Note Title

Krylov subspace methods.

This is a class of methods — symm. matrices
— non-symm. (non-normal matrices)

Idea: to project the n-dim'l problem to a lower dim'l subspace.

Typical situation/assumption: the matrix A is not available, but $A\cdot x$ can be computed for any vector x.

What information about A can be derived?
Start with some non-zero vector $b = y_1$ (say), $(b \in \mathbb{R}^n)$

$$\underset{=Ay_1}{y_2 = A\cdot b}, \; y_3 = A\cdot y_2, \; \ldots, \; y_n = A y_{n-1}; \; \ldots$$

Let $K = [y_1 | y_2 | \cdots | y_n]_{n\times n}$.

Then $A\cdot K = [Ay_1 | Ay_2 | \cdots | Ay_n]$

$$= [\underbrace{y_2 | y_3 | \cdots | y_n}_{\text{last } (n-1) \text{ columns of } K} | A^n y_1]$$

$$= K \underbrace{[e_2 | e_3 | \cdots | e_n | -c]}_{C}$$ (assuming that K is invertible)

$$AK = KC$$

$$\& \quad C = K^{-1}AK = \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_1 \\ 1 & 0 & & 0 & -c_2 \\ 0 & 1 & & \vdots & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & 1 & -c_n \end{bmatrix}$$

Notice that C is the companion matrix of

$$p(x) = x^n + \sum_{i=1}^{n} c_i x^{i-1}$$

- K could be singular
- K could be ill-conditioned.

The thing to do is take the QR factorization of K:

$$\text{if } K = QR, \text{ then } \underbrace{K^{-1}AK}_{C=} = R^{-1}Q^{-1}AQR$$

$$\therefore Q^{-1}AQ = RCR^{-1} = H = (h_{ij})$$

Notice that H is upper Hessenberg.

To compute the columns of Q, we use an idea similar to G-S.

Arnoldi's algorithm: for (partial reduction) to Hessenberg form.

(idea is similar to applying MGS to $\{y_1 = b, y_2 = Ab, \dots, y_n = Ay_{n-1}\}$.

$q_1 = b/\|b\|_2$

(suppose we wish to compute k columns of H).

for $j = 1$ to k
  $z = Aq_j$
  for $i = 1$ to j
    $h_{ij} = q_i^T z$
    $z = z - h_{ij} q_i$
  end
  $h_{j+1,j} = \|z\|$
  if $h_{j+1,j} = 0$ quit

Q ✓ $\quad H$ ✓

$r_{11} = \|q_1\|$

$r_{22} = \|y_2 - \sim\|$

$r_{ij} = q_i^T y_j$

$q_{j+1} = z / h_{j+1,j}$

end.

Now we have an orthogonal matrix Q s.t. $Q^{-1} A Q = H$.

Suppose we computed k columns of Q i.e.

$$Q = [\underbrace{q_1 | \cdots q_k}_{Q_k} | \underbrace{\cdots q_n}_{Q_u}]$$

$$\rightarrow H = Q^T A Q = [Q_k | Q_u]^T A [Q_k | Q_u]$$

$$= \left[\begin{array}{c|c} Q_k^T A Q_k & Q_k^T A Q_u \\ \hline Q_u^T A Q_k & Q_u^T A Q_u \end{array}\right] \leftarrow \text{upper Hessenberg.}$$

$$= \left[\begin{array}{c|c} H_k & H_{ku} \\ \hline H_{uk} & H_u \end{array}\right]$$

H_k is upper Hessenberg & is known.

☆ When A is symmetric, H is symmetric & tridiagonal; call it T.

$$T = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \ddots & \beta_{n-1} \\ & & \beta_{n-1} & \alpha_n \end{bmatrix}$$

$Q^T A Q = H$.

$Q^T A Q = T$.

$\underline{AQ} = \underline{QT}$.

Equating the j^{th} column of $AQ = QT$:

$$A q_j = \beta_{j-1} q_{j-1} + \alpha_j q_j + \beta_j q_{j+1}$$

Since columns of q are orthonormal,

$$q_j^T A q_j = \alpha_j$$

This gives a simplification of Arnoldi's algorithm:

Lanczos' algorithm:

$q_1 = b/\|b\|_2$, $\beta_0 = 0$, $q_0 = 0$.

for $j = 1$ to k

$\quad z = Aq_j$

$\quad \alpha_j = q_j^T z$

$\quad z = z - \alpha_j q_j - \beta_{j-1} q_{j-1}$

$\quad \beta_j = \|z\|_2$

$\quad$ if $\beta_j = 0$, quit.

$\quad q_{j+1} = z/\beta_j$

end.

$$T = Q^T A Q = [Q_k \; Q_u]^T A [Q_k \; Q_u]$$

$$= \left[\begin{array}{c|c} Q_k^T A Q_k & Q_k^T A Q_u \\ \hline Q_u^T A Q_k & Q_u^T A Q_u \end{array}\right]$$

$$= \begin{bmatrix} T_k & T_{ku} \\ T_{uk} & T_{uu} \end{bmatrix}.$$

T_k is known, T_{uk} is all zeroes except possibly on non-zero top right entry.

We will see foll. methods:

	$Ax=b$	$Ax=\lambda x$
$A=A^*$	CG.	Lanczos
$A\neq A^*$	GMRES	Arnoldi's

Defn: The Krylor subspace $K_k(A, b)$ is the span of $\{b, Ab, A^2b, \dots, A^{k-1}b\}$ $(1\le k\le n)$

$$K_k = Q_k R$$

Columns of Q_k form an orthonormal basis of K_k.

In general, we use Krylor subspaces in the foll. way: we look for a solution of $Ax=b$ in the K_k i.e. subspace spanned by Q_k

i.e. we look for $x_k = \sum_{j=1}^{k} z_j q_j \in \text{span } Q_k$.

We want x_k to be the "best" possible solution for $Ax=b$.

Let x denote the actual/true solution,

& let $r_k = b - Ax_k$ denote the residual.

Different definitions of 'what is best' give no diff. methods.

1 The 'best' x_k minimizes $\|x_k - x\|_2$, but this is not computable since x is unknown.

2. The 'best' x_k minimizes $\|r_k\|_2$ — this gives GMRES methods (Generalised Minimum RESidual)

3. The 'best' x_k makes $r_k \perp Q_k$ (orthogonal residual ppty).
(SYMMLQ, a variation of GMRES.)

4. When A is symmetric & pos. def. then it defines a norm on $\mathbb{R}^n$: $\|b\| = (b^T A b)^{1/2}$.

In general, when A is HPD $(n \times n)$, then A defines an inner product on $\mathbb{R}^n$: $\langle p, q \rangle = p^T A q$.