§7.6 The Conjugate Gradient Method (CG) for $A\mathbf{x} = \mathbf{b}$

Assumption: A is SYMMETRIC POSITIVE DEFINITE (SPD)

- $A^T = A$,
- $\mathbf{x}^T A\mathbf{x} \geq 0$ for any $\mathbf{x}$,
- $\mathbf{x}^T A\mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

§7.6 The Conjugate Gradient Method (CG) for $A\mathbf{x} = \mathbf{b}$

Assumption: A is SYMMETRIC POSITIVE DEFINITE (SPD)

- $A^T = A$,
- $\mathbf{x}^T A\mathbf{x} \geq 0$ for any $\mathbf{x}$,
- $\mathbf{x}^T A\mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

Thm: The vector $\mathbf{x}^*$ solves the SPD equations $A\mathbf{x} = \mathbf{b}$ if and only if it minimizes function

$$g(\mathbf{x}) \stackrel{def}{=} \mathbf{x}^T A\mathbf{x} - 2\,\mathbf{x}^T \mathbf{b}.$$

§7.6 The Conjugate Gradient Method (CG) for $A\mathbf{x} = \mathbf{b}$

Assumption: A is SYMMETRIC POSITIVE DEFINITE (SPD)

- $A^T = A$,
- $\mathbf{x}^T A\mathbf{x} \geq 0$ for any $\mathbf{x}$,
- $\mathbf{x}^T A\mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

Thm: The vector $\mathbf{x}^*$ solves the SPD equations $A\mathbf{x} = \mathbf{b}$ if and only if it minimizes function

$$g(\mathbf{x}) \stackrel{def}{=} \mathbf{x}^T A\mathbf{x} - 2\,\mathbf{x}^T \mathbf{b}.$$

Proof: Let $A\mathbf{x}^* = \mathbf{b}$. Then

$$\begin{aligned} g(\mathbf{x}) &= \mathbf{x}^T A\mathbf{x} - 2\,\mathbf{x}^T A\,\mathbf{x}^* \\ &= (\mathbf{x} - \mathbf{x}^*)^T A\,(\mathbf{x} - \mathbf{x}^*) - (\mathbf{x}^*)^T A\,(\mathbf{x}^*) \\ &= (\mathbf{x} - \mathbf{x}^*)^T A\,(\mathbf{x} - \mathbf{x}^*) + g(\mathbf{x}^*). \end{aligned}$$

Thus, $g(\mathbf{x}) \geq g(\mathbf{x}^*)$ for all $\mathbf{x}$; and $g(\mathbf{x}) = g(\mathbf{x}^*)$ iff $\mathbf{x} = \mathbf{x}^*$. □

CG for $A\mathbf{x} = \mathbf{b}$

Thm: The vector $\mathbf{x}^*$ solves the SPD equations $A\mathbf{x} = \mathbf{b}$ if and only if it minimizes function

$$g(\mathbf{x}) \stackrel{def}{=} \mathbf{x}^T A\mathbf{x} - 2\,\mathbf{x}^T\mathbf{b}.$$

The CG Idea: Starting from an initial vector $\mathbf{x}^{(0)}$, *quickly* compute new vectors $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(k)} \cdots$, with

$$g\left(\mathbf{x}^{(0)}\right) > g\left(\mathbf{x}^{(1)}\right) > g\left(\mathbf{x}^{(2)}\right) > \cdots > g\left(\mathbf{x}^{(k)}\right) > \cdots$$

so that the sequence $\{\mathbf{x}^{(k)}\}$ will converge to $\mathbf{x}^*$.

SEARCH DIRECTION and LINE SEARCH

- **The CG Idea:** Starting from an initial vector $\mathbf{x}^{(0)}$, *quickly* compute new vectors $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(k)} \cdots$, with

$$g\left(\mathbf{x}^{(0)}\right) > g\left(\mathbf{x}^{(1)}\right) > g\left(\mathbf{x}^{(2)}\right) > \cdots > g\left(\mathbf{x}^{(k)}\right) > \cdots$$

  so that the sequence $\{\mathbf{x}^{(k)}\}$ will converge to $\mathbf{x}^*$.

SEARCH DIRECTION and LINE SEARCH

- **The CG Idea:** Starting from an initial vector $\mathbf{x}^{(0)}$, *quickly* compute new vectors $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(k)} \cdots$, with

$$g\left(\mathbf{x}^{(0)}\right) > g\left(\mathbf{x}^{(1)}\right) > g\left(\mathbf{x}^{(2)}\right) > \cdots > g\left(\mathbf{x}^{(k)}\right) > \cdots$$

so that the sequence $\{\mathbf{x}^{(k)}\}$ will converge to $\mathbf{x}^*$.
- **Descent Method:** Assume a SEARCH DIRECTION $\mathbf{v}^{(k)}$ at iteration $\mathbf{x}^{(k-1)}$, next iteration with STEP-SIZE t_k

$$\mathbf{x}^{(k)} \stackrel{def}{=} \mathbf{x}^{(k-1)} + t_k \mathbf{v}^{(k)} \quad \text{minimizes} \quad g\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right).$$

SEARCH DIRECTION and LINE SEARCH

- **The CG Idea:** Starting from an initial vector $\mathbf{x}^{(0)}$, *quickly* compute new vectors $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(k)} \cdots$, with

$$g\left(\mathbf{x}^{(0)}\right) > g\left(\mathbf{x}^{(1)}\right) > g\left(\mathbf{x}^{(2)}\right) > \cdots > g\left(\mathbf{x}^{(k)}\right) > \cdots$$

  so that the sequence $\{\mathbf{x}^{(k)}\}$ will converge to $\mathbf{x}^*$.
- **Descent Method:** Assume a SEARCH DIRECTION $\mathbf{v}^{(k)}$ at iteration $\mathbf{x}^{(k-1)}$, next iteration with STEP-SIZE t_k

$$\mathbf{x}^{(k)} \stackrel{def}{=} \mathbf{x}^{(k-1)} + t_k\,\mathbf{v}^{(k)} \quad \text{minimizes} \quad g\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right).$$

- **Optimality Condition:**

$$\begin{aligned} 0 &= \frac{d}{d\,t} g\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right) = \left(\mathbf{v}^{(k)}\right)^T \nabla g\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right) \\ &= \left(\mathbf{v}^{(k)}\right)^T \left(2A\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right) - 2\,\mathbf{b}\right), \end{aligned}$$

SEARCH DIRECTION and LINE SEARCH

- **The CG Idea:** Starting from an initial vector $\mathbf{x}^{(0)}$, *quickly* compute new vectors $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(k)} \cdots$, with

$$g\left(\mathbf{x}^{(0)}\right) > g\left(\mathbf{x}^{(1)}\right) > g\left(\mathbf{x}^{(2)}\right) > \cdots > g\left(\mathbf{x}^{(k)}\right) > \cdots$$

so that the sequence $\{\mathbf{x}^{(k)}\}$ will converge to $\mathbf{x}^*$.
- **Descent Method:** Assume a SEARCH DIRECTION $\mathbf{v}^{(k)}$ at iteration $\mathbf{x}^{(k-1)}$, next iteration with STEP-SIZE t_k

$$\mathbf{x}^{(k)} \stackrel{def}{=} \mathbf{x}^{(k-1)} + t_k \mathbf{v}^{(k)} \quad \text{minimizes} \quad g\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right).$$

- **Optimality Condition:**

$$\begin{aligned} 0 &= \frac{d}{d\,t} g\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right) = \left(\mathbf{v}^{(k)}\right)^T \nabla g\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right) \\ &= \left(\mathbf{v}^{(k)}\right)^T \left(2A\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right) - 2\,\mathbf{b}\right), \\ t_k &= \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \stackrel{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}. \end{aligned}$$

SEARCH DIRECTION choices

For a small step-size t:

$$g\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right) \approx g\left(\mathbf{x}^{(k-1)}\right) + t\left(\mathbf{v}^{(k)}\right)^T \nabla g\left(\mathbf{x}^{(k-1)}\right).$$

- STEEPEST DESCENT: Greatest decrease in the value of $g\left(\mathbf{x}^{(k-1)} + t\mathbf{v}^{(k)}\right)$:

$$\mathbf{v}^{(k)} = -\nabla g\left(\mathbf{x}^{(k-1)}\right).$$

- A-**orthogonal** DIRECTIONS: non-zero vectors $\{\mathbf{v}^{(i)}\}_{i=1}^{n}$

$$\left(\mathbf{v}^{(i)}\right)^T \left(A\,\mathbf{v}^{(j)}\right) = 0 \quad \text{for all } i \neq j.$$

A-**orthogonal** vectors associated with the positive definite matrix A is linearly independent.

$A-$orthogonality Craft

$A-$orthogonality Craft

Thm: Let non-zero vectors $\{\mathbf{v}^{(k)}\}$ be A-orthogonal with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \stackrel{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}.$$

Then for $g(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} - 2\,\mathbf{x}^T\mathbf{b}$ and for $k = 1, \cdots, n$,

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1\mathbf{v}_1 + \cdots + \tau_k\mathbf{v}_k\right) = g\left(\mathbf{x}_0 + t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k\right).$$

$A-$orthogonality Craft

Thm: Let non-zero vectors $\{\mathbf{v}^{(k)}\}$ be A-orthogonal with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \stackrel{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}.$$

Then for $g(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} - 2\,\mathbf{x}^T \mathbf{b}$ and for $k = 1, \cdots, n$,

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1 \mathbf{v}_1 + \cdots + \tau_k \mathbf{v}_k\right) = g\left(\mathbf{x}_0 + t_1 \mathbf{v}_1 + \cdots + t_k \mathbf{v}_k\right).$$

Magic (I): $\mathbf{min}_{\tau_1} g\left(\mathbf{x}_0 + \tau_1 \mathbf{v}^{(1)}\right) = g\left(\mathbf{x}_0 + t_1 \mathbf{v}^{(1)}\right).$

$$\mathbf{min}_{\tau_1,\tau_2} g\left(\mathbf{x}_0 + \tau_1 \mathbf{v}^{(1)} + \tau_2 \mathbf{v}^{(2)}\right) = g\left(\mathbf{x}_0 + t_1 \mathbf{v}^{(1)} + t_2 \mathbf{v}^{(2)}\right).$$

$$\mathbf{min}_{\mathbf{x}} g(\mathbf{x}) =$$

$$\mathbf{min}_{\tau_1,\cdots,\tau_n} g\left(\mathbf{x}_0 + \tau_1 \mathbf{v}^{(1)} + \cdots + \tau_n \mathbf{v}^{(n)}\right) = g\left(\mathbf{x}_0 + t_1 \mathbf{v}^{(1)} + \cdots + t_n \mathbf{v}^{(n)}\right).$$

Thus $\mathbf{x} = \mathbf{x}_0 + t_1 \mathbf{v}^{(1)} + \cdots + t_n \mathbf{v}^{(n)}$ is solution to $A\mathbf{x} = \mathbf{b}$.

$A-$orthogonality Craft

$A-$orthogonality Craft

Thm: Let non-zero vectors $\{\mathbf{v}^{(k)}\}$ be A-orthogonal with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \stackrel{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}.$$

Then for $g(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} - 2\,\mathbf{x}^T\mathbf{b}$ and for $k = 1, \cdots, n$,

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1\mathbf{v}_1 + \cdots + \tau_k\mathbf{v}_k\right) = g\left(\mathbf{x}_0 + t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k\right).$$

$A-$orthogonality Craft

Thm: Let non-zero vectors $\{\mathbf{v}^{(k)}\}$ be A-orthogonal with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \stackrel{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}.$$

Then for $g(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} - 2\,\mathbf{x}^T \mathbf{b}$ and for $k = 1, \cdots, n$,

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1 \mathbf{v}_1 + \cdots + \tau_k \mathbf{v}_k\right) = g\left(\mathbf{x}_0 + t_1 \mathbf{v}_1 + \cdots + t_k \mathbf{v}_k\right).$$

Proof (I): Let $\mathbf{t} = (\tau_1, \cdots, \tau_k)$. Then

$$\begin{aligned}
& g\left(\mathbf{x}_0 + \tau_1 \mathbf{v}_1 + \cdots + \tau_k \mathbf{v}_k\right) = g\left(\mathbf{x}_0\right) \\
& + \mathbf{t}^T \left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)^T A \left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right) \mathbf{t} - 2\mathbf{t}^T \left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)^T \mathbf{r}^{(0)}, \\
& \nabla_{\mathbf{t}}\, g = 2\left(\left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)^T A \left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right) \mathbf{t} - \left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)^T \mathbf{r}^{(0)}\right)
\end{aligned}$$

$A-$orthogonality Craft

Thm: Let non-zero vectors $\{\mathbf{v}^{(k)}\}$ be A-orthogonal with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \overset{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}.$$

Then for $g(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} - 2\,\mathbf{x}^T\mathbf{b}$ and for $k = 1, \cdots, n$,

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1\mathbf{v}_1 + \cdots + \tau_k\mathbf{v}_k\right) = g\left(\mathbf{x}_0 + t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k\right).$$

Proof (I): Let $\mathbf{t} = (\tau_1, \cdots, \tau_k)$. Then

$$g\left(\mathbf{x}_0 + \tau_1\mathbf{v}_1 + \cdots + \tau_k\mathbf{v}_k\right) = g\left(\mathbf{x}_0\right)$$
$$+\mathbf{t}^T\left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)^T A\left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)\mathbf{t} - 2\mathbf{t}^T\left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)^T\mathbf{r}^{(0)},$$
$$\nabla_{\mathbf{t}}\, g = 2\left(\left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)^T A\left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)\mathbf{t} - \left(\mathbf{v}_1, \cdots, \mathbf{v}_k\right)^T\mathbf{r}^{(0)}\right)$$

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1\mathbf{v}_1 + \cdots + \tau_k\mathbf{v}_k\right) \iff \nabla_{\mathbf{t}}\, g = \mathbf{0}.$$

$A-$orthogonality Craft

$A-$orthogonality Craft

Thm: Let non-zero vectors $\{\mathbf{v}^{(k)}\}$ be A-orthogonal with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \stackrel{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}.$$

Then for $g(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} - 2\,\mathbf{x}^T \mathbf{b}$ and for $k = 1, \cdots, n$,

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1 \mathbf{v}^{(1)} + \cdots + \tau_k \mathbf{v}^{(k)}\right) = g\left(\mathbf{x}_0 + t_1 \mathbf{v}^{(1)} + \cdots + t_k \mathbf{v}^{(k)}\right).$$

$A-$orthogonality Craft

Thm: Let non-zero vectors $\{\mathbf{v}^{(k)}\}$ be A-orthogonal with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \stackrel{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}.$$

Then for $g(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} - 2\,\mathbf{x}^T\mathbf{b}$ and for $k = 1, \cdots, n$,

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1\mathbf{v}^{(1)} + \cdots + \tau_k\mathbf{v}^{(k)}\right) = g\left(\mathbf{x}_0 + t_1\mathbf{v}^{(1)} + \cdots + t_k\mathbf{v}^{(k)}\right).$$

Proof (II): Since vectors $\{\mathbf{v}^{(k)}\}$ are A-orthogonal

$$\nabla_{\mathbf{t}}\, g = 2\left(\mathbf{diag}\left(\left(\mathbf{v}^{(1)}\right)^T A\,\mathbf{v}^{(1)}, \cdots, \left(\mathbf{v}^{(k)}\right)^T A\,\mathbf{v}^{(k)}\right)\mathbf{t} - \left(\mathbf{v}^{(1)}, \cdots, \mathbf{v}^{(k)}\right)^T \mathbf{r}^{(0)}\right)$$

$$\nabla_{\mathbf{t}}\, g = \mathbf{0} \Longleftrightarrow \mathbf{t} = \begin{pmatrix} \frac{\left(\mathbf{v}^{(1)}\right)^T \left(\mathbf{r}^{(0)}\right)}{\left(\mathbf{v}^{(1)}\right)^T \left(A\,\mathbf{v}^{(1)}\right)} \\ \vdots \\ \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(0)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)} \end{pmatrix}.$$

$A-$orthogonality Craft

$A-$orthogonality Craft

Thm: Let non-zero vectors $\{\mathbf{v}^{(k)}\}$ be A-orthogonal with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \stackrel{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}.$$

Then for $g(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} - 2\,\mathbf{x}^T\mathbf{b}$ and for $k = 1, \cdots, n$,

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1\mathbf{v}^{(1)} + \cdots + \tau_k\mathbf{v}^{(k)}\right) = g\left(\mathbf{x}_0 + t_1\mathbf{v}^{(1)} + \cdots + t_k\mathbf{v}^{(k)}\right).$$

$A-$orthogonality Craft

Thm: Let non-zero vectors $\{\mathbf{v}^{(k)}\}$ be A-orthogonal with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{r}^{(k-1)} \stackrel{def}{=} \mathbf{b} - A\,\mathbf{x}^{(k-1)} \quad \text{(residual)}.$$

Then for $g(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} - 2\,\mathbf{x}^T\mathbf{b}$ and for $k = 1, \cdots, n$,

$$\mathbf{min}_{\tau_1,\cdots,\tau_k} g\left(\mathbf{x}_0 + \tau_1\mathbf{v}^{(1)} + \cdots + \tau_k\mathbf{v}^{(k)}\right) = g\left(\mathbf{x}_0 + t_1\mathbf{v}^{(1)} + \cdots + t_k\mathbf{v}^{(k)}\right).$$

Proof (III): Since

$$\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right) = \left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(0)} - \sum_{j=1}^{k-1} t_j\,A\,\mathbf{v}^{(j)}\right) = \left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(0)}\right), \quad \text{so}$$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(0)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)} = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}.$$

$A-$orthogonality vectors (I)

$A-$orthogonality vectors (I)

Thm: Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)}\,\mathbf{v}^{(j)}.$$

Assume that $\{\mathbf{v}^{(k)}\}$ are non-zero. Then they are $A-$orthogonal.

$A-$orthogonality vectors (I)

Thm: Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)}\,\mathbf{v}^{(j)}.$$

Assume that $\{\mathbf{v}^{(k)}\}$ are non-zero. Then they are $A-$orthogonal.
Induction Proof: For all $1 \leq i < k$,

$$\begin{aligned}
\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(i)}\right) &= -\left(\mathbf{r}^{(k-1)}\right)^T \left(A\,\mathbf{v}^{(i)}\right) \\
&\quad + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)} \left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(i)}\right) \\
&= -\left(\mathbf{r}^{(k-1)}\right)^T \left(A\,\mathbf{v}^{(i)}\right) + \left(\mathbf{v}^{(i)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right) = 0.
\end{aligned}$$

$A-$orthogonality vectors (II)

$A-$orthogonality vectors (II)

Thm: Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)} \mathbf{v}^{(j)}.$$

Let $\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + t_1 \mathbf{v}^{(1)} + \cdots + t_k \mathbf{v}^{(k)}$ and $\mathbf{r}^{(k)} = \mathbf{b} - A\,\mathbf{x}^{(k)}$. Then

$$\left(\mathbf{v}^{(j)}\right)^T \left(\mathbf{r}^{(k)}\right) = 0, \quad j = 1, \cdots, k; \quad \left(\mathbf{r}^{(j)}\right)^T \left(\mathbf{r}^{(k)}\right) = 0, \;\; j = 1, \cdots, k-1.$$

$A-$orthogonality vectors (II)

Thm: Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)} \mathbf{v}^{(j)}.$$

Let $\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + t_1\mathbf{v}^{(1)} + \cdots + t_k\mathbf{v}^{(k)}$ and $\mathbf{r}^{(k)} = \mathbf{b} - A\,\mathbf{x}^{(k)}$. Then

$$\left(\mathbf{v}^{(j)}\right)^T \left(\mathbf{r}^{(k)}\right) = 0, \quad j = 1, \cdots, k; \quad \left(\mathbf{r}^{(j)}\right)^T \left(\mathbf{r}^{(k)}\right) = 0, \quad j = 1, \cdots, k-1.$$

Proof: Due to optimality property of $\mathbf{x}^{(k)}$, for all τ and for $1 \le j \le k$,

$$\begin{aligned} g\left(\mathbf{x}^{(k)}\right) &\le g\left(\mathbf{x}^{(k)} + \tau\,\mathbf{v}^{(j)}\right) \\ &= g\left(\mathbf{x}^{(k)}\right) - 2\tau\left(\mathbf{r}^{(k)}\right)^T \mathbf{v}^{(j)} + \tau^2 \left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right). \end{aligned}$$

This is true only when $\left(\mathbf{r}^{(k)}\right)^T \mathbf{v}^{(j)} = 0.$

$A-$orthogonality vectors (II)

Thm: Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)}\,\mathbf{v}^{(j)}.$$

Let $\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + t_1\mathbf{v}^{(1)} + \cdots + t_k\mathbf{v}^{(k)}$ and $\mathbf{r}^{(k)} = \mathbf{b} - A\,\mathbf{x}^{(k)}$. Then

$$\left(\mathbf{v}^{(j)}\right)^T \left(\mathbf{r}^{(k)}\right) = 0, \quad j = 1, \cdots, k; \quad \left(\mathbf{r}^{(j)}\right)^T \left(\mathbf{r}^{(k)}\right) = 0, \ \ j = 1, \cdots, k-1.$$

Proof: Due to optimality property of $\mathbf{x}^{(k)}$, for all τ and for $1 \le j \le k$,

$$\begin{aligned} g\left(\mathbf{x}^{(k)}\right) &\le g\left(\mathbf{x}^{(k)} + \tau\,\mathbf{v}^{(j)}\right) \\ &= g\left(\mathbf{x}^{(k)}\right) - 2\tau\left(\mathbf{r}^{(k)}\right)^T \mathbf{v}^{(j)} + \tau^2\left(\mathbf{v}^{(j)}\right)^T\left(A\,\mathbf{v}^{(j)}\right). \end{aligned}$$

This is true only when $\left(\mathbf{r}^{(k)}\right)^T \mathbf{v}^{(j)} = 0$.

Residual vector orthogonality: $\mathbf{r}^{(j)} =$ linear combination of $\mathbf{v}^{(1)}, \cdots, \mathbf{v}^{(j+1)}$.

$A-$orthogonality vectors (III)

$A-$orthogonality vectors (III)

Thm: Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)} \mathbf{v}^{(j)}.$$

Let $\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + t_1\mathbf{v}^{(1)} + \cdots + t_k\mathbf{v}^{(k)}$ and $\mathbf{r}^{(k)} = \mathbf{b} - A\,\mathbf{x}^{(k)}$. Then

$$\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(j)}\right) = -\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right), \quad j = 1, \cdots, k-1.$$

$A-$orthogonality vectors (III)

Thm: Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)}\,\mathbf{v}^{(j)}.$$

Let $\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + t_1\mathbf{v}^{(1)} + \cdots + t_k\mathbf{v}^{(k)}$ and $\mathbf{r}^{(k)} = \mathbf{b} - A\,\mathbf{x}^{(k)}$. Then

$$\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(j)}\right) = -\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right), \quad j = 1, \cdots, k-1.$$

Proof (I): For $j = k - 1$,

$$\begin{aligned}\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right) &= \left(-\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)}\,\mathbf{v}^{(j)}\right)^T \left(\mathbf{r}^{(k-1)}\right) \\ &= -\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right).\end{aligned}$$

$A-$orthogonality vectors (III)

$A-$orthogonality vectors (III)

Thm: Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)} \mathbf{v}^{(j)}.$$

Let $\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + t_1\mathbf{v}^{(1)} + \cdots + t_k\mathbf{v}^{(k)}$ and $\mathbf{r}^{(k)} = \mathbf{b} - A\,\mathbf{x}^{(k)}$. Then

$$\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(j)}\right) = -\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right), \quad j = 1, \cdots, k-1.$$

$A-$orthogonality vectors (III)

Thm: Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + \sum_{j=1}^{k-1} \frac{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(j)}\right)^T \left(A\,\mathbf{v}^{(j)}\right)}\,\mathbf{v}^{(j)}.$$

Let $\mathbf{x}^{(k)} = \mathbf{x}^{(0)} + t_1\mathbf{v}^{(1)} + \cdots + t_k\mathbf{v}^{(k)}$ and $\mathbf{r}^{(k)} = \mathbf{b} - A\,\mathbf{x}^{(k)}$. Then

$$\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(j)}\right) = -\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right), \quad j = 1, \cdots, k-1.$$

Proof (II): For $j < k-1$

$$\begin{aligned}
\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(j)}\right) &= \left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right) + \left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(j)} - \mathbf{r}^{(k-1)}\right) \\
&= \left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right) + \left(\mathbf{v}^{(k)}\right)^T \left(\sum_{i=j+1}^{k-1} t_i A\,\mathbf{v}^{(i)}\right) \\
&= -\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right).
\end{aligned}$$

$A-$orthogonality: A Gift from Math God

$A-$orthogonality: A Gift from Math God

Set $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$, and for $k = 2, \cdots, n$, write

$$\mathbf{v}^{(k)} = \sum_{j=0}^{k-1} \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(j)}\right)}{\left(\mathbf{r}^{(j)}\right)^T \left(\mathbf{r}^{(j)}\right)} \mathbf{r}^{(j)}.$$

Then

$$\begin{aligned}
\mathbf{v}^{(k)} &= -\sum_{j=0}^{k-1} \frac{\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{r}^{(j)}\right)^T \left(\mathbf{r}^{(j)}\right)} \mathbf{r}^{(j)} \\
&= -\mathbf{r}^{(k-1)} - \frac{\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{r}^{(k-2)}\right)^T \left(\mathbf{r}^{(k-2)}\right)} \sum_{j=0}^{k-2} \frac{\left(\mathbf{r}^{(k-2)}\right)^T \left(\mathbf{r}^{(k-2)}\right)}{\left(\mathbf{r}^{(j)}\right)^T \left(\mathbf{r}^{(j)}\right)} \mathbf{r}^{(j)} \\
&= -\mathbf{r}^{(k-1)} + s_{k-1}\, \mathbf{v}^{(k-1)},
\end{aligned}$$

with $s_{k-1} = \dfrac{\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{r}^{(k-2)}\right)^T \left(\mathbf{r}^{(k-2)}\right)}$.

Thm: Let $\{\mathbf{v}^{(i)}\}_{i=1}^{n}$ be A-**orthogonal** with $\mathbf{v}^{(1)} = -\mathbf{r}^{(0)}$ and for $k = 1, \cdots, n$

$$t_k = \frac{\left(\mathbf{v}^{(k)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)} = -\frac{\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}, \quad \mathbf{x}^{(k)} \stackrel{def}{=} \mathbf{x}^{(k-1)} + t_k \mathbf{v}^{(k)}.$$

Then $A\,\mathbf{x}^{(n)} = b$ in *exact arithmetic*.

Conjugate Gradient Algorithm

Thm: For $k = 1, \cdots, n$, define,

$$\mathbf{v}^{(k)} = -\mathbf{r}^{(k-1)} + s_{k-1}\,\mathbf{v}^{(k-1)} \quad \text{with} \quad s_{k-1} = \frac{\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{r}^{(k-2)}\right)^T \left(\mathbf{r}^{(k-2)}\right)},$$

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + t_k\,\mathbf{v}^{(k)} \quad \text{with} \quad t_k = -\frac{\left(\mathbf{r}^{(k-1)}\right)^T \left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T \left(A\,\mathbf{v}^{(k)}\right)}.$$

Then vectors $\{\mathbf{v}^{(k)}\}$ are A-orthogonal and $A\,\mathbf{x}^{(n)} = b$ in *exact arithmetic*.

The CG Algorithm: C is for *Craft*, G is for *Gift*.

Algorithm 1 Conjugate Gradient Algorithm

Input: Symmetric positive definite $A \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^n$, initial guess $\mathbf{x}^{(0)} \in \mathbb{R}^n$, and tolerance $\tau > 0$.

Output: Approximate solution $\mathbf{x}$.

Algorithm:

Initialize: $\mathbf{r}^{(0)} = \mathbf{b} - A\,\mathbf{x}^{(0)}$, $\mathbf{v}^{(0)} = -\mathbf{r}^{(0)}$, $k = 1$

while $\left\|\mathbf{r}^{(k-1)}\right\|_2 \geq \tau$ **do**

$$t_k = -\frac{\left(\mathbf{r}^{(k-1)}\right)^T\left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T\left(A\,\mathbf{v}^{(k)}\right)}.$$

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + t_k\,\mathbf{v}^{(k)}$$

$$\mathbf{r}^{(k)} = \mathbf{r}^{(k-1)} - t_k\,A\,\mathbf{v}^{(k)}.$$

$$s_k = \frac{\left(\mathbf{r}^{(k)}\right)^T\left(\mathbf{r}^{(k)}\right)}{\left(\mathbf{r}^{(k-1)}\right)^T\left(\mathbf{r}^{(k-1)}\right)}.$$

$$\mathbf{v}^{(k+1)} = -\mathbf{r}^{(k)} + s_k\,\mathbf{v}^{(k)}\ .$$

$$k = k + 1.$$

end while

Preconditioned Conjugate Gradient: $A\mathbf{x} = \mathbf{b}$

Let $M = C\,C^T$ be <u>*some approximation*</u> to A. Define

$$\widehat{A} \stackrel{def}{=} C^{-1} \cdot A \cdot C^{-T}, \quad \widehat{\mathbf{x}} \stackrel{def}{=} C^T \cdot \mathbf{x}, \quad \widehat{\mathbf{b}} \stackrel{def}{=} C^{-1} \cdot \mathbf{b}.$$

Hope $\widehat{A}\,\widehat{\mathbf{x}} = \widehat{b}$ is easier to solve with CG than $A\,\mathbf{x} = \mathbf{b}$.

Algorithm 2 Conjugate Gradient Algorithm on $\widehat{A}\,\widehat{\mathbf{x}} = \widehat{b}$

Initialize: $\widehat{\mathbf{r}}^{(0)} = C^{-1}\left(\mathbf{b} - A \cdot C^{-T}\,\widehat{\mathbf{x}}^{(0)}\right)$, $\widehat{\mathbf{v}}^{(0)} = -\widehat{\mathbf{r}}^{(0)}$, $k = 1$

while $\left\|\mathbf{r}^{(k-1)}\right\|_2 \geq \tau$ **do**

$$t_k = -\frac{\left(\widehat{\mathbf{r}}^{(k-1)}\right)^T\left(\widehat{\mathbf{r}}^{(k-1)}\right)}{\left(\widehat{\mathbf{v}}^{(k)}\right)^T\left(C^{-1}\cdot A\cdot C^{-T}\,\widehat{\mathbf{v}}^{(k)}\right)}.$$

$$\widehat{\mathbf{x}}^{(k)} = \widehat{\mathbf{x}}^{(k-1)} + t_k\,\widehat{\mathbf{v}}^{(k)}$$

$$\widehat{\mathbf{r}}^{(k)} = \widehat{\mathbf{r}}^{(k-1)} - t_k\,C^{-1} \cdot A \cdot C^{-T}\,\widehat{\mathbf{v}}^{(k)}.$$

$$s_k = \frac{\left(\widehat{\mathbf{r}}^{(k)}\right)^T\left(\widehat{\mathbf{r}}^{(k)}\right)}{\left(\widehat{\mathbf{r}}^{(k-1)}\right)^T\left(\widehat{\mathbf{r}}^{(k-1)}\right)}.$$

$$\widehat{\mathbf{v}}^{(k+1)} = -\widehat{\mathbf{r}}^{(k)} + s_k\,\widehat{\mathbf{v}}^{(k)}\ .$$

$$k = k + 1.$$

end while

Set $\mathbf{v}^{(k)} = C^{-T}\,\widehat{\mathbf{v}}^{(k)}, \mathbf{x}^{(k)} = C^{-T}\,\widehat{\mathbf{x}}^{(k)}, \mathbf{r}^{(k)} = C\,\widehat{\mathbf{r}}^{(k)}$

Algorithm 3 Preconditioned Conjugate Gradient Algorithm

Initialize: $\mathbf{r}^{(0)} = \mathbf{b} - A\,\mathbf{x}^{(0)}$, $\mathbf{z}^{(0)} = M^{-1}\,\mathbf{r}^{(0)}$, $\mathbf{v}^{(1)} = -\mathbf{z}^{(0)}$, $k = 1$

while $\left\|\mathbf{r}^{(k-1)}\right\|_2 \geq \tau$ **do**

$$t_k = -\frac{\left(\mathbf{z}^{(k-1)}\right)^T\left(\mathbf{r}^{(k-1)}\right)}{\left(\mathbf{v}^{(k)}\right)^T\left(A\,\mathbf{v}^{(k)}\right)}.$$

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + t_k\,\mathbf{v}^{(k)}$$

$$\mathbf{r}^{(k)} = \mathbf{r}^{(k-1)} - t_k\,A\,\mathbf{v}^{(k)}.$$

$$\mathbf{z}^{(k)} = M^{-1}\,\mathbf{r}^{(k)}.$$

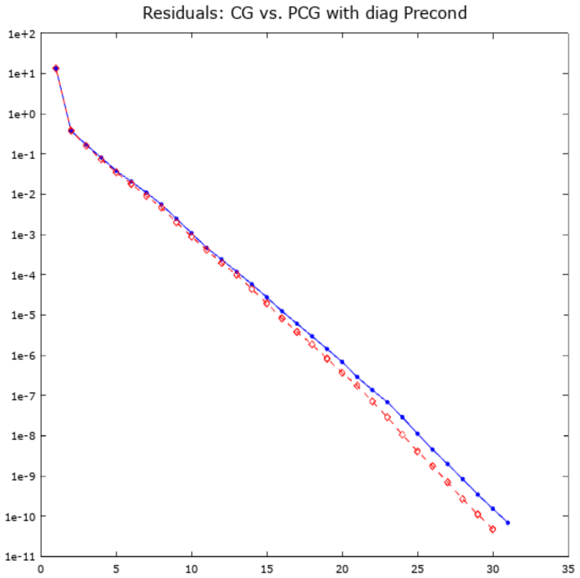
$$s_k = \frac{\left(\mathbf{z}^{(k)}\right)^T\left(\mathbf{r}^{(k)}\right)}{\left(\mathbf{z}^{(k-1)}\right)^T\left(\mathbf{r}^{(k-1)}\right)}.$$

$$\mathbf{v}^{(k+1)} = -\mathbf{z}^{(k)} + s_k\,\mathbf{v}^{(k)}\ .$$

$$k = k + 1.$$

end while

CG vs. PCG on 100×100 random SPD



§8.1 Discrete Least Squares Approximation

Example Problem:

- **Given:** Decennial census data since 1610 on US population.
- **Predict:** US population in next twenty years.

Real focus is on prediction

§8.1 Discrete Least Squares Approximation

Example Problem:

- **Given:** Decennial census data since 1610 on US population.
- **Predict:** US population in next twenty years.

Real focus is on prediction

Example Models:

- LINEAR MODEL:

$$\mathcal{P}\left(Year\right) \approx \alpha + \beta \times Year.$$

- LOG-LINEAR MODEL:

$$\mathcal{P}\left(Year\right) \approx \mathbf{exp}\left(\alpha + \beta \times Year\right).$$

§8.1 Discrete Least Squares Approximation

Example Problem:

- **Given:** Decennial census data since 1610 on US population.
- **Predict:** US population in next twenty years.

Real focus is on prediction

Example Models:

- LINEAR MODEL:

$$\mathcal{P}\left(Year\right) \approx \alpha + \beta \times Year.$$

- LOG-LINEAR MODEL:

$$\mathcal{P}\left(Year\right) \approx \mathbf{exp}\left(\alpha + \beta \times Year\right).$$

Model can not be exact, but could have predictive value.

US Population

Census year	Population
1610	350
1620	2,302
1630	4,646
1640	26,634
1650	50,368
1660	75,058
1670	111,935
1680	151,507
1690	210,372
1700	250,888
1710	331,711
1720	466,185
1730	629,445
1740	905,563
1750	1,170,760
1760	1,593,625
1770	2,148,076
1780	2,780,369
1790	3,929,214
1800	5,308,483
1810	7,239,881
1820	9,638,453
1830	12,866,020
1840	17,069,453
1850	23,191,876
1860	31,443,321
1870	38,558,371
1880	50,189,209
1890	62,979,766
1900	76,212,168
1910	92,228,496
1920	106,021,537
1930	123,202,624
1940	132,164,569
1950	151,325,798
1960	179,323,175
1970	203,211,926
1980	226,545,805
1990	248,709,873
2000	281,421,906
2010	308,745,538

Least Squares Model Solution

Given population data for $\text{year}_1, \cdots, \text{year}_n$. Define for $1 \leq i \leq n$, $x_i = \text{year}_i$,

$$y_i = \begin{cases} \mathcal{P}(\text{year}_i), & \textsc{Linear Model}, \\ \log(\mathcal{P}(\text{year}_i)), & \textsc{Log-linear Model}. \end{cases}$$

- **Least Squares Fit:**

$$\min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - (\alpha + x_i \beta))^2 .$$

- **Prediction:** For future year x, US population will be

$$\mathcal{P}(x) = \begin{cases} \alpha + x\beta, & \textsc{Linear Model}, \\ \exp(\alpha + x\beta), & \textsc{Log-linear Model}. \end{cases}$$

Least Squares Fit: $\mathbf{min}_{\alpha,\beta} \sum_{i=1}^{n} (y_i - (\alpha + x_i\beta))^2$

$$
\begin{aligned}
\sum_{i=1}^{n} (y_i - (\alpha + x_i\beta))^2 &= \left\| \begin{pmatrix} y_1 - (\alpha + x_1\beta) \\ \vdots \\ y_n - (\alpha + x_n\beta) \end{pmatrix} \right\|_2^2 \\
&= \left\| \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right\|_2^2 \\
&= \|\mathbf{b} - A\,\mathbf{x}\|_2^2, \quad \text{where}
\end{aligned}
$$

$$
\mathbf{b} \overset{def}{=} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad A \overset{def}{=} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{x} \overset{def}{=} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.
$$

Least Squares Fit becomes

$\mathbf{min}_{\mathbf{x}} \|\mathbf{b} - A\,\mathbf{x}\|_2^2$.

Least Squares Fit: $\min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - (\alpha + x_i\beta))^2$

$$
\begin{aligned}
\sum_{i=1}^{n} (y_i - (\alpha + x_i\beta))^2 &= \left\| \begin{pmatrix} y_1 - (\alpha + x_1\beta) \\ \vdots \\ y_n - (\alpha + x_n\beta) \end{pmatrix} \right\|_2^2 \\
&= \left\| \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right\|_2^2 \\
&= \|\mathbf{b} - A\,\mathbf{x}\|_2^2, \quad \text{where}
\end{aligned}
$$

$$
\mathbf{b} \stackrel{def}{=} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad A \stackrel{def}{=} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{x} \stackrel{def}{=} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.
$$

Least Squares Fit becomes

$\min_{\mathbf{x}} \|\mathbf{b} - A\,\mathbf{x}\|_2^2$. In general, $A \in \mathbb{R}^{n\times m}, \mathbf{x} \in \mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^n$; $n \geq m$.

Polynomial Least Squares Fit

Fitting data $\{(x_i, y_i), , i = 1, \cdots n\}$ with order $m-1$ polynomial:

$$\begin{aligned} & \mathbf{min}_{\alpha_0,\alpha_1,\cdots,\alpha_{m-1}} \sum_{i=1}^{n} \left(y_i - \left(\alpha_0 + \alpha_1 x_i + \cdots \alpha_{m-1} x_i^{m-1}\right)\right)^2 \\ = \quad & \mathbf{min}_{\mathbf{x}} \|\mathbf{b} - A\,\mathbf{x}\|_2^2, \quad \text{where} \end{aligned}$$

$$\mathbf{b} \stackrel{def}{=} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad A \stackrel{def}{=} \begin{pmatrix} 1 & x_1 & \cdots & x_1^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{m-1} \end{pmatrix}, \quad \mathbf{x} \stackrel{def}{=} \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{m-1} \end{pmatrix}.$$

Least Squares Solution: $\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_2^2$

$$\begin{aligned}
\text{Define} \quad g(\mathbf{x}) &= \|\mathbf{b} - A\mathbf{x}\|_2^2 = (\mathbf{b} - A\mathbf{x})^T (\mathbf{b} - A\mathbf{x}) \\
&= \mathbf{x}^T \left(A^T A\right) \mathbf{x} - 2\mathbf{x}^T \left(A^T \mathbf{b}\right) + \mathbf{b}^T \mathbf{b}. \\
\nabla g(\mathbf{x}) &= 2\left(A^T A\right) \mathbf{x} - 2\left(A^T \mathbf{b}\right) = 2A^T (A\mathbf{x} - \mathbf{b}).
\end{aligned}$$

- OPTIMALITY CONDITION:

$$\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_2^2 \iff \nabla g(\mathbf{x}) = \mathbf{0}.$$

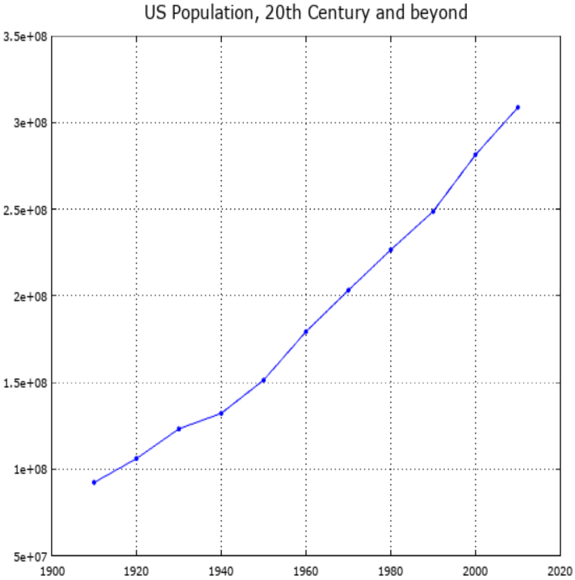
- NORMAL EQUATION:

$$A^T (A\mathbf{x} - \mathbf{b}) = \mathbf{0}.$$

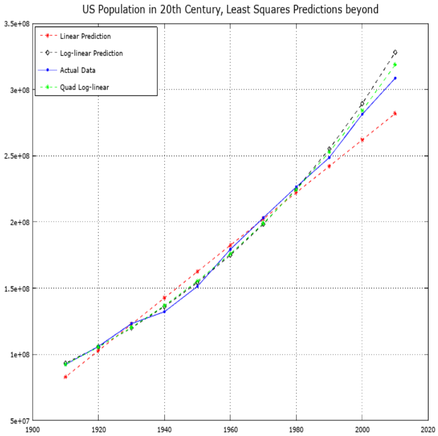
- LEAST SQUARES SOLUTION:

$$\mathbf{x} = \left(A^T A\right)^{-1} \left(A^T \mathbf{b}\right).$$

In general, A^{-1} does not exist, but $\left(A^T A\right)^{-1}$ does.


US Population, 20th Century and beyond
3.5e+08
3e+08
2.5e+08
2e+08
1.5e+08
1e+08
5e+07
1900
1920
1940
1960
1980
2000
2020

Predicting years 2000, 2010 with data through 1990



- Log-linear Model predicts year 2000 better than Linear Model.
- Quad Log-linear Model best.
- Predictions for year 2010 are worse.

§8.2 Orthogonal Polynomials/Least Squares Approximation

Suppose a polynomial $P_n(x) = \sum_{k=0}^{n} a_k x^k = a_0 + a_1 x + \cdots + a_n x^n$

is found to approximate a function $f \in C[a,b]$ by minimizing error

$$\mathcal{E} \stackrel{def}{=} \int_a^b \left(f(x) - \sum_{k=0}^{n} a_k x^k \right)^2 dx$$

§8.2 Orthogonal Polynomials/Least Squares Approximation

Suppose a polynomial $P_n(x) = \sum_{k=0}^{n} a_k x^k = a_0 + a_1 x + \cdots + a_n x^n$

is found to approximate a function $f \in C[a,b]$ by minimizing error

$$\begin{aligned}
\mathcal{E} &\stackrel{def}{=} \int_a^b \left(f(x) - \sum_{k=0}^{n} a_k x^k \right)^2 dx \\
&= \int_a^b f^2(x)\, dx - 2 \int_a^b f(x) \left(\sum_{k=0}^{n} a_k x^k \right) dx + \int_a^b \left(\sum_{k=0}^{n} a_k x^k \right)^2 dx \\
&= \int_a^b f^2(x)\, dx - 2 \sum_{k=0}^{n} a_k \int_a^b f(x)\, x^k\, dx + \sum_{j,k=0}^{n} a_j a_k \int_a^b x^{j+k}\, dx \\
&= \int_a^b f^2(x)\, dx - 2\, \mathbf{x}^T \mathbf{b} + \mathbf{x}^T H \mathbf{x}, \quad \text{where} \quad \mathbf{x} \stackrel{def}{=} (a_0, \cdots, a_n)^T,
\end{aligned}$$

$$\mathbf{b} \stackrel{def}{=} \left(\int_a^b f(x)\, x^0\, dx, \cdots, \int_a^b f(x)\, x^n\, dx \right)^T, \quad H \stackrel{def}{=} \left(\int_a^b x^{j+k}\, dx \right).$$

Normal Equation

$$\mathcal{E} = \int_a^b f^2(x)\,dx - 2\,\mathbf{x}^T\,\mathbf{b} + \mathbf{x}^T\,H\,\mathbf{x}; \quad \nabla\,\mathcal{E} = 2\,(H\,\mathbf{x} - \mathbf{b})\,.$$

- Optimality Condition:

$$\mathbf{min}_{\mathbf{x}}\,\mathcal{E} \iff \nabla\,\mathcal{E} = \mathbf{0}.$$

- Normal equation:

$$H\,\mathbf{x} = \mathbf{b}, \quad \text{where}$$

$$\begin{aligned}
\mathbf{b} &= \left(\int_a^b f(x)\,x^0\,dx, \cdots, \int_a^b f(x)\,x^n\,dx\right)^T, \\
H &= \left(\int_a^b x^{j+k}\,dx\right) = \left(\frac{b^{j+k+1} - a^{j+k+1}}{j+k+1}\right),
\end{aligned}$$

Normal Equation

$$\mathcal{E} \;=\; \int_a^b f^2(x)\, dx - 2\,\mathbf{x}^T \mathbf{b} + \mathbf{x}^T H\,\mathbf{x}; \quad \nabla\mathcal{E} = 2\,(H\,\mathbf{x} - \mathbf{b})\,.$$

- Optimality Condition:

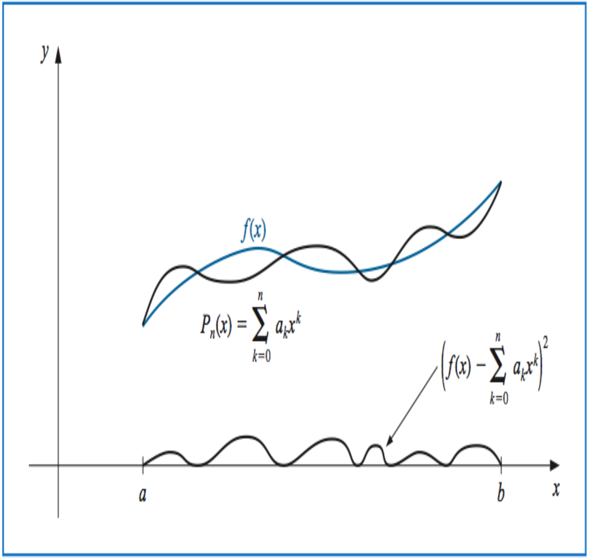
$$\mathbf{min}_{\mathbf{x}}\,\mathcal{E} \iff \nabla\mathcal{E} = \mathbf{0}.$$

- Normal equation:

$$H\,\mathbf{x} = \mathbf{b}, \quad \text{where}$$

$$\begin{aligned} \mathbf{b} &= \left(\int_a^b f(x)\, x^0\, dx, \cdots, \int_a^b f(x)\, x^n\, dx\right)^T, \\ H &= \left(\int_a^b x^{j+k}\, dx\right) = \left(\frac{b^{j+k+1} - a^{j+k+1}}{j+k+1}\right), \quad \text{ill-conditioned for large } n. \end{aligned}$$

$$\min_{a_0,\cdots,a_n} \int_a^b \left| f(x) - \sum_{k=0}^{n} a_k x^k \right|^2 dx$$



Ex: LS approximating quadratic polynomial

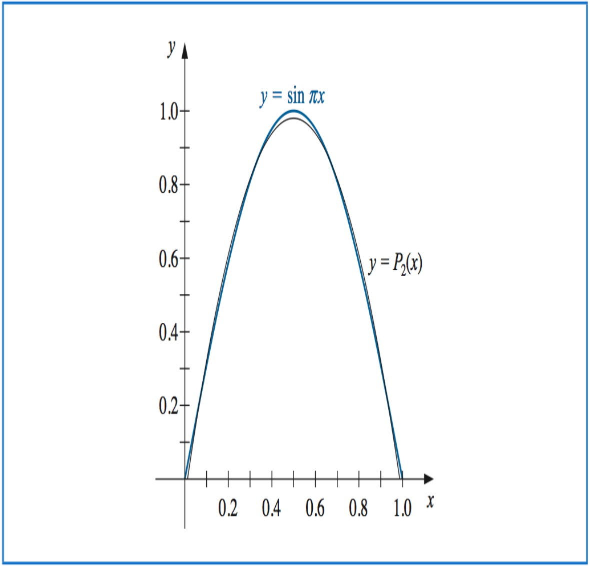
For the function $f(x) = \mathbf{sin}(\pi x) \in C[0,1]$.

$$\mathbf{min}_{a_0,a_1,a_2} \int_0^1 \left|\mathbf{sin}(\pi x) - \left(a_0 + a_1 x + a_2 x^2\right)\right|^2 dx.$$

Normal equation $H\mathbf{x} = \mathbf{b}$, where

$$\begin{aligned}
\mathbf{b} &= \left(\int_0^1 \mathbf{sin}(\pi x)\,dx, \int_0^1 \mathbf{sin}(\pi x)\,x\,dx, \int_0^1 \mathbf{sin}(\pi x)\,x^2\,dx\right)^T, \\
&= \left(\frac{2}{\pi}, \frac{1}{\pi}, \frac{\pi^2-4}{\pi^3}\right)^T, \\
H &= \left(\frac{1}{j+k+1}\right) = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}, \\
\mathbf{x} &= H^{-1}\mathbf{b} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \approx \begin{pmatrix} -0.050465 \\ 4.12251 \\ -4.12251 \end{pmatrix}.
\end{aligned}$$

$$\min_{a_0,a_1,a_2} \int_0^1 \left|\sin(\pi x) - \left(a_0 + a_1 x + a_2 x^2\right)\right|^2 dx$$



Def: Linearly Independent Functions

The set of functions $\{\phi_0, \cdots, \phi_n\}$ is LINEARLY INDEPENDENT on $[a, b]$ if, whenever

$$c_0\, \phi_0(x) + c_1\, \phi_1(x) + \cdots + c_n\, \phi_n(x) = 0, \quad \text{for all } x \in [a, b],$$

we have $c_0 = c_1 = \cdots = c_n = 0$. Otherwise the set of functions is LINEARLY DEPENDENT.

Def: Linearly Independent Functions

The set of functions $\{\phi_0, \cdots, \phi_n\}$ is LINEARLY INDEPENDENT on $[a, b]$ if, whenever

$$c_0\,\phi_0(x) + c_1\,\phi_1(x) + \cdots + c_n\,\phi_n(x) = 0, \quad \text{for all } x \in [a, b],$$

we have $c_0 = c_1 = \cdots = c_n = 0$. Otherwise the set of functions is LINEARLY DEPENDENT.

- **Ex:** Let $\phi_0(x) = 1$ and $\phi_j(x) = x^j +$ lower order terms for $j = 1, \cdots, n$. Then $\{\phi_0, \cdots, \phi_n\}$ is LINEARLY INDEPENDENT.

Def: Linearly Independent Functions

The set of functions $\{\phi_0, \cdots, \phi_n\}$ is LINEARLY INDEPENDENT on $[a, b]$ if, whenever

$$c_0\,\phi_0(x) + c_1\,\phi_1(x) + \cdots + c_n\,\phi_n(x) = 0, \quad \text{for all } x \in [a, b],$$

we have $c_0 = c_1 = \cdots = c_n = 0$. Otherwise the set of functions is LINEARLY DEPENDENT.

- **Ex:** Let $\phi_0(x) = 1$ and $\phi_j(x) = x^j$ + lower order terms for $j = 1, \cdots, n$. Then $\{\phi_0, \cdots, \phi_n\}$ is LINEARLY INDEPENDENT.
- **Ex:** Let $\phi_0(x) = 1$, $\phi_1(x) = x - 1$, and $\phi_2(x) = x^2 + x + 1$. Then any polynomial $Q(x) = a_0 + a_1\,x + a_2\,x^2$ is a LINEAR COMBINATION of $\{\phi_0(x), \phi_1(x), \phi_2(x)\}$: Set

$$Q(x) = c_0\,\phi_0(x) + c_1\,\phi_1(x) + c_2\,\phi_2(x), \quad \text{implying}$$

$$a_0 + a_1\,x + a_2\,x^2 = c_0 + c_1\,(x-1) + c_2\,\left(x^2 + x + 1\right).$$

Matching coefficients leads to

$$c_2 = a_2, \quad c_1 = a_1 - a_2, \quad c_0 = a_0 + a_1 - 2\,a_2.$$

Weighted Least Squares Approximation (WLS)

Def: An integrable function w is a WEIGHT FUNCTION on $[a, b]$ if $w(x) \geq 0$ for all $x \in [a, b]$, and w integrates to a positive value on any subinterval $I \subseteq [a, b]$.

Weighted Least Squares Approximation (WLS)

Def: An integrable function w is a WEIGHT FUNCTION on $[a, b]$ if $w(x) \geq 0$ for all $x \in [a, b]$, and w integrates to a positive value on any subinterval $I \subseteq [a, b]$.

Let the set of functions $\{\phi_0, \cdots, \phi_n\}$ be linearly independent on $[a, b]$ and let $w(x)$ be a weight function.

$$\mathcal{E} \stackrel{WLS}{=} \int_a^b w(x) \left(f(x) - \sum_{k=0}^{n} a_k \phi_k(x) \right)^2 dx$$

Weighted Least Squares Approximation (WLS)

Def: An integrable function w is a WEIGHT FUNCTION on $[a, b]$ if $w(x) \geq 0$ for all $x \in [a, b]$, and w integrates to a positive value on any subinterval $I \subseteq [a, b]$.

Let the set of functions $\{\phi_0, \cdots, \phi_n\}$ be linearly independent on $[a, b]$ and let $w(x)$ be a weight function.

$$\mathcal{E} \overset{WLS}{=} \int_a^b w(x) \left(f(x) - \sum_{k=0}^{n} a_k \phi_k(x) \right)^2 dx$$

$$= \int_a^b w(x)\, f^2(x)\, dx - 2\,\mathbf{x}^T \mathbf{b} + \mathbf{x}^T H \mathbf{x}, \quad \text{with} \quad \mathbf{x} \overset{def}{=} (a_0, \cdots, a_n)^T,$$

$$\mathbf{b} \overset{def}{=} \left(\int_a^b w(x)\, f(x)\, \phi_0(x)\, dx, \cdots, \int_a^b w(x)\, f(x)\, \phi_n(x)\, dx \right)^T,$$

$$H \overset{def}{=} \left(\int_a^b w(x)\, \phi_j(x)\, \phi_k(x)\, dx \right).$$

$$\mathcal{E} \;=\; \int_a^b w(x)\, f^2(x)\, dx - 2\,\mathbf{x}^T\mathbf{b} + \mathbf{x}^T H\,\mathbf{x}; \quad \nabla\,\mathcal{E} = 2\,(H\,\mathbf{x} - \mathbf{b})\,.$$

OPTIMALITY CONDITION and NORMAL EQUATION:

$$\mathbf{min}_{\mathbf{x}}\,\mathcal{E} \iff \nabla\,\mathcal{E} = \mathbf{0} \iff H\,\mathbf{x} = \mathbf{b},$$

$$\mathcal{E} = \int_a^b w(x)\, f^2(x)\, dx - 2\,\mathbf{x}^T\mathbf{b} + \mathbf{x}^T H\,\mathbf{x}; \quad \nabla\mathcal{E} = 2\,(H\,\mathbf{x} - \mathbf{b}).$$

OPTIMALITY CONDITION and NORMAL EQUATION:

$$\mathbf{min}_{\mathbf{x}}\,\mathcal{E} \iff \nabla\mathcal{E} = \mathbf{0} \iff H\,\mathbf{x} = \mathbf{b},$$

$$\mathbf{b} \stackrel{def}{=} \left(\int_a^b w(x)\, f(x)\, \phi_0(x)\, dx, \cdots, \int_a^b w(x)\, f(x)\, \phi_n(x)\, dx\right)^T,$$

$$H \stackrel{def}{=} \left(\int_a^b w(x)\, \phi_j(x)\, \phi_k(x)\, dx\right).$$

Def: ORTHOGONAL POLYNOMIALS

$$\int_a^b w(x)\, \phi_j(x)\, \phi_k(x)\, dx = 0 \iff \quad \text{for all } j \neq k.$$

$$\mathcal{E} = \int_a^b w(x)\, f^2(x)\, dx - 2\,\mathbf{x}^T \mathbf{b} + \mathbf{x}^T H \mathbf{x}; \quad \nabla \mathcal{E} = 2\,(H\mathbf{x} - \mathbf{b}).$$

Optimality Condition and Normal equation:

$$\mathbf{min}_{\mathbf{x}}\, \mathcal{E} \iff \nabla \mathcal{E} = \mathbf{0} \iff H\mathbf{x} = \mathbf{b},$$

$$\mathbf{b} \stackrel{def}{=} \left(\int_a^b w(x)\, f(x)\, \phi_0(x)\, dx, \cdots, \int_a^b w(x)\, f(x)\, \phi_n(x)\, dx \right)^T,$$

$$H \stackrel{def}{=} \left(\int_a^b w(x)\, \phi_j(x)\, \phi_k(x)\, dx \right).$$

Def: Orthogonal Polynomials

$$\int_a^b w(x)\, \phi_j(x)\, \phi_k(x)\, dx = 0 \iff \quad \text{for all } j \neq k.$$

WLS with orthogonal polynomials:

$$a_j = \frac{\int_a^b w(x)\, f(x)\, \phi_j(x)\, dx}{\int_a^b w(x)\, \phi_j^2(x)\, dx}, \quad j = 0, 1, \cdots, n.$$

Orthogonal Polynomials: Legendre polynomials

- weight function $w(x) \equiv 1$, on interval $[0, 1]$.
- $P_0(x) = 1,\ P_1(x) = x,\ P_2(x) = x^2 - \frac{1}{3}$,

  $(n+1) P_{n+1}(x) = (2n+1) x P_n(x) - n P_{n-1}(x), \quad n = 2, 3, \cdots$

Orthogonal Polynomials: Gram-Schmidt process

Thm: The set of polynomial functions $\{\phi_0, \cdots, \phi_n\}$ defined below on $[a, b]$ is orthogonal with respect to the weight function w.

$$\phi_0(x) = 1, \quad \phi_1(x) = x - B_1, \quad \text{and for } k \geq 2$$

$$\phi_k(x) = (x - B_k)\,\phi_{k-1}(x) - C_k\,\phi_{k-2}(x), \quad \text{with}$$

$$B_j = \frac{\int_a^b x\,w(x)\,\phi_{j-1}^2(x)\,dx}{\int_a^b w(x)\,\phi_{j-1}^2(x)\,dx}, \quad j = 1, 2, \cdots, n,$$

$$C_j = \frac{\int_a^b x\,w(x)\,\phi_{j-1}(x)\,\phi_{j-2}(x)\,dx}{\int_a^b w(x)\,\phi_{j-2}^2(x)\,dx}, \quad j = 2, 3, \cdots, n.$$