Consider the following algorithm, given $\phi$.

Let $\lambda = \phi(0)$.

Inserting the given bounds on $\phi''$ gives the desired result.

[Incidentally, if we are dealing with this function class, better complexity bounds are possible: basically, we can assure linear convergence, so the bound is $O(\ln(1/\epsilon))$, not $O(\sqrt{1/\epsilon})$.]
3. a) Suppose that $g$ is a convex function on $\mathbb{R}^n$. Show that, for every $z \in \mathbb{R}^n$ and $L > 0$, the problem

$$(P(z)) \quad \min g(x) + \frac{L}{2} \|x - z\|^2$$

has an optimal solution.

b) Assume that, for any $z \in \mathbb{R}^n$, you can solve $(P(z))$ efficiently. Show how you can solve the problem

$$\min f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{L}{2} \|x - x_k\|^2 + g(x)$$

efficiently. What is the solution if $g \equiv 0$?

c) Find the solution to $(P(z))$ explicitly if $g(x) := \|x\|_1$.

a) We are assuming that $g$ is a real-valued function on $\mathbb{R}^n$, so that the interior of its domain is $\mathbb{R}^n$. Thus it is continuous, and has a subgradient at every point. Let $h$ be a subgradient at $z$. Then for all $x \in \mathbb{R}^n$, $g(x) + (L/2)\|x - z\|^2 \geq g(z) + h^T (x - z) + (L/2)\|x - z\|^2$. This is greater than $g(z)$ for $\|x - z\| > (2/L)\|h\|$. Thus minimizing this function over $\mathbb{R}^n$ is equivalent to minimizing it over the ball centered at $z$ with radius $(2/L)\|h\|$. Since this set is compact and the function is continuous, it has a minimizer on this set, and hence on $\mathbb{R}^n$.

b) Note that $(L/2)\|x - (x_k - (1/L)\nabla f(x_k))\|^2 = (L/2)\|x - x_k\|^2 + \nabla f(x_k)^T (x - x_k) + \|\nabla f(x_k)\|^2/(2L)$. Hence the given function is a constant plus $g(x) + (L/2)\|x - z\|^2$ for $z := x_k - (1/L)\nabla f(x_k)$, so minimizing it is equivalent to $(P(z))$ for this $z$. If $g \equiv 0$, then $(P(z))$ is solved by $x = z$, so the solution then is $x_k - (1/L)\nabla f(x_k)$.

c) If $g(x) = \|x\|_1$, then $g(x) + (L/2)\|x - z\|^2 = \sum_j [x^{(j)}] + (L/2)(z^{(j)} - x^{(j)})^2]$, so we can minimize it by minimizing each term. But each term is a strictly convex function, so it has a unique minimizer at a point where 0 is a subgradient. By considering the three cases for the sign of $x^{(j)}$, we find the optimal $x^{(j)}$ is $z^{(j)} - 1/L$ if this is positive, $z^{(j)} + 1/L$ if this is negative, and 0 otherwise. (This can also be written $\text{mid}(z^{(j)} - 1/L, 0, z^{(j)} + 1/L)$, where $\text{mid}$ gives the median of its three arguments, a very useful function.)