Krylov Subspace Methods
– Linear Systems –

F. M. Faulstich

11/19/2024

Krylov Subspace Methods

- Depending on the problem different names arise

	$Ax = b$	$Ax = \lambda x$
$A = A^*$	CG	Lanczos
$A \neq A^*$	GMRES CGN BCG et al.	Arnoldi

Today: Conjugate Gradient CG

Quadratic Test function

- Consider the *quadratic test function*:
  Let $0 \prec \mathbf{A} \in \mathbb{H}_n(\mathbb{R})$ and $\mathbf{b}, \mathbf{x} \in \mathbb{R}^n$

$$\phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{b}$$

- The gradient of ϕ is given by

$$\nabla\phi(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$$

  Hence, at the critical point $\mathbf{x}_*$ we have

$$\nabla\phi(\mathbf{x}_*) = 0 \quad \Leftrightarrow \quad \mathbf{A}\mathbf{x}_* = \mathbf{b}$$

Quadratic Test function

- Consider the *quadratic test function*:
  Let $0 \prec \mathbf{A} \in \mathbb{H}_n(\mathbb{R})$ and $\mathbf{b}, \mathbf{x} \in \mathbb{R}^n$

$$\phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{b}$$

- The gradient of ϕ is given by

$$\nabla\phi(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$$

  Hence, at the critical point $\mathbf{x}_*$ we have

$$\nabla\phi(\mathbf{x}_*) = 0 \quad \Leftrightarrow \quad \mathbf{A}\mathbf{x}_* = \mathbf{b}$$

- Is this critical point unique?

Quadratic Test function

- Consider the *quadratic test function*:
  Let $0 \prec \mathbf{A} \in \mathbb{H}_n(\mathbb{R})$ and $\mathbf{b}, \mathbf{x} \in \mathbb{R}^n$

$$\phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{b}$$

- The gradient of ϕ is given by

$$\nabla\phi(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$$

  Hence, at the critical point $\mathbf{x}_*$ we have

$$\nabla\phi(\mathbf{x}_*) = 0 \quad \Leftrightarrow \quad \mathbf{A}\mathbf{x}_* = \mathbf{b}$$

- Is this critical point unique? – Yes!
  i) Note that

$$\nabla^2\phi(\mathbf{x}) = \mathbf{A} \succ \mathbf{0}$$

  $\Rightarrow \mathbf{x}_*$ is a minimum
  ii) $\nabla^2\phi(\mathbf{x})$ is constant! $\Rightarrow \phi$ is convex.

Line Search Methods

- *Line search methods* are iterative optimization method
- Idea:
  Start with an initial guess $\mathbf{x}_0$
  Update as
  $$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$$
  where $\mathbf{p}_k$ is the *search direction* and α_k is the *step length*

Steepest Descent

- Remember

$$\nabla\phi(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$$

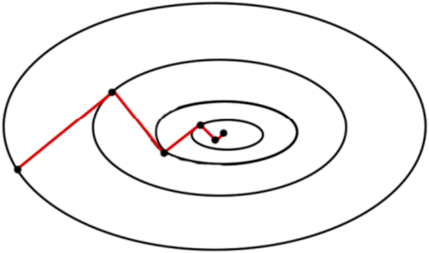
  points towards largest increase of ϕ in $\mathbf{x}$.
  $\Rightarrow$ Search direction should be $\mathbf{p}_k = -\nabla\phi(\mathbf{x}_k) = \mathbf{r}(\mathbf{x}_k)$
- What about the step length?

Steepest Descent

- Remember
$$\nabla\phi(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$$
points towards largest increase of ϕ in $\mathbf{x}$.
$\Rightarrow$ Search direction should be $\mathbf{p}_k = -\nabla\phi(\mathbf{x}_k) = \mathbf{r}(\mathbf{x}_k)$
- What about the step length?
Idea: Walk until we no longer descend!

$$0 \stackrel{!}{=} \partial_{\alpha_k}\phi(\mathbf{x}_{k+1}) \quad \Rightarrow \quad \alpha_k = \frac{\mathbf{r}_k^\top \mathbf{r}_k}{\mathbf{r}_k^\top \mathbf{A}\mathbf{r}_k}$$

Convergence



⇒ "ZigZag" convergence cannot be optimal!

Question: Can we use information from the previous iterations?

$\mathbf{A}$-conjugate direction

- We say a set of vectors $\{\mathbf{p}_1, ..., \mathbf{p}_k\}$ are conjugate w.r.t. the SPD matrix $\mathbf{A}$ iff
$$\mathbf{p}_i^\top \mathbf{A} \mathbf{p}_j = 0 \quad \forall i \neq j$$
- Claim: n $\mathbf{A}$-conjugate vectors form a basis of $\mathbb{R}^n$.
- Then
$$\mathbf{x}_* = \sum_{i=1}^{n} c_i \mathbf{p}_i \quad \Rightarrow \quad \mathbf{A}\mathbf{x}_* = \sum_{i=1}^{n} c_i \mathbf{A} \mathbf{p}_i$$
hence
$$\mathbf{p}_k^\top \mathbf{b} = \sum_{i=0}^{n-1} c_i \mathbf{p}_k^\top \mathbf{A} \mathbf{p}_i = c_k \mathbf{p}_k^\top \mathbf{A} \mathbf{p}_k \quad \Rightarrow \quad c_k = \frac{\mathbf{p}_k^\top \mathbf{b}}{\mathbf{p}_k^\top \mathbf{A} \mathbf{p}_k}$$
- If we have sequence of $\mathbf{A}$-conjugate vecorts we can solve for $\mathbf{x}_*$

A-conjugate vectors

How do we find the set of **A**-conjugate vectors?

$\mathbf{A}$-conjugate vectors

Zeroth Iteration:

- We start with $\mathbf{x}_0 = \mathbf{0} \in \mathbb{R}^n$
- Compute the residual

$$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$$

- Compute the search direction

$$\mathbf{p}_0 = -\nabla\phi(\mathbf{x}_0) = \mathbf{r}_0$$

- Compute the step length

$$\alpha_0 = \frac{\mathbf{p}_0^\top \mathbf{r}_0}{\mathbf{p}_0^\top \mathbf{A}\mathbf{p}_0}$$

- Update the iterate

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{p}_0$$

A-conjugate vectors

k*th* iteration:

- Compute the residual

$$\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k = -\nabla\phi(\mathbf{x}_k)$$

- Make the gradient conjugate to the previous $\{\mathbf{p}_0, ..., \mathbf{p}_{k-1}\}$

$$\mathbf{p}_k = \mathbf{r}_k - \sum_{i=0}^{k-1} \frac{\mathbf{p}_i^\top \mathbf{A}\mathbf{r}_k}{\mathbf{p}_i^\top \mathbf{A}\mathbf{p}_i}\mathbf{p}_i$$

- Compute the step length

$$\alpha_k = \frac{\mathbf{p}_k^\top \mathbf{r}_k}{\mathbf{p}_k^\top \mathbf{A}\mathbf{p}_k}$$

- Update the iterate

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

Computation of the search direction

- Let's take a closer look at the search direction

$$\mathbf{p}_k = \mathbf{r}_k - \sum_{i=0}^{k-1} \frac{\mathbf{p}_i^\top \mathbf{A} \mathbf{r}_k}{\mathbf{p}_i^\top \mathbf{A} \mathbf{p}_i} \mathbf{p}_i$$

- Better to impose conjugation explicitly:

$$\mathbf{p}_k = \mathbf{r}_k - \beta_k \mathbf{p}_{k-1}$$

Then $\mathbf{p}_{k-1}^\top \mathbf{A} \mathbf{p}_k = 0$ implies

$$\beta_k = \frac{\mathbf{p}_{k-1}^\top \mathbf{A} \mathbf{r}_k}{\mathbf{p}_{k-1}^\top \mathbf{A} \mathbf{p}_{k-1}}$$

Note that

$$\mathbf{r}_k^\top \mathbf{A} \mathbf{p}_{k-1} = -\frac{1}{\alpha_{k-1}} \mathbf{r}_k^\top \mathbf{r}_k \quad \text{and} \quad \mathbf{p}_k^\top \mathbf{A} \mathbf{p}_k = \frac{1}{\alpha_k} \mathbf{r}_k^\top \mathbf{r}_k$$

Hence

$$\beta_k = -\frac{\mathbf{r}_k^\top \mathbf{r}_k}{\mathbf{r}_{k-1}^\top \mathbf{r}_{k-1}}$$

Conjugate Gradient Algorithm

Input: $\mathbf{A}$
Output: $\mathbf{x}$ approximate solution to $\mathbf{A}\mathbf{x}_* = \mathbf{b}$

$\mathbf{x}_0 = \mathbf{0}$

$\mathbf{p}_0 = \mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

For k=1...

$\alpha_{k-1} = \frac{\mathbf{p}_{k-1}^\top \mathbf{r}_{k-1}}{\mathbf{p}_{k-1}^\top \mathbf{A}\mathbf{p}_{k-1}}$

$\mathbf{x}_k = \mathbf{x}_{k-1} + \alpha_{k-1}\mathbf{p}_{k-1}$

$\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$

$\beta_k = -\frac{\mathbf{r}_k^\top \mathbf{r}_k}{\mathbf{r}_{k-1}^\top \mathbf{r}_{k-1}}$

$\mathbf{p}_k = \mathbf{r}_k - \beta_k \mathbf{p}_{k-1}$

Why is this a Krylov subspace method?

- Claim: $\mathbf{x}_k \in \mathcal{K}_k = \text{span}(\mathbf{b}, \mathbf{Ab}, ..., \mathbf{A}^{k-1}\mathbf{b})$
- Claim: $\mathbf{r}_k \perp \mathcal{K}_k$
- Theorem:
  $\mathbf{x}_k \in \mathcal{K}_k$ is the unique point that minimizes $\|\mathbf{e}_k\|_A$ with $\mathbf{e}_k = \mathbf{x}_* - \mathbf{x}_k$ and $\|\mathbf{e}_k\|_\mathbf{A} \leq \|\mathbf{e}_{k-1}\|_\mathbf{A}$ and $\mathbf{e}_\ell = 0$ for some $\ell \geq n$.