LECTURE 34 THE CONJUGATE GRADIENT METHOD

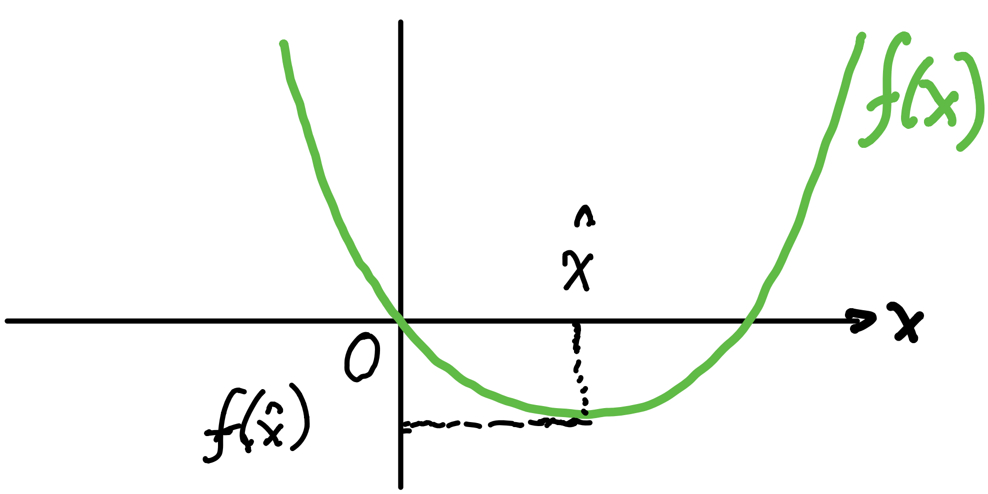
Applies to SPD matrices
(symmetric positive definite)

Assume $A \in \mathbb{R}^{m\times m}$, $x \in \mathbb{R}^m$ $b \in \mathbb{R}^m$.
We wish to find $x = \hat{x}$ such that

$$A\hat{x} = b \quad (A, b \text{ given}).$$

An optimization approach:

let $f(x) \equiv \frac{1}{2} x^T A x - x^T b$



So $\nabla_x f = Ax - b$

Direction of maximum Ascent

Note that $\nabla_x f = 0 \Rightarrow A\hat{x} = b$

$\hat{x}$ is the solution sought.

$\nabla_x^2 f = A$: since A is SPD, e'values are real and positive $\therefore f$ is CONVEX (concave up)

In optimization framework: find

$$\boxed{f(\hat{x}) = \min_{x \in \mathbb{R}^m} f(x)}$$

<u>Rmk:</u> We can recenter problem by a translation

let $\tilde{f} = \frac{1}{2}(x-\hat{x})^T A (x-\hat{x})$. Let $r = x - \hat{x}$

so $\tilde{f} = \frac{1}{2} r(x)^T A r(x)$ so $r(\hat{x}) = 0$ //

Introduce a ITERATIVE LINE SEARCH:

$$x^{k+1} = x^k + \alpha_{k+1} p_{k+1}, \quad k = 0, 1, \ldots$$

$\alpha \in \mathbb{R}$, $p_k \in \mathbb{R}^m$ $\begin{cases} p \text{ changes direction} \\ \alpha, \text{ its magnitude.} \end{cases}$

To generate a sequence $x_0, x_1, \ldots$ so that $\lim_{k \to \infty} x_k = \hat{x}$, starting with some initial guess x_0.

Find α_{k+1} & p_{k+1} so that the correction is $\perp$ to the current search direction:

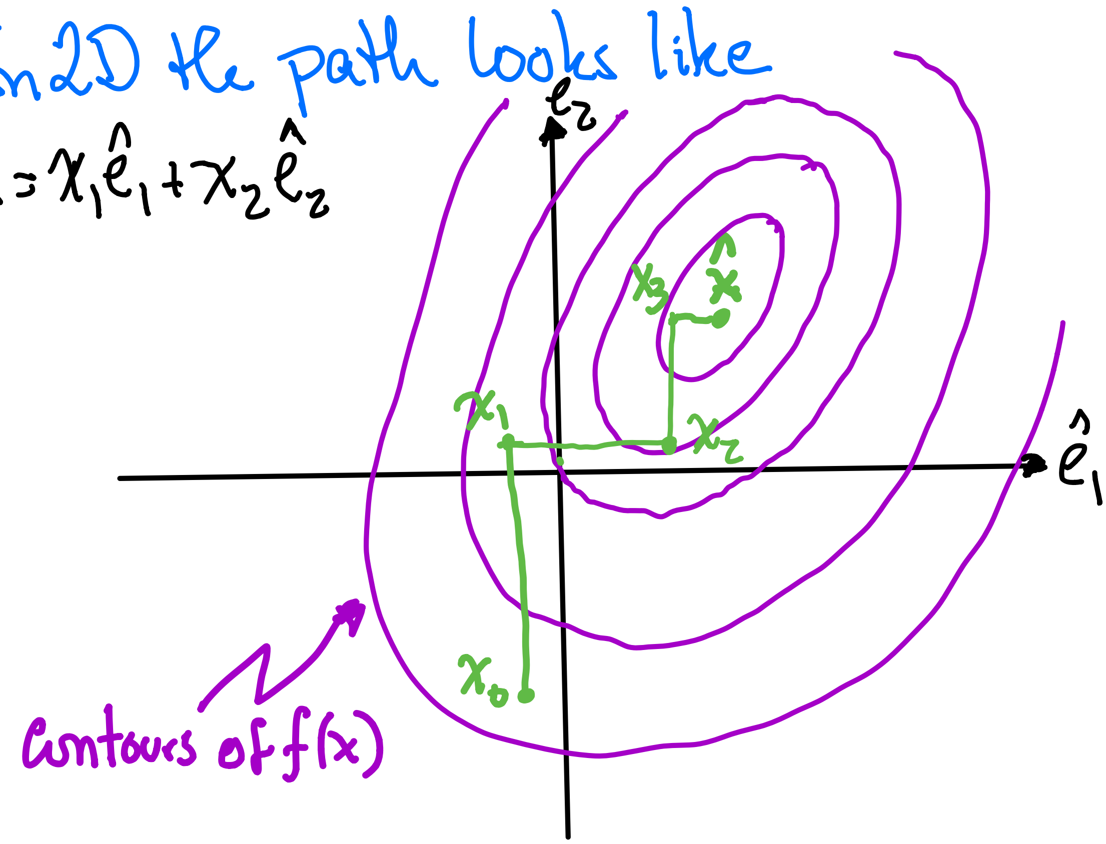
If $r_{k+1} = A x_{k+1} - b$

We will require that

$$r_{k+1}^T p_{k+1} = 0$$

In 2D the path looks like

$x = x_1 \hat{e}_1 + x_2 \hat{e}_2$



Now we present how α_{k+1} and p_{k+1} are designed:

let's assume we know p_{k+1}, focus on find α_{k+1}. Then, we'll figure out how to find p_{k+1}:

Take $f(x_{k+1}) = f(x_k + \alpha_{k+1} p_{k+1})$

differentiate wrt α_{k+1} and set to 0:

$$0 = \frac{d}{d\alpha_{k+1}} f(x_{k+1}) = \nabla_x f^T(x^{k+1}) \frac{d}{d\alpha_{k+1}}(x_k + \alpha_{k+1} p_{k+1})$$

$$0 = (Ax^{k+1} - b)^T \left(\frac{d}{d\alpha_{k+1}}(x_k + \alpha_{k+1} p_{k+1}) \right) = -r_{k+1}^T p_{k+1}$$

So if we design p_{k+1} so that

$$r_{k+1}^T p_{k+1} = 0$$

Since $r_{k+1} = b - A x_{k+1} = b - A(x_k + \alpha_{k+1} p_{k+1})$

$$= r_k - \alpha_{k+1} A p_{k+1}$$

then $p_{k+1}^T r_{k+1} = 0 = p_{k+1}^T r_k - \alpha_{k+1} p_{k+1}^T A p_{k+1}$

$$\therefore \boxed{\alpha_{k+1} = \frac{p_{k+1}^T r_k}{p_{k+1}^T A p_{k+1}}, \quad k = 0, 1, \dots}$$

In fact, we can design p_i's so that

$$(*) \quad r_{k+1}^T p_i = 0 \qquad 1 \le i \le k+1$$

Let

$$p_0 = -r_0$$

$$p_{k+1} = r_k + \beta_{k+1} p_k \qquad (\$)$$

So let's find β_{k+1}:

Propose that: $p_j^T A p_{k+1} = \delta_{ij} \quad (\star)$

$$1 \le i, j \le m$$

$$\delta_{ij} = \begin{cases} 0 & i \ne j \\ 1 & i = j \end{cases}$$

We say that the vectors p_j are "A-Conjugate" if $(\star)$ holds

Take $(\$)$ & multiply by $p_j^T A$

$$j = 0, 1, \ldots, k$$

$$\underset{0}{\underset{\shortparallel}{p_j^T A p_{k+1}}} = p_j^T A r_k + \beta_{k+1} p_j^T A p_k$$

$$\Rightarrow \quad \beta_{k+1} = -\frac{p_k^T A r_k}{p_k^T A p_k}$$

We notice that with $p_0 = -r_0$

$$p_{k+1} = r_k + \beta_{k+1} p_k$$

that $r_{k+1}^T p_i = 0 \quad 1 \le i \le k+1$

creates a sequence $p_0, p_1 \ldots$ that are A-conjugate.

Rmk: We can specify

$$\hat{x} = \sum_{i=1}^{m} \alpha_i p_i \quad \text{since } \{p_i\}_{i=1}^{m} \text{ spans } \mathbb{C}^m.$$

Hence $A\hat{x} = A(\alpha_1 p_1 + \cdots \alpha_m p_m)$

Multiply both sides by p_{k+1}^T:

$$p_j^T A\hat{x} = \alpha_{k+1} p_j^T A p_{k+1} = \alpha_{k+1} \delta_{j,k+1}$$

ALGORITHM:

We will use $\alpha_{k+1} = \dfrac{r_k^T r_k}{p_{k+1}^T A p_{k+1}}$

$$\beta_{k+1} = \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}}$$

Instead of

$$\alpha_{k+1} = \frac{p_{k+1}^T r_k}{p_{k+1}^T A\, p_{k+1}} ; \quad \beta_{k+1} = \frac{-p_k^T A r_k}{p_k^T A p_k}$$

We'll show, after presenting the algorithm that these are equivalent.

ALGORITHM

$r_0 = -b \;, \; x_0 = 0 \;, \; p_0 = b$

for $k = 1 : \text{ITMAX}$

$$\beta_{k+1} = \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}}$$

$$p_{k+1} = r_k + \beta_{k+1} p_k$$

$$\alpha_{k+1} = \frac{r_k^T r_k}{p_{k+1}^T A\, p_{k+1}}$$

$$x_{k+1} = x_k + \alpha_{k+1} p_{k+1}$$

$$r_{k+1} = b - A x_{k+1}$$

if $\|r_{k+1}\| < TOL_1$

and $\|x_{k+1} - x_k\| < TOL_2$

break, exit loop and report x_{k+1}

//

There are a couple of technical details that need to be addressed, so that all of the quantities are explained:

1

$$\alpha_{k+1} = \frac{p_{k+1}^T A r_k}{p_{k+1}^T A p_{k+1}} = \frac{r_k^T r_k}{p_{k+1}^T A p_{k+1}}$$

We know that

$$r_k = b - A x_k$$

$$= \underbrace{b - A x_{k-1}}_{r_{k-1}} - A \alpha_k p_k$$

$$\therefore \quad r_{k+1} = r_k - \alpha_{k+1} A p_{k+1} \qquad (*)$$

Multiply $(*)$ by r_k^T

$$\underset{\substack{\shortparallel \\ 0}}{r_k^T r_{k+1}} = r_k^T r_k - \alpha_{k+1} r_k^T A p_{k+1}$$

since they are $\perp$

$$\therefore \quad \alpha_{k+1} r_k^T A p_{k+1} = r_k^T r_k$$ reversing order & using $A = A^T$

or $$\alpha_{k+1} p_{k+1}^T A r_k = r_k^T r_k \quad (*)$$
(since $A = A^T$)

So (1) follows from $(*)$: $\frac{p_{k+1}^T A r_k}{p_{k+1}^T A p_{k+1}} = \frac{r_k^T r_k}{p_{k+1}^T A p_{k+1}}$ //

The other statement is:

(2) $$\beta_{k+1} = \frac{-p_k^T A r_k}{p_k^T A p_k} = \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}}$$

We show this next:

We know that

$$p_{k+1} = r_k + \beta_{k+1} p_k$$
$$r_{k+1} = r_k - \alpha_{k+1} A p_k$$

$$\underset{0}{\underset{\parallel}{r_k^T r_{k+1}}} = r_k^T r_k - \alpha_{k+1} r_k^T A p_{k+1}$$ transposing 2nd term:

$$\therefore \quad \alpha_{k+1} p_k^T A r_k = r_k^T r_k$$

$$\Rightarrow \quad p_k^T A r_k = \frac{1}{\alpha_{k+1}} r_k^T r_k$$

The numerator of β_{k+1}.

Now, the denominator:

$$p_k^T A p_k = \frac{1}{\alpha_{k+1}} r_{k-1}^T r_{k-1}$$

Since $r_k = r_{k-1} - \alpha_k A p_k$

$$\Rightarrow \alpha_k A p_k = r_{k-1} - r_k$$

$$\therefore \alpha_{k+1} p_k^T A p_k = \underbrace{\alpha_{k+1} p_k^T A}_{= -r_{k-1}^T} r_{k-1} + \alpha_{k+1} \beta_{k+1} \overbrace{p_k^T A p_{k-1}}^{=0}$$

So the ratio of

$$\frac{-p_k^T A r_k}{p_k^T A p_k} = \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}}$$

since the α_{k+1}'s cancel.

Another Useful fact: THEOREM 38.1

if $x_{k+1} = x_k + \alpha_{k+1} p_{k+1}$ (1)

$p_{k+1} = r_k + \beta_{k+1} p_k$ (2)

$r_{k+1} = r_k + \alpha_{k+1} A p_k$ (3)

Then for the problem $Ax = b$, as long as the iteration has not converged $r_k \neq 0$ the algorithm (CG) proceed w/o divisions by zero and the following subspaces are identical:

$$\langle x_1, x_2 \dots x_{k+1} \rangle = \langle p_0, p_1 \dots, p_k \rangle$$

$$= \langle r_0, r_1, \dots r_k \rangle = \langle b, Ab, \dots A^k b \rangle$$

The residuals are $\perp$:

$$r_{k+1}^T r_j = 0 \qquad j < k+1$$

and $$p_{k+1}^T A p_j = 0 \qquad j < k+1$$

Pf: Use induction. Set

Let $x_0 = 0$ $p_0 = b$ and $r_0 = b$

All of the following can be shown by applying 1 - 3:

Then by 1 we see that $x_{k+1} \in \langle p_0, p_1 \cdots p_k \rangle$

by 2 we see that $p_{k+1} \in \langle r_0, r_1 \cdots, r_k \rangle$

by 3 and $r_k = b - A x_k$ with $\begin{cases} x_0 = 0 \\ r_0 = b \end{cases}$

it follows that

$$\langle r_0, r_1 \ldots r_k \rangle = \langle b, Ab, \ldots A^k b \rangle$$

we already established that

$$r_{k+1}^T r_j = r_k^T r_j - \alpha_{k+1} p_k^T A r_j$$

that $r_{k+1}^T r_j = 0$, so long as $j < k+1$. True, since we can expand $r_j = \sum_{i=1}^{j} a_i p_i$,

and we assume $p_i^T A p_j = \delta_{ij}$

Finally since starting with $p_0 = r_0$: by construction:

$$p_{k+1}^T A p_j = 0 \quad j < k+1$$

//

Krylov Subspace $\mathcal{K}$:

The subspace defined as

$$\langle b, Ab, A^2b, \dots A^k b \rangle = \mathcal{K}_{k+1}$$

where $A \in \mathbb{C}^{m \times m}$, $b \in \mathbb{C}^m$ is a $k+1$-dimensional "Krylov Space" //

Rmk: since A is SPD in CG then

$$\mathcal{K}_m = \langle b, Ab, \dots A^m b \rangle$$

spans $\mathbb{C}^m$