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

$\vec{g}_k = \nabla f(\vec{x}_k)$

$$f(\vec{x}^k - \alpha_k \vec{g}^k) = \frac{1}{2}(\vec{x}^k - \alpha^k \vec{g}^k)^T A (\vec{x}^k - \alpha^k \vec{g}^k) - (\vec{x}^k - \alpha^k \vec{g}^k)^T \vec{b} + c$$

$$= \frac{1}{2}(\vec{x}_k^T - \alpha_k \vec{g}_k^T)(A\vec{x}_k - \alpha_k A\vec{g}_k) - (\vec{x}_k^T - \alpha_k \vec{g}_k^T)\vec{b} + c$$

$$= \frac{\vec{x}_k^T A \vec{x}_k}{2} - \frac{\alpha_k \vec{x}_k^T A \vec{g}_k}{2} - \frac{\alpha_k \vec{g}_k^T A \vec{x}_k}{2} + \frac{\alpha_k^2 \vec{g}_k^T A \vec{g}_k}{2} - \vec{x}_k^T \vec{b} + \alpha_k \vec{g}_k^T \vec{b} + c$$

$$\frac{\partial f(\vec{x}_k - \alpha_k \vec{g}_k)}{\partial \alpha_k} = -\frac{\vec{x}_k^T A \vec{g}_k}{2} - \frac{\vec{g}_k^T A \vec{x}_k}{2} + \alpha_k \vec{g}_k^T A \vec{g}_k + \vec{g}_k^T \vec{b} = 0$$

$$\alpha_k \vec{g}_k^T A \vec{g}_k = \frac{\vec{x}_k^T A \vec{g}_k}{2} + \frac{\vec{g}_k^T A \vec{x}_k}{2} - \vec{g}_k^T \vec{b}$$

$$= \frac{\vec{g}_k^T A^T \vec{x}_k}{2} + \frac{\vec{g}_k^T A \vec{x}_k}{2} - \vec{g}_k^T \vec{b}$$

$$= \vec{g}_k^T \left(\frac{A^T \vec{x}_k}{2} + \frac{A\vec{x}_k}{2} - \vec{b} \right)$$

$$\alpha_k = \frac{\vec{g}_k^T \left(\frac{A^T \vec{x}_k}{2} + \frac{A\vec{x}_k}{2} - \vec{b} \right)}{\vec{g}_k^T A \vec{g}_k}$$

$$\boxed{\alpha_k = \frac{\vec{g}_k^T \vec{g}_k}{\vec{g}_k^T A \vec{g}_k}}$$

Scanned by CamScanner

① ⓑ $\nabla f(\vec{x}) = A\vec{x} - \vec{b} = \dfrac{\partial f(\vec{x})}{\partial \vec{x}}$

$\vec{b}^T \vec{x}$ $\quad \vec{b}^T \partial \vec{x}$

$(\partial \vec{x})^T \vec{b}$

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

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

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

$= (\partial\vec{x})^T\left(\frac{1}{2}A\vec{x} + \frac{1}{2}A^T\vec{x} - \vec{b}\right)$

$\therefore$ by eqn 2.33 $\quad \nabla f(\vec{x}) = \vec{g}(\vec{x}) = \frac{1}{2}A\vec{x} + \frac{1}{2}A^T\vec{x} - \vec{b}$

① ⓐ show $\frac{1}{2}\|\vec{x}-A^{+}\vec{b}\|_A^2 + c - \frac{1}{2}\|\vec{b}\|_{A^{+}}^2 = \frac{1}{2}\vec{x}^T A\vec{x} - \vec{x}^T\vec{b} + c$

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

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

According to Defition 5, eqn 1.76, any matrix used for the weighted Euclidean Norm must be positive definite and symmetric. This means A^{+} is both square, full rank, and symmetric meaning

$$A^{+^T} = A^{+} = A^{-1}$$

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

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

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

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

(d) $f(\vec{x}_{k+1}) \le f(\vec{x}_k) + \langle \nabla f(\vec{x}_k), \vec{x}_{k+1} - \vec{x}_k \rangle + \frac{L}{2} \| \vec{x}_{k+1} - \vec{x}_k \|_2$

Lipschitz Continuous Gradient

$$f(\vec{x}) \le f(\vec{y}) + \langle \nabla f(\vec{y}), \vec{x} - \vec{y} \rangle + \frac{L}{2} \| \vec{x} - \vec{y} \|_2^2$$

$$f(\vec{x}_{k+1}) \le f(\vec{x}_k) + \langle \nabla f(\vec{x}_k), \vec{x}_{k-1} - \vec{x}_k \rangle + \frac{L}{2} \| \vec{x}_{k+1} - \vec{x}_k \|_2$$

$$= f(\vec{x}_k) - a_k \| \nabla f(\vec{x}^k) \| + \frac{a_k^2 L}{2} \| \nabla f(\vec{x}_k) \|_2^2$$

$$= f(\vec{x}_k) - a_k \left(1 - \frac{a_k L}{2}\right) \| \nabla f(\vec{x}^k) \|_2^2$$

$$\boxed{a_k \le \frac{2}{L}}$$

$$\vec{x}_{k+1} = \vec{x}_k - a_k \nabla f(\vec{x}_k)$$

$$a_k \nabla f(\vec{x}_k) = \vec{x}_k - \vec{x}_{k+1}$$