 + \cdots + \alpha_m A^m \right) V^{-1} \]
and hence

\[ p_m \left( B \right) = Vp_m \left( A \right) V^{-1} = V \text{diag} \left( p_m \left( \lambda_1 \right), \ldots, p_m \left( \lambda_n \right) \right) V^{-1} \]

So, the polynomial is applied to the eigenvalues individually.
This allows us to approximate solutions to linear systems, eigenvalue problems, and more general problems using polynomial approximation.
Approximation by Matrix Polynomials

Let $B = VAV^{-1}$, let $A(B) \subset \Omega \subset \mathbb{C}$.

If $p_m(t) \approx \frac{1}{t}$ for all $t \in \Omega$, then $p_m(B) \approx B^{-1}$.

Let $y = V\zeta$. Then $p_m(B)y = \sum_i v_i p_m(\lambda_i)\zeta_i \approx \sum_i v_i \frac{\zeta_i}{\lambda_i} = B^{-1}y$

Furthermore, let $\varepsilon \approx 0$ and $|\lambda_i - \lambda| > \delta$ (for some eigenvalue $\lambda$)

If $p_m(t) = \begin{cases} \varepsilon, & t \in \Omega \text{ and } |t - \lambda| > \delta, \\ 1, & t = \lambda, \end{cases}$

then $p_m(B)y \approx v_i \zeta_i$.

If we can construct such polynomials for modest $m$ we have an efficient linear solver or eigensolver.

Krylov Spaces

We have $M^{-1}N = M^{-1}(M - A) = I - M^{-1}A$.

So, $x_{k+1} - x_k = (I - M^{-1}A)^t M^{-1}r_0 \in K_{k+1}(M^{-1}A, M^{-1}r_0)$, and

$x_{k+1} = M^{-1}Nx_k + M^{-1}b$,

with fixed-point $\hat{x} = (I - M^{-1}A)\hat{x} + M^{-1}b \Leftrightarrow M^{-1}A\hat{x} = M^{-1}b$.

So, we solve the preconditioned problem $M^{-1}Ax = M^{-1}b$.

Preconditioning aims to improve the convergence of Richardson’s iteration

$x_{k+1} = (I - A)x_k + b$

However, with $\tilde{A} = M^{-1}A$, $\tilde{M} = I$, and $\tilde{b} = M^{-1}b$ our iteration becomes Richardson’s iteration, and we have Krylov spaces and polynomials based on $\tilde{A}$.

Hence, for simplicity we consider $A$ and $b$ as an explicitly preconditioned matrix and vector, and work with Krylov spaces in $A$ and $r_0$ (most of the time).
Approximations from Krylov Spaces

Richardson for $Ax = b$ seeks update $z_m \in K_m \left(I - A, r_0\right) = K_m \left(A, r_0\right)$

How to define an iteration that finds better approximation in same space?
Given $x_0$ and $r_0 = b -Ax_0$ find $z_m \in K_m \left(A, r_0\right)$ and set $x_m = x_0 + z_m$.

There are several possibilities. Two particularly important ones are:

1. Find $z_m \in K_m \left(A, r_0\right)$ such that $\| r_m \| = \| r_0 - Az_m \|$ is minimal.
2. Find $z_m \in K_m \left(A, r_0\right)$ such that $\| e_m \| = \| z - (x_m + z_m) \|$ is minimal.

The second one seems hard, but is possible in practice for special norms.
Further possibilities for optimal solutions exist, and for non-Hermitian matrices certain non-optimal solutions turn out to have advantages as well.

Inner Products

Many methods to select $z_m$ from the Krylov space are related to projections.

We call $f : S \times S \to \mathbb{R}$ an inner product over the real vector space $S$, if for all vectors $x, y, z$ and scalars $\alpha$:
1. $f(x, x) \geq 0$ and $f(x, x) = 0 \iff x = 0$
2. $f(\alpha x, z) = \alpha f(x, z)$
3. $f(x + y, z) = f(x, z) + f(y, z)$
4. $f(x, z) = f(z, x)$

For a complex inner product, $f : S \times S \to \mathbb{C}$, over a complex vector space $S$ we have instead of property (4): $f(x, z) = f(z, x)$.

Inner products are often written as $\langle x, y \rangle$, $\langle x, y \rangle_\alpha$, or $\langle x, y \rangle'_\alpha$, etc..
We say $x$ and $y$ are orthogonal (w.r.t $\alpha$-IP), $x \perp \alpha y$ if $\langle x, y \rangle_\alpha = 0$. 
Inner products and Norms

Each inner product defines, or induces, a norm: \( \|x\| = \sqrt{\langle x, x \rangle} \). (proof?)

Many norms are induced by inner products, but not all. Those norms that are have additional nice properties (that we’ll discuss soon).

An inner product and its induced norm satisfy: \( |\langle x, y \rangle| \leq \|x\| \|y\| \) (CS ineq)

A norm induced by an inner product satisfies the parallelogram equality:
\[ \|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \]

In this case we can find the inner product from the norm as well:

Real case: \( \langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2) \)

Complex case:
\[ \text{Re}(x, y) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2), \quad \text{Im}(x, y) = \frac{1}{4}(\|x + iy\|^2 - \|x - iy\|^2) \]

Minimum Residual Solutions

First, we consider minimizing the 2-norm of the residual:

Find \( z_m \in K_m(A, r_0) \) such that \( \|r_m\|_2 = \|r_0 - Az_m\|_2 \) is minimal.

The vector 2-norm is induced by the Euclidean inner product \( y^T x = \sum_{i=1}^n y_i x_i \).

It makes sense to minimize the residual, because the error is in general unknown, and we can only directly minimize special norms of the error (those that don’t require the error \( \ominus \)). Moreover, \( \|e_m\|_2 = \|A^{-1}r_m\|_2 \leq \|A^{-1}\|_2 \|r_m\|_2 \), so the norm of the error is bounded by a constant times the norm of the residual.

Finally, note that \( \|r_m\|_2 = \|e_m\|_2 \equiv \|Ae\|_2 \) (show this is a norm if \( A \) regular)

Theorem: \( z_m \) is the minimizer iff \( r_m = b - A(x_0 + z_m) \perp K_m(A, Ar_0) \).

Note that \( b - A(x_0 + z_m) = r_0 - Az_m \).
Minimum Residual Solutions

Proof: Let \( f(z) = \| r_0 - Az \|^2 \).

Then \( \hat{z} \in K_m(A, r_0) : \| r_0 - A\hat{z} \|^2 \) min, implies \( \hat{z} \) is a stationary point of \( f(z) \):

For any unit vector \( p \in K_m(A, r_0) \) we have \( \frac{\partial f(\hat{z})}{\partial p} = 0 \).

So, \( \lim_{\varepsilon \to 0} \frac{f(\hat{z} + \varepsilon p) - f(\hat{z})}{\varepsilon} = 0 \), which gives

\[
\lim_{\varepsilon \to 0} \frac{\| r_0 - A\hat{z} - \varepsilon A p \|^2 - \| r_0 - A\hat{z} \|^2}{\varepsilon} = 0
\]

\[
\lim_{\varepsilon \to 0} -\varepsilon p^\top A^\top (r_0 - A\hat{z}) - \varepsilon (r_0 - A\hat{z})^\top A p + \varepsilon^2 p^\top A^\top A p = 0 \quad \iff
\]

\[
p^\top A^\top (r_0 - A\hat{z}) - (r_0 - A\hat{z})^\top A p = 0 \quad \text{for any unit vector } p \in K_m(A, r_0).
\]

So, \( (r_0 - A\hat{z})^\top A p = 0 \iff (r_0 - A\hat{z}) \perp A p \) for any unit \( p \in K_m(A, r_0) \) (why?)

Since \( A p \in K_m(A, A r_0) \) we have \( (r_0 - A\hat{z}) \perp K_m(A, A r_0) \equiv AK_m(A, r_0) \).

---

Minimum Residual Solutions

Minimizing a norm (cont. function) is in general complicated. However, the orthogonality conditions lead naturally to a linear system of equations.

Let \( \{w_1, w_2, \ldots, w_m\} \) form a basis for \( K_m(A, r_0) \) and \( \{A w_i\}_{i=1}^m \) for \( K_m(A, A r_0) \).

Let \( W_m = [w_1, w_2 \ldots w_m] \). Then \( \hat{\zeta} = W_m \zeta \) (for some unknown \( \zeta \)).

Now the orthogonality conditions \( A w_i \perp (r_0 - A W_m \zeta) \)

yield the linear equations \( \sum_{j=1}^m (w_i^\top A w_j) \zeta_j = w_i^\top A r_0 \).

In matrix form: \( W_m^\top A W_m \zeta = W_m^\top A r_0 \) normal equations (accuracy problems)

LS problem: \( \min_\zeta \| r_0 - K_m \zeta \|_2 \) where \( K_m = AW_m \)

More accurate to solve LS problem using QR-decomposition.

Solving for \( \zeta \) requires only the solution of an \( m \times m \) system independent of \( n \).
Minimum Residual Solutions

We have seen that \( \min_i \| \tilde{r}_i \| - K_m \zeta \| \Leftrightarrow K_m \perp (r_0 - K_m \zeta) \)

Compute \( K_m = Q_m R_m \) (QR-decomposition) where
\[ Q_m \in \mathbb{C}^{n \times m} \left( \mathbb{R}^{n \times m} \right) \text{ s.t. } Q_m^T Q_m = I_m \text{ and } R_m \in \mathbb{C}^{n \times m} \text{ uppertriangular} \]

If rank\( (K_m) = m \), then \( R_m \) is nonsingular and range\( (Q_m) = \text{range}(K_m) \)

Now \( Q_m \perp (r_0 - K_m \zeta) \) gives \( Q_m^T r_0 - Q_m^T Q_m R_m \zeta = 0 \Leftrightarrow R_m \zeta = Q_m^T r_0 \) (easy solve)

Note that \( K_m \zeta = Q_m Q_m^T r_0 \perp r_0 - K_m \zeta = r_0 - Q_m Q_m^T r_0 \)

\( K_m \zeta \) is the orthogonal projection of \( r_0 \) onto range\( (K_m) = K_m (A, A r_0) \).

\( r_m = r_0 - K_m \zeta \) is the orthogonal projection of \( r_0 \) onto \( (K_m (A, A r_0))^\perp \).

\( Q_m Q_m^T \) and \( I - Q_m Q_m^T \) are orthogonal projectors.

\( P \) is projector if \( P^2 = P \), orthogonal projector if \( R(P) \perp N(P) \Leftrightarrow P^* = P \).

Minimum Residual Solutions

Iteration-wise the problem is solved in four steps:

1. Extend the Krylov spaces \( K_m (A, r_0) \) and \( K_m (A, A r_0) \) by adding the respective next vectors \( A^n r_0 \) and \( A^{n+1} r_0 \) (only 1 matvec)
2. Update orthogonal basis for \( K_m (A, A r_0) \): QR-decomp. of \( K_m \)
3. Update projected matrix and projection of \( r_0 \) (orthog) onto \( K_m (A, A r_0) \)
4. Solve the projected problem, e.g. \( R \zeta = Q_m^T r_0 \). Note that this problem is only \( m \times m \) or \( (m + 1) \times m \) irrespective of the size of the linear system.

These steps vary somewhat for different methods.
We would like to carry out these steps efficiently.

The GCR method (Generalized Conjugate Residuals) illustrates these steps well.
(Eisenstat, Elman, and Schulz 1983)
Minimum Residual Solutions: GCR

GCR: \( Ax = b \)

Choose \( x_0 \) (e.g. \( x_0 = 0 \)) and tolerance \( \varepsilon \); set \( r_0 = b - Ax_0; i = 0 \)

while \( \|r_i\| \geq \varepsilon \) do

\[
\begin{align*}
    i &= i + 1 \\
    r_i &= r_{i-1}; c_i = Au_i \\
    \text{for } j = 1, \ldots, i - 1 & \text{ do} \\
    u_i &= u_i - u_j c_j c_i \quad \text{(start QR decomposition)} \\
    c_i &= c_i - c_j c_i \quad \text{Orthogonalize } c_i \text{ against previous } c_j \text{ and} \\
    & \text{update } u_i \text{ such that } Au_i = c_i \text{ maintained} \\
    \text{end do} \\
    u_i &= u_i / \|c_i\|; c_i = c_i / \|c_i\| \quad \text{Normalize; (end QR decomposition)} \\
    x_i &= x_{i-1} + u_i c_i r_{i-1} \quad \text{Project new } c_i \text{ out of residual and update} \\
    r_i &= r_{i-1} - c_i c_i r_{i-1} \quad \text{solution accordingly}; \text{ note } r_i \perp c_j \text{ for } j \leq i \\
\end{align*}
\]

end do

What happens if \( c_i \perp r_{i-1} \)

Minimum Residual Solutions: GCR

From the algorithm we see that if \( c_j r_{j-1} = 0 \) for \( j = 1 \ldots i \):

\[
\begin{align*}
    u_i &= \nu r_0 \in \text{span}\{r_0\} = K_1(A, r_0), \quad \nu \text{ is a normalization constant} \\
    c_i &= \nu Ar_0 \in \text{span}\{Ar_0\} \in K_1(A, Ar_0) \\
    r_i &= r_0 - \alpha c_i = r_0 - \nu \alpha_1 Ar_0 \in K_2(A, r_0) \quad \text{also } Ar_0 \in \text{span}\{r_0, r_1\}
\end{align*}
\]

By induction (and the statements above) we can show

\[
\begin{align*}
    u_i &= r_{i-1} - \sum_{j<i} \beta_j u_j \in \text{span}\{r_0, \ldots, r_{i-1}\} = K_i(A, r_0) \\
    c_i &= Ar_{i-1} - \sum_{j<i} \beta_j c_j \in \text{span}\{Ar_0, \ldots, Ar_{i-1}\} = K_i(A, Ar_0) \quad \text{and also that} \\
    c_i &= Au_i, c_i \perp c_1, c_2, \ldots, c_{i-1}, \text{ and } c_i^t c_i = 1 \text{ (last by construction)} \\
    r_i &= r_{i-1} - \alpha c_i \in \text{span}\{r_0, Ar_0, Ar_1, \ldots, Ar_{i-1}\} = K_{i+1}(A, r_0)
\end{align*}
\]

These relations hold even if \( \{r_0, Ar_0, \ldots, A^m r_0\} \) are dependent for some \( m \).

In that case all the spaces have a maximum dimension of \( m + 1 \).
Minimum Residual Solutions: GCR

Theorem:
Let \( \epsilon_i r_{i-1} = 0 \) for \( i = 1, 2, \ldots, m \). Then \( \| b - Ax \|_2 \) for GCR is minimal.

We use our earlier theorem on the minimum residual. We have (previous slide)
\[ x_m = x_0 + \sum_{i=1}^{m} u_i \beta_i = x_0 + z_m \quad \text{and} \quad z_m \in K_m(A,r_0) \]
as required.

Now we only need to show that \( r_m \perp K_m(A,A r_0) \).
Note that our assumption implies \( r_{i-1} \neq 0 \) for \( i = 1, 2, \ldots, m \).

By construction we have \( r_i = r_0 - c_i c_i^* \) \( r_i = c_i r_0 - c_i c_i^* r_0 = 0 \)
Assume that for \( j = 1 \ldots i-1 \) we have \( r_{j+1} \perp c_1, c_2, \ldots, c_{j-1} \) (induc. hypo.).
From \( r_j = r_{j+1} - c_j c_j^* r_{j+1} \) we have \( r_j \perp c_j \).
For \( i < j \), \( c_i r_j = c_i^* r_{j+1} = 0 \) (from orthogonality \( c_i \) and induction hypothesis).

This proves the required orthogonality result since the \( c_i \) span \( K_m(A,A r_0) \).

Minimum Residual Solutions: GCR

Recapitulation of GCR after \( m \) it.s:
\[ \| r_m \|_2 = \min \{ \| r_0 - Ax \|_2 \mid z_m = U_m \zeta \} \]
\[ u_i \in K_m(A,r_0), \ c_i \in K_m(A,A r_0), \ r_i \in K_{m+1}(A,r_0) \] for \( i = 1, \ldots, m \ \& \ i = 0 \) for \( r_0 \).
This implies that \( K_{i+1}(A,r_0) = \text{span}\{r_0, \ldots, r_i\} \) for \( i = 0, \ldots, m \) (if \( c_i r_{i+1} \neq 0 \)).

Let \( U_m = [u_1 \ u_2 \ \cdots \ u_m] \) and \( C_m = [c_1 \ c_2 \ \cdots \ c_m] \)
Then \( AU_m = C_m, \ C^* C_m = I_m \), and \( \text{range}(U_m) = K_m(A,r_0) \).

For GCR the \textit{projected system} is the matrix for the normal equations (but computed implicitly):
\[ U_m^* A^* A U_m = (AU_m)^* (AU_m) = C_m^* C_m = I_m \]
The \textit{projected right hand side} is \( C_m^* r_0 \), and so we have \( \zeta = C_m^* r_0 \).

\[ z_m = U_m \zeta \] is given by condition \( C_m^* (r_0 - A U_m \zeta) = 0 \), which gives \( \zeta = C_m^* r_0 \).
This gives \( z_m = U_m C_m^* r_0, \ x_m = x_0 + z_m, \) and \( r_m = r_0 - C_m C_m^* r_0 \).
Minimum Residual Solutions: GCR

Note that $C_mC^*_m$ and $I - C_mC^*_m$ are both orthogonal projectors.

$C_mC^*_m$ is a projection since $C_mC^*_mC_mC^*_m = C_mC^*_m$.

$C_mC^*_m$ is an orthogonal projection since $C_mC^*_m$ is Hermitian.

$I - C_mC^*_m$ is a projection since

$$(I - C_mC^*_m)(I - C_mC^*_m) = I - C_mC^*_m - C_mC^*_m + C_mC^*_m = I - C_mC^*_m.$$ $I - C_mC^*_m$ is an orthogonal projection since $I - C_mC^*_m$ is Hermitian.

Model Problems

Discretize $-\left( pu_x \right)_x - \left( qu_y \right)_y + ru_x + su_y + tu = f$.

Integrate equality over box $V$. Use Gauss’ divergence theorem to get

$$\int_V \left( pu_x \right)_x + \left( qu_y \right)_y \, dx dy = \int_{\partial V} \left( pu_x \right)_n \cdot n \, ds$$

And approximate the line integral numerically.
Model Problems

Now we approximate the boundary integral \( \int_{\partial V} \left( \frac{pu}{qu} \right) \cdot n \, ds \).

We approximate the integrals over each side of box \( V \) using the midpoint rule and we approximate the derivatives using central differences.

\[
\int_{y}^{x} p_{x} n_{y} \, dy \approx \frac{\Delta y}{\Delta x} p_{i+1/2,j} (U_{i+1,j} - U_{i,j}) \quad \text{and so on for the other sides}
\]

We approximate the integrals over \( ru_y, su_y, tu, \) and \( f \) using the area of the box and the value at the midpoint of the box, where we use central differences for derivatives. So, \( u_x \approx \frac{U_{i+1,j} - U_{i-1,j}}{2\Delta x} \), and so on.

For various examples we will also do this while strong convection relative to the mesh size makes central differences a poor choice (as it gives interesting systems).

Model problems

This gives the discrete equations

\[
\begin{align*}
- \frac{\Delta y}{\Delta x} [p_{i+1/2,j} (U_{i+1,j} - U_{i,j})] - p_{i-1/2,j} (U_{i,j} - U_{i-1,j}) \\
- \frac{\Delta x}{\Delta y} [q_{i,j+1/2} (U_{i,j+1} - U_{i,j})] - q_{i,j-1/2} (U_{i,j} - U_{i,j-1}) \\
+ \left( \frac{\Delta y}{2} \right) r_{i,j} (U_{i+1,j} - U_{i-1,j}) + \left( \frac{\Delta x}{2} \right) s_{i,j} (U_{i,j+1} - U_{i,j-1}) \\
+ \Delta x \Delta y t_{i,j} U_{i,j} = \Delta x \Delta y f_{i,j}
\end{align*}
\]

Often we divide this result again by \( \Delta x \Delta y \).
Experiments with GCR

Solve on unit square with Dirichlet boundary conditions.

\[ 0 = n \nabla f \]

Note the modest deterioration of convergence rate as discretization becomes finer (so cost slightly worse than linear).
Experiments with GCR

Eigenvalues of the discrete Laplacian for three discretization sizes

For this problem, the smallest eigenvalues and the number of small eigenvalues matter most (as we shall consider later in more detail). All eigenvalues close to 0 for all three meshes.

Eigenvectors of the discrete Laplacian for three discretization sizes

Experiments with GCR
Experiments with GCR

- Convergence for 322 x 322 problem using GCR restarting every 200 it.s
- Restarting saves memory and time (from excessive orthogonalizations)
- Runtime: ~15 minutes on Laptop (Intel Core 2 Duo P8700 @ 2.53 GHz)
- No preconditioning

GCR(200)
Experiments with GCR

- Convergence for 322 x 322 problem using preconditioned GCR
- Restarting saves memory and time (from excessive orthogonalizations)
- Preconditioner: Incomplete ILU without fill-in

Runtimes: GCR(50) ~ 2 min, GCR (without restart) ~ 4 min.
Minimum Residual Solutions: GMRES

An alternative is to generate iteration-wise an orthogonal basis for $K_{m+1}(A, r_0)$. The Arnoldi algorithm (iteration) goes as follows:

Let $v_1 = r_0 / \| r_0 \|_2$;
for $k = 1 \ldots m$,
  $\tilde{v}_{k+1} = A v_k$;
  for $j = 1 \ldots k$,
    $h_{j,k} = v_j^T \tilde{v}_{k+1} ; v_{k+1} = \tilde{v}_{k+1} - h_{j,k} v_j$;
end
  $h_{k+1,k} = \| v_{k+1} \|_2 ; v_{k+1} = v_{k+1} / h_{k+1,k}$;
end

Note/show the following results:

- $A V_m = V_{m+1} H_m$ (Arnoldi recurrence)
- $V_{m+1} V_m = I_{m+1}$ (orthogonal),
- $H_m = V_{m+1}^T A V_m$ (upper Hessenberg)

Using $A V_m = V_{m+1} H_m$, we solve

$$\min_{\| z \|_2} \{ \| r_0 - A z \|_2 \mid z \in K_m(A, r_0) \}$$

as follows.

Let $z = V_m \zeta$, and minimize $\| r_0 - A V_m \zeta \|_2$ over all m-vectors $\zeta$.

Note that this is an $n \times m$ least squares problem (as before).

Now substitute $r_0 = V_{m+1} \eta_0$ and $A V_m = V_{m+1} H_m$. This gives

$$\| V_{m+1} \eta_0 \|_2 - V_{m+1} H_m \zeta = \| V_{m+1} (\eta_0 - H_m \zeta) \|_2 = \| \eta_0 \|_2 - H_m \zeta$$

The latter is a small $(m+1) \times m$ least squares problem we can solve by standard dense linear algebra techniques (e.g., using LAPACK).

We can exploit the structure of $H_m$ and the least squares problem to
1. do this efficiently.
2. compute the residual norm without computing the residual.
By construction $H_m$ has the following structure

$$H_m = \begin{bmatrix}
    h_{1,1} & h_{1,2} & h_{1,3} & \cdots & h_{1,m-1} & h_{1,m} \\
    h_{2,1} & h_{2,2} & h_{2,3} & \cdots & h_{2,m-1} & h_{2,m} \\
    h_{3,2} & h_{3,3} & \cdots & \cdots & h_{m-1,m-1} & h_{m-1,m} \\
    \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \\
    h_{4,3} & \cdots & h_{m-1,m-1} & h_{m-1,m} & h_{m,m} \\
    \vdots & \vdots & \cdots & h_{m,m} & h_{m+1,m} \\
\end{bmatrix}$$

(Upper Hessenberg)

Cheapest QR decompression is by Givens rotations to zero lower diagonal.

$$G_1^H H_m = \begin{bmatrix} c_1 & s_1 \\ -s_1 & c_1 \end{bmatrix} = \begin{bmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ h_{3,2} & \cdots & h_{3,m} \\
\vdots & \vdots & \ddots & \vdots \end{bmatrix}$$

Next step we compute:

$$G_2^H G_1^H H_m = \begin{bmatrix} 1 & c_2 & s_2 \\ -s_2 & c_2 \end{bmatrix} = \begin{bmatrix} * & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ h_{3,2} & h_{3,3} & \cdots & h_{3,m} \\
\vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix} = \begin{bmatrix} * & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ h_{4,3} & \cdots & h_{m,3} \\
\vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

After $m$ Givens rotations:

$$G_m^H \cdots G_1^H H_m = Q_m^H H_m = \begin{bmatrix} r_{1,1} & \cdots & r_{1,m} \\ 0 & r_{2,2} & \cdots \\ \vdots & 0 & \ddots \\ \vdots & \vdots & \ddots & r_{m,m} \end{bmatrix} = R_m$$
**GMRES**

Theorem: An unreduced \((m + 1) \times m\) Hessenberg matrix is nonsingular. (unreduced means no zeros on subdiagonal)

Proof: ?

---

So the least squares problem

\[
y_m = \arg \min \left\{ \| e_1 \| r_0 \|_2 - H_m y \|_2 : y \in \mathbb{C}^m \right\}
\]

can be solved by multiplying \(H_m y \approx e_1 \|r_0\|_2\) from left by \(R_m^{-1}Q_H^{-1}\):

\[
y_m = R_m^{-1}Q_H^{-1}e_1 \|r_0\|_2
\]

In practice:
Stepwise compute \(G_i^H(G_{i-1}^H \cdots G_1^H H_i)\) and \(G_i^H(G_{i-1}^H \cdots G_1^H e_1 \|r_0\|_2)\)
In Arnoldi step, update \(H_{i-1}\) with new column; then carry out previous Givens rotations on new column.
Compute new Givens rotation and update \(H_i\) and right hand side (of small least squares problem): \(G_i^H(G_{i-1}^H \cdots G_1^H e_1 \|r_0\|_2)\)
The least squares system now looks like \( R_i y_i = Q_{i+1}^H e_1 \|r_0\|_2 \).
We may assume \( R_i \) has no zeros on diagonal (see later)

Since bottom row of \( R_i \) is zero we can only solve for \((Q_{i+1}^H e_1 \|r_0\|_2)_{1...i}\) (first \( i \) coeff.s)

This is exactly what we do by solving \( R_i y_i = Q_{i+1}^H e_1 \|r_0\|_2 \)

Note LS residual norm equals the norm of the actual residual:
\[
\|r_i\|_2 = \|q_{i+1}^H e_1\|_2 \|r_0\|_2 (q_{i+1} \text{ since it changes with } i):
\]
\[
\|r_0 - AV_m y\|_2 = \|V_{m+1} e_1 \|_2 - V_{m+1} H_m y\|_2 = \|e_1 \|_2 - H_m y\|_2
\]
This way we can monitor convergence without actually computing updates to solution and residual (cheap).

---

**GMRES**

The least squares system now looks like \( R_i y_i = Q_{i+1}^H e_1 \|r_0\|_2 \).
We may assume \( R_i \) has no zeros on diagonal (see later)

Since bottom row of \( R_i \) is zero we can only solve for \((Q_{i+1}^H e_1 \|r_0\|_2)_{1...i}\) (first \( i \) coeff.s)

This is exactly what we do by solving \( R_i y_i = Q_{i+1}^H e_1 \|r_0\|_2 \)

Note LS residual norm equals the norm of the actual residual:
\[
\|r_i\|_2 = \|q_{i+1}^H e_1\|_2 \|r_0\|_2 (q_{i+1} \text{ since it changes with } i):
\]
\[
\|r_0 - AV_m y\|_2 = \|V_{m+1} e_1 \|_2 - V_{m+1} H_m y\|_2 = \|e_1 \|_2 - H_m y\|_2
\]
This way we can monitor convergence without actually computing updates to solution and residual (cheap).

---

**GMRES**

**GMRES:** \( Ax = b \)

**CHOOSE** \( x_0 \) (e.g. \( x_0 = 0 \)) AND \( tol \)

\( r_0 = b - Ax_0; \ k = 0; \ v_1 = r_0/\|r_0\|_2; \)

**WHILE** \( \|r_k\|_2 > tol \)

\( k = k + 1; \)

\( \tilde{v}_{k+1} = Av_k; \)

**FOR** \( j = 1 : k, \)

\( h_{j,k} = \tilde{v}_j^H v_{k+1}; \)

\( \tilde{v}_{k+1} = \tilde{v}_{k+1} - h_{j,k} v_j; \)

**END**

\( h_{k+1,k} = \|\tilde{v}_{k+1}\|_2; \)

\( v_{k+1} = \tilde{v}_{k+1}/h_{k+1,k}; \)

**UPDATE QR-DECOMP.:** \( H_k = Q_{k+1} R_k \)

\( \|r_k\|_2 = |q_{k+1}^H e_1| \|r_0\|_2 \)

**END**

\( y_k = R_k^{-1} Q_{k+1}^H e_1 \|r_0\|_2; \)

\( x_k = x_0 + V_k y_k; \)

\( r_k = r_0 - V_{k+1} H_k y_k = V_{k+1} \left( I - Q_k^H Q_k \right) e_1 \|r_0\|_2; \) (or simply \( r_k = b - Ax_k \))

---

30-31
09/04/07 / 2:15 AM
So we have generated the Krylov subspace (step 1), and we have an orthogonal basis for it (step 2, more or less). However, we do not have an orthogonal basis for the Krylov subspace (step 1), and we have generated the Krylov subspace.

We can drop last row of \( \bar{R}^{1+w} \bar{O} \) giving:

where \( \bar{R} \) is upper triangular and has last row entirely zero.

QR-decomposition (m Givens rotations),

\[
\bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H}
\]

\( (\dim w) \cdot \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

Obviously we don't want to orthogonalize \( w \bar{A} \bar{V}_m^w \) in addition.

\( (\dim w) \cdot \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

Obviously we don't want to orthogonalize \( w \bar{A} \bar{V}_m^w \) in addition.

\( (\dim w) \cdot \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)

\( \bar{w} \bar{R}^{1+w} \bar{O} = \bar{w}^{1+w} H \equiv \bar{w} \bar{H} \)
Using this QR-decomposition we have

\[
\begin{align*}
\bar{x} & = 0.1 \| I - \bar{\Omega}^{u} H \bar{\Omega}^{u \perp}^{1} R^{u \perp} \Lambda + 0 x = (0.1 - w \Lambda)_{\perp} V + 0 x = w x \\
\bar{x} & = 0.1 \| I - \bar{\Omega}^{u} H \bar{\Omega}^{u \perp}^{1} R^{u \perp} \bar{H}^{1+w \Lambda} - 0 \Lambda
\end{align*}
\]

and

\[
\begin{align*}
0.1 \| I - \bar{\Omega}^{u} H \bar{\Omega}^{u \perp}^{1} R^{u \perp} \bar{\Omega}^{1+w \Lambda} - 0 \Lambda = 0.1 \left( H \left( \bar{\Omega}^{1+w \Lambda} \right) \left( \bar{\Omega}^{1+w \Lambda} \right) - I \right) = w \Lambda
\end{align*}
\]

So for the cost of GMRES

\[
(0.1)_{\perp} V = (w \Lambda \perp \perp) \Lambda (0.1)_{\perp} V
\]

and

\[
\begin{align*}
\bar{x} & = 0.1 \| I - \bar{\Omega}^{u} H \bar{\Omega}^{u \perp}^{1} R^{u \perp} \bar{\Omega}^{1+w \Lambda} - 0 \Lambda = 0.1 \left( H \left( \bar{\Omega}^{1+w \Lambda} \right) \left( \bar{\Omega}^{1+w \Lambda} \right) - I \right) = w \Lambda
\end{align*}
\]

where \( \bar{\Omega}^{1+w \Lambda} \) is unitary and \( R^{u \perp} \) is upper triangular

\[
\begin{align*}
\bar{\Omega}^{u \perp}^{1+w \Lambda} = \bar{H}^{1+w \Lambda} = w \Lambda V
\end{align*}
\]

Using this QR-decomposition we have a QR-decomp. of

\[
\Lambda V
\]
the least squares solution:
\[ \bar{w} \bar{O} \bar{1} - \bar{R} = \lambda \]

So, multiplying an equation \( \lambda \) from the left by \( \bar{w} \bar{H} \) will give

\[ \bar{w} \bar{H} \bar{O} \bar{1} - \bar{R} \]

Note that \( \bar{H} \bar{O} \bar{1} - \bar{R} \) is the left inverse of \( \bar{H} \).

The solution to the least squares problem (in GCR) is given by

\[ \bar{w} \bar{H} \bar{O} \bar{1} = w \bar{R} \lambda \]

(\( \bar{w} = w \Lambda \bar{R} \lambda \) note the relation \( \Lambda \bar{R} \lambda \) and \( \bar{O} \bar{1} = w \Lambda \))

column with a unit scalar:

Comparing with GCR, we see that apart from possibly scaling each

GMRES
Minimum Residual Solutions: GMRES

GMRES: \( Ax = b \)
Choose \( x_0 \), tolerance \( \varepsilon \); set \( r_0 = b - Ax_0 \); \( v_1 = r_0 / \| r_0 \| \), \( k = 0 \).

while \( \| r_k \| \geq \varepsilon \) do
  \( k = k + 1 \)
  \( \tilde{v}_{k+1} = A v_k \);
  for \( j = 1 \ldots k \),
    \( h_{j,k} = v_j^T \tilde{v}_{k+1} \); \( \tilde{v}_{k+1} = \tilde{v}_{k+1} - h_{j,k} v_j \);
  end
  \( h_{k+1,k} = \| \tilde{v}_{k+1} \|_2 \); \( v_{k+1} = \tilde{v}_{k+1} / h_{k+1,k} \);
Solve LS \( \min || y_k - H_k \zeta || \) by construction
  (actually we update the solution rather than solve from scratch – see later)
end
\( x_k = x_0 + V_k \zeta_k \);
\( r_k = r_0 - V_k H_k \zeta_k = V_k ( \eta_k || r_k || - H_k \zeta_k ) \times \text{only } r_k = b - Ax_k \)

Convergence Restarted GCR

Test problem on unit square: 202 \( \times \) 202 grid points
Interior: \(-\nabla \cdot (\nabla u) = 0\) Boundary \( u = 1 \) for \( x = 0 \) and \( y = 1 \)
\( u = 0 \) elsewhere
Convergence restarted GMRES

Test problem on unit square: 202 x 202 grid points

Interior: \(-\nabla \cdot (\nabla u) = 0\)  
Boundary: \(u = 1\) for \(x = 0\) and \(y = 1\)  
\(u = 0\) elsewhere

GMRES vs GCR

<table>
<thead>
<tr>
<th>GMRES(m)</th>
<th>200 x 200 unknowns</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>time (s)</td>
<td>iterations</td>
<td>(\log_{10}(</td>
<td></td>
<td>r</td>
</tr>
<tr>
<td>full</td>
<td>72.888</td>
<td>587</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>40.256</td>
<td>1851</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>41.087</td>
<td>3043</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>63.604</td>
<td>6985</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>111.26</td>
<td>13761</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>199.42</td>
<td>27451</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rGCR(m)</th>
<th>200 x 200 unknowns</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>time (s)</td>
<td>iterations</td>
<td>(\log_{10}(</td>
<td></td>
<td>r</td>
</tr>
<tr>
<td>full</td>
<td>215.87</td>
<td>587</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>114.04</td>
<td>1851</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>97.89</td>
<td>3043</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>103.56</td>
<td>6985</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>131.69</td>
<td>13761</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>180.88</td>
<td>27451</td>
<td>-10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Optimal Krylov Subspace Methods with short recurrences

Multigrid Iterative Methods and
Consider again how GMRES builds an orthogonal basis for \( K_{m+1}^0(A,r_0) \).

Verify that the (Arnoldi) algorithm

\[ v_1 = r_0 / \| r_0 \| \]

for generates the following recurrence:

\[ v_{k+1} = Av_k \]

for \( j = 1 : m \),

\[ V_{m+1} = V_m + H_{m+1} \]

What does \( H \) look like?

\[ w \Lambda v_{k+1} = w \Lambda v_k - h_{j,k} v_j \]

Prove \( \Lambda^{I+1} H \Lambda = \Lambda^{I+1} H \Lambda \) is orthogonal.

Prove \( \Lambda^{I+1} H \Lambda = \Lambda^{I+1} H \Lambda \) is orthogonal.

\[ \end{end} \]

Note \( H \Lambda = \Lambda H \).

For \( j = 1 : m \),

\[ || v_j \| = \| v_j \| \]

Prove \( \Lambda \) is orthogonal.

Prove \( \Lambda \) is orthogonal.

\[ \end{end} \]

Note \( H \Lambda = \Lambda H \).

Prove \( \Lambda \) is orthogonal.

Prove \( \Lambda \) is orthogonal.

\[ \end{end} \]

Consider again how GMRES builds an orthogonal basis for \( K_{m+1}^0(A,r_0) \).

MINRES (1)
Now consider $A$ being Hermitian: $A^H = A$

Another way to write the recurrence relation from Arnoldi:

$$AV_m = V_{m+1}^H H_m = V_m H_m + v_{m+1} e_m^T h_{m+1,m},$$

where $H_m$ is the upper $m \times m$ part of $H_m$.

So, $V_m^H AV_m = V_m^H V_m H_m + V_m^H v_{m+1} e_m^T h_{m+1,m} = H_m$.

$$(V_m^H AV_m)^H = V_m^H A^H V_m = V_m^H AV_m$$ since $A^H = A$, and so $H_m$ must be Hermitian as well.

This has some important consequences ...
A Hermitian upper Hessenberg matrix is tridiagonal! This means that (in exact arithmetic) we need to orthogonalize each new vector $A v_i$ only against the vectors $v_1$ and $v_i - 1$. This means that (in exact arithmetic) we need to orthogonalize each new vector $A v_i$ only against the vectors $v_1$ and $v_i - 1$.

Then generated by the Arnoldi algorithm (so they span $K_m(A, v_1)$). Then span $A v_1, v_2, \ldots, v_i$ and so $A v_1, v_2, \ldots, v_i$ and so $A v_1, v_2, \ldots, v_i$.

**Theorem:** Let $A$ be Hermitian and let $v_1, v_2, \ldots, v_m$ be the vectors.

What is the computational cost of $m$ iterations of GMRES?

We could solve the least squares problem in the same way as for GMRES, except that we save on orthogonalizations (inner products and vector updates).

**Proof:**
Consider $Hv_H = 0$.

Proof:

We know $v_1 \in \text{span}(v_1, \ldots, v_{i-1}, v_i)$. Since $v_i \in \text{ker}(A, v_1)$, we have $Av_i \in \text{ker}(A, v_1)$. Hence $v_H = 0$. Therefore,$$Hv_H = 0.$$
The algorithm now proceeds as follows:

**Lanczos recurrence:** \( (T \text{ for tridiagonal}). \)

\[
AV_{m+1} \Lambda - 0 \mathbf{1} = w \mathbf{1} \Lambda \text{ and } w \Lambda^w \Lambda + 0 \mathbf{x} = w \mathbf{x}.
\]

At end we update our orthogonalization on previous 2 vectors.

Note that each step we only orthogonalize on previous 2 vectors.

\[
R_i y_i = Q_i H e_1 \| \mathbf{r}_0 \|_2^2.
\]

Every step we update the QR-decomposition of \( L \).

\[
\text{Solve least squares problem.}
\]

\[
\text{We have:}
\]

\[
\| \Lambda w A \mathbf{r} - \mathbf{r}_0 \|_2^2 \text{ just as in GMRES:}
\]

MINRES (5)

What would seem an obvious improvement. Can we do that here?

Lanczos is Arnoldi in the Hermitian case (2 orthogonalizations).

Lanczos recurrence: (for tridiagonal).

\[
\tilde{L}^{I+w} \Lambda = w \Lambda A
\]

MINRES (5)
Since we only orthogonalize on the previous two vectors, we would like to discard the other vectors. However, we need them for the update at the end. Can we update every step and discard the vectors $v_i$? The problem is that $R^w$ changes and hence $R^w v_i$ changes (in general). Hence we need all previous $v_i$. We need a trick.
A cunning plan:

Since changes completely every step, apply a change of variables

\[ y^m \]

Alternative for update:

\[ V^m y^m \]

Take and:

\[ W^m = V^m R^m - 1^m y^m \]

\[ = R^m y^m \]

\[ = R^m R^m - 1^m Q^m H^1 \epsilon r_0 \epsilon r_2 \]

Then and each iteration only the last component of

\[ W^m \]

changes. So we can update \( W^m \) without keeping all variables.

So looking at the last (=the new) column we have:

\[ w^m \Lambda = w^m R^m W^m \]

columns are found by solving each iteration.

up triangular with 2 upper diagonals.

from the Givens QR decomposition of a tridiagonal matrix is

Then \( W^m \Lambda = w^m \Lambda \)

Take for update \( W^m \Lambda = w^m \Lambda \)

Since changes completely every step, apply a change of variables.

A cunning plan:
What would be an update formula for \( y_m \)?

Do we need \( y_m \) to continue the iteration?

How many vectors do we need to keep (independent of # iterations)?

The update iteration-wise:

Since \( y_m \), contrary to \( y_m \), changes only in its last position we can do

Update solution:

\[
\begin{align*}
\overline{w}_m M + 0 x &= \overline{w}_m  A + 0 x = w x
\end{align*}
\]
END

\[ Ax = \ell x \]

\[ \ell = \ell_1 \]

\[ (\ell, 2x, 3x, \ldots, kx) \]

\[ 2 \parallel \ell_0 \parallel 1 \]

\[ \ell = \ell_1 \]

\[ (1, 2, \ldots, k) \]

\[ \ell + 1 = \ell \]

\[ \text{while } 0 \]

\[ \ell = \ell_1 \]

\[ 0 < 1 \]

\[ q = \ell_0 \]

\[ q = Ax \]

\[ q = x \]

MINRES (10)
**Conjugate Gradients (1)**

Hermitian matrices: Error minimization in the A-norm

We are solving $Ax = b$ with initial guess $x_0 \to r_0 = b - Ax_0$ and $\hat{x}$ is the solution to $Ax = b$.

The error at iteration $i$ is $\epsilon_i = \hat{x} - (x_0 + z_i)$, where $z_i \in K^i(A, r_0)$ is the $i$th update to the initial guess.

**Theorem:**

Let $A$ be Hermitian, then the vector $z_i \in K^i(A, r_0)$ satisfies

$$z_i = \arg\min\{ \|\hat{x} - (x_0 + z)\|_A : z \in K^i(A, r_0) \} \iff r_i = r_0 - Az_i$$

satisfies $r_i \perp K^i(A, r_0)$.

The most important algorithm of this class is the Conjugate Gradient Algorithm.

---

**Conjugate Gradients (2)**

Proof:

$$z_i = \arg\min\{ \|\hat{x} - (x_0 + z)\|_A : z \in K^i(A, r_0) \} \iff (\hat{x} - x_0) - z_i \perp_A K^i(A, r_0)$$

We know $K^i(A, r_0) = \text{span}\{r_0, r_1, \ldots, r_{i-1}\}$.

This gives $r_k \perp_A (\hat{x} - x_0 - z_i)$ for $k = 0, \ldots, i - 1$ \iff

$$\langle A(\hat{x} - x_0 - z_i), r_k \rangle \text{ for } k = 0, \ldots, i - 1 \iff$$

$$\langle b - Ax_0 - Az_i, r_k \rangle \text{ for } k = 0, \ldots, i - 1 \iff$$

$$\langle r_0 - Az_i, r_k \rangle \text{ for } k = 0, \ldots, i - 1 \iff$$

$$\langle r_i, r_k \rangle \text{ for } k = 0, \ldots, i - 1 \iff$$

$$r_i \perp K^i(A, r_0)$$
So we can minimize the error by choosing the update such that the new residual is orthogonal to all previous residuals. Hence, the name orthogonal residual methods.

(note the comparison between Orthomin(1) and Steepest Descent)

We can generate an orthogonal basis for the Krylov subspace using the Lanczos iteration, the 3-term recurrence version of the Arnoldi-iteration.

Lanczos iteration:

\[
\text{CHOOSE } q_1; \beta_0 = 0; q_0 = 0; \\
\text{FOR } i = 1, 2, \ldots \text{ DO} \\
\quad \tilde{q}_{i+1} = Aq_i; \\
\quad a_i = \langle Aq_i, q_i \rangle; \quad \tilde{q}_{i+1} = q_{i+1} - a_i q_i; \quad q_{i+1} = \tilde{q}_{i+1} - \beta_{i-1} q_{i-1}; \\
\quad \beta_i = \|\tilde{q}_{i+1}\|_2; \quad q_{i+1} = \tilde{q}_{i+1}/\beta_i; \\
\text{END}
\]

Show \( \tilde{q}_{i+1} = \tilde{q}_{i+1} - \beta_{i-1} q_{i-1} \) sets \( \tilde{q}_{i+1} \perp q_{i-1} \). (one argument is the symmetry of the Hessenberg matrix for Arnoldi, give another)

This algorithm generates the recurrence relation:

\[
AQ_i = Q_i T_i + \beta_i q_{i+1} e_i^T, \quad \text{where } Q_i = [q_1 \ q_2 \ \cdots \ q_i], \quad T_i = \begin{bmatrix}
    a_1 & \beta_1 & 0 & \cdots \\
    \beta_1 & a_2 & \beta_2 & \vdots \\
    0 & \beta_2 & \ddots & \ddots \\
    \vdots & \ddots & \ddots & \ddots
\end{bmatrix}.
\]
Conjugate Gradients (5)

Use Lanczos orthonormal basis for minimizing A-norm of error.

\[ z_i = \arg\min \{ \| \hat{x} - (x_0 + z) \|_A : z \in K^i(A, r_0) \} \]

iff \( r_i = r_0 - Az_i \) satisfies \( r_i \perp K^i(A, r_0) \).

\[ q_1 = r_0 / \| r_0 \|_2; \]
Lanczos method: \( AQ_i = Q_i T_i + \beta_i q_{i+1} e_i^T \)

Solve \( r_0 - AQ_i y_i \perp Q_i \Leftrightarrow Q_i^H (\| r_0 \|_2 q_1 - AQ_i y_i) = 0 \Leftrightarrow \]
\( Q_i^H (\| r_0 \|_2 q_1 - AQ_i y_i) = 0 \Leftrightarrow \| r_0 \|_2 e_1 - Q_i^H AQ_i y_i = 0 \)

Notice \( \text{range}(Q_i) = \text{span}\{r_0, r_1, \ldots, r_{i-1}\} \).
\( AQ_i = Q_i T_i + \beta_i q_{i+1} e_i^T \Rightarrow Q_i^H AQ_i = T_i \)

So we reduced our problem to solving \( \| r_0 \|_2 e_1 - T_i y_i = 0 \):

\[ y_i = T_i^{-1} e_1 \| r_0 \|_2 \]

Conjugate Gradients (6)

In order to update step-by-step we use same trick as in MINRES:

Let \( T_i = L_i D_i L_i^H \); then \( y_i = L_i^{-H} D_i^{-1} L_i^{-1} e_1 \| r_0 \|_2 \), where \( L_i \) is unit lower bi-diagonal with lower diagonal coeff.s, \( l_1, l_2, \ldots, l_{i-2} \) (index gives the column)

Change of variables: \( P_i = Q_i L_i^{-H} \) and \( \hat{y}_i = D_i^{-1} L_i^{-1} e_1 \| r_0 \|_2 \): \( Q_i y_i = P_i \hat{y}_i \)

Each iteration only the last component of \( \hat{y}_i \) changes. From \( P_i L_i^H = Q_i \) we get a recurrence for \( p_i \): \( p_i + l_{i-1} p_{i-1} = q_i \) \((p_1 = q_1)\)

So every new step we compute a new \( q_{i+1} \), we update the decomposition of \( T_i \) and from that \( \hat{y}_{i+1} \) and \( p_{i+1} \).

\( x_i = x_{i-1} + p_i \hat{y}_{i+1} \)
\( r_i = r_{i-1} - Ap_i \hat{y}_{i+1} = q_{i+1} \beta_i \hat{y}_{i+1} \) \((\text{where } \hat{y}_{i+1} \text{ is } i^{th} \text{ comp of vector } \hat{y}_i)\)
Conjugate Gradients (7)

This leads to the coupled two-term recurrence form of CG

CG algorithm: \( Ax = b \)

1. \( \text{CHOOSE} \, x_0 \rightarrow r_0 = b - Ax_0; \)
2. \( p_1 = r_0; \, i = 0 \)

\[ \text{WHILE} \, \| r_i \|_2 > \text{tol} \, \text{DO} \]

   1. \( i = i + 1; \)
   2. \( a_i = \frac{\langle r_{i-1}, r_{i-1} \rangle}{\langle p_{i-1}, Ap_{i-1} \rangle}; \)
   3. \( x_i = x_i + a_i p_i; \)
   4. \( r_i = r_{i-1} - a_i Ap_i; \)
   5. \( \beta_i = \frac{\langle r_i, r_i \rangle}{\langle r_{i-1}, r_{i-1} \rangle}; \)
   6. \( p_i = r_i - \beta_i p_{i-1}; \)

\[ \text{END} \]