

Conjugate Gradients II: CG and Variants

CS 205A:
Mathematical Methods for Robotics, Vision, and Graphics

Doug James (and Justin Solomon)

Idea

“Easy to apply,
hard to invert.”

$$\min_{\vec{x}} \left[\frac{1}{2} \vec{x}^\top A \vec{x} - \vec{b}^\top \vec{x} + c \right]$$

Line Search Along \vec{d} from \vec{x}

$$\min_{\alpha} g(\alpha) \equiv f(\vec{x} + \alpha \vec{d})$$

$$\alpha = \frac{\vec{d}^{\top} (\vec{b} - A\vec{x})}{\vec{d}^{\top} A\vec{d}} = \frac{\vec{d}^{\top} \vec{r}}{\vec{d}^{\top} A\vec{d}}$$

Conjugate Directions

A -Conjugate Vectors

Two vectors \vec{v} and \vec{w} are A -conjugate if $\vec{v}^\top A \vec{w} = 0$.

Corollary

Suppose $\{\vec{v}_1, \dots, \vec{v}_n\}$ are A -conjugate. Then, f is minimized in at most n steps by line searching in direction \vec{v}_1 , then direction \vec{v}_2 , and so on.

Another Clue

$$\vec{r}_k \equiv \vec{b} - A\vec{x}_k$$

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

Proposition

When performing gradient descent on f ,
 $\text{span} \{ \vec{r}_0, \dots, \vec{r}_k \} = \text{span} \{ \vec{r}_0, A\vec{r}_0, \dots, A^k\vec{r}_0 \}.$

Gradient Descent: Issue

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

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 (Greedyly) Want

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

What We (Greedy) Want

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

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

Method of Conjugate Directions

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

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

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 $\vec{v}_1, \dots, \vec{v}_{k-1}$

Small Adjustment

Only need a search direction

$$\vec{w} \in \text{span} \{ \vec{r}_0, \dots, A^{k-1} \vec{r}_0 \} \setminus \text{span} \{ \vec{r}_0, \dots, A^{k-2} \vec{r}_0 \}.$$

Small Adjustment

Only need a search direction

$$\vec{w} \in \text{span} \{ \vec{r}_0, \dots, A^{k-1} \vec{r}_0 \} \setminus \text{span} \{ \vec{r}_0, \dots, A^{k-2} \vec{r}_0 \}.$$

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

Small Adjustment

Only need a search direction

$$\vec{w} \in \text{span} \{ \vec{r}_0, \dots, A^{k-1} \vec{r}_0 \} \setminus \text{span} \{ \vec{r}_0, \dots, A^{k-2} \vec{r}_0 \}.$$

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

Small Adjustment

Only need a search direction

$$\vec{w} \in \text{span} \{ \vec{r}_0, \dots, A^{k-1} \vec{r}_0 \} \setminus \text{span} \{ \vec{r}_0, \dots, A^{k-2} \vec{r}_0 \}.$$

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

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

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

Remarkable Observation

Proposition

$$\langle \vec{r}_k, \vec{v}_\ell \rangle_A = 0 \text{ for all } \ell < k.$$

Warning: Technical.

Remarkable Observation

Proposition

$$\langle \vec{r}_k, \vec{v}_\ell \rangle_A = 0 \text{ for all } \ell < k.$$

Warning: Technical.

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

No memory!

Conjugate Gradients Algorithm

1. Update search direction:

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

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

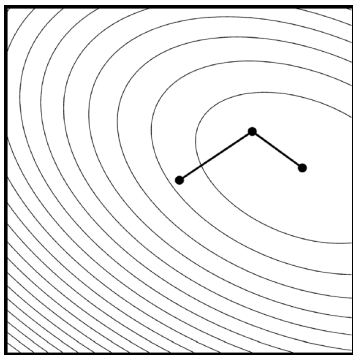
Properties

- ▶ \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

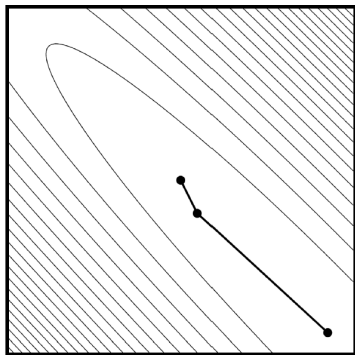
Properties

- ▶ \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



Well conditioned A



Poorly conditioned A

Nicer Form

See book.

- ▶ Update search direction: $\beta_k = \frac{\vec{r}_{k-1}^\top \vec{r}_{k-1}}{\vec{r}_{k-2}^\top \vec{r}_{k-2}}$

$$\vec{v}_k = \vec{r}_{k-1} + \beta_k \vec{v}_{k-1}$$

- ▶ 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!

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

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

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

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
 $\implies P^{-1} = EE^{\top}.$

Symmetrization

P symmetric and positive definite
 $\implies P^{-1} = EE^{\top}.$

Proposition

PA and $E^{-1}AE^{-\top}$ have the same eigenvalues.

Symmetrization

P symmetric and positive definite
 $\implies P^{-1} = EE^{\top}$.

Proposition

PA and $E^{-1}AE^{-\top}$ have the same eigenvalues.

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

Better-Conditioned CG

Update search direction: $\beta_k = \frac{\vec{r}_{k-1}^\top \vec{r}_{k-1}}{\vec{r}_{k-2}^\top \vec{r}_{k-2}}$

$$\vec{v}_k = \vec{r}_{k-1} + \beta_k \vec{v}_{k-1}$$

Line search: $\alpha_k = \frac{\vec{r}_{k-1}^\top \vec{r}_{k-1}}{\vec{v}_k^\top E^{-1} A E^{-\top} \vec{v}_k}$

Update estimate: $\vec{y}_k = \vec{y}_{k-1} + \alpha_k \vec{v}_k$

Update residual: $\vec{r}_k = \vec{r}_{k-1} - \alpha_k E^{-1} A E^{-\top} \vec{v}_k$

Preconditioned CG

Update search direction: $\beta_k = \frac{\tilde{r}_{k-1}^\top P \tilde{r}_{k-1}}{\tilde{r}_{k-2}^\top P \tilde{r}_{k-2}}$

$$\tilde{v}_k = P \tilde{r}_{k-1} + \beta_k \tilde{v}_{k-1}$$

Line search: $\alpha_k = \frac{\vec{r}_{k-1}^\top P \vec{r}_{k-1}}{\tilde{v}_k^\top A \tilde{v}_k}$

Update estimate: $\vec{x}_k = \vec{x}_{k-1} + \alpha_k \tilde{v}_k$

Update residual: $\tilde{r}_k = \tilde{r}_{k-1} - \alpha_k A \tilde{v}_k$

Common Preconditioners

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

Other Iterative Schemes I

- ▶ Splitting: $A = M - N \implies M\vec{x} = N\vec{x} + \vec{b}$
- ▶ 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 Schemes II

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

▶ Next