Today

Gradient Descent.
Minimizing a convex function.

$$\min f(x)$$

One dimension: go left or right with a magnitude.

What is $\alpha$?

If function $100000x^2$.

Should $\alpha$ be small or big?

small!

What do you want to do?

Get close!

$$f(x) - f(x^*) \leq \varepsilon.$$
Minimizing a convex function.

\[ \min f(x) \]

\[ f(y) \geq f(x) + (\nabla f(x))(y - x) \]

One dimension: go left or right with a magnitude.

What is \( \alpha \)?

If function \( 100000x^2 \).

Should \( \alpha \) be small or big?

small!

What do you want to do?

Get close!

\[ f(x) - f(x^*) \leq \varepsilon \]

Depends on function.
Minimizing a convex function.

\[
\begin{align*}
\min f(x) \\
f(y) & \geq f(x) + (\nabla f(x))(y - x) \\
x_{k+1} & = x_k - \alpha \nabla f(x).
\end{align*}
\]
Minimizing a convex function.

\[\min f(x)\]

\[f(y) \geq f(x) + (\nabla f(x))(y - x)\]

\[x_{k+1} = x_k - \alpha \nabla f(x).\]

One dimension: go left or right with a magnitude.
Minimizing a convex function.

\[
\min f(x)
\]

\[
f(y) \geq f(x) + (\nabla f(x))(y - x)
\]

\[
x_{k+1} = x_k - \alpha \nabla f(x).
\]

One dimension: go left or right with a magnitude.

What is \( \alpha \)?
Minimizing a convex function.

\[ \min f(x) \]

\[ f(y) \geq f(x) + (\nabla f(x))(y - x) \]

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

One dimension: go left or right with a magnitude.
What is \( \alpha \)?
If function 100000x^2.
Minimizing a convex function.

\[
\min f(x)
\]

\[
f(y) \geq f(x) + (\nabla f(x))(y - x)
\]

\[
x_{k+1} = x_k - \alpha \nabla f(x).
\]

One dimension: go left or right with a magnitude.

What is \(\alpha\)?

If function 100000\(x^2\).

Should \(\alpha\) be small or big?
Minimizing a convex function.

\[
\min f(x)
\]

\[
f(y) \geq f(x) + (\nabla f(x))(y - x)
\]

\[
x_{k+1} = x_k - \alpha \nabla f(x).
\]

One dimension: go left or right with a magnitude.

What is \(\alpha\)?

If function 100000\(x^2\).

Should \(\alpha\) be small or big? small!
Minimizing a convex function.

\[ \min f(x) \]

\[ f(y) \geq f(x) + (\nabla f(x))(y - x) \]

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

One dimension: go left or right with a magnitude.

What is \( \alpha \)?

If function 100000\(x^2\).

Should \( \alpha \) be small or big? small!

What do you want to do?
Minimizing a convex function.

\[ \min f(x) \]

\[ f(y) \geq f(x) + (\nabla f(x))(y - x) \]

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

One dimension: go left or right with a magnitude.

What is \( \alpha \)?

If function \( 100000x^2 \).

Should \( \alpha \) be small or big? small!

What do you want to do?

Get close!
Minimizing a convex function.

\[
\text{min } f(x)
\]

\[
f(y) \geq f(x) + (\nabla f(x))(y - x)
\]

\[
x_{k+1} = x_k - \alpha \nabla f(x).
\]

One dimension: go left or right with a magnitude.

What is \(\alpha\)?

If function 100000\(x^2\).

Should \(\alpha\) be small or big? small!

What do you want to do?

Get close! \(f(x) - f(x^*) \leq \varepsilon\).
Minimizing a convex function.

\[
\min f(x)
\]

\[
f(y) \geq f(x) + (\nabla f(x))(y - x)
\]

\[
x_{k+1} = x_k - \alpha \nabla f(x).
\]

One dimension: go left or right with a magnitude.

What is \( \alpha \)?

If function 1000000\(x^2 \).

Should \( \alpha \) be small or big? small!

What do you want to do?

Get close! \( f(x) - f(x^*) \leq \varepsilon \).

Depends on function.
Assume $\|\nabla f(x) - \nabla f(y)\| \leq L\|(x - y)\|$ for any $x, y$. 
Assume $\|\nabla f(x) - \nabla f(y)\| \leq L\|(x - y)\|$ for any $x, y$.

For $10000x^2$, what is $L$?
Assume $\|\nabla f(x) - \nabla f(y)\| \leq L \|(x - y)\|$ for any $x, y$.

For $10000x^2$, what is $L$?

For $10x^2 + 1000000x$, what is $L$?
Assume $\|\nabla f(x) - \nabla f(y)\| \leq L \|(x - y)\|$ for any $x, y$.

For $10000x^2$, what is $L$?

For $10x^2 + 1000000x$, what is $L$?

Intuitively, a bound on the second derivative.
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

\[ f(x_k) - f(x^*) \leq \frac{2L \|x_0 - x^*\|}{k}. \]
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

\[ f(x_k) - f(x^*) \leq \frac{2L\|x_0 - x^*\|}{k}. \]

Choose \( \alpha = \frac{1}{L} \).
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

\[ f(x_k) - f(x^*) \leq \frac{2L\|x_0 - x^*\|}{k}. \]

Choose \( \alpha = \frac{1}{L} \).

\[ x_{t+1} = x_t - \alpha \nabla f(x_t) \]
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

\[ f(x_k) - f(x^*) \leq \frac{2L\|x_0 - x^*\|}{k}. \]

Choose \( \alpha = \frac{1}{L} \).

\[ x_{t+1} = x_t - \alpha \nabla f(x_t) \]

Idea: \( \nabla f(x_t) \).
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

\[ f(x_k) - f(x^*) \leq \frac{2L \| x_0 - x^* \|}{k}. \]

Choose \( \alpha = \frac{1}{L} \).

\[ x_{t+1} = x_t - \alpha \nabla f(x_t) \]

Idea: \( \nabla f(x_t) \).

\[ |\nabla f(x_{t+1}) - \nabla f(x_t)| \leq L \| x_{t+1} - x_t \|^2 \]
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

\[ f(x_k) - f(x^*) \leq \frac{2L\|x_0 - x^*\|}{k}. \]

Choose \( \alpha = \frac{1}{L} \).

\[ x_{t+1} = x_t - \alpha \nabla f(x_t) \]

Idea: \( \nabla f(x_t) \).

\[ |\nabla f(x_{t+1}) - \nabla f(x_t)| \leq L\|x_{t+1} - x_t\|^2 \leq L(\frac{1}{2L})\nabla f(x_{t+1}) \]
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

\[ f(x_k) - f(x^*) \leq \frac{2L \|x_0 - x^*\|}{k}. \]

Choose \( \alpha = \frac{1}{L} \).

\[ x_{t+1} = x_t - \alpha \nabla f(x_t) \]

Idea: \( \nabla f(x_t) \).

\[ |\nabla f(x_{t+1}) - \nabla f(x_t)| \leq L \|x_{t+1} - x_t\|^2 \leq L(\frac{1}{2L})\nabla f(x_{t+1}) \leq \frac{1}{2} \nabla f(x_{t+1}). \]
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

\[ f(x_k) - f(x^*) \leq \frac{2L \|x_0 - x^*\|}{k}. \]

Choose \( \alpha = \frac{1}{L} \).

\[ x_{t+1} = x_t - \alpha \nabla f(x_t) \]

Idea: \( \nabla f(x_t) \).

\[ |\nabla f(x_{t+1}) - \nabla f(x_t)| \leq L \|x_{t+1} - x_t\|^2 \leq L \left( \frac{1}{2L} \right) \nabla f(x_{t+1}) \leq \frac{1}{2} \nabla f(x_{t+1}). \]

For \( \nabla f(y) > \nabla f(x_t)/2 \), for all \( y \in [x_t, x_{t+1}] \).
Convergence.

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

\[ f(x_k) - f(x^*) \leq \frac{2L\|x_0 - x^*\|}{k}. \]

Choose \( \alpha = \frac{1}{L} \).

\[ x_{t+1} = x_t - \alpha \nabla f(x_t) \]

Idea: \( \nabla f(x_t) \).

\[ |\nabla f(x_{t+1}) - \nabla f(x_t)| \leq L\|x_{t+1} - x_t\|^2 \leq L(\frac{1}{2L})\nabla f(x_{t+1}) \leq \frac{1}{2}\nabla f(x_{t+1}). \]

For \( \nabla f(y) > \nabla f(x_t)/2 \), for all \( y \in [x_t, x_{t+1}] \).

Thus, the function decreases by \( \|\nabla f(x_t)\|^2/4L \).
Convergence

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]
Convergence

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

Down by \( \| \nabla f(x_t) \|^2 / 4L \): \( f(x_{t+1}) \leq f(x_t) - \| \nabla f(x_t) \|^2 / 4L \).
Convergence

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

Down by \( \| \nabla f(x_t) \|^2 / 4L \): \( f(x_{t+1}) \leq f(x_t) - \| \nabla f(x_t) \|^2 / 4L. \)

Convexity: \( f(y) \geq f(x) + (\nabla f(x))(y - x) \) or \( f(y) \geq f(x) - (\nabla f(x))(x - y) \)
Convergence

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

Down by \( \| \nabla f(x_t) \|^2 / 4L \): \( f(x_{t+1}) \leq f(x_t) - \| \nabla f(x_t) \|^2 / 4L \).

Convexity: \( f(y) \geq f(x) + (\nabla f(x))(y - x) \) or \( f(y) \geq f(x) - (\nabla f(x))(x - y) \)

\[ \| \nabla f(x) \| \geq \frac{f(x) - f(x^*)}{\| x - x^* \|} \]
Convergence

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

Down by \( \| \nabla f(x_t) \|^2 / 4L \): \( f(x_{t+1}) \leq f(x_t) - \| \nabla f(x_t) \|^2 / 4L. \)

Convexity: \( f(y) \geq f(x) + (\nabla f(x))(y - x) \) or \( f(y) \geq f(x) - (\nabla f(x))(x - y) \)

\[ \| \nabla f(x) \| \geq \frac{f(x) - f(x^*)}{\| x - x^* \|} \quad \implies \quad f(x_{t+1}) \leq f(x_t) - \frac{1}{4L} \left( \frac{f(x_t) - f(x^*)}{\| x_t - x^* \|} \right)^2 \]
Convergence

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

Down by \( \|\nabla f(x_t)\|^2/4L: f(x_{t+1}) \leq f(x_t) - \|\nabla f(x_t)\|^2/4L. \)

Convexity: \( f(y) \geq f(x) + (\nabla f(x))(y - x) \) or \( f(y) \geq f(x) - (\nabla f(x))(x - y) \)

\[ \|\nabla f(x)\| \geq \frac{f(x) - f(x^*)}{\|x - x^*\|} \implies f(x_{t+1}) \leq f(x_t) - \frac{1}{4L} \left( \frac{f(x_t) - f(x^*)}{\|x_t - x^*\|} \right)^2 \]

Let \( |f(x_t) - f(x^*)| \in [\Delta, 2\Delta] \) for \( t \in [0, k] \)
Convergence

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

Down by \( \| \nabla f(x_t) \|^2 / 4L: f(x_{t+1}) \leq f(x_t) - \| \nabla f(x_t) \|^2 / 4L. \)

Convexity: \( f(y) \geq f(x) + (\nabla f(x))(y - x) \) or \( f(y) \geq f(x) - (\nabla f(x))(x - y) \)

\[ \| \nabla f(x) \| \geq \frac{f(x) - f(x^*)}{\| x - x^* \|} \implies f(x_{t+1}) \leq f(x_t) - \frac{1}{4L} \left( \frac{f(x_t) - f(x^*)}{\| x_t - x^* \|} \right)^2 \]

Let \(|f(x_t) - f(x^*)| \in [\Delta, 2\Delta] \) for \( t \in [0, k] \)

\[ f(x_k) \leq f(x_0) - \frac{k}{4L} \left( \frac{\Delta}{\| x_t - x^* \|} \right)^2 . \]
\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

Down by \( \|\nabla f(x_t)\|^2/4L \): \( f(x_{t+1}) \leq f(x_t) - \|\nabla f(x_t)\|^2/4L \).

Convexity: \( f(y) \geq f(x) + (\nabla f(x))(y - x) \) or \( f(y) \geq f(x) - (\nabla f(x))(x - y) \)

\[ \|\nabla f(x)\| \geq \frac{f(x) - f(x^*)}{\|x - x^*\|} \implies f(x_{t+1}) \leq f(x_t) - \frac{1}{4L} \left( \frac{f(x_t) - f(x^*)}{\|x_t - x^*\|} \right)^2 \]

Let \( |f(x_t) - f(x^*)| \in [\Delta, 2\Delta] \) for \( t \in [0, k] \)

\[ f(x_k) \leq f(x_0) - \frac{k}{4L} \left( \frac{\Delta}{\|x_t - x^*\|} \right)^2. \]

Choose \( k = \left( \frac{4L\|x_t - x^*\|^2}{\Delta} \right) \), makes contradiction: below 0
Convergence

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

Down by \( \| \nabla f(x_t) \|^2 / 4L \): \( f(x_{t+1}) \leq f(x_t) - \| \nabla f(x_t) \|^2 / 4L \).

Convexity: \( f(y) \geq f(x) + (\nabla f(x))(y - x) \) or \( f(y) \geq f(x) - (\nabla f(x))(x - y) \)

\[ \| \nabla f(x) \| \geq \frac{f(x) - f(x^*)}{\| x - x^* \|} \implies f(x_{t+1}) \leq f(x_t) - \frac{1}{4L} \left( \frac{f(x_t) - f(x^*)}{\| x_t - x^* \|} \right)^2 \]

Let \( |f(x_t) - f(x^*)| \in [\Delta, 2\Delta] \) for \( t \in [0, k] \)

\[ f(x_k) \leq f(x_0) - \frac{k}{4L} \left( \frac{\Delta}{\| x_t - x^* \|} \right)^2. \]

Choose \( k = \left( \frac{4L \| x_t - x^* \|^2}{\Delta} \right) \), makes contradiction: below 0

Can’t be in range for whole time.

\[ \implies \text{Error halves in } k \text{ iterations}. \]
Convergence

\[ x_{k+1} = x_k - \alpha \nabla f(x). \]

Down by \( \| \nabla f(x_t) \|^2 / 4L \): \( f(x_{t+1}) \leq f(x_t) - \| \nabla f(x_t) \|^2 / 4L \).

Convexity: \( f(y) \geq f(x) + (\nabla f(x))(y - x) \) or \( f(y) \geq f(x) - (\nabla f(x))(x - y) \)

\[ \| \nabla f(x) \| \geq \frac{f(x) - f(x^*)}{\| x - x^* \|} \quad \implies \quad f(x_{t+1}) \leq f(x_t) - \frac{1}{4L} \left( \frac{f(x_t) - f(x^*)}{\| x_t - x^* \|} \right)^2 \]

Let \( |f(x_t) - f(x^*)| \in [\Delta, 2\Delta] \) for \( t \in [0, k] \)

\[ f(x_k) \leq f(x_0) - \frac{k}{4L} \left( \frac{\Delta}{\| x_t - x^* \|} \right)^2. \]

Choose \( k = \left( \frac{4L\| x_t - x^* \|^2}{\Delta} \right) \), makes contradiction: below 0

Can’t be in range for whole time.

\( \implies \) Error halves in \( k \) iterations.

Geometric in \( \Delta \), so for arbitrary \( \varepsilon \):

\[ k = O\left( \frac{L\| x_t - x^* \|^2}{\varepsilon} \right). \]
Another Proof

∇f Lipchitz with constant $L$ $\implies$
Another Proof

\( \nabla f \) Lipchitz with constant \( L \) \( \implies \)

\[ f(y) \leq f(x) + \nabla f(x)(y - x) + \frac{L}{2}\|y - x\|^2 \text{ all } x, y. \]
Another Proof

$\nabla f$ Lipchitz with constant $L \implies$

$$f(y) \leq f(x) + \nabla f(x)(y - x) + \frac{L}{2} \|y - x\|^2 \text{ all } x, y.$$  
The last term comes from integrating $L\|y - x\|$ along $y - x$.  

Another Proof

∇f Lipchitz with constant $L$ $\implies$

$$f(y) \leq f(x) + \nabla f(x)(y - x) + \frac{L}{2} \|y - x\|^2 \text{ all } x, y.$$  

The last term comes from integrating $L\|y - x\|$ along $y - x$.

Plugging in $y = x_{t+1} = x_t - \alpha \nabla f(x)$. 
Another Proof

\[ \nabla f \text{ Lipchitz with constant } L \implies \]
\[ f(y) \leq f(x) + \nabla f(x)(y - x) + \frac{L}{2}\|y - x\|^2 \text{ all } x, y. \]
The last term comes from integrating \( L\|y - x\| \) along \( y - x \).

Plugging in \( y = x_{t+1} = x_t - \alpha \nabla f(x) \).

\[ f(x_{t+1}) \leq f(x_t) - \left(1 - \frac{L\alpha}{2}\right) \alpha \|\nabla f(x)\|^2. \]
Another Proof

$\nabla f$ Lipchitz with constant $L \implies$

$$f(y) \leq f(x) + \nabla f(x)(y - x) + \frac{L}{2}\|y - x\|^2 \text{ all } x, y.$$ 

The last term comes from integrating $L\|y - x\|$ along $y - x$.

Plugging in $y = x_{t+1} = x_t - \alpha \nabla f(x)$.

$$f(x_{t+1}) \leq f(x_t) - \left(1 - \frac{L\alpha}{2}\right)\alpha\|\nabla f(x)\|^2.$$ 

For $\alpha = 1/L$, using convexity: $f(x^*) + \nabla f(x_t)(x_t - x^*) \geq f(x_t)$
Another Proof

\[ \nabla f \text{ Lipchitz with constant } L \implies \]
\[ f(y) \leq f(x) + \nabla f(x)(y - x) + \frac{L}{2}\|y - x\|^2 \text{ all } x, y. \]

The last term comes from integrating \( L\|y - x\| \) along \( y - x \).

Plugging in \( y = x_{t+1} = x_t - \alpha \nabla f(x) \).

\[ f(x_{t+1}) \leq f(x_t) - \left(1 - \frac{L\alpha}{2}\right) \alpha \|\nabla f(x)\|^2. \]

For \( \alpha = 1/L \), using convexity: \( f(x^*) + \nabla f(x_t)(x_t - x^*) \geq f(x_t) \)

\[ f(x_{t+1}) \leq f(x^*) + \nabla f(x_t)(x_t - x^*) - \frac{1}{2L} \|\nabla f(x_t)\|^2 \]
Another Proof

$\nabla f$ Lipschitz with constant $L \implies$

$$f(y) \leq f(x) + \nabla f(x)(y - x) + \frac{L}{2}\|y - x\|^2 \text{ all } x, y.$$  

The last term comes from integrating $L\|y - x\|$ along $y - x$.

Plugging in $y = x_{t+1} = x_t - \alpha \nabla f(x)$.

$$f(x_{t+1}) \leq f(x_t) - \left(1 - \frac{L\alpha}{2}\right) \alpha\|\nabla f(x)\|^2.$$  

For $\alpha = 1/L$, using convexity: $f(x^*) + \nabla f(x_t)(x_t - x^*) \geq f(x_t)$

$$f(x_{t+1}) \leq f(x^*) + \nabla f(x_t)(x_t - x^*) - \frac{1}{2L}\|\nabla f(x_t)\|^2$$

$$\leq f(x^*) + \frac{1}{\alpha}(x_t - x^*)(x - x^*) - \frac{1}{2L}\|\nabla f(x_t)\|^2$$
Another Proof

\nabla f \text{ Lipchitz with constant } L \implies \nabla f \text{ Lipchitz with constant } L \implies

\begin{align*}
f(y) & \leq f(x) + \nabla f(x)(y - x) + \frac{L}{2} \|y - x\|^2 \text{ all } x, y. \\
\text{The last term comes from integrating } L\|y - x\| \text{ along } y - x. \\
\text{Plugging in } y = x_{t+1} = x_t - \alpha \nabla f(x). \\
f(x_{t+1}) & \leq f(x_t) - \left(1 - \frac{L \alpha}{2}\right) \alpha \|\nabla f(x)\|^2.
\end{align*}

For \( \alpha = 1/L \), using convexity: \( f(x^*) + \nabla f(x_t)(x_t - x^*) \geq f(x_t) \)

\begin{align*}
f(x_{t+1}) & \leq f(x^*) + \nabla f(x_t)(x_t - x^*) - \frac{1}{2L} \|\nabla f(x_t)\|^2 \\
& \leq f(x^*) + \frac{1}{\alpha} (x_t - x^*)(x - x^*) - \frac{1}{2L} \|\nabla f(x_t)\|^2 \\
& = f(x^*) + \frac{1}{2\alpha} (\|x_t - x^*\|^2 - \|x_{t+1} - x^*\|^2) \\
& = f(x^*) + \frac{L}{2} (\|x_t - x^*\|^2 - \|x_{t+1} - x^*\|^2)
\end{align*}
Convergence.

Sum over iterations.
Convergence.

Sum over iterations.

\[ \sum_{t=1}^{k} (f(x_t) - f(x^*)) \leq \frac{L}{2} (\|x_0 - x^*\|^2 - \|x_k - x^*\|^2) \]
Convergence.

Sum over iterations.

$$\sum_{t=1}^{k} (f(x_t) - f(x^*)) \leq \frac{L}{2} (\| x_0 - x^* \|^2 - \| x_k - x^* \|^2) \leq \frac{L}{2} (\| x_0 - x^* \|^2).$$
Convergence.

Sum over iterations.

$$\sum_{t=1}^{k} (f(x_t) - f(x^*)) \leq \frac{L}{2} (\|x_0 - x^*\|^2 - \|x_k - x^*\|^2)$$

$$\leq \frac{L}{2} (\|x_0 - x^*\|^2).$$

Since $f(x_t)$ is nonincreasing.
Convergence.

Sum over iterations.

\[ \sum_{t=1}^{k} (f(x_t) - f(x^*)) \leq \frac{L}{2} (\|x_0 - x^*\|^2 - \|x_k - x^*\|^2) \leq \frac{L}{2} (\|x_0 - x^*\|^2). \]

Since \( f(x_t) \) is nonincreasing.

\[ f(x_k) - f(x^*) \leq \frac{1}{k} \sum_{t=1}^{k} (f(x_t) - f(x^*)) \leq \frac{L\|x_0 - x^*\|^2}{2k}. \]
**Strong Convexity**

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$. 
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$.

$f(x) = 5x$?
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$.

$f(x) = 5x^2$?
Strong Convexity

Strong (strictly) Convexity: \( f(x) - m\|x\|^2 \) is convex for some \( m > 0 \).

\[ f(x) = 5x? \quad f(x) = 5x^2? \quad f(x) = 5x^3? \]
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$.

$f(x) = 5x$? $f(x) = 5x^2$? $f(x) = 5x^3$?

$f(x, y) = x^2 + y^2$?
Strong Convexity

Strong (strictly) Convexity: \( f(x) - m\|x\|^2 \) is convex for some \( m > 0 \).

\[
\begin{align*}
  f(x) &= 5x？ f(x) = 5x^2？ f(x) = 5x^3？ \\
  f(x, y) &= x^2 + y^2？ f(x, y) = 5x + 6y？
\end{align*}
\]
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$.

$f(x) = 5x$? $f(x) = 5x^2$? $f(x) = 5x^3$?

$f(x, y) = x^2 + y^2$? $f(x, y) = 5x + 6y$?

If $f(x)$ is twice differentiable.
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$.

- $f(x) = 5x$?
- $f(x) = 5x^2$?
- $f(x) = 5x^3$?

- $f(x, y) = x^2 + y^2$?
- $f(x, y) = 5x + 6y$?

If $f(x)$ is twice differentiable.

$$\nabla^2 f(x) \succeq mI$$ for all $x$. 

Strong Convexity

Strong (strictly) Convexity: \( f(x) - m\|x\|^2 \) is convex for some \( m > 0 \).

\[
f(x) = 5x? \quad f(x) = 5x^2? \quad f(x) = 5x^3? \\
f(x, y) = x^2 + y^2? \quad f(x, y) = 5x + 6y?
\]

If \( f(x) \) is twice differentiable.

\[
\nabla^2 f(x) \succeq ml \text{ for all } x.
\]

Hessian: \( \nabla^2 f(x) \) is matrix of \( \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \) evaluated at \( x \).
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$.

$f(x) = 5x$? $f(x) = 5x^2$? $f(x) = 5x^3$?

$f(x, y) = x^2 + y^2$? $f(x, y) = 5x + 6y$?

If $f(x)$ is twice differentiable.

$\nabla^2 f(x) \succeq ml$ for all $x$.

Hessian: $\nabla^2 f(x)$ is matrix of $\frac{\partial f(x)}{\partial x_i \partial x_j}$ evaluated at $x$.

Hessian for $f(x, y) = x^2 + y^2$.

\[
\begin{bmatrix}
2 & 0 \\
0 & 2
\end{bmatrix}
\]
Strong Convexity

Strong (strictly) Convexity: \( f(x) - m\|x\|^2 \) is convex for some \( m > 0 \).

\[
f(x) = 5x \? \quad f(x) = 5x^2 \? \quad f(x) = 5x^3 \?
\]

\[
f(x, y) = x^2 + y^2 \? \quad f(x, y) = 5x + 6y \?
\]

If \( f(x) \) is twice differentiable.

\[
\nabla^2 f(x) \succeq ml \text{ for all } x.
\]

Hessian: \( \nabla^2 f(x) \) is matrix of \( \frac{\partial f(x)}{\partial x_i \partial x_j} \) evaluated at \( x \).

Hessian for \( f(x, y) = x^2 + y^2 \).

\[
\begin{bmatrix}
2 & 0 \\
0 & 2
\end{bmatrix}
\]

Strictly convex with \( m = 2 \).
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$.

- $f(x) = 5x$?
- $f(x) = 5x^2$?
- $f(x) = 5x^3$?

- $f(x, y) = x^2 + y^2$?
- $f(x, y) = 5x + 6y$?

If $f(x)$ is twice differentiable.

- $\nabla^2 f(x) \succeq mI$ for all $x$.

Hessian: $\nabla^2 f(x)$ is matrix of $\frac{\partial f(x)}{\partial x_i \partial x_j}$ evaluated at $x$.

Hessian for $f(x, y) = x^2 + y^2$.

$$
\begin{bmatrix}
2 & 0 \\
0 & 2 \\
\end{bmatrix}
$$

Strictly convex with $m = 2$.

Hessian for $f(x, y) = x^2 + xy + y^2$.

$$
\begin{bmatrix}
2 & 1 \\
1 & 2 \\
\end{bmatrix}
$$

Strictly convex with $m = 1$. 
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$. 
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$. If $f(x)$ is twice differentiable.

Gradient descent: $x_{t+1} = x_t - \alpha \nabla f(x_t)$ with $\alpha = \frac{2m}{L}$ gets $f(x_k) - f(x^*) \leq \frac{c_k L^2}{2} \|x_0 - x^*\|^2$ by Lipschitz. $c_k = (1 - O(m/L))$. 
Strong Convexity

Strong (strictly) Convexity: \( f(x) - m\|x\|^2 \) is convex for some \( m > 0 \).

If \( f(x) \) is twice differentiable.

\[
\nabla^2 f(x) \succeq ml \quad \text{for all } x.
\]
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$.

If $f(x)$ is twice differentiable.

$\nabla^2 f(x) \succeq ml$ for all $x$.

Sharper lower bound than from convexity.
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$. If $f(x)$ is twice differentiable.

$$\nabla^2 f(x) \succeq ml \text{ for all } x.$$  

Sharper lower bound than from convexity.

$$f(y) \geq f(x) + \nabla f(x)(y-x) + \left(\frac{m}{2}\|y-x\|\right)^2 \text{ all } x, y.$$
Strong Convexity

Strong (strictly) Convexity: $f(x) - m\|x\|^2$ is convex for some $m > 0$. If $f(x)$ is twice differentiable.

$$\nabla^2 f(x) \succeq ml$$ for all $x$.

Sharper lower bound than from convexity.

$$f(y) \geq f(x) + \nabla f(x)(y - x) + \left(\frac{m}{2}\|y - x\|\right)^2$$ all $x, y$.

Gradient descent: $x_{t+1} = x_t - \alpha \nabla f(x)$ with $\alpha = \frac{2}{m+L}$ gets
Strong Convexity

Strong (strictly) Convexity: \( f(x) - m\|x\|^2 \) is convex for some \( m > 0 \).

If \( f(x) \) is twice differentiable.

\[ \nabla^2 f(x) \succeq mI \text{ for all } x. \]

Sharper lower bound than from convexity.

\[ f(y) \geq f(x) + \nabla f(x) (y - x) + \left( \frac{m}{2} \|y - x\| \right)^2 \text{ all } x, y. \]

Gradient descent: \( x_{t+1} = x_t - \alpha \nabla f(x) \) with \( \alpha = \frac{2}{m+L} \) gets

\[ f(x_k) - f(x^*) \leq c^k \frac{L}{2} \|x_0 - x^*\|^2 \text{ by Lipschitz.} \]
Strong Convexity

Strong (strictly) Convexity: \( f(x) - m\|x\|^2 \) is convex for some \( m > 0 \).

If \( f(x) \) is twice differentiable.

\[ \nabla^2 f(x) \succeq mI \text{ for all } x. \]

Sharper lower bound than from convexity.

\[ f(y) \geq f(x) + \nabla f(x)(y - x) + \left( \frac{m}{2} \|y - x\| \right)^2 \text{ all } x, y. \]

Gradient descent: \( x_{t+1} = x_t - \alpha \nabla f(x) \) with \( \alpha = \frac{2}{m + L} \) gets

\[ f(x_k) - f(x^*) \leq c^k \frac{L}{2} \|x_0 - x^*\|^2 \text{ by Lipschitz.} \]

\[ c = (1 - O(m/L)). \]
Convergence.

\[ x_{t+1} = x_t - \alpha \nabla f(x). \]
Convergence.

\[ x_{t+1} = x_t - \alpha \nabla f(x). \]
\[ \alpha = \frac{1}{2L}. \]
Convergence.

\[ x_{t+1} = x_t - \alpha \nabla f(x). \]
\[ \alpha = \frac{1}{2L}. \]

From before
Convergence.

\[ x_{t+1} = x_t - \alpha \nabla f(x). \]
\[ \alpha = \frac{1}{2L}. \]

From before

\[ \nabla (f(x)) \geq \frac{f(x) - f(x^*) + \frac{m}{2} \|x^* - x\|^2}{\|x - x^*\|} \geq \frac{m}{2} \|x^* - x\|. \]
\[ x_{t+1} = x_t - \alpha \nabla f(x). \]
\[ \alpha = \frac{1}{2L}. \]

From before
\[ \nabla(f(x)) \geq \frac{f(x) - f(x^*) + \frac{m}{2} \|x^* - x\|^2}{\|x - x^*\|} \geq \frac{m}{2} \|x^* - x\|. \]

Goes down by \( \frac{\alpha}{2} \|\nabla f(x_t)\|^2 \)
Convergence.

\[ x_{t+1} = x_t - \alpha \nabla f(x). \]
\[ \alpha = \frac{1}{2L}. \]

From before
\[ \nabla (f(x)) \geq \frac{f(x) - f(x^*) + \frac{m}{2} \| x^* - x \|^2}{\| x - x^* \|} \geq \frac{m}{2} \| x^* - x \|. \]

Goes down by \( \frac{\alpha}{2} \| \nabla f(x_t) \|^2 \)
\[ \alpha \frac{m}{2} \| x^* - x \|^2 \text{ in each step.} \]
Convergence.

\[ x_{t+1} = x_t - \alpha \nabla f(x). \]
\[ \alpha = \frac{1}{2L}. \]

From before
\[ \nabla (f(x)) \geq \frac{f(x) - f(x^*) + \frac{m}{2} \|x^* - x\|^2}{\|x - x^*\|} \geq \frac{m}{2} \|x^* - x\|. \]

Goes down by \( \frac{\alpha}{2} \|\nabla f(x_t)\|^2 \)
\[ \alpha \frac{m}{2} \|x^* - x\|^2 \text{ in each step.} \]

\( f(x) - f(x^*) \) is at most \( \frac{L}{2} \|x^* - x\|^2. \)

So decreases by \( (1 - \Theta(\frac{m^2}{L^2})) \) in each step.
Convergence.

\[ x_{t+1} = x_t - \alpha \nabla f(x). \]

\[ \alpha = \frac{1}{2L}. \]

From before

\[ \nabla(f(x)) \geq \frac{f(x) - f(x^*) + \frac{m}{2} \|x^* - x\|^2}{\|x - x^*\|} \geq \frac{m}{2} \|x^* - x\|. \]

Goes down by \( \alpha \frac{m}{2} \|x^* - x\|^2 \) in each step.

\( f(x) - f(x^*) \) is at most \( \frac{L}{2} \|x^* - x\|^2 \).

So decreases by \( (1 - \Theta(m^2/L^2)) \) in each step.

Better analysis: \( (1 - \Theta(m/L)) \) fraction in each step.
Next time.

Accelerated Gradient Descent.