7.6 The Conjugate Gradient Method

Assumption: $A\boldsymbol{x} = \boldsymbol{b}$, A is positive definite.

Definition: Inner production of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is $< \boldsymbol{x}, \boldsymbol{y} >= \boldsymbol{x}^t \boldsymbol{y}$.

Let A be positive definite, $< \boldsymbol{x}, A\boldsymbol{y} >= \boldsymbol{x}^t A \boldsymbol{y} = \boldsymbol{x}^t A^t \boldsymbol{y} = (A\boldsymbol{x})^t \boldsymbol{y} =< A\boldsymbol{x}, \boldsymbol{y} >$.

Definition: A quadratic form is a scale, quadratic function of a vector with the form

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

where A is a matrix, $\boldsymbol{x}$ and $\boldsymbol{b}$ are vectors, and c is a scalar constant.

Example. Let $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} 2 \\ -8 \end{bmatrix}$, $c = 0$. The solution $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{b}$ is $\boldsymbol{x} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$

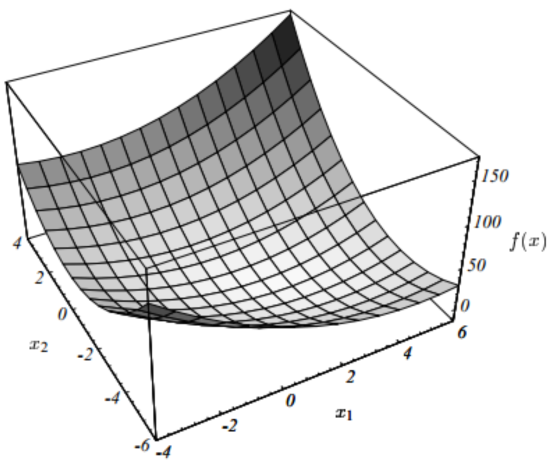

Figure: The graph of a quadratic form $f(\boldsymbol{x})$. The minimum point of this surface is the solution to $A\boldsymbol{x} = \boldsymbol{b}$.

The gradient of a quadratic form is defined to be

$$f'(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

$$f'(\boldsymbol{x}) = \frac{1}{2}A^t\boldsymbol{x} + \frac{1}{2}A\boldsymbol{x} - \boldsymbol{b}$$

If A is symmetric, $f'(\boldsymbol{x})$ reduces to $f'(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b}$.

Therefore, the solution to $A\boldsymbol{x} = \boldsymbol{b}$ is a critical point of $f(\boldsymbol{x})$.

Theorem. The vector $\boldsymbol{x}^*$ is a solution to the positive definite linear system $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $\boldsymbol{x}^*$ produces the minimal value of $g(\boldsymbol{x}) =< \boldsymbol{x}, A\boldsymbol{x} > -2 < \boldsymbol{x}, \boldsymbol{b} >$.

Proof Let $\boldsymbol{x}$ and $\boldsymbol{v} \neq 0$ be fixed vectors and t a real number variable.

$$g(\boldsymbol{x} + t\boldsymbol{v}) =< \boldsymbol{x} + t\boldsymbol{v}, A(\boldsymbol{x} + t\boldsymbol{v}) > -2 < \boldsymbol{x} + t\boldsymbol{v}, \boldsymbol{b} >$$

$$=< \boldsymbol{x}, A\boldsymbol{x} > -2 < \boldsymbol{x}, b > +2t < \boldsymbol{v}, A\boldsymbol{x} > -2t < \boldsymbol{v}, \boldsymbol{b} > +t^2 < \boldsymbol{v}, A\boldsymbol{v} >$$

So $\quad g(\boldsymbol{x} + t\boldsymbol{v}) = g(\boldsymbol{x}) - 2t < \boldsymbol{v}, \boldsymbol{b} - A\boldsymbol{x} > +t^2 < \boldsymbol{v}, A\boldsymbol{v} >$

Define $h(t) = g(\boldsymbol{x} + t\boldsymbol{v})$

Then $h(t)$ assumes a minimal value when $h'(t) = 0$.

$$h'(t) = -2 < \boldsymbol{v}, \boldsymbol{b} - A\boldsymbol{x} > +2t < \boldsymbol{v}, A\boldsymbol{v} >$$

The **minimum** occurs when $\hat{t} = \frac{<\boldsymbol{v},\boldsymbol{b}-A\boldsymbol{x}>}{<\boldsymbol{v},A\boldsymbol{v}>}$

$$h(\hat{t}) = g(\boldsymbol{x} + \hat{t}\boldsymbol{v}) = g(\boldsymbol{x}) - \frac{<\boldsymbol{v}, \boldsymbol{b} - A\boldsymbol{x}>^2}{<\boldsymbol{v}, A\boldsymbol{v}>}$$

For any vector $\boldsymbol{v} \neq \boldsymbol{0},$ we have $g(\boldsymbol{x} + \hat{t}\boldsymbol{v}) < g(\boldsymbol{x})$ unless $<\boldsymbol{v}, \boldsymbol{b} - A\boldsymbol{x}>= 0$.

Suppose $\boldsymbol{x}^*$ satisfies $A\boldsymbol{x}^* = \boldsymbol{b}$, then $<\boldsymbol{v}, \boldsymbol{b} - A\boldsymbol{x}^*>= 0$ for any $\boldsymbol{v}$. Thus $\boldsymbol{x}^*$ minimizes $g(\boldsymbol{x})$.

On the other hand, suppose that $\boldsymbol{x}^*$ is a vector minimizes $g(\boldsymbol{x})$. Then for any vector $\boldsymbol{v},$ $g(\boldsymbol{x}^* + \hat{t}\boldsymbol{v}) \geq g(\boldsymbol{x}^*)$. Thus $<\boldsymbol{v}, \boldsymbol{b} - A\boldsymbol{x}^*>= 0$. This implies that $\boldsymbol{b} - A\boldsymbol{x}^* = \boldsymbol{0}$.

The Method of Steepest Descent

a) Start with an arbitrary initial guess $\boldsymbol{x}^{(0)}$ to the solution $\boldsymbol{x}^*$ to $A\boldsymbol{x} = \boldsymbol{b}$

b) Let $\boldsymbol{v}^{(1)} = \boldsymbol{r}^{(0)} = \boldsymbol{b} - A\boldsymbol{x}^{(0)}$.

Compute

$$t_1 = \frac{<\boldsymbol{v}^{(1)}, \boldsymbol{b} - A\boldsymbol{x}^{(0)}>}{<\boldsymbol{v}^{(1)}, A\boldsymbol{v}^{(1)}>}$$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + t_1\boldsymbol{v}^{(1)}$$

Remark: the gradient of $g(\boldsymbol{x})$ is $\nabla g(\boldsymbol{x}) = 2(A\boldsymbol{x} - \boldsymbol{b}) = -2\boldsymbol{r}.$ The direction of greatest decrease in the value of $g(\boldsymbol{x})$ is $-\nabla g(\boldsymbol{x})$.

c) $\boldsymbol{v}^{(2)} = \boldsymbol{r}^{(1)} = \boldsymbol{b} - A\boldsymbol{x}^{(1)}$

$$t_2 = \frac{<\boldsymbol{v}^{(2)}, \boldsymbol{b} - A\boldsymbol{x}^{(1)}>}{<\boldsymbol{v}^{(2)}, A\boldsymbol{v}^{(2)}>}$$

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + t_2\boldsymbol{v}^{(2)}$$

d) Repeat the above process.

Remark: The Method of Steepest Descent does not lead to fastest convergence.

Definition: A set of nonzero vectors $\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(n)}\}$ that satisfy $< \boldsymbol{v}^{(i)}, A\boldsymbol{v}^{(j)} > = 0$, if $i \neq j$ is said to be A-**orthogonal**.

Theorem. Let $\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(n)}\}$ be an A-**orthogonal** set of nonzero vectors associated with the positive definite matrix A, and let $\boldsymbol{x}^{(0)}$ be arbitrary. Define

$$t_k = \frac{< \boldsymbol{v}^{(k)}, \boldsymbol{b} - A\boldsymbol{x}^{(k-1)} >}{< \boldsymbol{v}^{(k)}, A\boldsymbol{v}^{(k)} >}$$

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + t_1 \boldsymbol{v}^{(k)}$$

for $k = 1,2, \dots, n$. Then, assuming exact arithmetic, $A\boldsymbol{x}^{(n)} = b$.

Conjugate Gradient Method

Theorem. The residual vectors $\boldsymbol{r}^{(k)}$, where $k = 1,2, \dots, n$, for a conjugate direction method, satisfy the equation $< \boldsymbol{r}^{(k)}, \boldsymbol{v}^{(j)} >= 0$, for each $j = 1,2, \dots, k$.

Algorithm:

a) Start with an arbitrary initial guess $\boldsymbol{x}^{(0)}$ to the solution $\boldsymbol{x}^{*}$ to $A\boldsymbol{x} = \boldsymbol{b}$

Set

$$\boldsymbol{r}^{(0)} = \boldsymbol{b} - A\boldsymbol{x}^{(0)}$$

$$\boldsymbol{v}^{(1)} = \boldsymbol{r}^{(0)}$$

b) for $k = 1,2,3, \ldots$

$$t_k = \frac{< \boldsymbol{r}^{(k-1)}, \boldsymbol{r}^{(k-1)} >}{< \boldsymbol{v}^{(k)}, A\boldsymbol{v}^{(k)} >}$$

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

$$\boldsymbol{r}^{(k)} = \boldsymbol{r}^{(k-1)} - t_k A\boldsymbol{v}^{(k)}$$

$$s_k = \frac{< \boldsymbol{r}^{(k)}, \boldsymbol{r}^{(k)} >}{< \boldsymbol{r}^{(k-1)}, \boldsymbol{r}^{(k-1)} >}$$

$$\boldsymbol{v}^{(k+1)} = \boldsymbol{r}^{(k)} + s_k \boldsymbol{v}^{(k)}$$

Preconditioning

Motivation: When A is ill-conditioned, conjugate gradient method is highly sensitive to rounding errors.

Instead of solving $A\boldsymbol{x} = \boldsymbol{b}$ directly, consider to solve$\tilde{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}$, where $\tilde{A} = C^{-1}A(C^{-1})^t$, $\tilde{\boldsymbol{x}} = C^t\boldsymbol{x}$ and $\tilde{\boldsymbol{b}} = C^{-1}\boldsymbol{b}$**.**

Here C is a nonsingular matrix.

To see $A\boldsymbol{x} = \boldsymbol{b}$ and $\tilde{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}$ are equivalent,

$$\tilde{A}\tilde{\boldsymbol{x}} = C^{-1}A(C^{-1})^t C^t \boldsymbol{x} = C^{-1}A\boldsymbol{x} = C^{-1}\boldsymbol{b}.$$

Preconditioned Conjugate Gradient Method

Since $\tilde{\boldsymbol{x}}^{(k)} = C^t\boldsymbol{x}^{(k)}$ and $\tilde{\boldsymbol{r}}^{(k)} = \tilde{\boldsymbol{b}} - \tilde{A}\tilde{\boldsymbol{x}}^{(k)} = C^{-1}\boldsymbol{r}^{(k)}$

Let $\tilde{\boldsymbol{v}}^{(k)} = C^t\boldsymbol{v}^{(k)}$ and $\boldsymbol{w}^{(k)} = C^{-1}\boldsymbol{r}^{(k)}$

A) Start with an arbitrary initial guess $\boldsymbol{x}^{(0)}$ to the solution $\boldsymbol{x}^*$ to $A\boldsymbol{x} = \boldsymbol{b}$

Set

$$\boldsymbol{r}^{(0)} = \boldsymbol{b} - A\boldsymbol{x}^{(0)}$$

$$\boldsymbol{w}^{(0)} = C^{-1}\boldsymbol{r}^{(0)}$$

$$\boldsymbol{v}^{(1)} = \boldsymbol{r}^{(0)}$$

B) for $k = 1,2,3, \ldots$

$$\tilde{t}_k = \frac{< \boldsymbol{w}^{(k-1)}, \boldsymbol{w}^{(k-1)} >}{< \boldsymbol{v}^{(k)}, A\boldsymbol{v}^{(k)} >}$$

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + \tilde{t}_k\boldsymbol{v}^{(k)}$$

$$\boldsymbol{r}^{(k)} = \boldsymbol{r}^{(k-1)} - \tilde{t}_k A\boldsymbol{v}^{(k)}$$

$$\tilde{s}_k = \frac{< \boldsymbol{w}^{(k)}, \boldsymbol{w}^{(k)} >}{< \boldsymbol{w}^{(k-1)}, \boldsymbol{w}^{(k-1)} >}$$

$$\boldsymbol{v}^{(k+1)} = C^{-1}\boldsymbol{w}^{(k)} + \tilde{s}_k\boldsymbol{v}^{(k)}$$