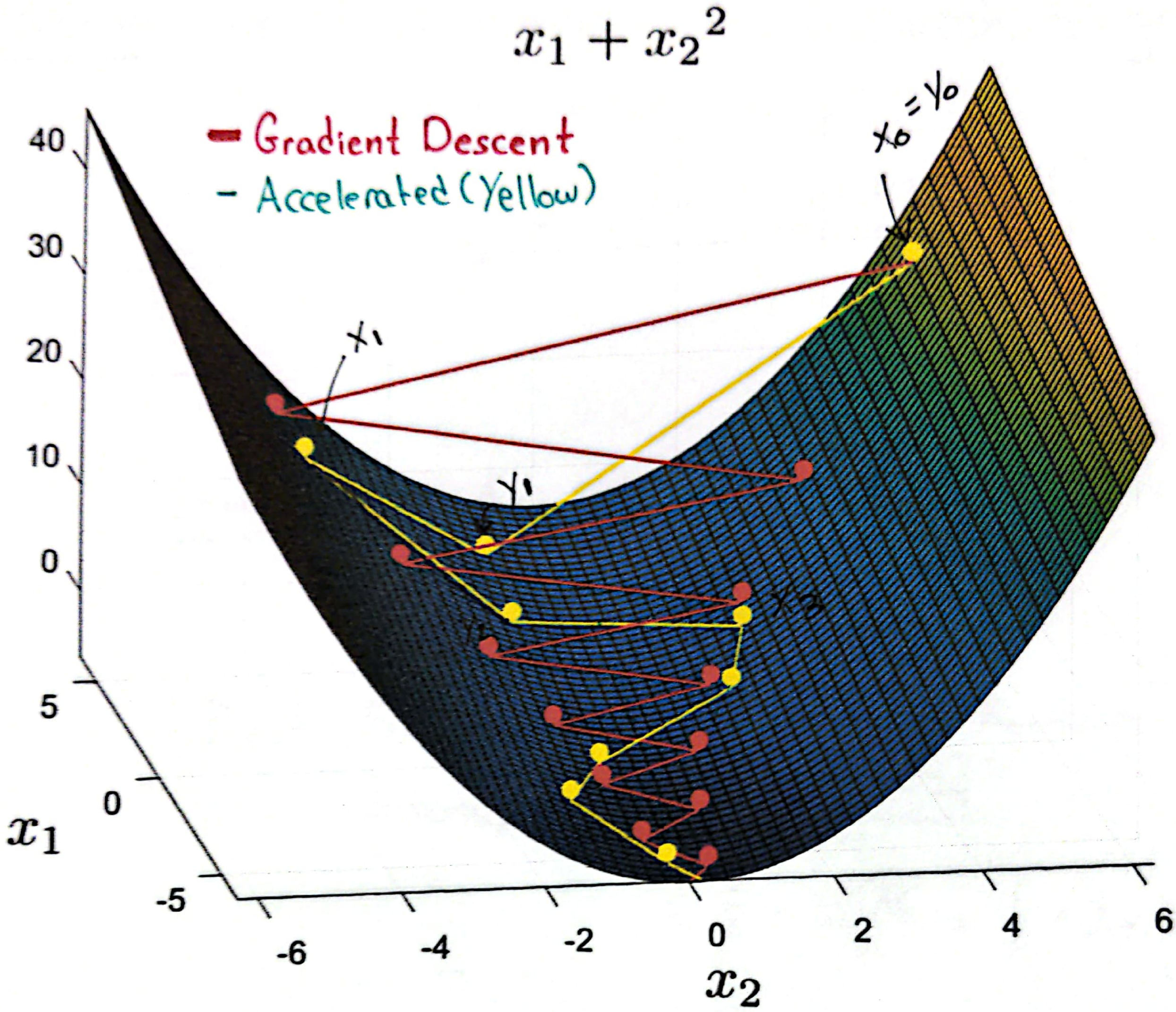
$x_1 + x_2{}^2$
Gradient Descent
- Accelerated (Yellow)
$x_0 = y_0$
x_1
y_1
x_2
40
30
20
10
0
5
0
-5
x_1
-6
-4
-2
0
2
4
6
x_2

Theorem Let f be convex with L-Lipschitz grad. Then for any minimizer x^*,

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

Starting 9/23

Proof. Lemma 1 The λ_k sequence has $\lambda_{k+1}^2 - \lambda_{k+1} = \lambda_k$ and for any $k \geq 1$, $\lambda_k \geq \frac{k+1}{2}$

Proof. First statement by Quadratic Formula.

Second, $\lambda_{k+1} = \dfrac{1 + \sqrt{1 + 4\lambda_k^2}}{2}$

$$\geq \frac{1}{2} + \frac{\sqrt{4\lambda_k^2}}{2} = \frac{1}{2} + \lambda_k. \quad \square$$

Checking $\dfrac{\lambda_k - 1}{\lambda_{k+1}} \approx \dfrac{\frac{k+1}{2} - 1}{\frac{k+2}{2}} = \dfrac{k-1}{k+2} \approx \dfrac{k}{k+3}$

Lemma 2 For $u, v \in \mathbb{R}^d$, we have

$$f(u - \tfrac{1}{L}\nabla f(u)) - f(v) \leq -\frac{1}{2L}\|\nabla f(u)\|_2^2 + \nabla f(u)^T(u - v).$$

Proof. Use convexity and then Descent Lemma:

$$f(u - \tfrac{1}{L}\nabla f(u)) - f(v) \leq f(u - \tfrac{1}{L}\nabla f(u)) - (f(u) + \nabla f(u)^T(v - u))$$

$$\leq -\frac{1}{2L}\|\nabla f(u)\|_2^2 + \nabla f(u)^T(u - v).$$

$\square$

We want a reccurrence on $\delta_k = f(y_k) - f(x^*)$ in terms $1/\lambda_k^2$.

Applying Lemma 2 with $u = x_k$, $v = y_k$ gives

$$\delta_{k+1} - \delta_k = f(y_{k+1}) - f(y_k) \le -\frac{1}{2L}\|\nabla f(x_k)\|_2^2 + \nabla f(x_k)^T(x_k - y_k)$$

(using $\nabla f(x_k) = -L(y_{k+1} - x_k)$)
$\Leftrightarrow y_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$)

$$= -\frac{L}{2}\|y_{k+1} - x_k\|_2^2 - L(y_{k+1} - x_k)^T(x_k - y_k) \quad (1)$$

Applying Lemma with $u = x_k$, $v = x^*$ gives

$$\delta_{k+1} = f(y_{k+1}) - f(x^*) \le -\frac{1}{2L}\|\nabla f(x_k)\|_2^2 + \nabla f(x_k)^T(x_k - x^*) \quad (2)$$

$$= -\frac{L}{2}\|y_{k+1} - x_k\|_2^2 - L(y_{k+1} - x_k)^T(x_k - x^*).$$

Summing $(\lambda_k - 1)(1) + (2)$ gives

$$\lambda_k \delta_{k+1} - (\lambda_k - 1)\delta_k \le -\frac{L\lambda_k}{2}\|y_{k+1} - x_k\|_2^2 - L(y_{k+1} - x_k)^T(\lambda_k x_k - (\lambda_k - 1)y_k - x^*)$$

Multiplying by λ_k gives

$$\lambda_k^2 \delta_{k+1} - \underbrace{(\lambda_k^2 - \lambda_k)}_{= \lambda_{k-1}^2}\delta_k \le -\frac{L}{2}\Big(\|\lambda_k(y_{k+1} - x_k)\|_2^2 + 2\lambda_k(y_{k+1} - x_k)^T(\lambda_k x_k - (\lambda_k - 1)y_k - x^*)$$

"completing the square"
$$= -\frac{L}{2}\left(\|\lambda_k y_{k+1} - (\lambda_k - 1) y_k - x^*\|_2^2 - \|\lambda_k x_k - (\lambda_k - 1) y_k - x^*\|_2^2\right)$$

$$\left(\text{Note } x_{k+1} = y_{k+1} + \frac{\lambda_k - 1}{\lambda_{k+1}}(y_{k+1} - y_k) \Leftrightarrow \lambda_{k+1} x_{k+1} - (\lambda_{k+1} - 1) y_{k+1} = \lambda_k y_{k+1} - (\lambda_k - 1) y_k \right)$$

$$\left(\lambda_k^2 \delta_{k+1} - \lambda_{k-1}^2 \delta_k \le \right) = -\frac{L}{2}\left(\|\underbrace{\lambda_{k+1} x_{k+1} - (\lambda_k - 1) y_{k+1} - x^*}_{=: u_{k+1}}\|_2^2 - \|\underbrace{\lambda_k x_k - (\lambda_k - 1) y_k - x^*}_{=: u_k}\|_2^2\right)$$

Summing from $k = 1, \dots, T$, gives

$$\lambda_{T-1}^2 \delta_T - \overset{=0}{\lambda_0^2} \delta_1 \le -\frac{L}{2}\left(\overset{\ge 0}{\|u_T\|_2^2} - \|u_1\|_2^2\right)$$

$$\Rightarrow \frac{T^2}{4}\delta_T \le \lambda_{T-1}^2 \delta_T \le +\frac{L}{2}\|\lambda_1 x_1 - (\overset{=0}{\lambda_1 - 1}) y_1 - x^*\|_2^2$$

$$= +\frac{L}{2}\|x_1 - x^*\|_2^2$$

$$\Rightarrow \delta_T \le \frac{2L\|x_1 - x^*\|_2^2}{T^2}. \quad \square$$

How can we make a single smooth convex func that is bad for every gradient method (under Assumption 1)?

WLOG, $x_0 = 0$ by shifting $\bar{f}(x) = f(x + x_0)$.

Lets design a function f_k s.t. x_k is all zeros after the k^{th} coordinate.

The "worst-function in the world"

$$f_k(x) = \frac{L}{4}\left(\frac{1}{2}(x^{(1)})^2 + \sum_{i=1}^{k-1}\frac{1}{2}(x^{(i+1)} - x^{(i)})^2 + \frac{1}{2}(x^{(k)})^2 - x^{(1)}\right)$$

1st component of the x vector

$$\nabla^2 f_k(x) = \frac{L}{4} A_k = \frac{L}{4}\begin{bmatrix} \left.\begin{matrix} 2 & -1 & & & & 0 \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & -1 & 2 & -1 & \\ 0 & & & & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{matrix}\right\} k \text{ lines} & 0_{k,\,d-k} \\ 0_{d-k,\,k} & 0_{d-k,\,d-k} \end{bmatrix}$$

$$\nabla f_k(x) = \frac{L}{4}(A_k x - e_1)$$