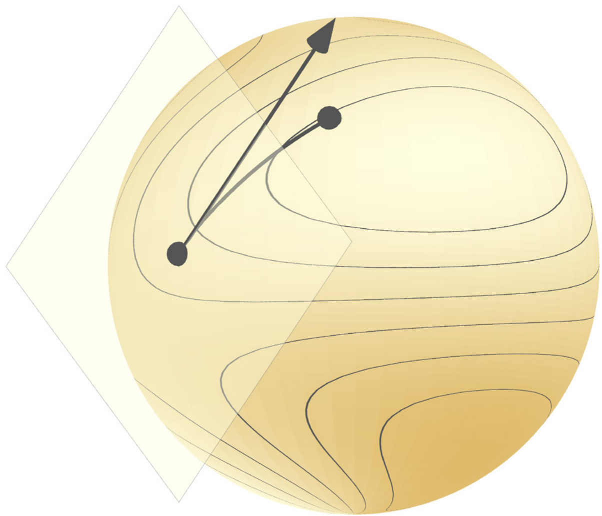
209

More about conjugate gradients

Spring 2023

Optimization on manifolds, MATH 512 @ EPFL

Instructor: Nicolas Boumal



Why do people really care about CG?

It is not because of finite termination:
that fails numerically, and it is irrelevant in high dimension.

It is because, in practice, it converges much faster than GD,
for essentially the same cost per iteration.

We understand very well *why*, and the proof is beautiful.

Conjugate gradients to minimize $g(v) = \frac{1}{2}\langle v, Hv\rangle_x - \langle b, v\rangle_x$ on $\mathrm{T}_x\mathcal{M}$:

Initialize $v_0 = 0, r_0 = b, p_0 = r_0$

For n in 1, 2, 3, …

- Compute Hp_{n-1}
- $\alpha_n = \frac{\|r_{n-1}\|_x^2}{\langle p_{n-1}, Hp_{n-1}\rangle_x}$
- $v_n = v_{n-1} + \alpha_n p_{n-1}$
- $r_n = r_{n-1} - \alpha_n H p_{n-1}$
- **If** $\|r_n\|_x \leq \mathrm{tol} \cdot \|b\|_x$, **output** v_n
- $\beta_n = \frac{\|r_n\|_x^2}{\|r_{n-1}\|_x^2}$
- $p_n = r_n + \beta_n p_{n-1}$

$$v_n = v_{n-1} + \alpha_n p_{n-1}$$
$$r_n = r_{n-1} - \alpha_n H p_{n-1}$$
$$p_n = r_n + \beta_n p_{n-1}$$

The three sequences of CG

The iterates $v_0, v_1, v_2, \ldots$ converge to the minimizer of g.

The residues $r_0, r_1, r_2, \ldots$ converge to zero.

The basis vectors $p_0, p_1, p_2, \ldots$ are H-conjugate directions.

$$v_n = v_{n-1} + \alpha_n p_{n-1}$$
$$r_n = r_{n-1} - \alpha_n H p_{n-1}$$
$$p_n = r_n + \beta_n p_{n-1}$$

The Krylov space $\mathcal{K}_n$

Fact: $\mathcal{K}_n = \mathrm{span}(p_0, \dots, p_{n-1}) = \mathrm{span}(b, Hb, H^2 b, \dots, H^{n-1} b)$

Proof. It's clear for $n = 1$. Now proceed by induction.

Note: $\dim \mathrm{span}(b, \dots, H^n b) \leq n + 1 = \dim \mathrm{span}(p_0, \dots, p_n)$.

By induction, we know $\mathrm{span}(p_0, \dots, p_{n-1}) \subseteq \mathrm{span}(b, \dots, H^n b)$.

Exercise: show p_n is in $\mathrm{span}(b, \dots, H^n b)$ and conclude.

Detour: the H-norm property of v_n

We already know v_n minimizes $g(v) = \frac{1}{2}\langle v, Hv\rangle_x - \langle b, v\rangle_x$ over $\mathcal{K}_n$.

Fact: v_n also minimizes the H-norm distance to the global min.

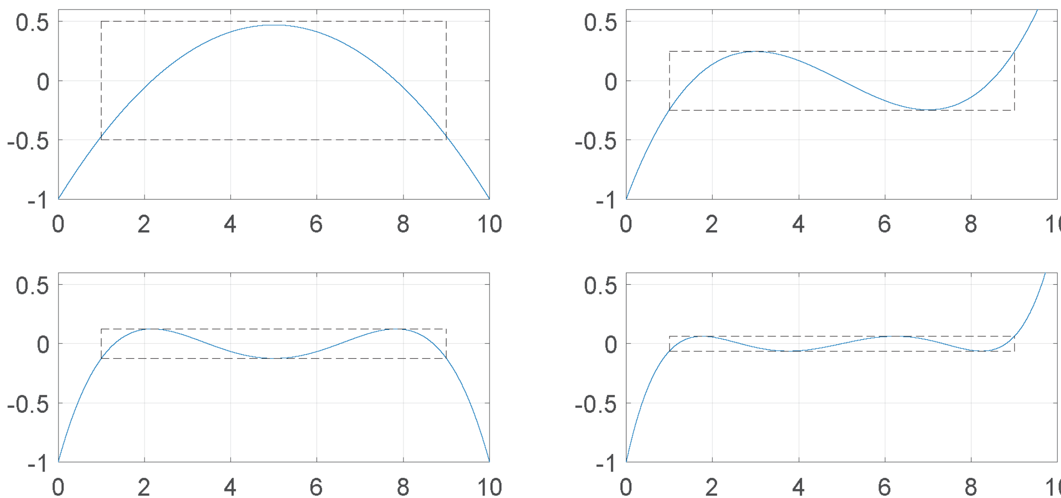
$$v_n = v_{n-1} + \alpha_n p_{n-1}$$
$$r_n = r_{n-1} - \alpha_n H p_{n-1}$$
$$p_n = r_n + \beta_n p_{n-1}$$

The polynomial perspective

$$v_n = \operatorname*{argmin}_{v \in \mathcal{K}_n} g(v) \quad \text{with} \quad \mathcal{K}_n = \operatorname{span}(b, Hb, H^2 b, \dots, H^{n-1} b)$$

Theorem. If the eigenvalues of H are in $[\lambda_{\min}, \lambda_{\max}]$ with $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$, then:

$$\|v_n - s\|_H \leq \|s\|_H \cdot 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n \leq \|s\|_H \cdot 2e^{-n/\sqrt{\kappa}}.$$



For each degree n, one can find a polynomial in Q_n with maximal absolute value less than $2\left(\frac{\sqrt{9}-1}{\sqrt{9}+1}\right)^n$ over the interval $[1,9]$: see Fig. 6.1 in book for details.