Iterative methods for linear systems: conjugate gradient and GMRES
Calderon preconditioning

Stefanie Müller

11.11.2010

Index

Direct methods vs iterative methods

Full matrix of order n:

- direct method: costs about $\frac{2}{3}n^3$
- iterarive method: costs about n^2 for every iteration

conjugate gradient

- Solve the system $A\mathbf{x} = \mathbf{b}$, where A is a symmetric positive definite matrix.
- We want to find $\mathbf{x}^k$ recursively:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{p}^k$$

- We define $\mathbf{r}^k = \mathbf{b} - A\mathbf{x}^k$

conjugate gradient

We define

$$\Phi(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T A\mathbf{y} - \mathbf{y}^T\mathbf{b}$$

Theorem

We have that:

$$\mathbf{x} \text{ solution of } A\mathbf{x} = \mathbf{b} \Leftrightarrow \mathbf{x} \text{ minimum point of } \Phi(\mathbf{y})$$

$\Rightarrow$ We want to find the minimum point of the function Φ, starting from a point $\mathbf{x}^0$

conjugate gradient

Given the direction $\mathbf{p}^k$, we can find α_k that minimizes $\Phi(\mathbf{x}^{k+1}) = \Phi(\mathbf{x}^k + \alpha_k \mathbf{p}^k)$
We obtain

$$\alpha_k = \frac{{\mathbf{p}^k}^T \mathbf{r}^k}{{\mathbf{p}^k}^T A \mathbf{p}^k}$$

How to find $\mathbf{p}^k$?

conjugate gradient

Definition

A solution $\mathbf{x}^k$ is said to be optimal with respect to a direction $\mathbf{p} \neq \mathbf{0}$ if

$$\Phi(\mathbf{x}^k) \leq \Phi(\mathbf{x}^k + \lambda \mathbf{p}) \quad \forall \lambda \in \mathbb{R}$$

If $\mathbf{x}^k$ is optimal w. r. t. all directions of a vector space V, $\mathbf{x}^k$ is said to be optimal w. r. t. V.

Theorem

If $\mathbf{x}^k$ is optimal with respect to $\mathbf{p}$, $\mathbf{p}$ is orthogonal to $\mathbf{r}^k$.

conjugate gradient

- We look for directions which conserve the optimality of the iterates.
- Suppose to have $\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{q}$, with $\mathbf{x}^k$ optimal with respect to a direction $\mathbf{p}$ (i.e. $\mathbf{r}^k \perp \mathbf{p}$).
- Impose $\mathbf{x}^{k+1}$ optimal with respect to $\mathbf{p}$ (i.e. $\mathbf{r}^{k+1} \perp \mathbf{p}$).
  We obtain that

$$\mathbf{p}^T A\mathbf{q} = 0$$

That is, the directions are A-orthogonal, or A-conjugate.

conjugate gradient

How to find these directions?

- Set $\mathbf{p}^0 = \mathbf{r}^0$
- $\mathbf{p}^{k+1} = \mathbf{r}^{k+1} - \beta_k \mathbf{p}^k$ for $k = 0, 1, \ldots$
  where β_k is defined such that ${\mathbf{p}^j}^T A\mathbf{p}^{k+1} = 0$ for $j = 0, 1, \ldots, k$
- We get for β_k:

$$\beta_k = \frac{(A\mathbf{p}^k)^T \mathbf{r}^{k+1}}{(A\mathbf{p}^k)^T \mathbf{p}^k}$$

conjugate gradient

Summarizing, we get the method of the conjugate gradient:
Choose $\mathbf{x}^0$, set $\mathbf{r}^0 = \mathbf{b} - A\mathbf{x}^0$, $\mathbf{p}^0 = \mathbf{r}^0$
Iterate over k=0,1, ...

$$
\begin{aligned}
\alpha_k &= \frac{{\mathbf{p}^k}^T \mathbf{r}^k}{{\mathbf{p}^k}^T A\mathbf{p}^k} \\
\mathbf{x}^{k+1} &= \mathbf{x}^k + \alpha_k \mathbf{p}^k \\
\mathbf{r}^{k+1} &= \mathbf{r}^k - \alpha_k A\mathbf{p}^k \\
\beta_k &= \frac{(A\mathbf{p}^k)^T \mathbf{r}^{k+1}}{(A\mathbf{p}^k)^T \mathbf{p}^k} \\
\mathbf{p}^{k+1} &= \mathbf{r}^{k+1} - \beta_k \mathbf{p}^k
\end{aligned}
$$

conjugate gradient

Remark

One can show that:

- $\alpha_k = \frac{\|\mathbf{r}^k\|_2^2}{{\mathbf{p}^k}^T A\mathbf{p}^k}$
- $\beta_k = \frac{\|\mathbf{r}^{k+1}\|_2^2}{\|\mathbf{r}^k\|_2^2}$
- $A\mathbf{r}^k = -\frac{1}{\alpha_k}\mathbf{r}^{k+1} + (\frac{1}{\alpha_k} - \frac{\beta_{k-1}}{\alpha_{k-1}})\mathbf{r}^k + \frac{\beta_{k-1}}{\alpha_{k-1}}\mathbf{r}^{k-1}$

conjugate gradient

Theorem

Let A be a symmetric, positive definite matrix, $n \times n$. The method of conjugate gradient for the system $A\mathbf{x} = \mathbf{b}$ converges at most in n steps. Moreover, the error $\mathbf{e}^k$ is orthogonal to $\mathbf{p}^j$ for $j = 0, 1, ..., k-1$ and

$$\|\mathbf{e}^k\|_A \leq \frac{2c^k}{1+c^{2k}}\|\mathbf{e}^0\|_A \quad \text{where } c := \frac{\sqrt{\kappa_2(A)}-1}{\sqrt{\kappa_2(A)}+1}$$

Remark

To have a better convergence, we want $\kappa_2(A)$ small, where $\kappa_2(A) := \|A\|_2\|A^{-1}\|_2 = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$

preconditioned conjugate gradient

We have seen that to have a faster convergence $\kappa_2(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$ should be as small as possible.
So if $\kappa_2(A) >> 1$, we can write the system in the form:

$$P^{-\frac{1}{2}} A P^{-\frac{1}{2}} \mathbf{y} = P^{-\frac{1}{2}} \mathbf{b} \ \text{ with } \mathbf{y} = P^{\frac{1}{2}} \mathbf{x}$$
$$\text{i.e. } P^{-\frac{1}{2}} A \mathbf{x} = P^{-\frac{1}{2}} \mathbf{b}$$

preconditioned conjugate gradient

We obtain the method of preconditioned conjugate gradient:
Given $\mathbf{x}^0$, set $\mathbf{r}^0 = \mathbf{b} - A\mathbf{x}^0$, $\mathbf{z}^0 = P^{-1}\mathbf{r}^0$, $\mathbf{p}^0 = \mathbf{z}^0$.
Iterate over $k = 0, 1, \dots$

$$\alpha_k = \frac{{\mathbf{p}^k}^T \mathbf{r}^k}{(A\mathbf{p}^k)^T \mathbf{p}^k}$$

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{p}^k$$

$$\mathbf{r}^{k+1} = \mathbf{r}^k - \alpha_k A\mathbf{p}^k$$

$$P\mathbf{z}^{k+1} = \mathbf{r}^{k+1}$$

$$\beta_k = \frac{(A\mathbf{p}^k)^T \mathbf{z}^{k+1}}{{\mathbf{p}^k}^T A\mathbf{p}^k}$$

$$\mathbf{p}^{k+1} = \mathbf{z}^{k+1} - \beta_k \mathbf{p}^k$$

preconditioned conjugate gradient

Remark

- The estimation of the errors is the same as in the CG, substituting A with $P^{-1}A$.
- The implementation of PCG does not request to compute $P^{\frac{1}{2}}$ or $P^{-\frac{1}{2}}$.
- Solving $P\mathbf{z}^{k+1} = \mathbf{r}^{k+1}$ increases the computational cost w.r.t. the CG.
- We need to find a preconditioning matrix P such that:
  - It is easy to solve the linear system $P\mathbf{z}^{k+1} = \mathbf{r}^{k+1}$
  - $\kappa_2(P^{-1}A)$ should be near to 1, to decrease the number of steps necessary to get a good convergence

iterative methods in Krylov's spaces

Consider the Richardson's method $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{r}^k$
We have that:

$$\mathbf{r}^k = \prod_{j=0}^{k-1} (I - \alpha_j A)\mathbf{r}^0 \tag{1}$$

So $\mathbf{r}^k = p_k(A)\mathbf{r}^0$, where $p_k(A)$ is a polynom in A of degree k.

Definition

We define the Krylov's space of order m as:

$$K_m(A, \mathbf{v}) = \mathrm{span}\left\{\mathbf{v}, A\mathbf{v}, ..., A^{m-1}\mathbf{v}\right\}$$

It is a subspace of the space generated by all vectors $\mathbf{u} \in \mathbb{R}^n$ of the form $\mathbf{u} = p_{m-1}(A)\mathbf{v}$, where p_{m-1} is a polynom in A of degree $\leq m - 1$.

iterative methods in Krylov's spaces

Remark

(1) implies that $\mathbf{r}^k \in K_{k+1}(A, \mathbf{r}^0)$

We can observe that

$$\mathbf{x}^k = \mathbf{x}^0 + \sum_{j=0}^{k-1} \alpha_j \mathbf{r}^j$$

where $\sum_{j=0}^{k-1} \alpha_j \mathbf{r}^j$ is a polynom in A of degree $\leq k-1$, and so

$$\mathbf{x}^k \in W_k := \left\{ \mathbf{v} = \mathbf{x}^0 + \mathbf{y} : \mathbf{y} \in K_k(A, \mathbf{r}^0) \right\}$$

That is, we are looking for a solution approximating $\mathbf{x}$ in the space W_k

iterative methods in Krylov's spaces

In general, we have methods of the form:

$$\mathbf{x}^k = \mathbf{x}^0 + q_{k-1}(A)\mathbf{r}^0$$

where $q_{k-1}(A)$ is a polynom choosen such that $\mathbf{x}^k$ is the best approximation of $\mathbf{x}$ in W_k.

Definition

Such methods are called Krylov's methods.

Property

Let $A \in \mathbb{R}^{nxn}, \mathbf{v} \in \mathbb{R}^n$. The Krylov's subspace $K_m(A, \mathbf{v})$ has dimension m if and only if the degree of $\mathbf{v}$ with respect to A, $deg_A(\mathbf{v})$, is not smaller than m, being the degree of $\mathbf{v}$ w. r. t. A the minimum degree of a monic non-zero polynomial p in A, for which $p(A)\mathbf{v} = 0$.

iterative methods in Krylov's spaces

Fixed m, we can compute an orthonormal basis for $K_m(A, \mathbf{v})$, using Arnoldi's algorithm, based on Gram-Schmidt's algoritm. Applying Gram-Schmidt we would get:

$$\mathbf{v}_1 = \frac{\mathbf{v}}{\|\mathbf{v}\|_2}$$

$$\mathbf{w}_{k+1} = A^k \mathbf{v} - \sum_{i=1}^{k} h_{ik} \mathbf{v}_i$$

$$\mathbf{v}_{k+1} = \frac{\mathbf{w}_{k+1}}{\|\mathbf{w}_{k+1}\|_2}$$

where h_{ik}'s are choosen imposing the orthogonalaty of $\mathbf{w}_{k+1}$.

iterative methods in Krylov's spaces

Applying Arnoldi's algorithm we get:

$$\mathbf{v}_1 = \frac{\mathbf{v}}{\|\mathbf{v}\|_2}$$

$$h_{ik} = \mathbf{v}_i^T A\mathbf{v}_k \quad i = 1, 2, \dots k$$

$$\mathbf{w}_{k+1} = A\mathbf{v}_k - \sum_{i=1}^{k} h_{ik}\mathbf{v}_i$$

$$h_{k+1,k} = \|\mathbf{w}_{k+1}\|_2$$

$$\mathbf{v}_{k+1} = \frac{\mathbf{w}_{k+1}}{\|\mathbf{w}_{k+1}\|_2}$$

iterative methods in Krylov's spaces

$\mathbf{v}_1, \ldots, \mathbf{v}_m$ build an orthonormal basis for $K_m(A, \mathbf{v})$.
Defining $V_m = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$, we have that

$$V_m^T A V_m =: H_m$$
$$V_{m+1}^T A V_m =: \hat{H}_m$$

where $\hat{H}_m$ superior Hessenberg matrix with entries h_{ij} from above.

Remark

The algorithm stops at an intermediate step $k < m$ if and only if $deg_A(\mathbf{v}_1) = k$.

Now we can apply a Krylov's method of type

$$\mathbf{x}^k = \mathbf{x}^0 + q_{k-1}(A)\mathbf{r}^0$$

to solve the system $A\mathbf{x} = \mathbf{b}$.

iterative methods in Krylov's spaces

How to find $\mathbf{x}^k$? We have two possibilities:

- $\mathbf{x}^k \in W_k$ such that $\mathbf{r}^k$ is orthogonal to every vector in $K_k(A, \mathbf{r}^0)$, that is

$$\mathbf{x}^k \in W_k \text{ such that } \mathbf{v}^T(\mathbf{b} - A\mathbf{x}^k) = 0 \ \forall \mathbf{v} \in K_k(A, \mathbf{r}^0)$$

  $\Rightarrow$ FOM (= Full Orthogonalization Method)

- $\mathbf{x}^k \in W_k$ such that it minimizes the Euclidean norm of the residual $\|\mathbf{r}^k\|_2$, that is

$$\|\mathbf{b} - A\mathbf{x}^k\|_2 = \min_{\mathbf{v} \in W_k} \|\mathbf{b} - A\mathbf{v}\|_2$$

  $\Rightarrow$ GMRES (= Generalized Minimum RESiduals)

GMRES

We build a basis for $K_k(A, \mathbf{r}^0)$ with Arnoldi's algorithm, setting $\mathbf{v}_1 = \frac{\mathbf{r}^0}{\|\mathbf{r}^0\|_2}$, and we find $V_k = (\mathbf{v}_1, ..., \mathbf{v}_k)$.
We can compute $\mathbf{x}^k = \mathbf{x}^0 + V_k \mathbf{z}^k$.
How to choose $\mathbf{z}^k$?

$$\begin{aligned}
\mathbf{x}^k &= \mathbf{x}^0 + V_k \mathbf{z}^k \\
\mathbf{r}^k &= \mathbf{r}^0 - AV_k \mathbf{z}^k = \mathbf{v}_1 \|\mathbf{r}^0\|_2 - AV_k \mathbf{z}^k \\
V_{k+1}^T \mathbf{r}^k &= \mathbf{e}_1 \|\mathbf{r}^0\|_2 - \hat{H}_k \mathbf{z}^k \\
\mathbf{r}^k &= V_{k+1}(\mathbf{e}_1 \|\mathbf{r}^0\|_2 - \hat{H}_k \mathbf{z}^k)
\end{aligned}$$

So choose $\mathbf{z}^k$ such that $\| \|\mathbf{r}^0\|_2 \mathbf{e}_1 - \hat{H}_k \mathbf{z}_k \|_2$ is minimum.

GMRES

Remark

The GMRES stops at most after n iterations, giving the exat solution.

Remark

GMRES solves at every step a minimum squares problems, which requires many computations.
$\Rightarrow$ GMRES useful if convergence is reached in a small number of steps.

Calderon preconditioning

- Let A_h^{BEM} be the stiffness matrix obtained with the Galerkin approximation.
- A_h^{BEM} is a symmetric, positive definite matrix, hence it holds $\kappa_2(A_h^{BEM}) = \frac{\lambda_{max}(A_h^{BEM})}{\lambda_{min}(A_h^{BEM})}$

Lemma

It holds:

- $\lambda_{max}(A_h^{BEM}) \leq Ch^2$
- $\lambda_{min}(A_h^{BEM}) \geq C'h^3$

Calderon preconditioning

It follows that:

$$\kappa_2(A_h^{BEM}) \leq \tilde{C}\frac{1}{h}$$

Note that if we halve the mesh size we get

$$\kappa_2(A_{h/2}^{BEM}) \leq 2\tilde{C}\frac{1}{h}$$

$\Rightarrow \kappa_2(A_{h/2}^{BEM}) \approx 2\kappa_2(A_h^{BEM})$

Calderon preconditioning

Problem:

- we want a small mesh h
- we want a small conditioning number for A_h^{BEM}
- mesh decreases $\Rightarrow$ conditioning number increases

$\Rightarrow$ We need a preconditioning matrix!

Calderon preconditioning

Recall the Calderon projection

$$\begin{pmatrix} \gamma_D u \\ \gamma_N u \end{pmatrix} = \begin{pmatrix} \frac{1}{2} I - K_0 & V_0 \\ W_0 & \frac{1}{2} I + K_0' \end{pmatrix} \begin{pmatrix} \gamma_D u \\ \gamma_N u \end{pmatrix}$$

where the Calderon projector

$$C = \begin{pmatrix} \frac{1}{2} I - K_0 & V_0 \\ W_0 & \frac{1}{2} I + K_0' \end{pmatrix}$$

has the property $C = C^2$.

Calderon preconditioning

We get

$$V_0 W_0 = (\frac{1}{2}I + K_0)(\frac{1}{2}I - K_0) = \frac{1}{4}I - K_0^2$$
$$W_0 V_0 = (\frac{1}{2}I + K_0')(\frac{1}{2}I - K_0') = \frac{1}{4}I - K_0'^2$$

- We know that $\kappa_2(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$
- We know that K_0 , K_0' are compact operators
- The eigenvalues of a compact operator are finite or they are a sequence converging to zero
- Adding the identity to a compact operator, we can avoid that $\lambda_{min} = 0$
- This way the conditioning number can be controlled