TANA15/Lecture 10 - Contents

Projection Methods

- Definition. Optimality Results. The Minimal Residual Method.
- Krylov Subspaces. GMRES and CG.

Applications

- Image Deblurring.
- Sparse Least Squares and Eigenvalues.

Projection Methods for Linear Systems

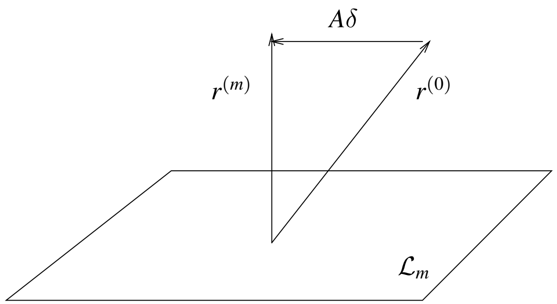
Definition Let $\mathcal{K}_m$ and $\mathcal{L}_m$ be two m–dimensional subspaces. The *projection method* computes an approximate solution to $Ax = b$ by finding

$$x^{(m)} \in x^{(0)} + \mathcal{K}_m \text{ such that } r^{(m)} = b - Ax^{(m)} \perp \mathcal{L}_m.$$

Questions Existance and Uniqueness?

How to compute $x^{(m+1)}$ from $x^{(m)}$? Convergence properties?

Illustration



Find a $\delta \in \mathcal{K}_m$ so that $x^{(0)} + \delta$ has a residual orthogonal to $\mathcal{L}_m$.

Solution Let $V = (v_1, \ldots, v_m)$ be a basis for $\mathcal{K}_m$ and $W = (w_1, \ldots, w_m)$ be a basis for $\mathcal{L}_m$. The solution can be written,

$$x^{(m)} = x^{(0)} + Vy.$$

The orthogonality relation $r^{(m)} \perp \mathcal{L}_m$ leads to

$$W^T AVy = W^T r^{(0)} \quad \Longleftrightarrow \quad x^{(m)} = x^{(0)} + V(W^T AV)^{-1} W^T r^{(0)}.$$

Remark The matrix $W^T AV$ is of size $m \times m$ and the projection step is well defined for $W^T AV$ non-singular.

Lemma Let A be symmetric and positive definite and chose $\mathcal{L}_m = \mathcal{K}_m$. Then the projection method is well-defined.

Lemma Let A be non-singular and chose $\mathcal{L}_m = A\mathcal{K}_m$. Then the projection method is well-defined.

Remark In both cases $W^T AV$ is non-singular.

Two classes of methods. For matrices that are either symmetric and positive definite, or for general non-singular matrices.

Example Let $\mathcal{K}_1 = \mathcal{L}_1 = \text{span}(r^{(k)})$ and suppose A is symmetric and positive definite. Derive the projection method, and write down the algorithm, for this case.

What about convergence?

Optimality results

Let $x^{(m)}$ be given by the *projection method* with $\mathcal{L}_m$, $\mathcal{K}_m$ and $x^{(0)}$.

Definition If A is symmetric and positive definite the *energy norm* is $\|x\|_A^2 = x^T Ax$.

Proposition If A symmetric, positive definite and $\mathcal{L}_m = \mathcal{K}_m$. Then

$$\|x^{(m)} - x^*\|_A = \min_{x \in x^{(0)} + \mathcal{K}_m} \|x - x^*\|_A.$$

Remark The error in the *energy norm* is minimized.

Let $x^{(m)}$ be given by the *projection method* with $\mathcal{L}_m$, $\mathcal{K}_m$ and $x^{(0)}$.

Proposition If A is non-singular and $\mathcal{L}_m = A\mathcal{K}_m$. Then

$$\|b - Ax^{(m)}\|_2 = \min_{x \in x^{(0)} + \mathcal{K}_m} \|b - Ax\|_2.$$

Remark The *residual* is minimized methods.

Observation If $A \in \mathbb{R}^{n \times n}$ and $m = n$ then the projection method produces the exact solution.

Example Let A non-singular. In the *Minimal Residual Iteration* we use $V = (r^{(k)})$ and $W = AV = (Ar^{(k)})$.

Algorithm Let $x^{(0)}$ be the starting vector.

for $k = 1, 2, \ldots$
- $r^{(k-1)} = b - Ax^{(k-1)}$.
- $z^{(k-1)} = Ar^{(k-1)}$.
- $\alpha_k = (z^{(k-1)}, r^{(k-1)})/\|z^{(k-1)}\|_2^2$.
- $x^{(k)} = x^{(k-1)} + \alpha_k r^{(k-1)}$.

end

What about the convergence?

Krylov subspaces

Definition The *Krylov subspace* $\mathcal{K}_m(A, r^{(0)})$ is

$$\mathcal{K}_m(A, r^{(0)}) = \text{span}(r^{(0)}, Ar^{(0)}, \ldots, A^{m-1}r^{(0)}).$$

Definition A *Krylov subspace* method computes an approximate solution of the form

$$x^{(m)} \in x^{(0)} + \mathcal{K}_m(A, r^{(0)}).$$

Remark A basis for $K_m(A, r^{(0)})$ can be computed while only accessing the matrix through matrix-vector multiply.

Observation The Krylov vectors $A^j r^{(0)}$ are close to linearly dependent for large j. How can this be a good subspace?

Remark Different choices for $\mathcal{L}_m$ gives rise to different *Krylov subspace methods*.

If A is non-singular and $\mathcal{L}_m = A\mathcal{K}_m(A, r^{(0)})$ then we have The *Generalized Minimal Residual* method (GMRES).

If A is symmetric and positive definite and $\mathcal{L}_m = \mathcal{K}_m(A, r^{(0)})$ then we have the *Conjugate gradient* method (CG).

The Arnoldi process

Algorithm Find an orthonormal basis for $\mathcal{K}_m(A, v_1)$ by

for $j = 1, 2, \ldots, m$ **do**
- $w_j = Av_j$
- **for** $i = 1, 2, \ldots, j$ **do**
  - $h_{ij} = (w_j, v_i)$.
  - $w_j := w_j - h_{ij}v_j$.
- **end**
- $h_{j+1,j} = \|w_j\|_2$. **if** $h_{j+1,j} = 0$ **then** *stop*.
- $v_{j+1} = w_j/h_{j+1,j}$.

end

Remark Only matrix-vector multiply. The Gram-Schmidt process makes the new vector w_j orthogonal to the previous basis vectors $v_1, v_2, \ldots, v_j$.

Break down

Definition If $Ax \in \mathcal{K}$ for all $x \in \mathcal{K}$ then $\mathcal{K}$ is an *invariant subspace*.

If the dimension of the Krylov subspace $\mathcal{K}_m(A, r^{(0)})$ is n for $m \geq n$. Then $\mathcal{K}_m(A, r^{(0)})$ is an invariant subspace.

Lemma If $\mathcal{K}_n(A, r^{(0)})$ is an invariant subspace then $A^{-1}b \in x^{(0)} + \mathcal{K}_n(A, r^{(0)})$.

Remark Break down in Arnoldi's method is good.

Proposition Let H_m be the Hessenberg matrix and V_m be the orthogonal basis computed by the Arnoldi method. Then

$$AV_m = V_m H_m + w_m e_m^T, \quad \text{and,} \quad V_m^T A V_m = H_m.$$

Remark If the dimension n is large then storing the matrix V_m can require alot of memory. Often only a small dimension m for the Krylov subspace $\mathcal{K}_m(A, r^{(0)})$ is used.

The Generalized Minimal Redisual Method

Definition The *Generalized Minimal Redisual Method* (GMRES) is the *projection method* using the subspaces $\mathcal{K}_m = \mathcal{K}_m(A, r^{(0)})$ and $\mathcal{L}_m = A\mathcal{K}_m$.

Corollary The solution $x^{(m)} = x^{(0)} + V_m y$ minimize the residual, i.e.

$$\min_{x^{(m)} \in x^{(0)} + \mathcal{K}_m(A, r^{(0)})} \|b - Ax^{(m)}\|_2 = \min_{y \in \mathbb{R}^m} \|r^{(0)} - AV_m y\|_2,$$

where V_m is an orthogonal basis for $\mathcal{K}_m(A, r^{(0)})$.

Remark As long as $\mathcal{K}_m$ has full column rank $x^{(m)} \to x^*$ as $m \to n$. If *break down* occurs then we have an exact solution.

Implementation of GMRES

Algorithm Solve $Ax = b$ by the steps

Compute $r^{(0)} = b - Ax^{(0)}$, $\beta = \|r^{(0)}\|_2$, and $v_1 = r^{(0)}/\beta$.
Let $\bar{H} = (h_{ij}) \in \mathbb{R}^{(m+1)\times m}$ and set $\bar{H} = 0$.
for $j = 1, 2, \ldots, m$ **do**
  Compute $w_j := Av_j$.
  for $i = 1, 2, \ldots, j$ **do**
    $h_{ij} := (w_j, v_i)$.
    $w_j := w_j - h_{ij} v_j$.
  end
  $h_{j+1,j} = \|w_j\|_2$. **if** $h_{j+1,j} = 0$ **then** set $m := j$ *break*.
  $v_{j+1} = w_j / h_{j+1,j}$.
end
Solve $\|\beta e_1 - \bar{H}_m y\|_2$ and set $x^{(m)} = x^{(0)} + V_m y$.

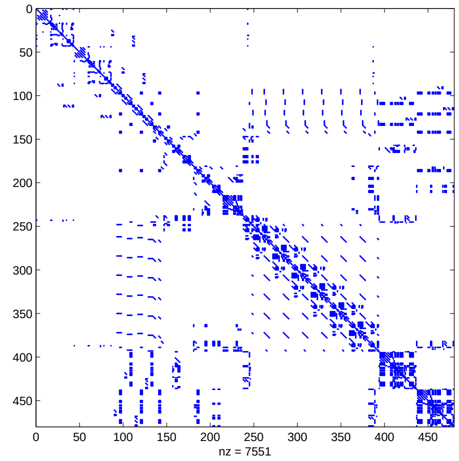
Remarks Can modify the algorithm so that $\|\beta e_1 - \bar{H}_j y\|$ is solved in each step.

Expensive to store V_m for large m. Instead use a small m and *restart* using $x_{k+1}^{(0)} := x_k^{(m)}$.

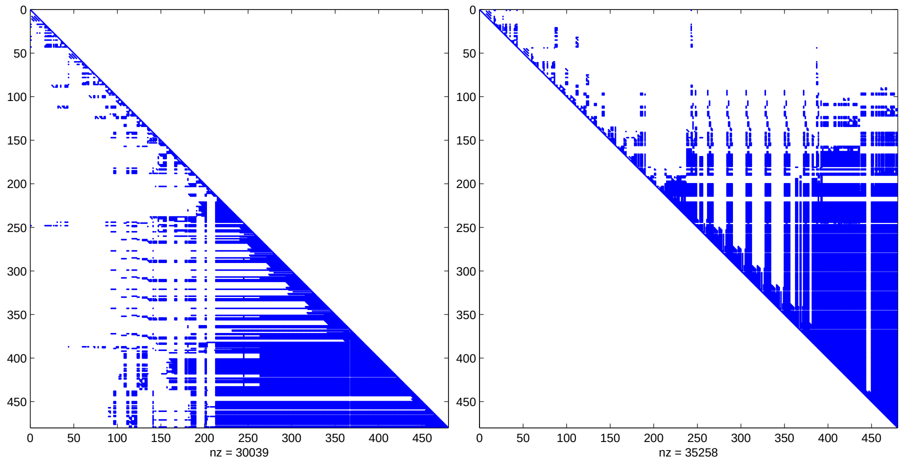
Matlab Solve $Ax = b$ in Matlab using

```
>> x = gmres(A,b, 10 , 10^-12 );
```

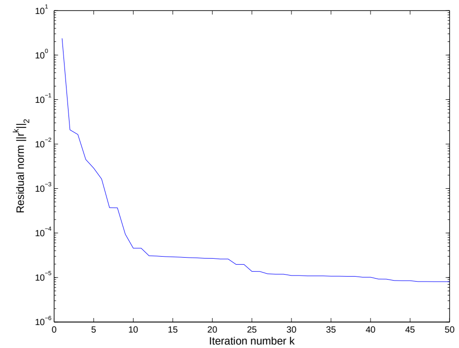
The Arnoldi process stops after a Krylov space of dimension $m = 10$ is obtained and the algorithm restarts with a new residual.



Example Test matrix `West0479`. Dimension $N = 479$. Symmetric and positive definite. 7551 non-zero elements.



The matrices L (left) and U (right) in the decomposition $LU = PA$. A total of 65297 non-zero elements to store.



Results Residual norm $\|b - Ax^{(k)}\|_2$ for the first 50 GMRES iterations. Restart after 8 steps in the inner loop. Around 10–15 iterations is enough for a good solution.

Remark The residual is monotonically decreasing.

Symmetric Matrices

Definition The *conjugate gradient method* is the projection method using $\mathcal{K}_m = \mathcal{L}_m = \mathcal{K}_m(A, r^{(0)})$.

Definition The vectors x and y are *conjugate*, or A–orthogonal, if they are orthogonal with respect to $(x, y)_A = x^T A y$.

Definition If A is symetric positive and definite the A–norm is

$$\|x\|_A = x^T A x = \|Rx\|_2, \quad A = R^T R.$$

Proposition If A symmetric and positive definite the *conjugate gradient method* finds a solution $x^{(m)}$ that satisfies,

$$\|x^* - x^{(m)}\|_A = \min_{x \in x^{(0)} + \mathcal{K}_m} \|x^* - x\|_A.$$

Remark The norm $\|\cdot\|_A$ is often called the *energy norm* and is natural to consider in applications, e.g. finite elements.

Question How to implement? The Arnoldi process computes a basis for $\mathcal{K}_m$ but doesn't take advantage of symmetry.

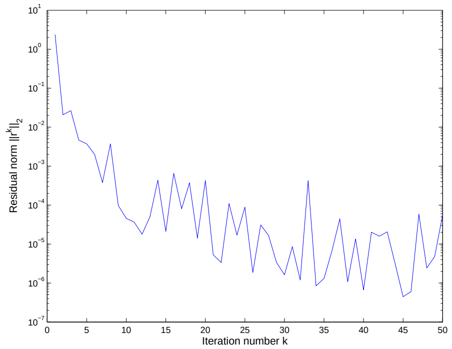
The alternative for symmetric matrices is called the *Lanzcos* process.

The Conjugate Gradient Method

Algorithm Suppose A is symmetric and Positive Definite. Solve $Ax = b$ by the steps

Let $r^{(0)} = b - Ax^{(0)}$, $p_0 := r^{(0)}$.
for $j = 1, 2, \ldots$ **do**
$\alpha_j = (r^{(j)}, r^{(j)})/(Ap_j, p_j)$.
$x^{(j+1)} := x^{(j)} + \alpha_j p_j$.
$r^{(j+1)} := r^{(j)} - \alpha_j A p_j$.
$\beta_j := \|r^{(j+1)}\|_2 / \|r^{(j)}\|_2$.
$p_{j+1} := r^{(j+1)} + \beta_j p_j$.
end

Remark Need to store 4 vectors (x, p, Ap, and r). In Matlab `pcg` implements this.



Results Residual norm $\|b - Ax^{(k)}\|_2$ for the first 50 CG iterations. Around 10 iterations is enough for a good solution.

Remark The error $\|x^{(k)} - x^*\|_A$ is monotically decreasing but not nessecarily the residual.

Sparse Least Squares problems

Problem Want to minimize $\|Ax - b\|_2$ where A is large and sparse.

Lemma The solution of the least squares problem $\min \|Ax - b\|_2$ is obtained by solving the normal equations

$$A^T Ax = A^T b.$$

Remark The matrix $A^T A$ is symmetric and positive definite. In Matlab

```
>> [x]=cgls(A,b);
```

The product $A^T A$ is not sparse. Compute $y = (A^T A)x = A^T(Ax)$.

Application: Image Deblurring

Example An out-of-focus photo appears blurred.

Optics gives an explicit expression for the blurring operator A. We have $I_{obs} = A \cdot I_{exact}$. The image is 260×260 pixels.

Method The matrix is *ill-conditioned*. Compute the solution by solving

$$\min \|AI_\lambda - I_{obs}\|_2^2 + \lambda^2 \|I_\lambda\|_2^2,$$

for a proper value of λ. This is called *Tikhonovs method*.

Lemma The Tikhonov solution I_λ can be computed by solving the *modified normal equations*

$$(A^T A + \lambda I)I_\lambda = A^T I_{obs}.$$

Remark This is a symmetric linear system that can be solved using the CG method (i.e. `pcg` in Matlab). Avoid computing $C = A^T A$.

Results Use $\lambda = 10^{-5}$ and type

```
>> bfun = @(I) A'*(A*I)+lambda^2*I
>> Itik=pcg( bfun , A'*Ib , 10^-4 , 200 );
```

The restored image is obtained after 29 CG iterations.

Proposition Suppose A is non-singular and V_m is the orthogonal basis computed by the Arnoldi method. Then

$$V_m^T A V_m = H_m \in \mathbb{R}^{m \times m}.$$

Remark If $m = n$ then the matrix V_m is orthogonal and this is an similarity transformation and $\lambda(A) = \lambda(H_m)$.

Run the Arnoldi process but keep only the vectors $v_j, v_{j-1}, \ldots, v_{j-k}$. Then $\lambda_{1,\ldots,k}(A) \approx \lambda_{1,\ldots,k}(H_m)$.

Matlab The function `eigs` implements this.