Mathematics Methods for Computer Science

Instructor: Xubo Yang

SJTU-SE DALAB

Lecture

Conjugate Gradients II: CG and Variants

Gradient Descent: Issue

$$\vec{x}_k - \vec{x}_0 \neq \arg\min_{\vec{v} \in \operatorname{span}\left\{\vec{r}_0, A\vec{r}_0, \ldots, A^{k-1}\vec{r}_0\right\}} f\left(\vec{x}_0 + \vec{v}\right)$$

But if this did hold ...

Convergence in n steps!

Outline for CG

1. (Somehow) generate search direction $\vec{v}_k$ (initialize to $\vec{r}_0$)
2. Line search: $\alpha_k = \frac{\vec{v}_k^\top \vec{r}_{k-1}}{\vec{v}_k^\top A \vec{v}_k}$
3. Update estimate: $\vec{x}_k = \vec{x}_{k-1} + \alpha_k \vec{v}_k$
4. Update residual: $\vec{r}_k = \vec{r}_{k-1} + \alpha_k A \vec{v}_k$

What We (Greedily) Want

1. Easy way to generate n conjugate directions $\{\vec{v}_1, \ldots, \vec{v}_n\}$
2. $\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_n\} = \operatorname{span}\left\{\vec{r}_0, A\vec{r}_0, \ldots, A^{k-1}\vec{r}_0,\right\}$ for all k

What We (Greedily) Want

1. Easy way to generate n conjugate directions $\{\vec{v}_1, \ldots, \vec{v}_n\}$
2. $\text{span}\,\{\vec{v}_1, \ldots, \vec{v}_n\} = \text{span}\,\{\vec{r}_0, A\vec{r}_0, \ldots, A^{k-1}\vec{r}_0,\}$ for all k

---

1. Converge in n steps
2. Always better than gradient descent

Method of Conjugate Directions

$$\vec{v}_k \leftarrow A^{k-1}\vec{r}_0 - \sum_{i<k} \frac{\left\langle A^{k-1}\vec{r}_0, \vec{v}_i \right\rangle_A}{\left\langle \vec{v}_i, \vec{v}_i \right\rangle_A} \vec{v}_i$$

$(A-)$ Gram-Schmidt on $\vec{r}_0, \ldots, A^{k-1}\vec{r}_0$

Problems with Method of Conjugate Directions

- $A^{k-1}\vec{r}_0$ looks like largest eigenvector of A, so projection is poorly conditioned

- Have to store $\vec{v}_1, \ldots, \vec{v}_{k-1}$

Small Adjustment

Only need a search direction

$$\vec{w} \in \operatorname{span}\left\{\vec{r}_0, \ldots, A^{k-1}\vec{r}_0,\right\} \setminus \operatorname{span}\left\{\vec{r}_0, \ldots, A^{k-2}\vec{r}_0,\right\}$$

$$\vec{r}_k = \vec{r}_{k-1} - \alpha_k A \vec{v}_k$$

Small Adjustment

Only need a search direction

$$\vec{w} \in \operatorname{span}\left\{\vec{r}_0, \ldots, A^{k-1}\vec{r}_0,\right\} \setminus \operatorname{span}\left\{\vec{r}_0, \ldots, A^{k-2}\vec{r}_0,\right\}$$

$\vec{r}_k = \vec{r}_{k-1} - \alpha_k A\vec{v}_k \quad \rightarrow \quad$ Take $\vec{v}_k = \vec{r}_{k-1}$

Small Adjustment

Only need a search direction

$$\vec{w} \in \operatorname{span}\left\{\vec{r}_0, \ldots, A^{k-1}\vec{r}_0,\right\} \setminus \operatorname{span}\left\{\vec{r}_0, \ldots, A^{k-2}\vec{r}_0,\right\}$$

$$\vec{r}_k = \vec{r}_{k-1} - \alpha_k A\vec{v}_k \quad \rightarrow \quad \text{Take } \vec{v}_k = \vec{r}_{k-1}$$

$$\vec{v}_k \leftarrow \vec{r}^{k-1} - \sum_{i<k} \frac{\langle \vec{r}_{k-1}, \vec{v}_i \rangle_A}{\langle \vec{v}_i, \vec{v}_i \rangle_A} \vec{v}_i$$

Small Adjustment

Only need a search direction

$$\vec{w} \in \operatorname{span}\left\{\vec{r}_0, \ldots, A^{k-1}\vec{r}_0,\right\} \setminus \operatorname{span}\left\{\vec{r}_0, \ldots, A^{k-2}\vec{r}_0,\right\}$$

$$\vec{r}_k = \vec{r}_{k-1} - \alpha_k A\vec{v}_k \quad \rightarrow \quad \text{Take } \vec{v}_k = \vec{r}_{k-1}$$

$$\vec{v}_k \leftarrow \vec{r}^{k-1} - \sum_{i<k} \frac{\left\langle \vec{r}_{k-1}, \vec{v}_i \right\rangle_A}{\left\langle \vec{v}_i, \vec{v}_i \right\rangle_A} \vec{v}_i$$

Fixed conditioning, but still have to store $\vec{v}_i$'s

Remarkable Observation

Proposition

$\langle \vec{r}_k, \vec{v}_l \rangle_A = 0$ for all $l < k$.

Warning: Technical.

Remarkable Observation

Proposition

$\langle \vec{r}_k, \vec{v}_l \rangle_A = 0$ for all $l < k$.

Warning: Technical.

$$\boxed{\vec{v}_k \leftarrow \vec{r}_{k-1} - \frac{\langle \vec{r}_{k-1}, \vec{v}_{k-1} \rangle_A}{\langle \vec{v}_{k-1}, \vec{v}_{k-1} \rangle_A} \vec{v}_{k-1}}$$

No memory!

即每次只需要计算当前k对应的值就行了，而不需要重复计算前面n-1个值！

Conjugate Gradient Algorithm

1. Update search direction:
$$\vec{v}_k \leftarrow \vec{r}_{k-1} - \frac{\langle \vec{r}_{k-1}, \vec{v}_{k-1} \rangle_A}{\langle \vec{v}_{k-1}, \vec{v}_{k-1} \rangle_A} \vec{v}_{k-1}$$
2. Line search: $\alpha_k = \frac{\vec{v}_k^\top \vec{r}_{k-1}}{\vec{v}_k^\top A \vec{v}_k}$
3. Update estimate: $\vec{x}_k = \vec{x}_{k-1} + \alpha_k \vec{v}_k$
4. Update residual: $\vec{r}_k = \vec{r}_{k-1} - \alpha_k A \vec{v}_k$

Properties

- $\vec{x}_k$ optimal in subspace spanned by first k directions $\vec{v}_i$

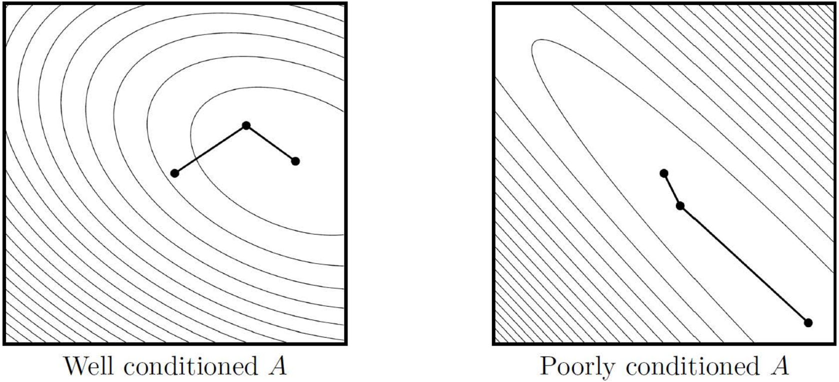
Properties

- $\vec{x}_k$ optimal in subspace spanned by first k directions $\vec{v}_i$
- $f(\vec{x}_k)$ upper-bounded by f of k-th iterate of gradient descent

Properties

- $\vec{x}_k$ optimal in subspace spanned by first k directions $\vec{v}_i$
- $f(\vec{x}_k)$ upper-bounded by f of k-th iterate of gradient descent
- Converges in n steps

Visualization



Nicer Form

See book.

- Update search direction: $\beta_k = \frac{\vec{r}_{k-1}, \vec{r}_{k-1}}{\vec{r}_{k-2}, \vec{r}_{k-2}}$
  $\vec{v}_k = \vec{r}_{k-1} + \beta_k \vec{v}_{k-1}$
- Line search: $\alpha_k = \frac{\vec{r}_{k-1}^\top \vec{r}_{k-1}}{\vec{v}_k^\top A \vec{v}_k}$
- Update estimate: $\vec{x}_k = \vec{x}_{k-1} + \alpha_k \vec{v}_k$
- Update residual: $\vec{r}_k = \vec{r}_{k-1} - \alpha_k A \vec{v}_k$

Typical Stopping Conditions

Don't run to completion!

$$\frac{\|\vec{r}_k\|_2}{\|\vec{r}_0\|_2} < \epsilon$$

Convergence

$$\frac{f(\vec{x}_k)-f(\vec{x}^*)}{f(\vec{x}_0)-f(\vec{x}^*)} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k$$

Convergence

$$\frac{f(\vec{x}_k) - f(\vec{x}^*)}{f(\vec{x}_0) - f(\vec{x}^*)} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k$$

Still depends on condition number!

Observation

$$\operatorname{cond} A \neq \operatorname{cond} PA$$

But we can solve

$$PA\vec{x} = P\vec{b}$$

Preconditioning

Solve $PA\vec{x} = P\vec{b}$ for $P \approx A^{-1}$

Two Issues

- PA may not be symmetric or positive definite
- Need to find an "easy" P

Symmetrization

P symmetric and positive definite

$$\Rightarrow P^{-1} = EE^\top$$

Symmetrization

$$P \text{ symmetric and positive definite}$$
$$\Rightarrow P^{-1} = EE^{\top}$$

Proposition

PA and $E^{-1}AE^{-T}$ have the same eigenvalues

Symmetrization

$$P \text{ symmetric and positive definite}$$
$$\Rightarrow P^{-1} = EE^{\top}$$

Proposition

PA and $E^{-1}AE^{-T}$ have the same eigenvalues

1. Solve $E^{-1}AE^{-T}\vec{y} = E^{-1}\vec{b}$
2. Solve $\vec{x} = E^{-1}\vec{y}$

Better-Conditioned CG

$$
\begin{aligned}
\text{Update search direction:}\quad \beta_k &= \frac{\vec{r}_{k-1}^{\vec{r}}\vec{r}_{k-1}}{\vec{r}_{k-2}^{-1}\vec{r}_{k-2}} \\
\vec{v}_k &= \vec{r}_{k-1} + \beta_k \vec{v}_{k-1} \\
\text{Line search:}\quad \alpha_k &= \frac{\vec{r}_{k-1}^{\top}\vec{r}_{k-1}}{\vec{v}_k^{T} E^{-1} A E^{-\top} \vec{v}_k} \\
\text{Update residual:}\quad \vec{r}_k &= \vec{r}_{k-1} - \alpha_k E^{-1} A E^{-\top} \vec{v}_k
\end{aligned}
$$

Preconditioned CG

$$
\begin{aligned}
\text{Update search direction:} \quad & \beta_k = \frac{\tilde{r}_{k-1}^\top P \tilde{r}_{k-1}}{\tilde{r}_{k-2}^{-1} P \tilde{r}_{k-2}} \\
& \tilde{v}_k = P\tilde{r}_{k-1} + \beta_k \tilde{v}_{k-1} \\
\text{Line search:} \quad & \alpha_k = \frac{\vec{r}_{k-1}^\top P \vec{r}_{k-1}}{\tilde{v}_k^\top A \tilde{v}_k} \\
\text{Update estimate:} \quad & \vec{x}_k = \vec{x}_{k-1} + \alpha_k \tilde{v}_k \\
\text{Update residual:} \quad & \tilde{r}_k = \tilde{r}_{k-1} - \alpha_k A \tilde{v}_k
\end{aligned}
$$

Common Preconditioners

- Diagonal ("Jacobi") $A \approx D$
- Sparse approximate inverse
- Incomplete Cholesky $A \approx L_* L_*^\top$
- Domain decomposition

Other Iterative Schemes I

- Splitting: $A = M - N \Rightarrow M\vec{x} = N\vec{x} + \vec{b}$
- Conjugate gradient normal equation residual (CGNR): $A^\top A\vec{x} = A^\top \vec{b}$
- Conjugate gradient normal equation error (CGNE): $AA^\top \vec{y} = \vec{b}$; $\vec{x} = A^\top \vec{y}$
- MINRES, SYMLQ: $g(\vec{x}) \equiv ||\vec{b} - A\vec{x}||_2^2$ for symmetric A
- LSQR, LSMR: Normal equations, same g

Other Iterative Scheme II

- GMRES, QMR, BiCG, CGS, BiCGStab: Any invertible A
- Fletcher-Reeves, Polak-Ribière: Nonlinear problems; replace residual with ∇f and add back line search