Conjugate gradient

$Q \in \mathbb{R}^{m \times m}, \quad v, x \in \mathbb{R}^m$

$Q \succ 0$

$$f(x) = \frac{1}{2} x^T Q x - v^T x + \text{const}$$

$$x_* = \arg\min f(x) \iff Qx = v$$

$$\frac{1}{2}(x - x_*)^T Q (x - x_*) = \frac{1}{2} x^T Q x - \underbrace{\frac{1}{2} \underbrace{x_*^T Q}_{v^T} x - \frac{1}{2} x^T \underbrace{Q x_*}_{v}} + \text{const}$$

$$= \frac{1}{2} \|x - x_*\|_Q^2$$

$$= \frac{1}{2} x^T Q x - v^T x + \text{const} = f(x) + \text{const}$$

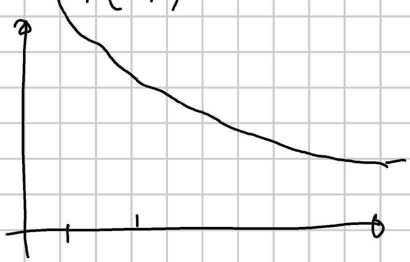
Def: given $w \in \mathbb{R}^m$, we call the Q-norm of w

$$\|w\|_Q = \sqrt{w^T Q w}$$

This is also a norm, provided Q is pos. def.

With CG, the objective function decreases monotonically

$f(x_k)$

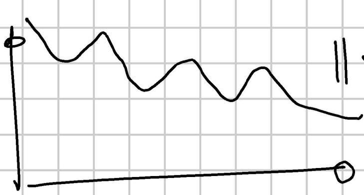


k

decreases monotonically

k

$\|x_k - x_*\|_Q$ decreases monotonically



$\|x_k - x_*\|$ might increase from one step to the other

Conjugate gradient

$r_k = v - Qx_k = -\text{gradient}$

$x_0 = 0 \quad r_0 = v \quad d_0 = v$

for $j = 1 \dots n$

$$\alpha_j = (r_{j-1}^T r_{j-1}) / (d_{j-1}^T \boxed{Q d_{j-1}})$$

$$x_j = x_{j-1} + \alpha_j d_{j-1}$$

$$r_j = r_{j-1} - \alpha_j \boxed{Q d_{j-1}}$$

$$\beta_j = (r_j^T r_j) / (r_{j-1}^T r_{j-1})$$

$$d_j = r_j + \beta_j d_{j-1}$$

end

Cost per step: 1 matvec $+ O(m)$ operations

Def: The Krylov space $K_n(Q, v)$ is

$$K_n(Q,v) = \text{span}\{v, Qv, Q^2 v, \dots Q^{n-1} v\}$$

$$= \{w : p(Q)v : \quad p(t) \text{ polynomial of degree} < n\}$$

$$= \{vc_0 + Qvc_1 + Q^2 v c_2 + \dots + Q^{n-1} v c_{n-1} : c_0, \dots, c_{n-1} \in \mathbb{R}\}$$

Assume for now that $v, Qv, \dots Q^{n-1} v$ are linearly independent. Then, $c_0, c_1, \dots c_{n-1}$ are uniquely defined.

Given $w \in K_n(Q,v)$, its degree w.r.t. Q, v is uniquely defined:

$5v$ degree 0

$v + Qv$ degree 1

$Q^5 v$ degree 5

Obs: The iterates of CG have well-defined degrees

Assume $\alpha_j \neq 0$ for all j

$\cancel{x_0 = 0}$ $\quad r_0 = v \quad d_0 = v$ degree 0

$x_1 = x_0 + \alpha_1 d_0$ degree 0

$r_1 = r_0 - \alpha_1 Q d_0$ degree 1

$d_1 = r_1 + \beta_1 d_0$ degree 1

$x_2 = x_1 + \alpha_2 d_1$ degree 1

$r_2 = r_1 - \alpha_2 Q d_1$ degree 2

$d_2 = r_2 + \beta_2 d_1$ degree 2

$\vdots \qquad \vdots$

By induction: x_j, d_{j-1}, r_{j-1} have degree $j-1$ and belong to $K_j(Q,v)$

Moreover, since the degrees are exactly $0,1,2,\dots j-1$, we have that:

$\{x_1, x_2, \dots x_j\}$ is a basis of $K_j(Q,v)$

$\{r_0, r_1, \dots r_{j-1}\}$ " " " "

$\{d_0, d_1, \dots d_{j-1}\}$ " " " "

There is more:

Theorem: $r_i^T r_j = 0$ for all $i \neq j$

$d_i^T Q d_j = 0$ for all $i \neq j$

Proof: We assume $i < j$, and work by induction on j

We assume that $r_i^T r_{j-1} = 0$ for $i < j-1$

and $d_i^T Q d_{j-1} = 0$ for $i < j-1$.

→ Let us prove that $r_i^T r_j = 0$ for all $i < j$

(we skip proving $d_i^T Q d_j = 0$ because it is similar)

If $i < j-1$,

$$r_i^T r_j = r_i^T \left(r_{j-1} - \alpha_j Q d_{j-1} \right) = \underbrace{r_i^T r_{j-1}}_{=0} - \alpha_j \left(\underbrace{r_i^T Q d_{j-1}}_{} \right)$$

$r_i^T Q d_{j-1}$ is also zero

Note that $r_i \in K_{i+1}(Q, v)$

so $r_i = \eta_0 d_0 + \eta_1 d_1 + \eta_2 d_2 + \dots + \eta_i d_i$

and $r_i^T Q d_{j-1} = \left(\eta_0 d_0 + \dots + \eta_i d_i \right)^T Q d_{j-1} = 0$

For $i = j-1$

$$r_{j-1}^T r_j = r_{j-1}^T \left(r_{j-1} - \alpha_j Q d_{j-1} \right) = r_{j-1}^T r_{j-1} - \alpha_j r_{j-1}^T Q d_{j-1}$$

This is 0 because of the value of α_j:

$$\alpha_j = \frac{r_{j-1}^T r_{j-1}}{r_{j-1}^T Q d_{j-1}}$$

Recall that α_j is defined as $\alpha_j = \dfrac{r_{j-1}^T r_{j-1}}{d_{j-1}^T Q d_{j-1}}$

But

$d_{j-1} = r_{j-1} + \beta_{j-1} d_{j-2}$ so

$$d_{j-1}^T Q d_{j-1} = r_{j-1}^T Q d_{j-1} + \beta_{j-1} \underbrace{d_{j-2}^T Q d_{j-1}}_{=0}$$

□

We have made 2 assumptions in the past:

1) $v, Qv, Q^2 v, \dots Q^{n-1} v$ are lin. independent

2) $\alpha_j \neq 0$ for all j

If one of these assumptions breaks down, it means you reached a point where $r_j = 0$ ("lucky breakdown")

2) holds because $\alpha_j = \dfrac{r_{j-1}^T r_{j-1}}{\dots} = 0 \quad \|r_{j-1}\| = 0$

1) holds because we can still prove that $r_i^T r_j = 0$ for $i \neq j$, and if

$\underbrace{v, Qv, Q^2 v, \dots, Q^{n-1} v,}\ Q^n v$ are lin. dependent,

$K_n(Q, v) = K_{n+1}(Q, v)$

and $r_n \in K_{n+1}(Q, v)$ and $r_i^T r_n = 0$ for $i < n$,

and $r_0, r_1, r_2, \dots r_{n-1}$ are a basis of $K_n(Q, v) = K_n(Q, v)$

so r_n is orthogonal to the whole $K_n(Q, v)$

in particular $r_n^T r_n = 0$.

CG converges faster than the gradient method:

- The gradient method produces iterates $x_k \in K_k(Q, v)$
- CG also produces iterates $x_k \in K_k(Q, v)$
  $= \text{span}\{d_0, d_1, \dots d_{k-1}\}$
- CG is optimal:

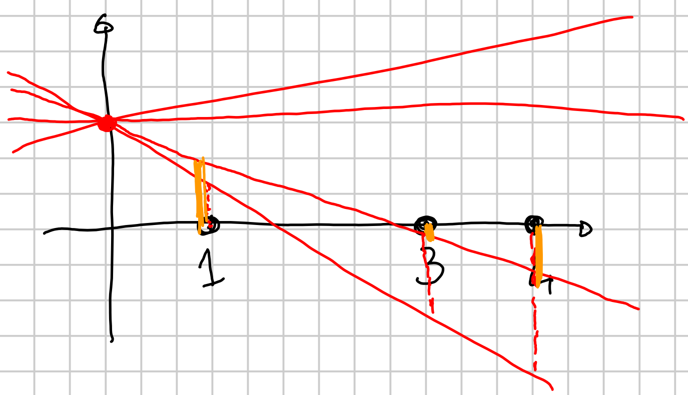
  $x_k = \arg\min f(x)$ over $\text{span}\{d_0, d_1, \dots d_{k-1}\}$

Theorem: for CG,

$$\frac{\|x_n - x_*\|_Q}{\|x_0 - x_*\|_Q} \leq \min_{r(t)} \max_{\substack{\lambda_1, \ldots \lambda_m \\ \text{eigenvalues of } Q}} |r(\lambda_i)|$$

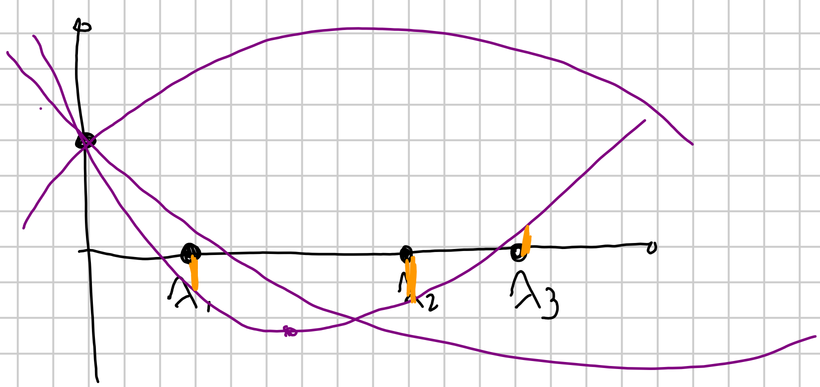
The minimum is over all polynomials $r(t)$ of degree $\leq n$, normalized so that $r(0)=1$.

Ex: $n=1$ Q has eigenvalues $\lambda_1 = 1$, $\lambda_2 = 3$, $\lambda_3 = 4$



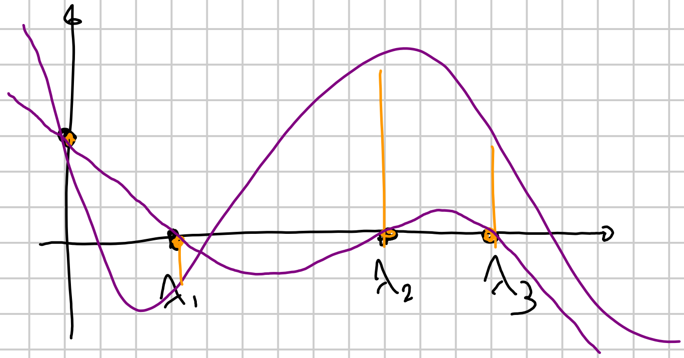
$$\frac{\|x_1 - x_*\|_Q}{\|x_0 - x_*\|_Q} \leq \text{length of the largest of the orange segments}$$

$n=2$



$$\frac{\|x_2 - x_*\|_Q}{\|x_0 - x_*\|_Q} \leq \text{length of largest orange segment}$$

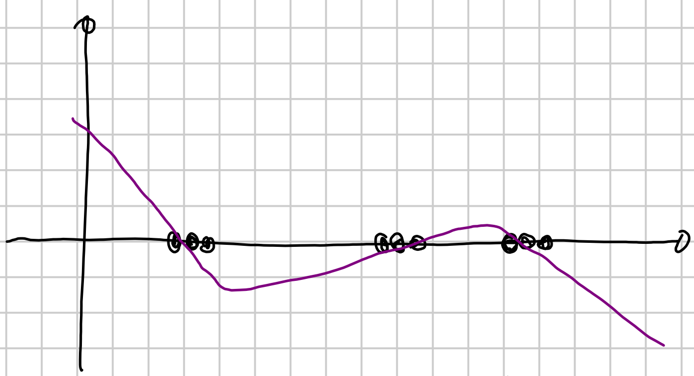
$n=3$



If I have only $n=3$ distinct eigenvalues, then there is a polynomial through these 3 points, and

$$\frac{\|x_3 - x_*\|_Q}{\|x_0 - x_*\|_Q} \leq 0$$

Also applies if $Q \in \mathbb{R}^{m \times n}$ with $m > 3$ but with 3 distinct eigenvalues



If Q has eigenvalues clustered close to n distinct values, then

$$\frac{\|x_n - x_*\|_Q}{\|x_0 - x_*\|_Q} \text{ is small}$$

Worst-case bound:

$$\frac{\|x_n - x_*\|_Q}{\|x_n - x_*\|_Q} \leq \left(\frac{\sqrt{\lambda_{max}} - \sqrt{\lambda_{min}}}{\sqrt{\lambda_{max}} + \sqrt{\lambda_{min}}} \right)^n$$

$$= \left(\frac{\sqrt{K} - 1}{\sqrt{K} + 1} \right)^{n}, \quad \text{where} \quad K = \frac{\lambda_{max}}{\lambda_{min}}$$