Lanczos = symmetric Arnoldi

If A is symmetric ($A = A^T$), then H_n is symmetric for each n and hence it is tridiagonal.

```
A = bucky(); b = randn(size(A, 1), 1);
[Q, H] = arnoldi(bucky, b, 40);
spy(H);
```

Indeed, $H_{ji} = q_j^T A q_i = (q_j^T A q_i)^T = q_i^T A q_j = H_{ij}$.

So we can shorten the orthogonalization loop:

```
w = A * Q(:, j);
for i = j-1:j % only 2 vectors to check
    %and actually we already know H(j-1,j)=H(j,j-1)
    H(i,j) = Q(:, i)' * w;
    w = w - H(i,j) * Q(:, i);
end
```

$v \mapsto A \cdot v$

Reduces the cost to n matrix products $+ O(mn)$.

Lanczos iteration, used to compute eigenvalues/vectors.

$$Aq_j = \cancel{\beta_1 q_1} + \cancel{\beta_2 q_2} + \cdots + \cancel{\beta_{j-2} q_{j-2}} + \beta_{j-1} q_{j-1} + \beta_j q_j + \beta_{j+1} q_{j+1}$$

"short recurrence"

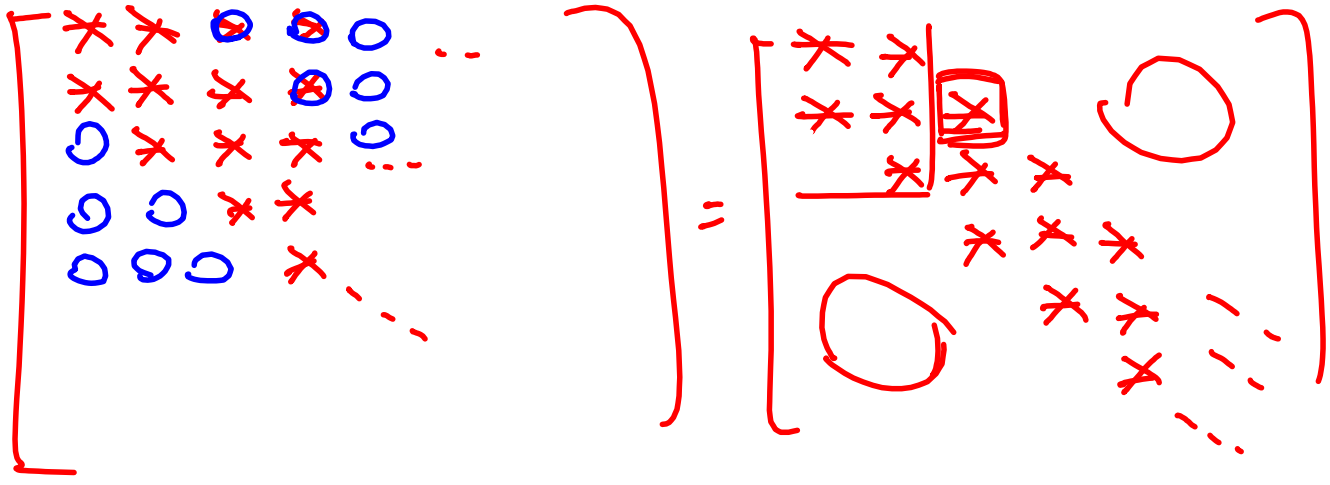
H_{ij} contains $B_i = q_i^T A q_j$

If $A = A^T$, then $H_{ij} = q_i^T A q_j =$

$= \left(q_i^T A q_j\right)^T = q_j^T A^T q_i = q_j^T A q_i = H_{ji}$

because it's 1×1

$A = A^T$



Conjugate gradient

AF, optimization

Suppose $A = A^T$ is positive definite. Then, we can find the solution to $Ax = b$ by minimizing the (strictly convex) function

$$f(x) = \frac{1}{2} x^T A x - b^T x.$$

Surprisingly, conjugate gradient on this problem can be interpreted as a Krylov subspace method.

Three ingredients: current iterate x_k, residual $r_k = b - Ax_k = -\nabla f(x_k)$, and search direction d_k.

CG iteration

$x_0 = 0$, $r_0 = d_0 = b$;
for $k = 1{:}n$ **do**
  $\alpha_k = (r_{k-1}^T r_{k-1})/(d_{k-1}^T A d_{k-1})$;
  $x_k = x_{k-1} + \alpha_k d_{k-1}$;
  $r_k = r_{k-1} - \alpha_k A d_{k-1}$; // check it!
  $\beta_k = (r_k^T r_k)/(r_{k-1}^T r_{k-1})$;
  $d_k = r_k + \beta_k d_{k-1}$; // A-orthogonalize search dir
end

solution update
residual/gradient update

$d_{k-1}^T A d_k = 0$

1)

$$x_k = x_{k-1} + \alpha_k d_{k-1}$$

We can choose α by imposing for instance

$$v^T (Ax_k - b) = 0 \quad \text{for a certain vector } v,$$

or by minimizing $f(x_{k-1} + \alpha d_{k-1})$

2) Once we update x_{k-1} to $x_k = x_{k-1} + \alpha_k d_{k-1}$

we need to compute $r_k = b - Ax_k$

$$r_k = b - Ax_k = b - A(x_{k-1} + \alpha_k d_{k-1}) =$$

$$= \underbrace{b - Ax_{k-1}}_{r_{k-1}} - \alpha_k A d_{k-1}$$

CG: choose search direction d_k as

$$d_k = r_k - \beta_k d_{k-1}$$

Choose β_k so that the search direction d_k is "A-conjugate" to d_{k-1}, i.e.

$$d_{k-1}^T A d_k = 0$$

$$f(x) = \frac{1}{2} x^T A x - b^T x$$

$$\nabla f(x) = Ax - b$$

$\nabla^2 f = A$, if we assume A posdef it's strictly convex

one global minimum in $\mathbb{R}^n$, where $\nabla f(x) = Ax - b = 0$

CG — Considerations

Storage: 3 vectors: x_k, r_k, d_k. No need to keep whole 'history'.
Cost: n mat-vec products (Ad_{k-1}) plus $O(nm)$.

Where are the Krylov subspaces? At each step k,

$$K_k(A, b) = \text{span}(x_1, x_2, \ldots, x_k) = \text{span}(d_0, d_1, \ldots, d_{k-1}) = \text{span}(r_0, r_1, \ldots, r_{k-1}).$$

(We sketch what happens in the first steps.)

$x_0 = 0 \quad d_0 = r_0 = b$

$\text{span}(d_0) = \text{span}(r_0) = b$

$x_1 = x_0 + \alpha d_0$

$\text{span}(x_1) = b$

$r_0 = b - Ax_0$

$r_1 = b - Ax_1$

$r_1 - r_0 = -A(x_1 - x_0) = A\alpha d_0$

$r_1 = r_0 - \alpha A d_0 \quad \in \text{span}(b, Ab)$

$d_1 = r_1$

CG — Orthogonality

Where are the orthogonal bases?

At each step $r_k^T r_i = d_k^T A d_i = 0$ for all $i < k$.

I.e., the r_i are orthogonal, and the d_i are A-orthogonal.
Proof (sketch) Suppose (induction) it holds up to step $k-1$.
From $r_k = r_{k-1} - \alpha A d_{k-1}$ it follows that

$$r_i^T r_k = r_i^T r_{k-1} - \alpha_k r_i^T A d_{k-1}.$$

If $i < k-1$, then $r_i^T r_{k-1}$ is zero by induction and
$r_i \in \text{span}(\text{d}_0, \text{d}_1, \ldots, \text{d}_\text{i})$ is A-orthogonal to d_{k-1}.
If $i = k-1$, then the RHS is zero provided $\alpha_k = \frac{r_{k-1}^T r_{k-1}}{r_{k-1}^T A d_{k-1}}$. The formula we gave was slightly different: at the denominator we had $d_{k-1}^T A d_{k-1}$ instead. But since $d_{k-1} = r_{k-1} + \beta_{k-1} d_{k-2}$, the difference between the two denominators is $\beta_{k-1} d_{k-2}^T A d_{k-1} = 0$ (by induction). The second part is similar.

CG — orthogonal bases

The r_i are an orthogonal (but not orthonormal!) basis of $K_k(A, b)$ at each step k. In particular, $\frac{1}{\|r_i\|} r_i$ coincides (up to a sign) with the q_i obtained with Arnoldi.

Last ingredient: what do we solve at each step? Not $\min_{z \in K_k(A,b)} \|Az - b\|$ like in GMRES, but

$$x_k \text{ is the solution of } \min_{z \in K_k(A,b)} (x - z)^T A (x - z).$$

where x is the true solution of $Ax - b$.

Indeed, let $z = x_k + y$, and note that $A(x - x_k) = b - Ax_k = r_k$ is orthogonal to each $y \in K_k(A, b)$:

$$\begin{aligned}(x - z)^T A (x - z) &= (x - x_k)^T A (x - x_k) + (x - x_k)^T A y + y^T A (x - x_k) \\ &\quad + y^T A y \\ &= (x - x_k)^T A (x - x_k) + y^T A y \geq (x - x_k)^T A (x - x_k).\end{aligned}$$

Convergence of CG

First result: let $\lambda_{\max}$, $\lambda_{\min}$ be the maximum/minimum eigenvalue of A; then, CG converges with rate

$$\|x - x_k\| \leq \left(\frac{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \right)^k \|x - x_0\|.$$

(we won't see a proof).

We can rewrite it in terms of a more familiar quantity: for a positive definite matrix, eigenvalues and singular values coincide, hence

$$\frac{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} = \frac{\sqrt{\sigma_1} - \sqrt{\sigma_m}}{\sqrt{\sigma_1} + \sqrt{\sigma_m}} = \frac{\sqrt{\frac{\sigma_1}{\sigma_m}} - 1}{\sqrt{\frac{\sigma_1}{\sigma_m}} + 1} = \frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1}.$$

For large values of $\kappa(A)$, this is approximately $1 - \frac{2}{\sqrt{\kappa(A)}}$.

Convergence of CG

Again, that is a worst-case result — usually, convergence is faster and depends on the eigenvalues of A (like in GMRES).

Let $\|v\|_A := v^T A v$ "A-norm" of a vector; the quantity that is minimized at each step. It makes sense if A is posdef.

$x_n \in K_n(A, b) \iff x_n = p(A)b$ for a polynomial p of degree $< n$.

$$\begin{aligned} \min_{x_n \in K_n(A,b)} \|x - x_n\|_A &= \min_{p(x)} \|x - p(A)Ax\|_A \\ &= \min_{\substack{q(x)=1-xp(x) \\ \text{of degree} \le n}} \|q(A)x\|_A. \end{aligned}$$

As for GMRES, if A has only n different eigenvalues, then this minimum reaches 0 after n steps. If the eigenvalues of A are 'clustered', one can construct polynomials such that $q(\lambda)$ is small for each $\lambda \implies$ fast convergence.