1 March 11th, 2021

1.1 Properties of Conjugate Gradient

Going back to the projection framework, we know that CG is derived from 2-dim projection method by choosing $K = \text{span}\{r_k, d_{k-1}\}$.

Theorem 1.1

CG is a K-dim projection method at step K.

Since

$$x_{k+1} = \underset{x \in x_k + \text{span}\{r_k, d_{k-1}\}}{\arg\min} \|x_* - x\|_A.$$

The residue vector must be orthogonal to the subspace, meaning:

$$\begin{aligned} &\langle x_* - x_{k+1}, v \rangle = 0 \quad \forall v \in \text{span}\{r_k, d_{k-1}\} \\ \iff &\langle r_{k+1}, v \rangle = 0. \end{aligned}$$

Therefore:

$$\langle r_{k+1}, r_k \rangle = 0, \quad \langle r_{k+1}, d_{k-1} \rangle = 0, \quad \langle r_{k+1}, d_k \rangle = 0.$$

Thus, with $\alpha_k \neq 0$ (i.e. $r_k \neq 0$), β_k is optimal in the sense that:

$$\begin{aligned} \beta_k &= \underset{\beta \in \mathbb{R}}{\arg\min} \|x_k + \alpha_k(r_k + \beta d_{k-1}) - x_*\|_A \\ \iff d_k &= \underset{d \in r_k + \text{span}\{d_{k-1}\}}{\arg\min} \|x_k + \alpha_k d - x_*\|_A \\ \iff d_k &= \underset{d \in r_k + \text{span}\{d_{k-1}\}}{\arg\min} \left\| d - \frac{1}{\alpha_k}(x_* - x_k) \right\|_A. \end{aligned}$$

Thus d_k is the projection of $\frac{1}{\alpha_k}(x_* - x_k)$ onto the 1-dim subspace $r_k + \text{span}\{d_{k-1}\}$. As such, we have:

$$\begin{aligned} &\left\langle d_{k-1}, d_k - \frac{1}{\alpha_k}(x_* - x_k) \right\rangle_A = 0 \\ &\langle d_{k-1}, d_k \rangle_A = \frac{1}{\alpha_k} \langle d_{k-1}, x_* - x_k \rangle_A = \frac{1}{\alpha_k} \langle d_{k-1}, r_k \rangle = 0. \end{aligned}$$

since $\langle r_{k+1}, d_k \rangle = 0$. As such:

$$\langle d_{k-1}, d_k \rangle_A = 0, \quad if r_k \neq 0$$

which means that each d_k is orthogonal from d_{k-1}.

Remark 1.2 — If $r_k = 0$, then the algorithm stops, since we have achieved x_*.

In general $a \perp b, b \perp c \;\not\Longrightarrow\; a \perp c$, since orthogonality is not transitive. However, the orthogonality of vector produced by CG is transitive.

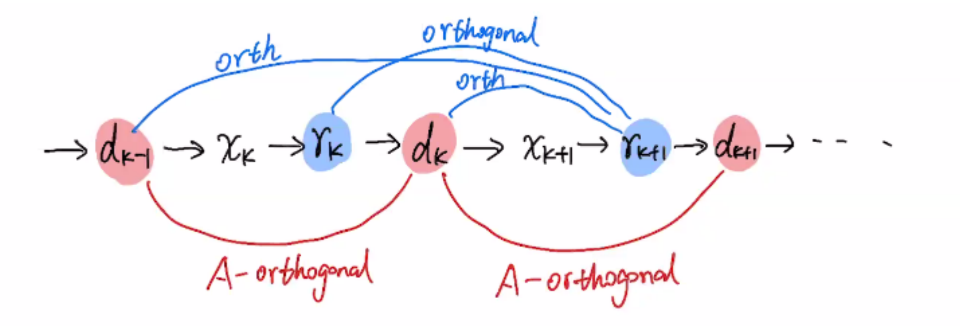

Figure 1: Diagram showing Orthogonality between d_k and r_k

Theorem 1.3

Assume A is SPD. Assume $r_0, r_1, r_2, \ldots, r_{i-1} \neq 0$. Then:

1. $\langle r_j, r_j \rangle = 0$ for all $j \leq i-1$ (meaning $\{r_0, r_1, \ldots, r_i\}$ are orthogonal)
2. (a) $\langle r_i, d_j \rangle = 0$ for all $j \leq i-1$
   (b) $\langle r_i, d_j \rangle_A = 0$ for all $j \leq i-2$
   (c) $\langle d_i, r_j \rangle_A = 0$ for all $j \leq i-1$
3. $\langle d_i, d_j \rangle_A = 0$ for all $j \leq i-1$ ($\{d_0, d_1, \ldots, d_i\}$ are A-orthogonal)

Proof. By Induction. Check notes. □

In matrix form, this is equivalent to:

1. $\iff$ Let $R_i = \begin{bmatrix} r_0 & r_1 & \ldots & r_i \end{bmatrix} \in \mathbb{R}^{n\times(i+1)}$. Then:$R_i^T R_i$ is diagonal.
2. $\iff$ Let $D_i = \begin{bmatrix} d_0 & d_1 & \ldots & d_i \end{bmatrix} \in \mathbb{R}^{n\times(i+1)}$. Then:

   (a) $R_i^T D_i$ is $\begin{bmatrix} \times & \ldots & \times \\ & \ddots & \vdots \\ 0 & & \times \end{bmatrix}$, i.e. upper triangular.

   (b) $R_i^T A D_i$ is $\begin{bmatrix} \times & & & & \\ \times & \times & & & \\ & \times & \ddots & & \\ & & \ddots & \ddots & \\ & & & \times & \times \end{bmatrix}$, i.e. upper triangular.

3. $\iff$ $D_i^T A D_i$ is diagonal.

Theorem 1.4

$\{x_k\}$ generated by CG satisfies:

$$\langle x_* - x_k, v\rangle_A = 0 \quad \forall v \in K_k$$

where K_k is the Krylov subspace. As a result:

$$x_k = \underset{x \in x_0 + K_k}{\arg\min} \|x_* - x\|_A$$

Definition 1.5 (**Krylov Subspace**).

$$K_k := \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{k-1} r_0\}$$

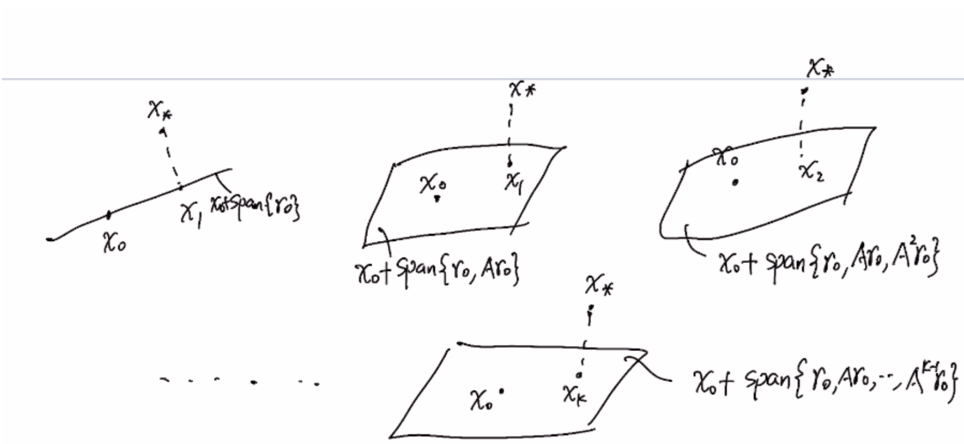

Figure 2: Pictoral Representation of Theorem 1.4

Proof. □

Corollary 1.6

If we run CG for N steps, it is equivalent to projecting to $\mathbb{R}^n$, which is x_*, thus meaning that CG is optimal.