$$f(x) = \frac{1}{2} x^T A x - b^T x + c$$

$$\nabla f(x) = \frac{1}{2}(A + A^T)x - b =$$

$$= Ax - b \quad (\text{if } A = A^T)$$

$$\boxed{-\nabla f(x) = b - Ax = r} = \text{residual (of } Ax = b)$$

1D line search along $\vec{s}_k$

$$\phi(\alpha_k) = f(\vec{x}_k + \alpha_k \vec{s}_k)$$

find min by setting $\frac{d\phi}{d\alpha} = 0$ and solving for α_k. We can do this analytically for quadratic f:

$$\frac{d\phi}{d\alpha}(\alpha_k) \underset{\text{chain rule}}{=} \underbrace{\nabla f(x_k + \alpha_k s_k)^T s_k}_{\text{directional derivative of } f \text{ along } s_k} = \nabla f(x_{k+1})^T s_k = 0$$

$$\Rightarrow \boxed{\vec{r}_{k+1}^{\,T} \vec{s}_k = 0}$$

Note: residual at $\vec{x}_{k+1}$ is orthogonal to search direction $\vec{s}_k$

$$\nabla f(x_{k+1})^T s_k = (Ax_{k+1} - b)^T s_k$$

$$= (A(x_k + \alpha_k s_k) - b)^T s_k$$

$$= (Ax_k - b + \alpha_k A s_k)^T s_k$$

$$= (-r_k + \alpha_k A s_k)^T s_k$$

$$\Rightarrow \boxed{\alpha_k = \frac{s_k^T r_k}{s_k^T A s_k}}$$

Proof: When s_k are A-orthogonal, only need n iterations.

s_i linearly indep.
Express e_0 in terms:

$$e_0 = \sum_{i=0}^{n-1} a_i s_i$$

$$s_k^T A e_0 = \sum_{i=0}^{n-1} a_i s_k^T A s_i = (*)$$

if $s_k^T A s_j = 0$ for $j \neq k$, then

$$(*) = a_k s_k^T A s_k$$

$$\Rightarrow \boxed{a_k = \frac{s_k^T A e_0}{s_k^T A s_k}} \quad \boxed{a_k = \frac{s_k^T r_0}{s_k^T A s_k}}$$

$$s_k^T A e_0 = s_k^T A \big(\underbrace{e_0 + \sum_{i=0}^{k-1} \alpha_i s_i}_{e_k}\big) = s_k^T A e_k$$

$$\Rightarrow a_k = \frac{s_k^T A e_k}{s_k^T A s_k} = \boxed{-\frac{s_k^T r_k}{s_k^T A s_k}} = a_k$$

$$\Rightarrow \alpha_k = -a_k$$

$$e_k = e_0 + \sum_{j=0}^{k-1} \alpha_j s_j$$

$$= \sum_{j=0}^{n-1} -\alpha_j s_j + \sum_{j=0}^{k-1} \alpha_j s_j$$

$$= -\sum_{j=k}^{n-1} \alpha_j s_j$$

$$e_n = 0$$

If we express the error e_0 in terms of A-conjugate s_k, then we see its components are same as steps

Properties

$$\begin{cases} x_{i+1} = x_i + \alpha_i s_i \\ e_i = x_i - x^* \\ r_i = b - Ax_i \end{cases} \qquad e_0 = \sum_{k=0}^{n-1} a_k s_k$$

already showed $s_i^T r_{i-1} = 0$

① $-s_i^T r_j$ for $j > i$

$= s_i^T A e_j$

$= s_i^T A \left(\sum_{k=j}^{n-1} a_k s_k \right) = 0$ ✓.

$\cancel{-s_i^T A e_j}$
$\cancel{-s_i^T A \left[\sum_{k=j}^{n-1} a_k s_k\right]}$
$\cancel{= 0}$
$\cancel{\text{if } j > i}$

$= \boxed{s_i^T r_j = 0 \text{ for } j > i}$

residuals are orthogonal

search directions are A-orthogonal.

$\Rightarrow s_i = r_i + \sum_{k=0}^{i-1} \beta_{ik} s_k$

② $\Rightarrow \left(r_i + \sum_{k=0}^{i-1} \beta_{ik} s_k \right)^T r_j = 0$.

$r_i^T r_j + \cancel{\sum_{k=0}^{i-1} \beta_{ik} s_k^T r_j} = 0 \Rightarrow \boxed{r_i^T r_j = 0 \quad j > i}$ ✓

③ $r_{i+1} = -Ae_{i+1}$

$= -A(e_i + \alpha_i s_i)$

$r_{i+1} = r_i - \alpha_i A s_i$

$$\boxed{A s_i = \frac{1}{\alpha_i}(r_i - r_{i+1})}$$

What happens when we use residuals as the $\{u_i\}$?

$$r_{i+1} = -A e_{i+1}$$
$$= -A(e_i + \alpha_i s_i)$$
$$\boxed{r_{i+1} = r_i - \alpha_i A s_i} \quad (*)$$

Recall: $\quad \beta_{ik} = -\dfrac{s_k^T A r_i}{s_k^T A s_k}, \quad k = 0, \dots, i-1$

$(u_i = r_i)$

$$s_k^T A r_i = r_i^T A s_k = \langle 3 \rangle \qquad \overset{0}{=} \text{ by } (2)$$
$$= -\frac{r_i^T}{\alpha_k}\left[r_{k+1} - r_k \right] = \frac{-1}{\alpha_k}\left[r_i^T r_{k+1} - r_i^T r_k \right]$$
$$= \begin{cases} \dfrac{-1}{\alpha_k} r_i^T r_{k+1} & k = i-1 \\ 0 & k < i-1 \quad \text{by } (2) \end{cases}$$

only non-zero β is $\Rightarrow \beta_{i,i-1} = \dfrac{-s_{i-1}^T A r_i}{s_{i-1}^T A s_{i-1}} = +\dfrac{1}{\alpha_{i-1}} \dfrac{r_i^T r_i}{s_{i-1}^T A s_{i-1}} \triangleq \beta_i$

Recall $\quad \alpha_{i-1} = +\dfrac{s_{i-1}^T r_{i-1}}{s_{i-1}^T A s_{i-1}} = +\dfrac{r_{i-1}^T r_{i-1}}{s_{i-1}^T A s_{i-1}}$ by (1), and $s_{i-1} = r_{i-1} + \sum_{k=0}^{i-2} \beta_{i-1,k} s_k$

(s_{i-1} and r_{i-1} differ by search directions prior to $i-1$)

$$\Rightarrow \beta_i = \frac{r_i^T r_i}{s_{i-1}^T r_{i-1}} = \frac{r_i^T r_i}{r_{i-1}^T r_{i-1}}$$

Method of Conjugate Gradients

$$S_0 = r_0 = b - A x_0$$ (steepest descent direction)

for $i = 0, 1, 2, \ldots$

$$\alpha_i = \frac{r_i^T r_i}{s_i^T A s_i}$$

$$x_{i+1} = x_i + \alpha_i s_i$$

$$r_{i+1} = r_i - \alpha_i A s_i$$

$$\beta_{i+1} = \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i}$$

$$s_{i+1} = r_{i+1} + \beta_{i+1} s_i$$

end.

Space + time requirements have been reduced from $O(n^2) \rightarrow O(m)$.