Krylov subspace methods

- Introduction to Krylov subspace techniques
- FOM, GMRES, practical details.
- Symmetric case: Conjugate gradient
- See Chapter 6 of text for details.

**Motivation**

- Common feature of one-dimensional projection techniques:
  \[ x_{\text{new}} = x + \alpha d \]
  where \( d \) = a certain direction.
- \( \alpha \) is defined to optimize a certain function.
- Equivalently: determine \( \alpha \) by an orthogonality constraint
- Example:
  In MR:
  \[ x(k+1) = x + \alpha r \]
  where \( r = b - Ax \), with \( d = b - Ax \).
- \( \min_{\alpha} \| b - Ax(\alpha) \|_2 \) reached iff \( b - Ax(\alpha) \perp r \)
- One-dimensional projection methods are greedy methods. They are 'short-sighted'.

**Example**

Recall in Steepest Descent: New direction of search \( \tilde{r} \) is \( \perp \) to old direction of search \( r \).

\[ r \leftarrow b - Ax, \quad \alpha \leftarrow (r, r)/(Ar, r) \]
\[ x \leftarrow x + \alpha r \]

**Question:** can we do better by combining successive iterates?
- Yes: Krylov subspace methods.

**Krylov subspace methods: Introduction**

- Consider MR (or steepest descent). At each iteration:
  \[ r_{k+1} = b - A(x^{(k)} + \alpha_k r_k) = r_k - \alpha_k Ar_k = (I - \alpha_k A)r_k \]
- In the end:
  \[ r_{k+1} = (I - \alpha_k A)(I - \alpha_{k-1} A) \cdots (I - \alpha_0 A)r_0 = p_{k+1}(A)r_0 \]
  where \( p_{k+1}(t) \) is a polynomial of degree \( k + 1 \) of the form
  \[ p_{k+1}(t) = 1 - tq_k(t) \]
- Show that:
  \[ x^{(k+1)} = x^{(0)} + q_k(A)r_0 \], with \( \deg(q_k) = k \)
- Krylov subspace methods: iterations of this form that are 'optimal' [from \( m \)-dimensional projection methods]
**Krylov subspace methods**

**Principle:** Projection methods on Krylov subspaces:

\[ K_m(A, v_1) = \text{span}\{v_1, Av_1, \ldots, A^{m-1}v_1\} \]

- The most important class of iterative methods.
- Many variants exist depending on the subspace \( L \).

**Simple properties of \( K_m \)**

- Notation: \( \mu = \text{deg. of minimal polynomial of } v \). Then:
  - \( K_m = \{p(A)v | p = \text{polynomial of degree } \leq m - 1\} \)
  - \( K_m = K_\mu \) for all \( m \geq \mu \). Moreover, \( K_\mu \) is invariant under \( A \).
  - \( \dim(K_m) = m \) iff \( \mu \geq m \).

---

**A little review: Gram-Schmidt process**

**Goal:** given \( X = [x_1, \ldots, x_m] \) compute an orthonormal set \( Q = [q_1, \ldots, q_m] \) which spans the same subspace.

**ALGORITHM : 1. Classical Gram-Schmidt**

1. For \( j = 1, \ldots, m \) Do:
2. \( \hat{q}_j := x_j \)
3. For \( i = 1, \ldots, j - 1 \) Do
4. \( r_{ij} = (\hat{q}_j, q_i) \)
5. \( \hat{q}_j := \hat{q}_j - r_{ij}q_i \)
6. EndDo
7. \( r_{jj} = \|\hat{q}_j\|_2 \). If \( r_{jj} == 0 \) exit
8. \( q_j := \hat{q}_j / r_{jj} \)
9. EndDo

**Result:**

\[ X = QR \]
Arnoldi’s algorithm

Goal: to compute an orthogonal basis of \( K_m \).

Input: Initial vector \( v_1 \), with \( \|v_1\|_2 = 1 \) and \( m \).

For \( j = 1, \ldots, m \) Do:

Compute \( w := Av_j \)

For \( i = 1, \ldots, j \) Do:

\[ h_{i,j} := (w, v_i) \]

\[ w := w - h_{i,j}v_i \]

EndDo

Compute: \( h_{j+1,j} = \|w\|_2 \) and \( v_{j+1} = w / h_{j+1,j} \)

EndDo

Result of orthogonalization process (Arnoldi):

1. \( V_m = [v_1, v_2, \ldots, v_m] \) orthonormal basis of \( K_m \).
2. \( AV_m = V_{m+1}H_m \)
3. \( V_m^TAV_m = H_m \equiv \tilde{H}_m - \) last row.

Arnoldi’s Method for linear systems \((L_m = K_m)\)

From Petrov-Galerkin condition when \( L_m = K_m \), we get

\[ x_m = x_0 + V_mH_m^{-1}V_m^T r_0 \]

Select \( v_1 = r_0 / \|r_0\|_2 \equiv r_0/\beta \) in Arnoldi’s. Then

\[ x_m = x_0 + \beta V_mH_m^{-1}e_1 \]

What is the residual vector \( r_m = b - Ax_m \)?

Several algorithms mathematically equivalent to this approach:


* Also Conjugate Gradient method [see later]

Minimal residual methods \((L_m = AK_m)\)

When \( L_m = AK_m \), we let \( W_m \equiv AV_m \) and obtain relation

\[ x_m = x_0 + V_m[W_m^TAV_m]^{-1}W_m^T r_0 \]

\[ = x_0 + V_m[(AV_m)^TAV_m]^{-1}(AV_m)^T r_0. \]

Use again \( v_1 := r_0 / (\beta := \|r_0\|_2) \) and the relation

\[ AV_m = V_{m+1}H_m \]

\[ x_m = x_0 + V_m[\tilde{H}_m^T \tilde{H}_m]^{-1}\tilde{H}_m^T \beta e_1 = x_0 + V_m y_m \]

where \( y_m \) minimizes \( \|\beta e_1 - \tilde{H}_m y\|_2 \) over \( y \in \mathbb{R}^m \).
Gives the Generalized Minimal Residual method (GMRES) ([Saad-Schultz, 1986]):

\[ x_m = x_0 + V_m y_m \]
\[ y_m = \min_y \| \beta e_1 - \bar{H}_m y \|_2 \]

Several Mathematically equivalent methods:
- Axelsson’s CGLS
- Orthomin (1980)
- Orthodir
- GCR

A few implementation details: GMRES

**Issue 1**: How to solve the least-squares problem?

**Issue 2**: How to compute residual norm (without computing solution at each step)?

Several solutions to both issues. Simplest: use Givens rotations.

Recall: We want to solve least-squares problem
\[ \min_y \| \beta e_1 - \bar{H}_m y \|_2 \]

Transform the problem into upper triangular one.

Rotation matrices of dimension \( m + 1 \). Define (with \( s_i^2 + c_i^2 = 1 \)):

\[ \Omega_i = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \\ -s_i & c_i & 1 \\ c_i & s_i & \end{bmatrix} \]

\( \bar{H}_m \) and right-hand side \( \bar{g}_0 \equiv \beta e_1 \) by a sequence of such matrices from the left. \( s_i, c_i \) selected to eliminate \( h_{i+1,i} \)

1-st Rotation:

\[ \Omega_1 = \begin{bmatrix} c_1 & s_1 & 1 \\ -s_1 & c_1 & 1 \\ 1 & \cdots & 1 \\ \end{bmatrix} \]

with:

\( s_1 = \frac{h_{21}}{\sqrt{h_{11}^2 + h_{21}^2}} \)
\( c_1 = \frac{h_{11}}{\sqrt{h_{11}^2 + h_{21}^2}} \)
\[
\tilde{H}_{m}^{(1)} = \begin{bmatrix}
h_{11}^{(1)} & h_{12}^{(1)} & h_{13}^{(1)} & h_{14}^{(1)} & h_{15}^{(1)} \\
(1) & (1) & (1) & (1) & (1) \\
h_{22} & h_{23} & h_{24} & h_{25} & \\
(1) & (1) & (1) & (1) & \\
h_{33} & h_{34} & h_{35} & & \\
(1) & (1) & (1) & & \\
h_{44} & h_{45} & & & \\
(1) & (1) & & & \\
h_{55} & & & & \\
(1) & & & & \\
h_{65} & & & & \\
0 & & & & \\
\end{bmatrix}, \quad \bar{g}_1 = \begin{bmatrix}
c_1 \beta \\
s_1 \beta \\
0 \\
0 \\
0 \\
\end{bmatrix}
\]

Repeat with \(\Omega_2, \ldots, \Omega_5\).

Result:

\[
\tilde{H}_{5}^{(5)} = \begin{bmatrix}
h_{11}^{(5)} & h_{12}^{(5)} & h_{13}^{(5)} & h_{14}^{(5)} & h_{15}^{(5)} \\
(1) & (1) & (1) & (1) & (1) \\
h_{22} & h_{23} & h_{24} & h_{25} & \\
(1) & (1) & (1) & (1) & \\
h_{33} & h_{34} & h_{35} & & \\
(1) & (1) & (1) & & \\
h_{44} & h_{45} & & & \\
(1) & (1) & & & \\
h_{55} & & & & \\
(1) & & & & \\
h_{65} & & & & \\
0 & & & & \\
\end{bmatrix}, \quad \bar{g}_5 = \begin{bmatrix}
\gamma_1 \\
\gamma_2 \\
\gamma_3 \\
\vdots \\
\gamma_6 \\
\end{bmatrix}
\]

Define

\[
Q_m = \Omega_m \Omega_{m-1} \ldots \Omega_1, \\
\tilde{R}_m = \tilde{H}_m^{(m)} = Q_m \tilde{H}_m, \\
\bar{g}_m = Q_m (\beta e_1) = (\gamma_1, \ldots, \gamma_{m+1})^T.
\]

\[\Longrightarrow\] Since \(Q_m\) is unitary,

\[
\min \|\beta e_1 - \tilde{H}_m y\|_2 = \min \|\bar{g}_m - \tilde{R}_m y\|_2.
\]

\[\Longrightarrow\] Delete last row and solve resulting triangular system.

\[
\tilde{R}_m y_m = \bar{g}_m
\]

**Proposition:**

1. The rank of \(AV_m\) is equal to the rank of \(R_m\). In particular, if \(r_{mm} = 0\) then \(A\) must be singular.

2. The vector \(y_m\) that minimizes \(\|\beta e_1 - \tilde{H}_m y\|_2\) is given by

\[
y_m = R_m^{-1} g_m.
\]

3. The residual vector at step \(m\) satisfies

\[
b - Ax_m = V_{m+1} [\beta e_1 - \tilde{H}_m y_m] \\
= V_{m+1} Q_m^T (\gamma_{m+1} e_{m+1})
\]

4. As a result, \(\|b - Ax_m\|_2 = |\gamma_{m+1}|\).