Theorem 1. *Let* $\mathbf{f}(\mathbf{x})$ *be a strongly convex function on a set* S *with convexity parameter* μ*, then the term* $\|\mathbf{x}-\mathbf{x}^*\|$*, which is the distance between any point and optimal point in the set* S *is upper bounded by* $\frac{\mathbf{2}\|\nabla\mathbf{f}\|_\mathbf{2}}{\mu}$.

Proof. $\mathbf{f}(\mathbf{x})$ is a strongly convex function implies (1) for any $x, y \in S$

$$\mathbf{f}(\mathbf{y}) \geq \mathbf{f}(\mathbf{x}) + \nabla^\intercal \mathbf{f}(\mathbf{x})(\mathbf{y}-\mathbf{x}) + \frac{\mu\|\mathbf{x}-\mathbf{y}\|^\mathbf{2}}{2} \tag{1}$$

The righthand side is a convex quadratic function of y (for a fixed x). Setting the gradient with respect to y equal to zero, we find that $\tilde{\mathbf{y}} = \mathbf{x} - \frac{\nabla\mathbf{f}(\mathbf{x})}{\mu}$ minimizes the RHS of (1). Therefore, we have

$$\begin{aligned} \mathbf{f}(\mathbf{y}) &\geq \mathbf{f}(\mathbf{x}) + \langle\nabla\mathbf{f}(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \frac{\mu}{2}\|\mathbf{y}-\mathbf{x}^2\| \\ &\geq \mathbf{f}(\mathbf{x}) + \langle\nabla\mathbf{f}(\mathbf{x}), \tilde{\mathbf{y}}-\mathbf{x}\rangle + \frac{\mu}{2}\|\tilde{\mathbf{y}}-\mathbf{x}^2\| \\ &= \mathbf{f}(\mathbf{x}) - \frac{\|\tilde{y}-x\|^2}{2\mu} \end{aligned} \tag{2}$$

Since this holds for any $y \in S$, $\mathbf{f}^* \geq \mathbf{f}(\mathbf{x}) - \frac{1}{2\mu}\|\nabla\mathbf{f}(\mathbf{x})\|^2$

Here $\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \mathbf{f}(\mathbf{x})$ Similarly, one can derive a bound on $\|\mathbf{x}-\mathbf{x}^*\|$, the distance between $\mathbf{x}$ and any optimal point $\mathbf{x}^*$ in terms of $\|\nabla\mathbf{f}(\mathbf{x})\|$. First apply (1) with $\mathbf{y} = \mathbf{x}^*$ to obtain:

$$\begin{aligned} \mathbf{f}^* = \mathbf{f}(\mathbf{x}^*) &\geq \mathbf{f}(\mathbf{x}) + \langle\nabla\mathbf{f}(\mathbf{x}), \mathbf{x}^*-\mathbf{x}\rangle + \frac{\mathbf{m}}{\mathbf{2}}\|\mathbf{x}-\mathbf{x}^*\|^\mathbf{2} \\ &\geq \mathbf{f}(\mathbf{x}) - \|\nabla\mathbf{f}(\mathbf{x})\|\|\mathbf{x}^*-\mathbf{x}\| + \frac{\mu}{\mathbf{2}}\|\mathbf{x}-\mathbf{x}^*\|^\mathbf{2} \end{aligned} \tag{3}$$

Here we use Cauchy-Schwarz inequality in the second part of (3). Since $\mathbf{f}^* \leq \mathbf{f}(\mathbf{x})$, one must have,

$$-\|\nabla\mathbf{f}(\mathbf{x})\|\|\mathbf{x}^*-\mathbf{x}\| + \frac{\mu}{\mathbf{2}}\|\mathbf{x}-\mathbf{x}^*\|^\mathbf{2} \leq \mathbf{0} \tag{4}$$

$$\|\mathbf{x}-\mathbf{x}^*\| \leq \frac{\mathbf{2}\|\nabla\mathbf{f}(\mathbf{x})\|}{\mu} \quad \square$$

$\square$