Special Assignment #1

Andrius Buinovskij - 18-940-270

Assignment 1. *Suppose that f is smooth with parameter L, i.e. for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2$. Show that*

$$d_{t+1}^2 - d_t^2 \leq -\frac{h_t}{2L} \qquad and \qquad \frac{1}{T}\sum_{t=0}^{T-1} h_t \leq \frac{2Ld_0^2}{T}.$$

Proof. First of all $d_t = \|\mathbf{x}_t - \mathbf{x}^*\|_2$, $h_t = f(\mathbf{x}_t) - f(\mathbf{x}^*)$, so our first goal is

$$d_{t+1}^2 - d_t^2 \leq -\frac{h_t}{2L} \tag{1}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq -\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \tag{2}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq \frac{f(\mathbf{x}^*) - f(\mathbf{x}_t)}{2L} \tag{3}$$

Well, I suppose a thing to try is to just plug stuff into smoothness. We need terms involving $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$, so We'll use smoothness twice to get:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}_t) + \nabla f(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{4}$$

and

$$f(\mathbf{x}^*) \leq f(\mathbf{x}_{t+1}) + \nabla f(\mathbf{x}_{t+1})^\top (\mathbf{x}^* - \mathbf{x}) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_{t+1}\|_2^2 \tag{5}$$

Hmm well when We subtract the two above terms We will lose the left hand side. We are trying to get rid of $f(\mathbf{x}_{t+1})$, so We should go for something more like

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}^* - \mathbf{x}_{t+1}) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{6}$$

and

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{7}$$

Since We are given that $\mathbf{x}^*$ exists and We are not restricted to any subset of $\mathbb{R}^d$, We know that

$$\nabla f(\mathbf{x}^*) = 0$$

This changes our previous equations to

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{8}$$

and

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{9}$$

Subtracting them yields

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{10}$$

Alright this will not work. I know that 3.6 in the notes seems pretty close to what I want.

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \gamma_t \mathbf{g}_t - \mathbf{x}^*\|_2^2 \tag{11}$$
$$= (\mathbf{x}_t - \gamma_t \mathbf{g}_t - \mathbf{x}^*)^\top (\mathbf{x}_t - \gamma_t \mathbf{g}_t - \mathbf{x}^*) \tag{12}$$
$$= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \|\gamma_t \mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top (\gamma_t \mathbf{g}_t) \tag{13}$$
$$= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top (\gamma_t \mathbf{g}_t) \tag{14}$$

Cool. By convexity of f We also have

$$f(\mathbf{y}) \geq f(\mathbf{x}) + g(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \implies \tag{15}$$
$$f(\mathbf{x}^*) \geq f(\mathbf{x}_t) + g(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{16}$$
$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq g(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{17}$$
$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq g(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{18}$$
$$\gamma_t(f(\mathbf{x}_t) - f(\mathbf{x}^*)) \leq \gamma_t g(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{19}$$
$$\gamma_t g(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}^*) \geq \gamma_t(f(\mathbf{x}_t) - f(\mathbf{x}^*)) \tag{20}$$
$$\gamma_t g(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}^*) \geq \gamma_t(f(\mathbf{x}_t) - f^*) \tag{21}$$

Where $f^* = f(\mathbf{x}^*)$. Using the above in 14 We get

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top (\gamma_t \mathbf{g}_t) \tag{22}$$
$$\leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t(f(\mathbf{x}_t) - f^*) \tag{23}$$

Now We use substitute in Polyak stepsize:

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t(f(\mathbf{x}_t) - f^*) \tag{24}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}\right)^2 \|\mathbf{g}_t\|_2^2 - 2\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}(f(\mathbf{x}_t) - f^*) \tag{25}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} - 2\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*)^2)}{\|\mathbf{g}_t\|_2^2} \tag{26}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} \tag{27}$$

Therefore:

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t(f(\mathbf{x}_t) - f^*) \tag{28}$$

$$\leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} \tag{29}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq -\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} \tag{30}$$

So, again, no good. I mean it's actually pretty close, but the terms are squared. Furthermore the denominator is a squared gradient, which is cool and all, but We want our bound in terms of L, and L is more concerned with the second order gradient. We don't have the fact that f is twice differentiable, so We can't, like, bound it either.

So, maybe We could try with smoothness again?

So! We have

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top(\gamma_t \mathbf{g}_t) \tag{31}$$

And We want to mess around with that last term. Now that squared step size seems to be a problem now that I think of it - it'll always have some annoying squared stuff. Maybe there is some way to get rid of it.

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{32}$$

$$f(\mathbf{x}^*) \leq f(\mathbf{x}_t) + \nabla f(\mathbf{x}_t)^\top(\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{33}$$

$$-\mathbf{g}^\top(\mathbf{x}^* - \mathbf{x}_t) \leq f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{34}$$

$$\mathbf{g}^\top(\mathbf{x}^* - \mathbf{x}_t) \geq f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{35}$$

Okay okay okay. None of this seems to be working out. This is a fact:

3

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top (\gamma_t \mathbf{g}_t) \tag{36}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{37}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{38}$$

Then We go back to smoothness:

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \tag{39}$$

$$f(\mathbf{x}^*) \leq f(\mathbf{x}_t) + \nabla f(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{40}$$

$$-\mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \leq f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{41}$$

$$\mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \geq f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{42}$$

Okay then We use the above:

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{43}$$

$$\geq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t (f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2) \tag{44}$$

None of that seems right. Expanding these those terms You'll get some nightmare term on the far right.
I guess it's back to smoothness?
Recall our goal is

$$d_{t+1}^2 - d_t^2 \leq \frac{h_t}{2L} \tag{45}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \tag{46}$$

Start again
So, the idea is that equation 3.6 is just too close to the solution to not be a key for it. The second idea is that 3.6 was arrived to through just straight convexity, and maybe that can be improved.
So We start with the fact that

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top (\gamma_t \mathbf{g}_t) \tag{47}$$

$$= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{48}$$

$$= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{49}$$

4

And that is plain true. Then by smoothness We get

$$f(\mathbf{y}) \le f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{50}$$

$$f(\mathbf{x}^*) \le f(\mathbf{x}_t) + \nabla f(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{51}$$

$$-\nabla f(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}_t) \le f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{52}$$

$$\nabla f(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}^*) \le f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{53}$$

$$\nabla f(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}_t) \ge f(\mathbf{x}^*) - f(\mathbf{x}_t) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{54}$$

So We just gained that rightmost term, which tightens the bound I think - well doesn't it loosen it? The term of left is smaller than the term on right. The smaller the term on right the closer it gets to the term. Well wait a second, what the hell.

Okay so the original derivation from the lecture slides used a bound from below by convexity. A bound from above here does us fuck all. This approach is wrong. Or is it lol.

I mean, the original approach is using some property of f to replace the cross term in the cosine law. We're doing the same just with a "stronger" condition I suppose.

Anyway, We do know that the term on the right is strictly positive.

I wanna know if the right side is strictly positive actually.

Well, I mean, We know

$$\nabla f(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}^*) \le f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{55}$$

$$-\nabla f(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}^*) \ge -\left(f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2\right) \tag{56}$$

$$\nabla f(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}_t) \ge f(\mathbf{x}^*) - f(\mathbf{x}_t) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{57}$$

Which doesn't tell me shit, really. Both of those are negative terms, and We know that the left side is greater, but that does not guarantee that the left side is positive. Just above some negative term. Maybe instead We can do by convexity

$$f(\mathbf{x}^*) \ge f(\mathbf{x}_t) + \nabla f(\mathbf{x}_t)(\mathbf{x}^* - \mathbf{x}_t) \tag{58}$$
$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \ge \nabla f(\mathbf{x}_t)(\mathbf{x}^* - \mathbf{x}_t) \tag{59}$$

Oh so it's strictly negative? The right hand side is negative or 0, and the left hand side is smaller. Oooookay.

So We have

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{60}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \geq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \left(f(\mathbf{x}^*) - f(\mathbf{x}_t) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \right) \tag{61}$$

Cool so now We have to try to simplify

$$\gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \left(f(\mathbf{x}^*) - f(\mathbf{x}_t) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \right) \tag{62}$$

$$\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \right)^2 \|\mathbf{g}_t\|_2^2 + 2\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \right) \left(f(\mathbf{x}^*) - f(\mathbf{x}_t) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \right) \tag{63}$$

$$\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} + 2\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \right) \left(f(\mathbf{x}^*) - f(\mathbf{x}_t) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \right) \tag{64}$$

$$\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} + 2\left(\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} \right) - L\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \cdot \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \tag{65}$$

$$3\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} - L\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \cdot \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \tag{66}$$

Plugging that back in gives

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \geq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \left(f(\mathbf{x}^*) - f(\mathbf{x}_t) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \right) \tag{67}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \geq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + 3\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} - L\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \cdot \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \tag{68}$$

Then You could drop the strictly positive term and maybe try to prove that one quantity is less than another. No way this is the correct way of going about it.
Start again
So now what. Using smoothness may work. What the hell, give it a shot:

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{69}$$

Cool. Now We have choices. Any time We invoke smoothness, it will need to have $\mathbf{x}^*$ as one of the endpoints. Hmm, or will it? We can split up $\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2$ into it's components. Yeah go on then:

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_{t+1} - \mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{70}$$

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{71}$$

Aight, now We use

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{72}$$

To get

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}^*) + \frac{L}{2}\left(\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t)\right) \tag{73}$$

Which does not look promising. I feel like wherever gamma shows up We get screwed.

Start again

Okay so this involves smoothness. It is the only expression We have that involves L. Is there any other way to involve L?

Well, a difference between two gradient terms can be upper bounded by it. Oh hey maybe that can be used...

I guess let's just try smoothness.

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{74}$$

Then We can do

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{75}$$

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{76}$$

and

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_{t+1} - \mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{77}$$

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{78}$$

Subtracting We get

$$f(\mathbf{x}_t) - f(\mathbf{x}_{t+1}) \leq \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{79}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}_{t+1}) \leq \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{L}{2}\left(\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t)\right) \tag{80}$$

I don't think so.

Start again

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \tag{81}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 + 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{82}$$

Start again
Man the exercise solution is so close. Maybe I can just bound it somehow? Let's rederive it.

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \gamma \mathbf{g}_t - \mathbf{x}^*\|_2^2 \tag{83}$$

$$= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \|\gamma_t \mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{84}$$

$$= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{85}$$

Cool. Then to mess with that We establish by convexity that

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{86}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{87}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{88}$$

So $\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*)$ is more than or equal to a non-negative quantity on the right, and so $-2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*)$ is negative.
However We have that $\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*)$ is a larger positive quantity, so We *can* say:

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{89}$$

$$\leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \big(f(\mathbf{x}_t) - f(\mathbf{x}^*) \big) \tag{90}$$

Then You stick in Polyak's step size and out comes 3.6.
So, not messing about with smoothness We simplify

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \big(f(\mathbf{x}_t) - f(\mathbf{x}^*) \big) \tag{91}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*}{\|\mathbf{g}_t\|_2^2} \right)^2 \|\mathbf{g}_t\|_2^2 - 2 \left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*}{\|\mathbf{g}_t\|_2^2} \right) \big(f(\mathbf{x}_t) - f(\mathbf{x}^*) \big) \tag{92}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} - 2 \cdot \frac{f((\mathbf{x}_t) - f(\mathbf{x}^*)^2}{\|\mathbf{g}_t\|_2^2} \tag{93}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} \tag{94}$$

Plugging the above simplified expression back into the original context yields:

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \le \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \big(f(\mathbf{x}_t) - f(\mathbf{x}^*)\big) \tag{95}$$

$$\le \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} \tag{96}$$

I guess We can try to work with that last term?
The numerator is involved in a bunch of equations. Not sure what to do. Just dick around I suppose.

$$f(\mathbf{x}^*) \le f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{97}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \le \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{98}$$

Well working with the term is not going anywhere.
Start again
Could use smoothness to get expressions for distances?

$$f(\mathbf{y}) \le f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \tag{99}$$

Then We can do

$$f(\mathbf{x}_t) \le f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{100}$$

$$f(\mathbf{x}_t) \le f(\mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{101}$$

$$\frac{2(f(\mathbf{x}_t) - f(\mathbf{x}^*))}{L} \le \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{102}$$

and

$$f(\mathbf{x}_{t+1}) \le f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_{t+1} - \mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{103}$$

$$f(\mathbf{x}_{t+1}) \le f(\mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{104}$$

$$\frac{2(f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*))}{L} \le \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{105}$$

subtracting yields

$$\frac{2(f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*))}{L} - \frac{2(f(\mathbf{x}_t) - f(\mathbf{x}^*))}{L} \le \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{106}$$

and ya don't get an $f(\mathbf{x}^*)$ in there so no bueno.

9

start again

So

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \tag{107}$$

Plugging stuff in We get

$$f(\mathbf{x}^*) \leq f(\mathbf{x}_t) + \nabla f(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{108}$$

and

$$f(\mathbf{x}^*) \leq f(\mathbf{x}_{t+1}) + \nabla f(\mathbf{x}_{t+1})^\top (\mathbf{x}^* - \mathbf{x}_{t+1}) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_{t+1}\|_2^2 \tag{109}$$

Well those gradient terms ain't going nowhere. And We still lose $f(\mathbf{x}^*)$. Maybe something like

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_{t+1} - \mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{110}$$

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{111}$$

$$\tag{112}$$

Subtracting yields

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*) \leq \tag{113}$$

$$f(\mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - f(\mathbf{x}_t) + \nabla f(\mathbf{x}_t)^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{114}$$

Which does not look promising

Start again

It must be the original formulation.

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \gamma_t \mathbf{g}_t - \mathbf{x}^*\|_2^2 \tag{115}$$

$$= (\mathbf{x}_t - \gamma_t \mathbf{g}_t - \mathbf{x}^*)^\top (\mathbf{x}_t - \gamma_t \mathbf{g}_t - \mathbf{x}^*) \tag{116}$$

$$= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \|\gamma_t \mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top (\gamma_t \mathbf{g}_t) \tag{117}$$

$$= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top (\gamma_t \mathbf{g}_t) \tag{118}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top (\gamma_t \mathbf{g}_t) \tag{119}$$

So, what can We do with that. Plug in stepsize is one thing:

$$\gamma_t^2\|\mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top(\gamma_t\mathbf{g}_t) \tag{120}$$

$$\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}\right)^2\|\mathbf{g}_t\|_2^2 - 2\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}\right)\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \tag{121}$$

$$\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} - 2\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}\right)\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \tag{122}$$

What does smoothness give me? It gives me a bound on the rate of change of the gradient. That would point toward taking the difference of gradients.
Okay so the original bound consists of breaking down the vector into components, substituting the gradient term and plugging in the stepsize.
First step is unambiguous.
Second step may allow us to leverage smoothness but probably not
Third step unambiguous.
So two possible avenues are using smoothness instead of convexity yeah I don't see it.
start again
Back to convexity ah fuck okay

$$f(\mathbf{y}) \le f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{123}$$

Alright so We can try

$$f(\mathbf{x}_{t+1}) \le f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top(\mathbf{x}_{t+1} - \mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^2\|_2^2 \tag{124}$$

$$f(\mathbf{x}_{t+1}) \le f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{125}$$

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*) \le \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{126}$$

$$\frac{2(f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*))}{L} \le \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{127}$$

Okay cool. We can split up the right side?

$$\frac{2(f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*))}{L} \le \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{128}$$

$$\frac{2(f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*))}{L} \le \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2\|\mathbf{g}_t\|_2^2 - 2(\mathbf{x}_t - \mathbf{x}^*)^\top(\gamma_t\mathbf{g}_t) \tag{129}$$

Nah doesn't work.
Back to combining smoothness terms?

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^2\|_2^2 \tag{130}$$

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{131}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{132}$$

$$\frac{2(f(\mathbf{x}_t) - f(\mathbf{x}^*))}{L} \leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{133}$$

Also doesn't seem to work.

start again

Brute force perhaps. Useful note yet again:

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{134}$$

$$\leq \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \big(f(\mathbf{x}_t) - f(\mathbf{x}^*)\big) \tag{135}$$

We start with

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \tag{136}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{137}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{138}$$

Okay, now We need to show that

$$\gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq -\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \tag{139}$$

So, first We have:

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{140}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{141}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{142}$$

So that dot product is positive. Cool. Alright smoothness on the other hand gets us

12

$$f(\mathbf{x}^*) \leq f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2$$

$$(143)$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \leq \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{144}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \geq \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{145}$$

$$-2\gamma_t \left(f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \right) \leq -2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{146}$$

We are allowed to use this substitution in the following inequality since We are making a negative term larger, therefore making the left side smaller, therefore loosening the bound:

$$\gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \geq \gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t \left(f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \right)$$

$$(147)$$

This doesn't seem to be working out. Let's look again at our options:

$$\gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{148}$$

This is what We've got. There is a simplification which can take place if We use convexity:

$$\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \right)^2 \|\mathbf{g}_t\|_2^2 - 2\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \right) \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{149}$$

$$\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} - 2\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \right) \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{150}$$

$$\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \left(f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \right) \tag{151}$$

I mean, not super inclined to fuck with the first term in the bracket since of the form We want.

Well, the coefficient is positive, so We have that. We also have

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{152}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{153}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{154}$$

And We also have

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \geq \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{155}$$

Quick aside to satisfy my curiosity

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{156}$$

$$-\frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \leq f(\mathbf{x}) - f(\mathbf{y}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \tag{157}$$

$$\frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \geq f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \tag{158}$$

Now We plug in our values:

$$\frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \geq f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \tag{159}$$

$$\frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \geq f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*) - \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_{t+1} - \mathbf{x}^*) \tag{160}$$

$$\frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \geq f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*) \tag{161}$$

Similarly

$$\frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \geq f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \tag{162}$$

$$\frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \geq f(\mathbf{x}_t) - f(\mathbf{x}^*) - \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{163}$$

$$\frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \geq f(\mathbf{x}_t) - f(\mathbf{x}^*) \tag{164}$$

$$-\frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq f(\mathbf{x}^*) - f(\mathbf{x}_t) \tag{165}$$

Can't subtract though.
start again
We start with

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \tag{166}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{167}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{168}$$

Well. We could do something like

$$\frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 = \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 \tag{169}$$

$$\frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 - \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \le \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 + f(\mathbf{x}^*) - f(\mathbf{x}_t) \tag{170}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \le \|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 + \frac{2}{L}\left(f(\mathbf{x}^*) - f(\mathbf{x}_t)\right) \tag{171}$$

$$\tag{172}$$

Nah.
Start again
We start with

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \tag{173}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{174}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{175}$$

We have a positive part and a negative part, and the negative part is negative due to the minus sign.
Polyak decrease:

$$f(\mathbf{x}_{t+1}) \le f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}_{t+1} - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 \tag{176}$$

$$\le f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}_t - \gamma_t \mathbf{g}_t - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 \tag{177}$$

$$\le f(\mathbf{x}_t) + \gamma_t \mathbf{g}_t^\top (\mathbf{g}_t) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 \tag{178}$$

$$\le f(\mathbf{x}_t) + \gamma_t\|\mathbf{g}_t\|_2^2 + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 \tag{179}$$

$$\le f(\mathbf{x}_t) + \left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}\right)\|\mathbf{g}_t\|_2^2 + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 \tag{180}$$

$$\le f(\mathbf{x}_t) + f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 \tag{181}$$

$$\le 2f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 \tag{182}$$

Start again
So We really just have convexity, differentiability and smoothness. There really is not much to be played around with.

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \tag{183}$$

$$\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{184}$$

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{185}$$

A positive term that We can replace with larger positive terms, and a negative term whose coefficient We can make smaller.
Now, the goal is:

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq \frac{f(\mathbf{x}^*) - f(\mathbf{x}_t)}{2L} \tag{186}$$

And so the goal can be rephrased as:

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq \frac{f(\mathbf{x}^*) - f(\mathbf{x}_t)}{2L} \tag{187}$$

$$\gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq \frac{f(\mathbf{x}^*) - f(\mathbf{x}_t)}{2L} \tag{188}$$

By smoothness We have

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2 \tag{189}$$

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 \leq L^2\|\mathbf{x} - \mathbf{y}\|_2^2 \tag{190}$$

Simplify stuff first

$$\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}\right)^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{191}$$

$$\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} - 2\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}\right)\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{192}$$

And this is it. We can either expand positive terms or shrink negative terms.

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{193}$$

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{194}$$

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{195}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{196}$$

Well, maybe to guide our journey We should write:

16

$$\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} - 2\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}\right)\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \leq -\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \quad (197)$$

$$\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} - 2\left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2}\right)\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) + \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \leq 0$$
$$(198)$$

$$\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} - 2\frac{\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} + \frac{1}{2L} \leq 0 \quad (199)$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \quad (200)$$

Okay! We can expand the positive terms or shrink the negative terms and still show this inequality. I suppose We can do the simple

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) \quad (201)$$
$$f(\mathbf{x}^*) \geq f(\mathbf{x}_t) + \nabla f(\mathbf{x}_t)^\top(\mathbf{x}^* - \mathbf{x}_t) \quad (202)$$
$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \mathbf{g}_t^\top(\mathbf{x}^* - \mathbf{x}_t) \quad (203)$$
$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \quad (204)$$
$$-2(f(\mathbf{x}_t) - f(\mathbf{x}^*)) \geq -2\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \quad (205)$$

Yeah We can't use that because it would make the coefficient of the negative term smaller oh wait - it would make the coefficient of the negative terms smaller therefore making it easier for us okay

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2(f(\mathbf{x}_t) - f(\mathbf{x}^*)) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \quad (206)$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \quad (207)$$

We can loosen the positive term again?

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) + \frac{L^2\|\mathbf{x}_t - \mathbf{x}^*\|_2^2}{2L} \leq 0 \quad (208)$$

$$-(f(\mathbf{x}_t) - f(\mathbf{x}^*)) + \frac{L\|\mathbf{x}_t - \mathbf{x}^*\|_2^2}{2} \leq 0 \quad (209)$$

Recall We also had

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \quad (210)$$

$$-(f(\mathbf{x}_t) - f(\mathbf{x}^*)) \geq -\frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \quad (211)$$

No good. The bound has been relaxed too much. Close though. **Second attempt**.

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{212}$$

Smoothness perhaps

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{213}$$

$$f(\mathbf{x}^*) \leq f(\mathbf{x}_t) + \nabla f(\mathbf{x}_t)^\top(\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{214}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \leq \nabla f(\mathbf{x}_t)^\top(\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{215}$$

Not that way.

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) - \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \leq \nabla f(\mathbf{x}_t)^\top(\mathbf{x}^* - \mathbf{x}_t) \tag{216}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \geq \mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \tag{217}$$

$$2f(\mathbf{x}^*) - 2f(\mathbf{x}_t) - L\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \geq -2\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \tag{218}$$

$$\tag{219}$$

So

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{220}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2f(\mathbf{x}^*) - 2f(\mathbf{x}_t) - L\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{221}$$

$$-3f(\mathbf{x}^*) - f(\mathbf{x}_t) - L\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{222}$$

Nah I don't think so. **Third attempt:**

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{223}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - \mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) - \mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{224}$$

$$\tag{225}$$

Smoothness again

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \le \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{226}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) - \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \le \frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{227}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \ge -\frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{228}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \ge -\frac{L}{2}\|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{229}$$

Nope. Can't replace with a smaller term.

$$f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \tag{230}$$

$$f(\mathbf{x}^*) \ge f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{231}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) - \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \ge 0 \tag{232}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \le 0 \tag{233}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \le 0 \tag{234}$$

And We can use that

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) - \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \le 0 \tag{235}$$

$$-\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \le 0 \tag{236}$$

$$-\cos(\theta)\|\mathbf{g}_t\|_2\|\mathbf{x}_t - \mathbf{x}^*\|_2 + \frac{\|\mathbf{g}_t\|_2^2}{2L} \le 0 \tag{237}$$

$$-\cos(\theta)\|\mathbf{x}_t - \mathbf{x}^*\|_2 + \frac{\|\mathbf{g}_t\|_2}{2L} \le 0 \tag{238}$$

$$-\cos(\theta)\|\mathbf{x}_t - \mathbf{x}^*\|_2 + \frac{L\|\mathbf{x}_t - \mathbf{x}^*\|_2}{2L} \le 0 \tag{239}$$

$$-\cos(\theta) + \frac{1}{2} \le 0 \tag{240}$$

$$\tag{241}$$

Close but no cigar. **Fourth attempt**
I can try to bound θ? lol

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{242}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{243}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{244}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \cos(\theta) \|\mathbf{g}_t\|_2 \|\mathbf{x}_t - \mathbf{x}^*\|_2 \tag{245}$$

$$\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2 \|\mathbf{x}_t - \mathbf{x}^*\|_2} \leq \cos(\theta) \tag{246}$$

Aight. Any approximation We introduce seems to fuck things up.

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{247}$$

What can We even do with this? One idea was to involve as many terms together as possible to get tighter stuff. I feel like there must be a way to use smoothness here that is not just bounding the gradient

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \frac{L}{2} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{248}$$

And so this gives us a lower bound on the smooth term. Haha, maybe We can bound L.

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L \|\mathbf{x} - \mathbf{y}\|_2 \tag{249}$$

$$\frac{\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2}{\|\mathbf{x} - \mathbf{y}\|_2} \leq L \tag{250}$$

$$\frac{\|\nabla f(\mathbf{x}_t) - \nabla f(\mathbf{x}^*)\|_2}{\|\mathbf{x}_t - \mathbf{x}^*\|_2} \leq L \tag{251}$$

$$\tag{252}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{253}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2} \cdot \frac{\|\mathbf{x}_t - \mathbf{x}^*\|_2}{\|\mathbf{g}_t - \nabla f(\mathbf{x}^*)\|_2} \leq 0 \tag{254}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2} \cdot \frac{\|\mathbf{x}_t - \mathbf{x}^*\|_2}{\|\mathbf{g}_t\|_2} \leq 0 \tag{255}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2 \cdot \|\mathbf{x}_t - \mathbf{x}^*\|_2}{2} \leq 0 \tag{256}$$

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \qquad (257)$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) - \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \geq 0 \qquad (258)$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \leq 0 \qquad (259)$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq 0 \qquad (260)$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq -\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \qquad (261)$$

$$-\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2 \cdot \|\mathbf{x}_t - \mathbf{x}^*\|_2}{2} \leq 0 \qquad (262)$$

$$-2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \|\mathbf{g}_t\|_2^2 \cdot \|\mathbf{x}_t - \mathbf{x}^*\|_2 \leq 0 \qquad (263)$$

$$-2\frac{\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*)}{\|\mathbf{g}_t\|_2 \cdot \|\mathbf{x}_t - \mathbf{x}^*\|_2} + \|\mathbf{g}_t\|_2 \leq 0 \qquad (264)$$

$$-2\frac{\cos(\theta)\|\mathbf{g}_t\| \cdot \|\mathbf{x}_t - \mathbf{x}^*\|)}{\|\mathbf{g}_t\|_2 \cdot \|\mathbf{x}_t - \mathbf{x}^*\|_2} + \|\mathbf{g}_t\|_2 \leq 0 \qquad (265)$$

$$-2\cos(\theta) + \|\mathbf{g}_t\|_2 \leq 0 \qquad (266)$$

Start again

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \qquad (267)$$

$$(268)$$

Smoothness:

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \tag{269}$$

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}_{t+1} - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 \tag{270}$$

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}_t - \gamma \mathbf{g}_t - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}_t - \gamma \mathbf{g}_t - \mathbf{x}_t\|_2^2 \tag{271}$$

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}_t) + \mathbf{g}_t^\top (-\gamma \mathbf{g}_t) + \frac{L}{2} \|\gamma \mathbf{g}_t\|_2^2 \tag{272}$$

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}_t) \leq -\gamma_t \|\mathbf{g}_t\|_2^2 + \frac{L\gamma_t^2}{2} \|\mathbf{g}_t\|_2^2 \tag{273}$$

$$\tag{274}$$

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}_t) \leq -\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \|\mathbf{g}_t\|_2^2 + \left(\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \right)^2 \frac{L}{2} \|\mathbf{g}_t\|_2^2 \tag{275}$$

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}_t) \leq -\left(f(\mathbf{x}_t) - f(\mathbf{x}^*) \right) + \frac{\left(f(\mathbf{x}_t) - f(\mathbf{x}^*) \right)^2}{\mathbf{g}_t\|_2^2} \frac{L}{2} \tag{276}$$

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}_t) \leq f(\mathbf{x}^*) - f(\mathbf{x}_t) + \frac{\left(f(\mathbf{x}_t) - f(\mathbf{x}^*) \right)^2}{\|\mathbf{g}_t\|_2^2} \frac{L}{2} \tag{277}$$

$$\tag{278}$$

Back to the roots

$$\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) = \frac{1}{\gamma_t} (\mathbf{x}_t - \mathbf{x}_{t+1})^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{279}$$

No problems there. Oh and We have

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \gamma_t \mathbf{g}_t - \mathbf{x}^*\|_2^2 \tag{280}$$

$$= \|(\mathbf{x}_t - \mathbf{x}^*) - \gamma_t \mathbf{g}_t\|_2^2 \tag{281}$$

$$= \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \|\gamma_t \mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{282}$$

And of course as a consequence We have

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \|\gamma_t \mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{283}$$

$$2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) = \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \|\gamma_t \mathbf{g}_t\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{284}$$

$$\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) = \frac{1}{2} \left(\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 + \gamma_t^2 \|\mathbf{g}_t\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \right) \tag{285}$$

That seems to be most of the roots.
Our goal:

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 = \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq \frac{f(\mathbf{x}^*) - f(\mathbf{x}_t)}{2L} \tag{286}$$

$$\gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq \frac{f(\mathbf{x}^*) - f(\mathbf{x}_t)}{2L} \tag{287}$$

$$\gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \leq 0 \tag{288}$$

And then We can do a little simplification:

$$\frac{\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right)^2}{\|\mathbf{g}_t\|_2^2} - 2\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \leq 0 \tag{289}$$

$$\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} - \frac{2}{\|\mathbf{g}_t\|_2^2} \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{1}{2L} \leq 0 \tag{290}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{291}$$

Cool.
First attempt

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{292}$$

$$\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) = \frac{1}{\gamma_t} (\mathbf{x}_t - \mathbf{x}_{t+1})^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{293}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\frac{1}{\gamma_t} (\mathbf{x}_t - \mathbf{x}_{t+1})^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{294}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\frac{\|\mathbf{g}_t\|_2^2}{f(\mathbf{x}_t) - f(\mathbf{x}^*)} (\mathbf{x}_t - \mathbf{x}_{t+1})^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{295}$$

aside

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq -\frac{(f(\mathbf{x}_t) - f(\mathbf{x}^*))^2}{\|\mathbf{g}_t\|_2^2} \tag{296}$$

$$\leq -\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right)\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \tag{297}$$

What can I do? I can make the right side larger, and if it is true in the larger case, it is true in the smaller case. To make the right side larger, I would need

to reduce the coefficient of the negative term. That means I can reduce the nominator or increase the denominator.

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq -\big(f(\mathbf{x}_t) - f(\mathbf{x}^*)\big)\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \tag{298}$$

Or rather, maybe, the goal can be phrased as

$$-\big(f(\mathbf{x}_t) - f(\mathbf{x}^*)\big)\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} \leq -\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \tag{299}$$

$$-\big(f(\mathbf{x}_t) - f(\mathbf{x}^*)\big)\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} + \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \leq 0 \tag{300}$$

Alright. **First attempt**

$$-\big(f(\mathbf{x}_t) - f(\mathbf{x}^*)\big)\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} + \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \leq 0 \tag{301}$$

$$-\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} + \frac{1}{2L} \leq 0 \tag{302}$$

$$-\big(f(\mathbf{x}_t) - f(\mathbf{x}^*)\big) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{303}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \leq 0 \tag{304}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) + \frac{L^2\|\mathbf{x}_t - \mathbf{x}^*\|_2^2}{2L} \leq 0 \tag{305}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) + \frac{L\|\mathbf{x}_t - \mathbf{x}^*\|_2^2}{2} \leq 0 \tag{306}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \leq -\frac{L\|\mathbf{x}_t - \mathbf{x}^*\|_2^2}{2} \tag{307}$$

$$f(\mathbf{x}_t) \leq f(\mathbf{x}^*) + \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{308}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{309}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq -\frac{L}{2}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{310}$$

Second attempt

$$- \left(f(\mathbf{x}_t) - f(\mathbf{x}^*) \right) \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} + \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \le 0 \tag{311}$$

$$- \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{\|\mathbf{g}_t\|_2^2} + \frac{1}{2L} \le 0 \tag{312}$$

$$- \left(f(\mathbf{x}_t) - f(\mathbf{x}^*) \right) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \le 0 \tag{313}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \le 0 \tag{314}$$

$$f(\mathbf{x}^*) \le f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{315}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \le \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}^* - \mathbf{x}_t\|_2^2 \tag{316}$$

$$f(\mathbf{x}^*) \le f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{317}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \le \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{318}$$

God knows what anymore

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) - 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{\|\mathbf{g}_t\|_2^2}{2L} \le 0 \tag{319}$$

Not a lot of things left to try, really. I suppose We can try Polyak decrease one last time

$$f(\mathbf{x}_{t+1}) \le f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}_{t+1} - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 \tag{320}$$

$$f(\mathbf{x}_{t+1}) \le f(\mathbf{x}_t) + \mathbf{g}_t^\top (-\gamma_t \mathbf{g}_t) + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 \tag{321}$$

$$f(\mathbf{x}_{t+1}) \le f(\mathbf{x}_t) - \gamma_t \|\mathbf{g}_t\|_2^2 + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 \tag{322}$$

Thievin
Lemma 4:
Claim 1:

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \le \frac{L}{2} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{323}$$

Proof.

$$f(\mathbf{x}_t) \le f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{324}$$

$$f(\mathbf{x}_t) \le f(\mathbf{x}^*) + \frac{L}{2} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{325}$$

$\square$

Claim 2:

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \geq \frac{1}{L} \|\mathbf{g}_t\|_2^2 \tag{326}$$

Proof.

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \geq f(\mathbf{x}_t) - f(\mathbf{x}_{t+1}) \tag{327}$$

Using smoothness We have

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}_t) + \mathbf{g}_t^\top (\mathbf{x}_{t+1} - \mathbf{x}_t) + \frac{L}{2} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|_2^2 \tag{328}$$

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}_t) + \mathbf{g}_t^\top (-\gamma_t \mathbf{g}_t) + \frac{L}{2} \| -\gamma_t \mathbf{g}_t\|_2^2 \tag{329}$$

$$f(\mathbf{x}_{t+1}) \leq f(\mathbf{x}_t) - \gamma_t \|\mathbf{g}_t\|_2^2 + \gamma_t^2 \frac{L}{2} \|\mathbf{g}_t\|_2^2 \tag{330}$$

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}_t) \leq -\gamma_t \|\mathbf{g}_t\|_2^2 + \gamma_t^2 \frac{L}{2} \|\mathbf{g}_t\|_2^2 \tag{331}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}_{t+1}) \geq \gamma_t \|\mathbf{g}_t\|_2^2 - \gamma_t^2 \frac{L}{2} \|\mathbf{g}_t\|_2^2 \tag{332}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}_{t+1}) \geq \left(\frac{1}{2L} \|\mathbf{g}_t\|_2^2 \right) \left(\gamma_t \cdot 2L - L^2 \gamma_t^2 \right) \tag{333}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}_{t+1}) \geq \left(\frac{1}{2L} \|\mathbf{g}_t\|_2^2 \right) \left(\gamma_t \cdot L(2 - L\gamma_t) \right) \tag{334}$$

So We have the restriction that both γ_t and L are strictly positive.

$$\frac{d}{d\gamma_t} 2L\gamma_t - L^2 \gamma_t^2 = 2L - 2L^2 \gamma_t \tag{335}$$

$$\frac{d^2}{d\gamma_t^2} 2L\gamma_t - L^2 \gamma_t^2 = -2L^2 \tag{336}$$

$$\frac{d}{dL} 2L\gamma_t - L^2 \gamma_t^2 = 2\gamma_t - 2L\gamma_t^2 \tag{337}$$

$$\frac{d^2}{dL^2} 2L\gamma_t - L^2 \gamma_t^2 = -2\gamma_t^2 \tag{338}$$

So at (0, 0) the derivative is 0, the second derivative is strictly negative, and the coefficient itself is 0 at (0, 0), so it's only going down from here, so We can replace it with a positive number.
Well fuck it let's find a max

26

$$2L - 2L^2\gamma_t = 0 \tag{339}$$

$$-2L^2\gamma_t = -2L \tag{340}$$

$$\gamma_t = \frac{1}{L} \tag{341}$$

$$\tag{342}$$

$$2\gamma_t - 2L\gamma_t^2 = 0 \tag{343}$$

$$2\frac{1}{L} - 2L\frac{1}{L^2} = 0 \tag{344}$$

$$2\frac{1}{L} - 2\frac{1}{L} = 0 \tag{345}$$

So the maximum is 1 and it is obtain when L is the reciprocal of γ_t and vice versa.
Basically We have

$$1 \geq \big(\gamma_t \cdot L(2 - L\gamma_t)\big) \tag{346}$$

So You can't fucking replace it with 1 can You.

$$f(\mathbf{x}) \leq f(\mathbf{y}) + \nabla f(\mathbf{y})^\top(\mathbf{x} - \mathbf{y}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2 \tag{347}$$

Okay We are just going to assume all of these

$$\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\| \leq L\|\mathbf{y} - \mathbf{x}\| \tag{348}$$

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|_2^2 \tag{349}$$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}) + \frac{1}{2L}\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_2^2 \tag{350}$$

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})^\top(\mathbf{x} - \mathbf{y}) \geq \frac{1}{L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2 \tag{351}$$

Paper: https://arxiv.org/pdf/1903.08688.pdf textbook: First order methods, chapter 5, page 110
Anyway, using

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{1}{2L} \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_2^2 \tag{352}$$

$$f(\mathbf{x}_t) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}_t - \mathbf{x}^*) + \frac{1}{2L} \|\nabla f(\mathbf{x}_t) - \nabla f(\mathbf{x}^*)\|_2^2 \tag{353}$$

$$f(\mathbf{x}_t) \geq f(\mathbf{x}^*) + \frac{1}{2L} \|\nabla f(\mathbf{x}_t)\|_2^2 \tag{354}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \geq \frac{1}{2L} \|\nabla f(\mathbf{x}_t)\|_2^2 \tag{355}$$

$$\tag{356}$$

$$\square$$

Recall We had

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq -\frac{\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right)^2}{\|\mathbf{g}_t\|_2^2} \tag{357}$$

So then We take our new result

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \geq \frac{1}{2L} \|\nabla f(\mathbf{x}_t)\|_2^2 \tag{358}$$

$$\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right)\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \geq \frac{1}{2L} \|\nabla f(\mathbf{x}_t)\|_2^2 \left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \tag{359}$$

$$-\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right)\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \leq -\frac{1}{2L} \|\nabla f(\mathbf{x}_t)\|_2^2 \left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \tag{360}$$

$$-\frac{\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right)}{\|\nabla f(\mathbf{x}_t)\|_2^2}\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \leq -\frac{1}{2L}\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \tag{361}$$

$$\tag{362}$$

Therefore

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq -\frac{1}{2L}\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \tag{363}$$

Next We want to show that

$$\frac{1}{T} \sum_{t=0}^{T-1} h_t \geq \frac{2L d_0^2}{T} \tag{364}$$

Rewriting our previous result

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \leq -\frac{1}{2L}\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \tag{365}$$

$$2L\left(\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2\right) \geq f(\mathbf{x}_t) - f(\mathbf{x}^*) \tag{366}$$

28

Taking the sum yields

$$\sum_{t=0}^{T-1} f(\mathbf{x}_t) - f(\mathbf{x}^*) \le 2L \sum_{t=0}^{T-1} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2^2 \tag{367}$$

$$\sum_{t=0}^{T-1} f(\mathbf{x}_t) - f(\mathbf{x}^*) \le 2L \left(\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 - \|\mathbf{x}_T - \mathbf{x}^*\|_2^2 \right) \tag{368}$$

$$\sum_{t=0}^{T-1} f(\mathbf{x}_t) - f(\mathbf{x}^*) \le 2L \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 \tag{369}$$

$$\frac{1}{T} \sum_{t=0}^{T-1} f(\mathbf{x}_t) - f(\mathbf{x}^*) \le \frac{2L \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2}{T} \tag{370}$$

Where the second step used telescoping and the third step dropped a strictly negative term to loosen the bound. $\square$

Assignment 2. *Now suppose that f is both smooth with parameter L and strongly convex with parameter μ, i.e. for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ We have $\frac{\mu}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 \le f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \le \frac{L}{2} \|\mathbf{x} - \mathbf{y}\|_2^2$. Show that*

$$d_{t+1}^2 - d_t^2 \le -\frac{\mu}{4L} d_t^2 \qquad and \qquad h_T \le \frac{L}{2} d_0^2 \left(1 - \frac{\mu}{4L} \right)^\top.$$

Proof. Start with convexity I suppose.
Recall We had

$$d_{t+1}^2 - d_t^2 \le -\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \tag{371}$$

$$f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{\mu}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \tag{372}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \ge \frac{\mu}{2} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{373}$$

$$- \left(f(\mathbf{x}_t) - f(\mathbf{x}^*) \right) \le -\frac{\mu}{2} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{374}$$

$$- \frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \le -\frac{\mu}{4L} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{375}$$

So We then have

$$d_{t+1}^2 - d_t^2 \le -\frac{f(\mathbf{x}_t) - f(\mathbf{x}^*)}{2L} \le -\frac{\mu}{4L} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{376}$$

$$d_{t+1}^2 - d_t^2 \le -\frac{\mu}{4L} d_t^2 \tag{377}$$

29

Brilliant.

Now the observation here is that We have shown a relation between d_{t+1} and d_t:

$$d_{t+1}^2 - d_t^2 \leq -\frac{\mu}{4L} d_t^2 \tag{378}$$

$$d_{t+1}^2 \leq -\frac{\mu}{4L} d_t^2 + d_t^2 \tag{379}$$

$$d_{t+1}^2 \leq d_t^2 \left(1 - \frac{\mu}{4L}\right) \tag{380}$$

Using smoothness We have

$$f(\mathbf{x}_T) - f(\mathbf{x}^*) \leq \frac{L}{2} d_T \tag{381}$$

$$f(\mathbf{x}_T) - f(\mathbf{x}^*) \leq \frac{L}{2} d_0^2 \left(1 - \frac{\mu}{4L}\right)^T \tag{382}$$

Where the second line is arrived at by recursively replacing d_{t+1} with $d_t^2 \left(1 - \frac{\mu}{4L}\right)$ until We reach d_0.

$\square$

Assignment 3. *Next, We consider the setting in which We only know a lower bound $\tilde{f} \leq f(\mathbf{x}^*)$ and thus adopt a step size $\tilde{\gamma}_t := \frac{f(\mathbf{x}_t) - \tilde{f}}{2\|\mathbf{g}_t\|_2^2}$. Let $\bar{x} = \mathrm{argmin}_{\mathbf{x}_t : 0 \leq t \leq T} f(\mathbf{x}_t)$, show that if f is both smooth with parameter L and strongly convex with parameter μ, then one of the following holds:*

$$f(\bar{\mathbf{x}}) - f(\mathbf{x}^*) \leq \min\left\{\frac{4Ld_0^2}{T}, \frac{L}{2} d_0^2 \left(1 - \frac{\mu}{8L}\right)^T\right\}$$

or

$$0 \leq f(\mathbf{x}^*) - \frac{f(\bar{\mathbf{x}}) + \tilde{f}}{2} \leq \frac{f(\mathbf{x}^*) - \tilde{f}}{2}$$

Proof. Well We have adjusted the stepsize, so our analysis needs to be redone.

$$d_{t+1} - d_t = \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{383}$$

$$= \left(\frac{f(\mathbf{x}_t) - \tilde{f}}{2\|\mathbf{g}_t\|_2^2}\right)^2 \|\mathbf{g}_t\|_2^2 - 2\frac{f(\mathbf{x}_t) - \tilde{f}}{2\|\mathbf{g}_t\|_2^2} \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{384}$$

Then We use convexity:

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_t) + \mathbf{g}_t^\top(\mathbf{x}^* - \mathbf{x}_t) \tag{385}$$

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_t) + \mathbf{g}_t^\top(\mathbf{x}^* - \mathbf{x}_t) \tag{386}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \mathbf{g}_t^\top(\mathbf{x}^* - \mathbf{x}_t) \tag{387}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \mathbf{g}_t^\top(\mathbf{x}^* - \mathbf{x}_t) \tag{388}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \mathbf{g}_t^\top(\mathbf{x}^* - \mathbf{x}_t) \tag{389}$$

Well, We can't use convexity.
There's a problem here -

$$\tilde{f} \leq f^* \tag{390}$$

$$-\tilde{f} \geq -f^* \tag{391}$$

$$f(\mathbf{x}_t) - \tilde{f} \geq f(\mathbf{x}_t) - f^* \tag{392}$$

$$\left(f(\mathbf{x}_t) - \tilde{f}\right)^2 \geq \left(f(\mathbf{x}_t) - f^*\right)^2 \tag{393}$$

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{2\|\mathbf{g}_t\|_2^2} \geq \frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{394}$$

$$-\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{2\|\mathbf{g}_t\|_2^2} \leq -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{395}$$

Where We are allowed to square both sides since We know that both sides are positive terms.
The results of previous exercises can be summed up as:

$$-\frac{\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right)^2}{\|\mathbf{g}_t\|_2^2} \leq -\frac{1}{2L}\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \leq -\frac{\mu}{4L}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{396}$$

$$-\frac{\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right)^2}{2\|\mathbf{g}_t\|_2^2} \leq -\frac{1}{4L}\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \leq -\frac{\mu}{8L}\|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \tag{397}$$

Ah okay so the above holds if $\tilde{f} = f^*$. In particular We then have

$$d_{t+1}^2 - d_t^2 \leq -\frac{1}{4L}\left(f(\mathbf{x}_t) - f(\mathbf{x}^*)\right) \tag{398}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq 4L\left(d_t^2 - d_{t+1}^2\right) \tag{399}$$

$$\sum_{t=0}^{T-1} f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq 4L\sum_{t=0}^{T-1}\left(d_t^2 - d_{t+1}^2\right) \tag{400}$$

$$\frac{\sum_{t=0}^{T-1} f(\mathbf{x}_t) - f(\mathbf{x}^*)}{T} \leq \frac{4Ld_0^2}{T} \tag{401}$$

31

Where We skipped a few steps but it was just telescoping and dropping the negative $-d_T^2$.

Similarly, We can also establish the following relationship:

$$d_{t+1}^2 - d_t^2 \leq -\frac{\mu}{8L}d_t^2 \tag{402}$$

$$d_{t+1}^2 \leq d_t^2\left(1 - \frac{\mu}{8L}\right) \tag{403}$$

And then by smoothness We can do

$$f(\mathbf{x}_T) - f(\mathbf{x}^*) \leq \frac{L}{2}\|\mathbf{x}_T - \mathbf{x}^*\|_2^2 \tag{404}$$

$$\leq \frac{L}{2}d_0^2\left(1 - \frac{\mu}{8L}\right)^T \tag{405}$$

With the same logic as before of recursively replacing d_{t+1} with $\left(1 - \frac{\mu}{8L}\right)d_t$.

Well, it's more like so long as Your choice of $\tilde{f}$ preserves

Now in the case of $\tilde{f} < f^*$.

So, here is the idea. In the original case We had:

$$d_{t+1} - d_t = \gamma_t^2\|\mathbf{g}_t\|_2^2 - 2\gamma_t\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \tag{406}$$

$$= \left(\frac{f(\mathbf{x}_t) - \tilde{f}}{2\|\mathbf{g}_t\|_2^2}\right)^2\|\mathbf{g}_t\|_2^2 - 2\frac{f(\mathbf{x}_t) - \tilde{f}}{2\|\mathbf{g}_t\|_2^2}\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \tag{407}$$

$$= \frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|\mathbf{g}_t\|_2^2} - \frac{f(\mathbf{x}_t) - \tilde{f}}{\|\mathbf{g}_t\|_2^2}\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \tag{408}$$

And now We need to perform a substitution, previously established via convexity, to move forward. Convexity won't save us now unfortunately.

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \mathbf{g}_t^\top(\mathbf{x}^* - \mathbf{x}_t) \tag{409}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \tag{410}$$

So strictly positive quantity. This will end up being wedged from below by the equality and from above by the convexity or whatever. Well the maximal bound mai

Look lad for this analysis to hold We need

$$d_{t+1}^2 - d_t^2 = \frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|\mathbf{g}_t\|_2^2} - \frac{f(\mathbf{x}_t) - \tilde{f}}{\|\mathbf{g}_t\|_2^2}\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \leq? \leq -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{411}$$

32

Where the question mark is our substitution. When $\tilde{f} = f^*$ We can use convexity. So

Okay so this is all very confused, and I think I am looking at the wrong quantities.

Redo

So We restart our analysis with a new stepsize:

$$d_{t+1} - d_t = \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{412}$$

$$= \left(\frac{f(\mathbf{x}_t) - \tilde{f}}{2\|\mathbf{g}_t\|_2^2}\right)^2 \|\mathbf{g}_t\|_2^2 - 2 \frac{f(\mathbf{x}_t) - \tilde{f}}{2\|\mathbf{g}_t\|_2^2} \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{413}$$

$$= \frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|\mathbf{g}_t\|_2^2} - \frac{f(\mathbf{x}_t) - \tilde{f}}{\|\mathbf{g}_t\|_2^2} \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{414}$$

And so now what.

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq \mathbf{g}_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{415}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{416}$$

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq -\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{417}$$

Well so long as We have

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|\mathbf{g}_t\|_2^2} - \frac{f(\mathbf{x}_t) - \tilde{f}}{\|\mathbf{g}_t\|_2^2} \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{418}$$

We're good

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4} - \left(f(\mathbf{x}_t) - \tilde{f}\right)\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2} \tag{419}$$

$$\frac{f(\mathbf{x}_t) - \tilde{f}}{4} - \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \leq -\frac{f(\mathbf{x}_t) - f^*}{2} \frac{f(\mathbf{x}_t) - f^*}{f(\mathbf{x}_t) - \tilde{f}} \tag{420}$$

$$\tag{421}$$

I guess the problem is that ultimately I don't know the relation between $\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*)$ and $f(\mathbf{x}_t) - \tilde{f}$

Ah what the fuck

$$0 \leq f(\mathbf{x}^*) - \frac{f(\bar{\mathbf{x}}) + \tilde{f}}{2} \leq \frac{f(\mathbf{x}^*) - \tilde{f}}{2} \tag{422}$$

$$0 \leq 2f(\mathbf{x}^*) - f(\bar{\mathbf{x}}) + \tilde{f} \leq f(\mathbf{x}^*) - \tilde{f} \tag{423}$$

$$\tilde{f} \leq 2f(\mathbf{x}^*) - f(\bar{\mathbf{x}}) + 2\tilde{f} \leq f(\mathbf{x}^*) \tag{424}$$

$$\tilde{f} \leq 2(f(\mathbf{x}^*) - \tilde{f}) - f(\bar{\mathbf{x}}) + \leq f(\mathbf{x}^*) \tag{425}$$

33

Comparing values:

$$\frac{f(\mathbf{x}_t) - \tilde{f}}{\|\mathbf{g}_t\|^2} \geq \frac{f(\mathbf{x}_t) - f^*}{\|\mathbf{g}_t\|_2^2} \tag{426}$$

So that's true, but We don't know if

$$\frac{f(\mathbf{x}_t) - \tilde{f}}{2\|\mathbf{g}_t\|^2} \geq \frac{f(\mathbf{x}_t) - f^*}{\|\mathbf{g}_t\|_2^2} \tag{427}$$

And for the above to hold We would need

$$f(\mathbf{x}_t) - \tilde{f} \geq 2f(\mathbf{x}_t) - 2f^* \tag{428}$$

$$\tilde{f} - f(\mathbf{x}_t) \leq 2f^* - 2f(\mathbf{x}_t) \tag{429}$$

Alright alright alright fuck.

There are two parts to this analysis - forward and backward. The "forward" bit We showed holds just fine without any difficulty.

The "backward" bit is the tricky one. Christ alive We need

$$d_{t+1}^2 - d_t^2 = \frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|\mathbf{g}_t\|_2^2} - \frac{f(\mathbf{x}_t) - \tilde{f}}{\|\mathbf{g}_t\|_2^2}\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \leq ? \leq -\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{430}$$

Okay so that means

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|\mathbf{g}_t\|_2^2} - \frac{f(\mathbf{x}_t) - \tilde{f}}{\|\mathbf{g}_t\|_2^2}\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \overset{?}{\leq} -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{431}$$

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|\mathbf{g}_t\|_2^2} - \frac{f(\mathbf{x}_t) - \tilde{f}}{\|\mathbf{g}_t\|_2^2}\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \overset{?}{\leq} -\frac{2\left(f(\mathbf{x}_t) - f^*\right)^2}{4\|\mathbf{g}_t\|_2^2} \tag{432}$$

$$\frac{3\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|\mathbf{g}_t\|_2^2} - \frac{f(\mathbf{x}_t) - \tilde{f}}{\|\mathbf{g}_t\|_2^2}\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \overset{?}{\leq} 0 \tag{433}$$

$$\frac{3}{4}\left(f(\mathbf{x}_t) - \tilde{f}\right)^2 - \left(f(\mathbf{x}_t) - \tilde{f}\right)\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \leq 0 \tag{434}$$

$$\frac{3}{4}\left(f(\mathbf{x}_t) - \tilde{f}\right) - \mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \overset{?}{\leq} 0 \tag{435}$$

Well fuck. Strong convexity? No.

Okay the only moving part in here is $\tilde{f}$. The question is is there a choice for $\tilde{f}$ that fucks the inequalities.

And I don't really know if there is TBH.

Maybe compare the gammas.

$$d_{t+1} - d_t = \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{436}$$

$$d_{t+1} - d_t = \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{437}$$

$$\tilde{\gamma}_t^2 \|\mathbf{g}_t\|_2^2 - 2\tilde{\gamma}_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \overset{?}{\leq} \gamma_t^2 \|\mathbf{g}_t\|_2^2 - 2\gamma_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{438}$$

$$\left(\tilde{\gamma}_t^2 - \gamma_t^2\right) \|\mathbf{g}_t\|_2^2 - \left(\tilde{\gamma}_t - \gamma_t\right) 2\mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \overset{?}{\leq} 0 \tag{439}$$

$$\left(\tilde{\gamma}_t^2 - \gamma_t^2\right) \|\mathbf{g}_t\|_2^2 - \left(\tilde{\gamma}_t - \gamma_t\right) 2(f(\mathbf{x}_t) - f^*) \overset{?}{\leq} 0 \tag{440}$$

$$\left(\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|g_t\|_2^4} - \frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{\|g_t\|_2^4} \right) \|\mathbf{g}_t\|_2^2 - \left(\tilde{\gamma}_t - \gamma_t\right) 2(f(\mathbf{x}_t) - f^*) \overset{?}{\leq} 0 \tag{441}$$

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|g_t\|_2^2} - \frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{\|g_t\|_2^2} - \left(\tilde{\gamma}_t - \gamma_t\right) 2(f(\mathbf{x}_t) - f^*) \overset{?}{\leq} 0 \tag{442}$$

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|g_t\|_2^2} - \frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{\|g_t\|_2^2} - \frac{f(\mathbf{x}_t) - \tilde{f} - 2f(\mathbf{x}_t) + 2f^*}{2\|g_t\|_2^2} 2(f(\mathbf{x}_t) - f^*) \overset{?}{\leq} 0 \tag{443}$$

$$\tilde{\gamma}_t^2 \|\mathbf{g}_t\|_2^2 - 2\tilde{\gamma}_t \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \overset{?}{\leq} -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{444}$$

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|g_t\|_2^2} - 2\frac{f(\mathbf{x}_t) - \tilde{f}}{2\|g_t\|_2^2} \mathbf{g}_t^\top (\mathbf{x}_t - \mathbf{x}^*) \overset{?}{\leq} -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{445}$$

Now We can do

$$f(\mathbf{x}^*) - f(\mathbf{x}_t) \geq g_t^\top (\mathbf{x}^* - \mathbf{x}_t) \tag{446}$$

$$f(\mathbf{x}_t) - f(\mathbf{x}^*) \leq g_t^\top (\mathbf{x}_t - \mathbf{x}^*) \tag{447}$$

And We also have

$$\tilde{f} \leq f^* \tag{448}$$

$$-\tilde{f} \geq -f^* \tag{449}$$

$$f(\mathbf{x}_t) - \tilde{f} \geq f(\mathbf{x}_t) - f^* \tag{450}$$

$$f(\mathbf{x}_t) - f^* \leq f(\mathbf{x}_t) - \tilde{f} \tag{451}$$

So We can do

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|g_t\|_2^2} - 2\frac{f(\mathbf{x}_t) - \tilde{f}}{2\|g_t\|_2^2}\mathbf{g}_t^\top(\mathbf{x}_t - \mathbf{x}^*) \overset{?}{\leq} -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{452}$$

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4\|g_t\|_2^2} - 2\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|g_t\|_2^2} \overset{?}{\leq} -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2\|\mathbf{g}_t\|_2^2} \tag{453}$$

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4} - \left(f(\mathbf{x}_t) - f^*\right)^2 \overset{?}{\leq} -\frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2} \tag{454}$$

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4} \overset{?}{\leq} \frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2} \tag{455}$$

$$\frac{\left(f(\mathbf{x}_t) - \tilde{f}\right)^2}{4} \overset{?}{\leq} \frac{\left(f(\mathbf{x}_t) - f^*\right)^2}{2} \tag{456}$$

$$\left(f(\mathbf{x}_t) - \tilde{f}\right)^2 \overset{?}{\leq} 2\left(f(\mathbf{x}_t) - f^*\right)^2 \tag{457}$$

$$f(\mathbf{x}_t) - \tilde{f} \overset{?}{\leq} \sqrt{2}f(\mathbf{x}_t) - \sqrt{2}f^* \tag{458}$$

$$-\tilde{f} \overset{?}{\leq} \sqrt{2}f(\mathbf{x}_t) - f(\mathbf{x}_t) - \sqrt{2}f^* \tag{459}$$

$$\tilde{f} \overset{?}{\geq} -\sqrt{2}f(\mathbf{x}_t) + f(\mathbf{x}_t) + \sqrt{2}f^* \tag{460}$$

$\square$

Assignment 5. *Prove that f is convex.*

Proof.

$$f(\mathbf{x}) := \frac{1}{K}\log\left(\sum_{i=1}^m \exp\left(K\left(\mathbf{a}_i^\top\mathbf{x} - \mathbf{b}_i\right)\right)\right) \tag{461}$$

We know that

$$K\left(\mathbf{a}_i^\top\mathbf{x} - \mathbf{b}_i\right) \tag{462}$$

Is convex, since it is a linear function.
Furthermore, by exercise set 1, exercise 8, We know that log-sum-exp is a convex function. Lastly, by lemma 1.18 part ii), the composition of a convex function and a linear function is itself convex. Therefore the composition of of our linear function and log-sum-exp is also convex.

$\square$

Assignment 6. *Prove that f is smooth with parameter $\|A\|$.*

Proof. Alright well the most promising definition seems to be

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \tag{463}$$

So We need a derivative of log-sum-exp. If We can show that the spectral norm of it is equal to the spectral norm of A, then We're good.

Let $w = \sum_{j=1}^{m} \exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right) \ v_j = \exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right)$

$$\frac{d}{d\mathbf{x}_i} \frac{1}{K} \log\left(\sum_{j=1}^{m} \exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right)\right) \tag{464}$$

$$= \frac{1}{K} \frac{1}{w} \frac{d}{d\mathbf{x}_i} \sum_{j=1}^{m} \exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right) \tag{465}$$

$$= \frac{1}{K} \frac{1}{w} \sum_{j=1}^{m} \frac{d}{d\mathbf{x}_i} \exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right) \tag{466}$$

$$= \frac{1}{K} \frac{1}{w} \sum_{j=1}^{m} \exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right) \frac{d}{d\mathbf{x}_i} K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right) \tag{467}$$

$$= \frac{1}{K} \frac{1}{w} \sum_{j=1}^{m} v_j \frac{d}{d\mathbf{x}_i} K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right) \tag{468}$$

$$= \frac{1}{K} \frac{1}{w} \sum_{j=1}^{m} v_j K \mathbf{a}_{j,i} \tag{469}$$

$$= \frac{1}{w} \sum_{j=1}^{m} v_j \mathbf{a}_{j,i} \tag{470}$$

$$= \frac{1}{\sum_{j=1}^{m} v_j} \sum_{j=1}^{m} v_j \mathbf{a}_{j,i} \tag{471}$$

Well there it is, as far as I can tell. We essentially sum over column i such that the summation overall well it's weird. It's not just summation, You scaled it.

Well I think that that is the correct derivative.

We can probably write it in terms of A - We are summing over columns of A.

Let $\mathbf{v}^\mathbf{x}$ be a vector such that

$$\mathbf{v}_j^\mathbf{x} = \frac{\exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right)}{\sum_{j=1}^{m} \exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right)} \tag{472}$$

$$\mathbf{v}_j^\mathbf{x} = \frac{v_j}{w} \tag{473}$$

Let's explain this notation a little bit - the $\mathbf{x}$ in $\mathbf{v}^\mathbf{x}$ specifies the underlying vector used in calculations of $\mathbf{v}$. So, for example, as opposed $\mathbf{v}^\mathbf{x}$ defined above, $\mathbf{v}^\mathbf{y}$ would look like:

$$\mathbf{v}_j^\mathbf{y} = \frac{\exp\left(K\left(\mathbf{a}_j^\top \mathbf{y} - \mathbf{b}_j\right)\right)}{\sum_{j=1}^{m} \exp\left(K\left(\mathbf{a}_j^\top \mathbf{y} - \mathbf{b}_j\right)\right)} \tag{474}$$

Anyway the derivative can then be written as:

$$\nabla f(\mathbf{x}) = A^\top \mathbf{v^x} \tag{475}$$

Cool so then

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| = \|A^\top \mathbf{v^x} - A^\top \mathbf{v^y}\| \tag{476}$$
$$= \|A^\top (\mathbf{v^x} - \mathbf{v^y})\| \tag{477}$$
$$= \|A^\top\|(\mathbf{v^x} - \mathbf{v^y})\| \tag{478}$$
$$\leq \|A^\top\| \cdot \|\mathbf{v^x} - \mathbf{v^y}\| \tag{479}$$
$$\leq \|A\| \cdot \|\mathbf{v^x} - \mathbf{v^y}\| \tag{480}$$

Now if We can show that

$$\|\mathbf{v^x} - \mathbf{v^y}\| \leq \|\mathbf{x} - \mathbf{y}\| \tag{481}$$

We're good.

$$(\mathbf{v^x} - \mathbf{v^y})_i = \mathbf{v}_i^\mathbf{x} - \mathbf{v}_i^\mathbf{y} \tag{482}$$
$$= \frac{\exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right)}{\sum_{j=1}^m \exp\left(K\left(\mathbf{a}_j^\top \mathbf{x} - \mathbf{b}_j\right)\right)} - \frac{\exp\left(K\left(\mathbf{a}_j^\top \mathbf{y} - \mathbf{b}_j\right)\right)}{\sum_{j=1}^m \exp\left(K\left(\mathbf{a}_j^\top \mathbf{y} - \mathbf{b}_j\right)\right)} \tag{483}$$

$$\square$$