GMRES: a Krylov method for linear systems

Goal: solving a large-scale linear system $Ax = b$, $A \in \mathbb{R}^{m\times m}$.

Idea we look for 'the closest thing to solution' inside $K_n(A, b)$.

$$\min\|Ax - b\|, \quad x = Q_n y.$$

I.e., $\min_{y\in\mathbb{R}^n}\|AQ_n y - b\|$: a least-squares problem.

Can be reduced to a smaller-size problem:

$$\|AQ_n y - b\| = \|Q_{n+1}\underline{H}_n y - b\| = \left\|\begin{bmatrix} Q_{n+1} & \widehat{Q}\end{bmatrix}^T (Q_{n+1}\underline{H}_n y - b)\right\|$$
$$= \left\|\begin{bmatrix} \underline{H}_n y - \|b\|e_1 \\ \\ 0 \end{bmatrix}\right\| = \left\|\underline{H}_n y - \|b\|e_1\right\|.$$

A LS problem of size $(n+1)\times n$.

Implementation details

- `qr(H)` can be computed in $O(n^2)$ using the special shape of H (Hessenberg matrix).
  Not such a big deal because Arnoldi costs more anyway.
- Instead of doing a QR at the end, we can compute QRs of $\underline{H}_1, \underline{H}_2, \ldots$ and update them at each step.
  This allows us to compute residuals $\|Ax_n - b\|$ (that we can use as stopping criterion).

We won't see an implementation with all these details.
Matlab has `gmres(A, b)` (and Python has `scipy.sparse.linalg.gmres`).

Example

```
m = 1000;
A = 10*speye(m) + sprandn(m, m, 0.01);
b = randn(size(A,1), 1);
[Q, H] = arnoldi(A, b, 50);
r = nan(50,1);
% tests several 'slices' of Q,H
for n = 1:50
    y = H(1:n+1, 1:n) \ eye(n+1, 1) * norm(b);
    x = Q(:, 1:n) * y;
    r(n) = norm(A*x - b);
end
semilogy(r)
```

See how changing that factor 10 affects convergence speed.

GMRES convergence

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

$$\min_{x \in K_n(A,b)} \|Ax - b\| = \min_{\substack{q(x)=1-xp(x) \\ \text{of degree} \le n}} \|q(A)b\|.$$

If $A = V\Lambda V^{-1}$ diagonalizable, then

$$q(A) = V \begin{bmatrix} q(\lambda_1) & & \\ & \ddots & \\ & & q(\lambda_m) \end{bmatrix} V^{-1}.$$

- If A has very few distinct eigenvalues ($k \le n$ of them), then we can find q such that $q(\lambda_i) = 0$ for all i.
- If A has few 'clusters' of eigenvalues, we can find q such that $q(\lambda_i)$ is small for all i.

Example Repeat the previous experiment with `A=bucky()`.