Convex–Concave Backtracking for Inertial Bregman Proximal Gradient Algorithms in Non-Convex Optimization

Peter Ochs

Mathematical Optimization Group
Saarland University

— 19.09.2019 —

joint work: M.C. Mukkamala, T. Pock and S. Sabach

A Simple Illustrative Setting

Smooth minimization problem:

$$\min_{x\in\mathbb{R}^N} f(x)\,, \qquad \nabla f \text{ is Lipschitz}\,.$$

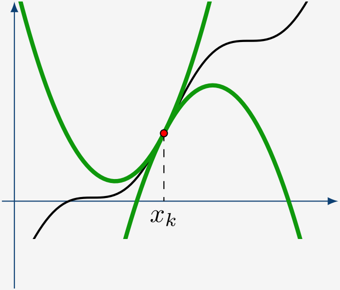
- **Gradient Descent**: $(\tau > 0)$

$$x_{k+1} = x_k - \tau\nabla f(x_k) \iff x_{k+1} = \operatorname*{argmin}_x \langle \nabla f(x_k), x - x_k\rangle + \frac{1}{2\tau}\|x-x_k\|^2\,.$$

- **Descent Lemma**:

∇f is L-Lipschitz

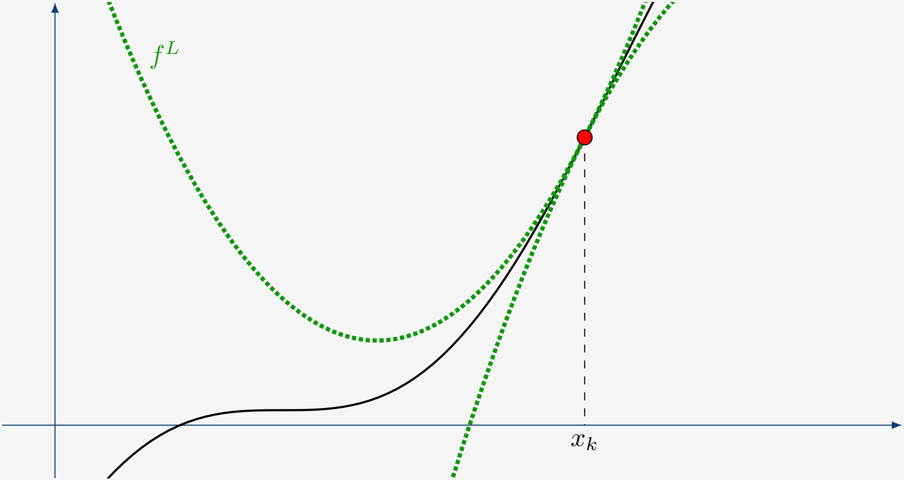
$$\implies |f(x) - f(x_k) - \langle \nabla f(x_k), x - x_k\rangle| \leq \frac{L}{2}\|x-x_k\|^2$$



$$\iff \quad -\frac{L}{2}\|x-x_k\|^2 \leq f(x) - f(x_k) - \langle \nabla f(x_k), x - x_k\rangle \leq \frac{L}{2}\|x-x_k\|^2$$

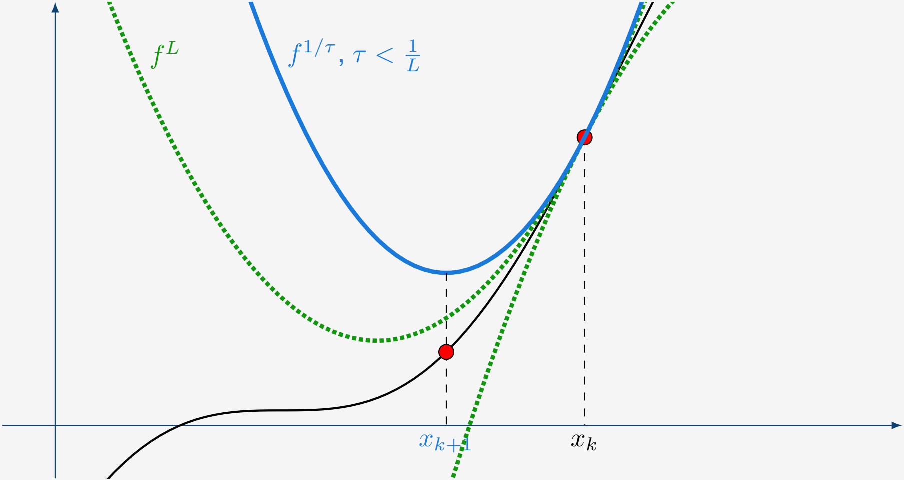
Gradient Descent: Sufficient decrease

$$x_{k+1} = \operatorname*{argmin}_x \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\tau}\|x - x_k\|^2}_{=: f^{1/\tau}} \,.$$



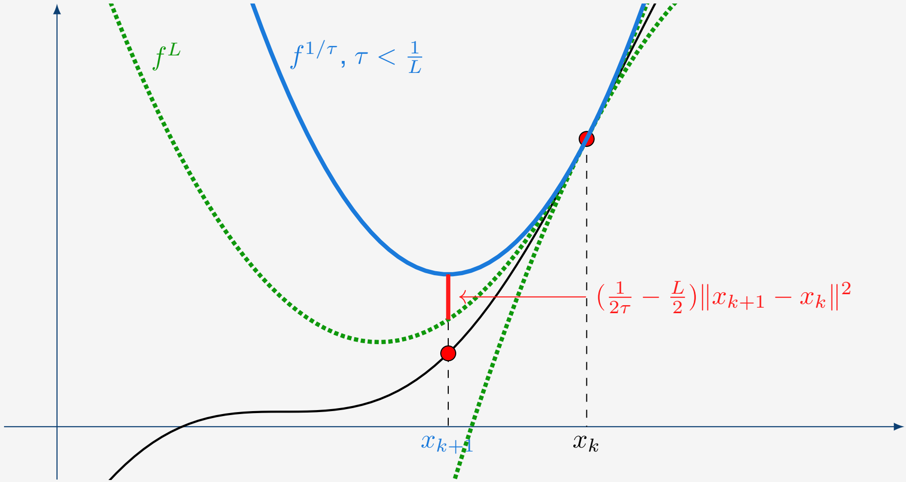
Gradient Descent: Sufficient decrease

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\tau}\|x - x_k\|^2}_{=: f^{1/\tau}} \ .$$



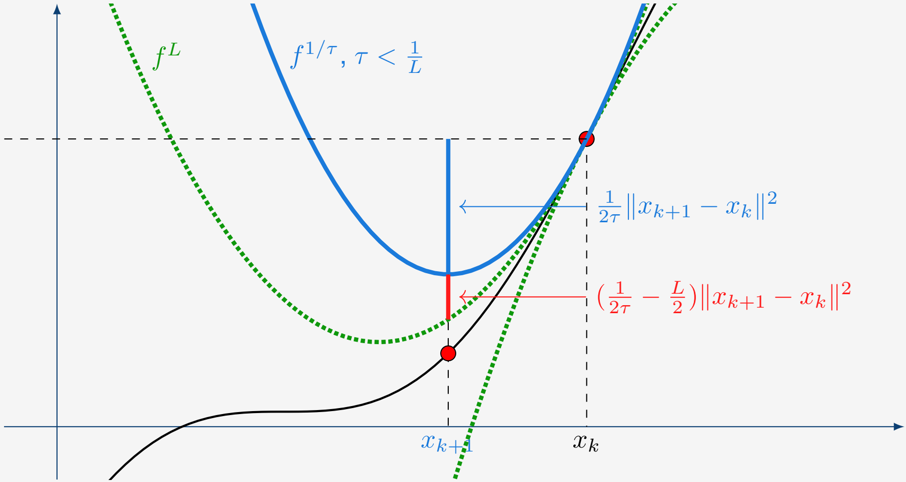
Gradient Descent: Sufficient decrease

$$x_{k+1} = \operatorname*{argmin}_x \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\tau}\|x - x_k\|^2}_{=: f^{1/\tau}} \, .$$



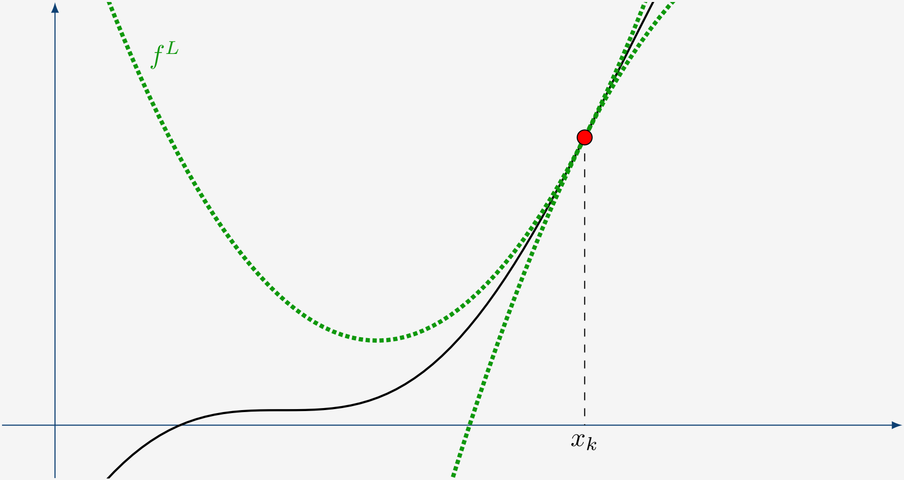
Gradient Descent: Sufficient decrease

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \; \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\tau}\|x - x_k\|^2}_{=: f^{1/\tau}} \; .$$



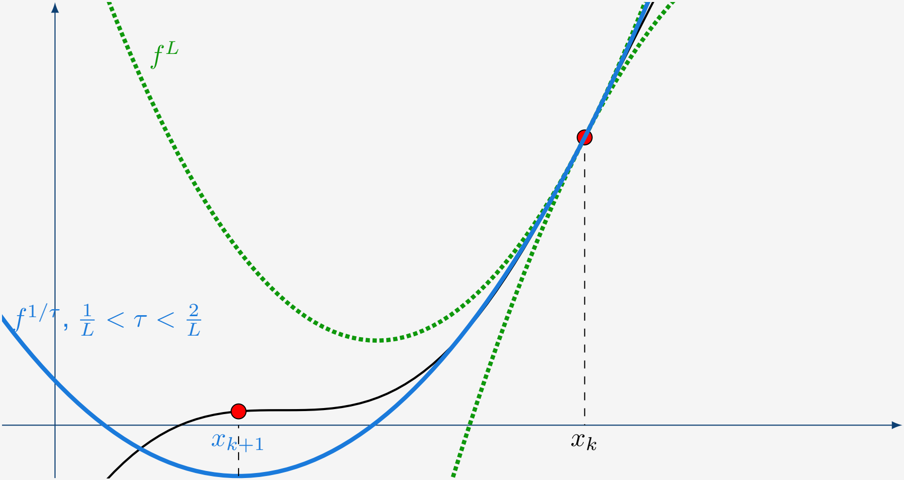
Gradient Descent: Sufficient decrease

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\tau}\|x - x_k\|^2}_{=: f^{1/\tau}} .$$



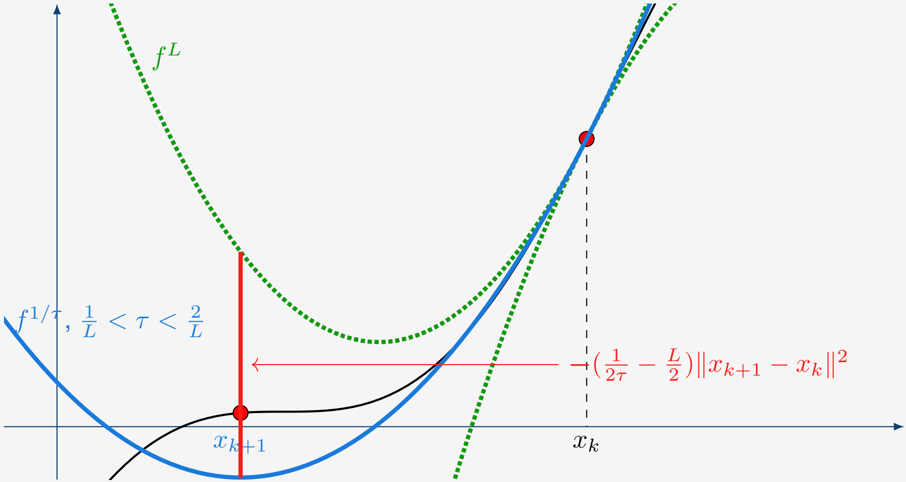
Gradient Descent: Sufficient decrease

$$x_{k+1} = \operatorname*{argmin}_x \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\tau}\|x - x_k\|^2}_{=: f^{1/\tau}} \, .$$



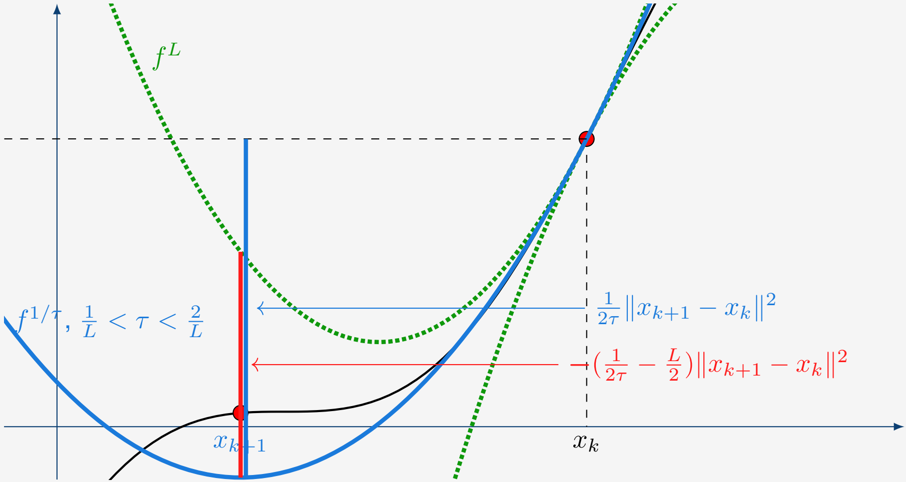
Gradient Descent: Sufficient decrease

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \; \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\tau}\|x - x_k\|^2}_{=: f^{1/\tau}} \; .$$



Gradient Descent: Sufficient decrease

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\tau}\|x - x_k\|^2}_{=: f^{1/\tau}} \, .$$



Proof of Convergence for Gradient Descent

Proof of Sufficient Decrease Condition:

$$f(x_{k+1}) \overset{\text{quad. upper bound}}{\leq} f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$$

$$\leq f(x_k) + \underbrace{\langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{1}{2\tau}\|x_{k+1} - x_k\|^2}_{(\tau^{-1}\text{-strongly convex}) \quad \leq -\frac{1}{2\tau}\|x_{k+1}-x_k\|^2} - \left(\frac{1}{2\tau} - \frac{L}{2}\right)\|x_{k+1} - x_k\|^2$$

$$\leq f(x_k) - \underbrace{\left(\frac{1}{\tau} - \frac{L}{2}\right)}_{=:a>0}\|x_{k+1} - x_k\|^2 \quad \rightsquigarrow \text{condition: } \tau < \frac{2}{L}$$

Proof of Relative Error Condition:

$$\|\nabla f(x_{k+1})\| \leq \underbrace{\|\nabla f(x_k)\|}_{=\|x_{k+1}-x_k\|/\tau} + \|\nabla f(x_{k+1}) - \nabla f(x_k)\| \leq \underbrace{\left(\frac{1}{\tau} + L\right)}_{=:b>0}\|x_{k+1} - x_k\|\,.$$

Abstract descent algorithms [Attouch et al. 2013]

Abstract descent algorithms: [Attouch et al. 2013]

$$\min_{z\in\mathbb{R}^N} F(z)\,, \quad F\colon \mathbb{R}^N \to (-\infty,\infty] \text{ proper, lsc.}$$

Let $a, b > 0$ fixed. $(z_k)_{k\in\mathbb{N}}$ is a *gradient-like sequence*, if ....

(h1) (**Sufficient decrease condition**). For each $k \in \mathbb{N}$,

$$F(z_{k+1}) + a\|z_{k+1} - z_k\|^2 \le F(z_k)\,;$$

(h2) (**Relative error condition**). For each $k \in \mathbb{N}$,

$$\|\partial F(z_{k+1})\|_- \le b\|z_{k+1} - z_k\|\,;$$

(h3) (**Continuity condition**). There exists $K \subset \mathbb{N}$ and $\tilde{z}$ such that

$$z_k \to \tilde{z} \quad \text{and} \quad F(z_k) \to F(\tilde{z}) \quad \text{as } k \underset{k\in K}{\to} \infty\,.$$

An abstract convergence theorem

Theorem: [Attouch et al. 2013]

If $(z_k)_{k\in\mathbb{N}}$ is

- gradient-like sequence w.r.t. F and
- F has Kurdyka-Łojasiewicz-property at $\tilde{z}$,

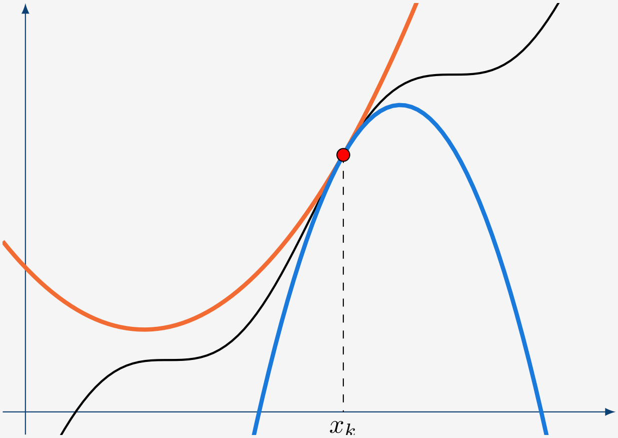
then

- $(z_k)_{k\in\mathbb{N}}$ converges to $\tilde{z}$
- $\tilde{z}$ is a critical point of F, i.e., $0 \in \partial F(\tilde{z})$, and
- $(z_k)_{k\in\mathbb{N}}$ has a finite length, i.e.,

$$\sum_{k=0}^{\infty} \|z_{k+1} - z_k\| < +\infty\,.$$

$\rightsquigarrow$ **nearly any function in practice** (*excludes many pathological cases.*) [Łojasiewicz '63], [Kurdyka '98], [Bolte, Daniilidis, Lewis, Shiota 2007], [Attouch, Bolte, Redont, Soubeyran 2010]

Different Constants for Tight Upper and Lower Quadratic Bounds:



$$-\frac{\underline{L}}{2}\|x - x_k\|^2 \leq f(x) - f(x_k) - \langle \nabla f(x_k), x - x_k \rangle \leq \frac{\overline{L}}{2}\|x - x_k\|^2$$

Proof of Sufficient Decrease Condition: (revisited)

$$f(x_{k+1}) \overset{\text{quad. upper bound}}{\leq} f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{\overline{L}}{2} \|x_{k+1} - x_k\|^2$$

$$\leq f(x_k) + \underbrace{\langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{1}{2\tau} \|x_{k+1} - x_k\|^2}_{(\tau^{-1}\text{-strongly convex}) \quad \leq -\frac{1}{2\tau}\|x_{k+1}-x_k\|^2} - \Big(\frac{1}{2\tau} - \frac{\overline{L}}{2}\Big) \|x_{k+1} - x_k\|^2$$

$$\leq f(x_k) - \underbrace{\Big(\frac{1}{\tau} - \frac{\overline{L}}{2}\Big)}_{=:a>0} \|x_{k+1} - x_k\|^2 \quad \rightsquigarrow \text{condition: } \tau < \frac{2}{\overline{L}}$$

Proof of Relative Error Condition: $L = \max\{\underline{L}, \overline{L}\}$

$$\|\nabla f(x_{k+1})\| \leq \underbrace{\|\nabla f(x_k)\|}_{=\|x_{k+1}-x_k\|/\tau} + \|\nabla f(x_{k+1}) - \nabla f(x_k)\| \leq \underbrace{\Big(\frac{1}{\tau} + L\Big)}_{=:b>0} \|x_{k+1} - x_k\| \,.$$

Conclusion

Conclusion for Gradient Descent:

- need Lipschitz continuity of ∇f to prove convergence
- larger steps possible if $\overline{L} > \underline{L}$
- lower bound has no influence (on the step size), even if

$$\underline{L} = 0 \quad \rightsquigarrow \quad f \text{ convex}$$

- lower bound matters for convergence rate (dictated by the problem)

Goal: explore lower bound algorithmically

$\rightsquigarrow$ **inertial algorithms**

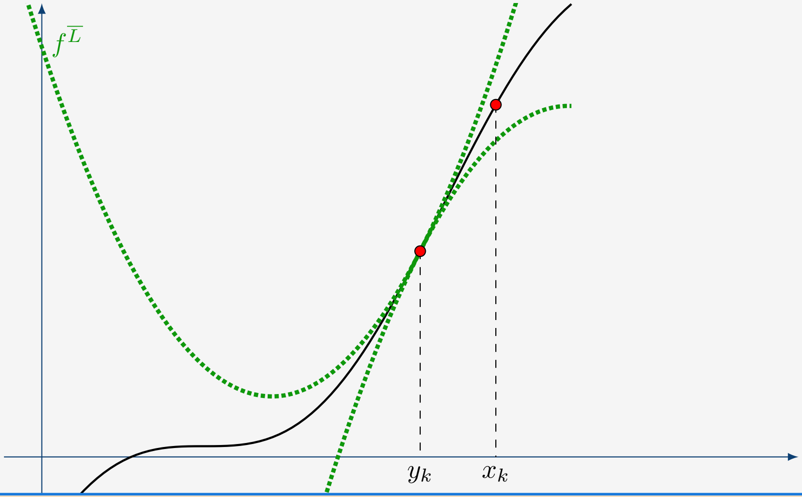
Gradient Descent with Nesterov Extrapolation:

$$y_k = x_k + \gamma_k (x_k - x_{k-1})$$

$$x_{k+1} = y_k - \tau \nabla f(y_k) = \operatorname*{argmin}_x \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2\tau} \|x - y_k\|^2$$

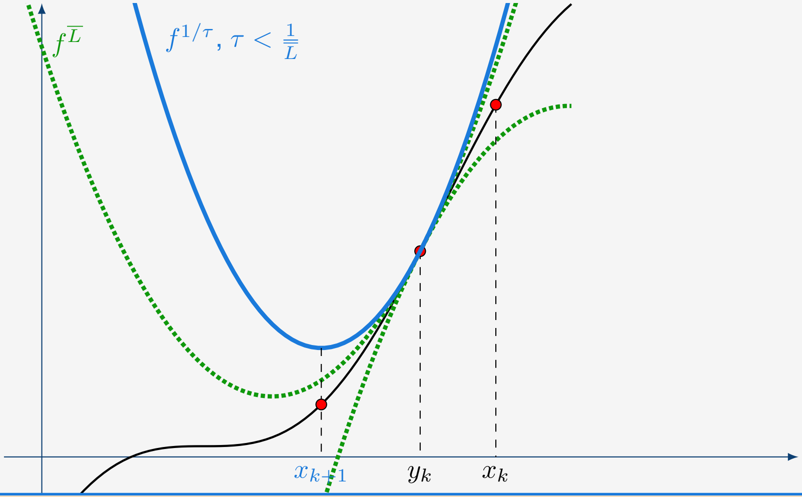
Nesterov's Extrapolation: Sufficient decrease

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \underbrace{f(y_k) + \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2\tau}\|x - y_k\|^2}_{=: f^{1/\tau}}, \quad y_k = x_k + \gamma_k(x_k - x_{k-1})$$



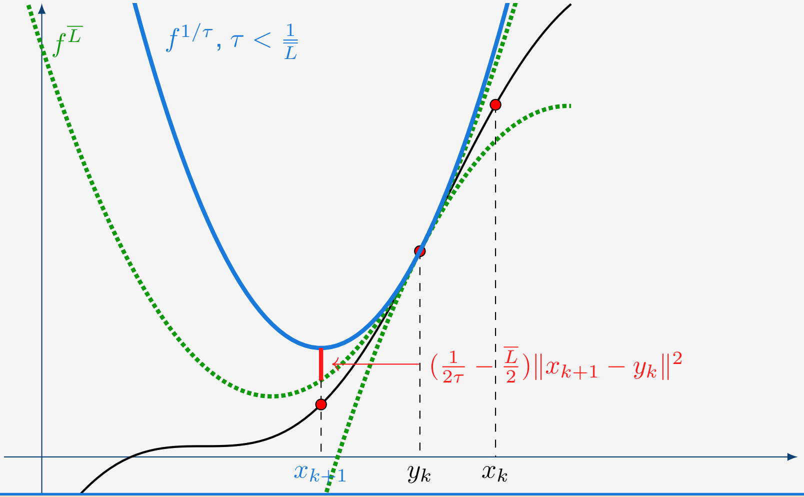
Nesterov's Extrapolation: Sufficient decrease

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \underbrace{f(y_k) + \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2\tau}\|x - y_k\|^2}_{=: f^{1/\tau}}, \quad y_k = x_k + \gamma_k(x_k - x_{k-1})$$



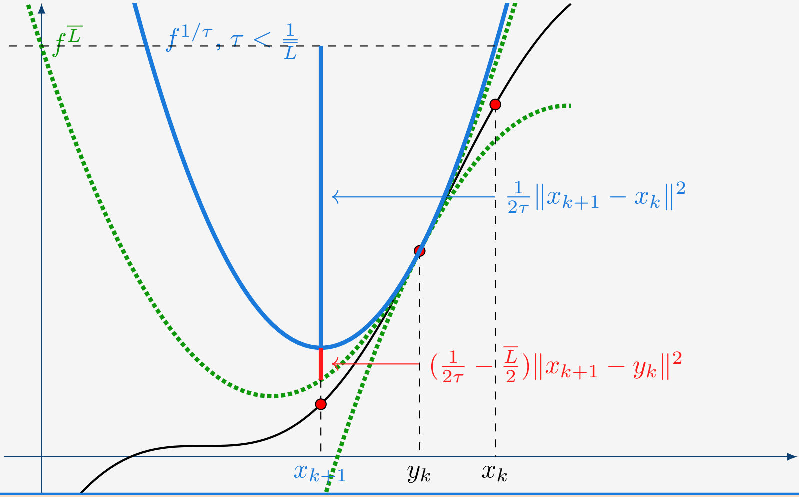
Nesterov's Extrapolation: Sufficient decrease

$$x_{k+1} = \underset{x}{\text{argmin}} \underbrace{f(y_k) + \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2\tau}\|x - y_k\|^2}_{=: f^{1/\tau}}, \quad y_k = x_k + \gamma_k(x_k - x_{k-1})$$



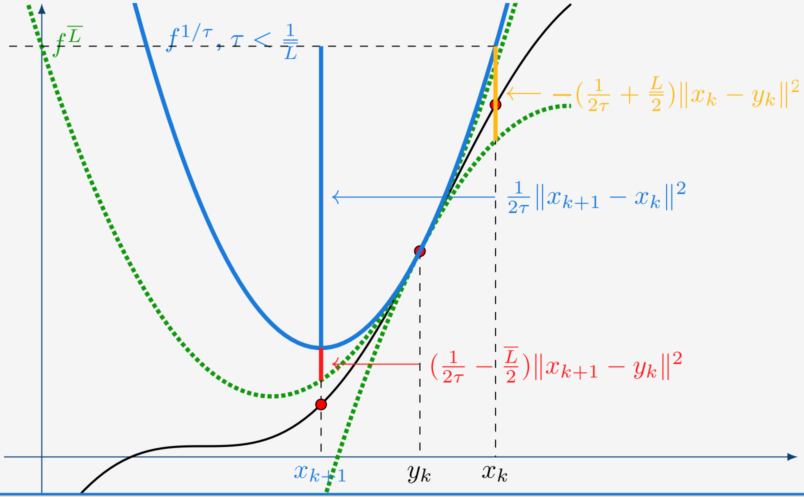
Nesterov's Extrapolation: Sufficient decrease

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \underbrace{f(y_k) + \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2\tau} \|x - y_k\|^2}_{=: f^{1/\tau}}, \quad y_k = x_k + \gamma_k (x_k - x_{k-1})$$



Nesterov's Extrapolation: Sufficient decrease

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \underbrace{f(y_k) + \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2\tau}\|x - y_k\|^2}_{=: f^{1/\tau}}, \quad y_k = x_k + \gamma_k (x_k - x_{k-1})$$



Proof of Convergence for Nesterov's Extrapolation

Gradient Descent with Nesterov Extrapolation:

$$y_k = x_k + \gamma_k(x_k - x_{k-1})$$
$$x_{k+1} = y_k - \tau\nabla f(y_k) = \operatorname*{argmin}_x \, \langle \nabla f(y_k), x - y_k\rangle + \frac{1}{2\tau}\|x - y_k\|^2$$

Proof of Sufficient Descrease Condition:

$$f(x_{k+1}) \overset{\text{quad. upper bound}}{\leq} f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k\rangle + \frac{\overline{L}}{2}\|x_{k+1} - y_k\|^2$$
$$\leq f(y_k) + \underbrace{\langle \nabla f(y_k), x_{k+1} - y_k\rangle + \frac{1}{2\tau}\|x_{k+1} - y_k\|^2}_{\leq \langle \nabla f(y_k), x_k - y_k\rangle + \frac{1}{2\tau}\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_k - x_{k+1}\|^2} - \underbrace{\left(\frac{1}{2\tau} - \frac{\overline{L}}{2}\right)}_{\geq 0}\|x_{k+1} - y_k\|^2$$
$$\overset{\text{quad. lower bound}}{\leq} f(x_k) + \left(\frac{\underline{L}}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_k - x_{k+1}\|^2$$

- From previous slide:

$$f(x_{k+1}) \leq f(x_k) + \left(\frac{L}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_k - x_{k+1}\|^2 .$$

- Define $F_\tau(x_{k+1}, x_k) := f(x_{k+1}) + \frac{1}{2\tau}\|x_{k+1} - x_k\|^2$, then this is equivalent to

$$F_\tau(x_{k+1}, x_k) \leq F_\tau(x_k, x_{k-1}) + \left(\frac{L}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_{k-1} - x_k\|^2$$

Proof of Convergence for Nesterov's Extrapolation

- From previous slide:

$$f(x_{k+1}) \leq f(x_k) + \left(\frac{L}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_k - x_{k+1}\|^2 .$$

- Define $F_\tau(x_{k+1}, x_k) := f(x_{k+1}) + \frac{1}{2\tau}\|x_{k+1} - x_k\|^2$, then this is equivalent to

$$F_\tau(x_{k+1}, x_k) \leq F_\tau(x_k, x_{k-1}) + \left(\frac{L}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_{k-1} - x_k\|^2$$

Condition: $\left(\frac{L}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_{k-1} - x_k\|^2 \leq -\frac{\varepsilon}{2}\|x_k - x_{k-1}\|^2$

Proof of Convergence for Nesterov's Extrapolation

- From previous slide:

$$f(x_{k+1}) \le f(x_k) + \left(\frac{L}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_k - x_{k+1}\|^2 .$$

- Define $F_\tau(x_{k+1}, x_k) := f(x_{k+1}) + \frac{1}{2\tau}\|x_{k+1} - x_k\|^2$, then this is equivalent to

$$F_\tau(x_{k+1}, x_k) \le F_\tau(x_k, x_{k-1}) + \left(\frac{L}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_{k-1} - x_k\|^2$$

Condition: $\left(\frac{L}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_{k-1} - x_k\|^2 \le -\frac{\varepsilon}{2}\|x_k - x_{k-1}\|^2$

⇝ **Sufficient Decrease Condition:**

$$F(z_{k+1}) \le F(z_k) - a\|z_{k+1} - z_k\|^2 , \qquad z_{k+1} = (x_{k+1}, x_k)$$

Proof of Convergence for Nesterov's Extrapolation

Condition: $\left(\frac{\underline{L}}{2}+\frac{1}{2\tau}\right)\|x_k-y_k\|^2-\frac{1}{2\tau}\|x_{k-1}-x_k\|^2 \le -\frac{\varepsilon}{2}\|x_k-x_{k-1}\|^2$

Example: $(\varepsilon = 0)$

- For $y_k = x_k + \gamma_k(x_k - x_{k-1})$, i.e., $\|x_k - y_k\|^2 = \gamma_k^2\|x_k - x_{k-1}\|^2$, this is simply

$$\sup_k \gamma_k^2\left(\frac{\underline{L}}{2}+\frac{1}{2\tau}\right) < \frac{1}{2\tau} \iff \sup_k \gamma_k^2 < \frac{\tau^{-1}}{\underline{L}+\tau^{-1}} \overset{\boxed{\tau^{-1}=\overline{L}}}{=} \frac{\overline{L}}{\underline{L}+\overline{L}} \overset{\boxed{\underline{L}=\overline{L}}}{=} \frac{1}{2}.$$

[Wen, Chen, Pong 2017]

- For f convex, i.e., $\underline{L} = 0$, we obtain convergence for $\sup_{k\in\mathbb{N}} \gamma_k < 1$.

Proof of Convergence for Nesterov's Extrapolation

Proof of Relative Error Condition: $L = \max\{\underline{L}, \overline{L}\}$

$$
\begin{aligned}
\|\nabla F_\tau(x_{k+1}, x_k)\| &\leq \|\nabla f(x_{k+1})\| + \frac{2}{\tau}\|x_{k+1} - x_k\| \\
&\leq \|\nabla f(y_k)\| + \|\nabla f(x_{k+1}) - \nabla f(y_k)\| + \frac{2}{\tau}\|x_{k+1} - x_k\| \\
&\leq \frac{1}{\tau}\|x_{k+1} - y_k\| + L\|x_{k+1} - y_k\| + \frac{2}{\tau}\|x_{k+1} - x_k\| \\
&\leq \left(\frac{1}{\tau} + L\right)\Big(\|x_{k+1} - x_k\| + \gamma_k\|x_k - x_{k-1}\|\Big) + \frac{2}{\tau}\|x_{k+1} - x_k\| \\
&\leq b\|z_{k+1} - z_k\|
\end{aligned}
$$

$\rightsquigarrow$ Conclude **convergence to a stationary point** using [Attouch et al. 2013].

Convex–Convace Inertial Backtracking (CoCaIn):

- **Goal:** Choose τ_k and γ_k adaptively (and sequentially)

$$y_k = x_k + \gamma_k(x_k - x_{k-1})$$

$$x_{k+1} = y_k - \tau_k \nabla f(y_k) = \operatorname*{argmin}_x \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2\tau_k}\|x - y_k\|^2$$

such that the following holds

$$f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{\overline{L}_k}{2}\|x_{k+1} - y_k\|^2$$

$$f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle - \frac{\underline{L}_k}{2}\|x_k - y_k\|^2$$

and

$$\left(\frac{\underline{L}_k}{2} + \frac{1}{2\tau}\right)\|x_k - y_k\|^2 - \frac{1}{2\tau}\|x_{k-1} - x_k\|^2 \leq -\frac{\varepsilon}{2}\|x_k - x_{k-1}\|^2$$

Convex–Convace Inertial Backtracking (CoCaIn):

- **Goal:** Choose τ_k and γ_k adaptively (and sequentially)

$$y_k = x_k + \gamma_k(x_k - x_{k-1})$$

$$x_{k+1} = y_k - \tau_k \nabla f(y_k) = \underset{x}{\operatorname{argmin}} \ \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2\tau_k} \|x - y_k\|^2$$

such that the following holds

$$f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{\overline{L}_k}{2} \|x_{k+1} - y_k\|^2$$

$$f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle - \frac{\underline{L}_k}{2} \|x_k - y_k\|^2$$

and (for simplicity, set $\tau_k = \overline{L}_k^{-1}$)

$$\|x_k - y_k\|^2 \leq \frac{\overline{L}_k - \varepsilon}{\overline{L}_k + \underline{L}_k} \|x_k - x_{k-1}\|^2$$

Convex–Convace Inertial Backtracking (CoCaIn):

- **Goal:** Choose τ_k and γ_k adaptively (and sequentially)

$$y_k = x_k + \gamma_k(x_k - x_{k-1})$$
$$x_{k+1} = y_k - \tau_k \nabla f(y_k) = \operatorname*{argmin}_x \langle \nabla f(y_k), x - y_k \rangle + \frac{1}{2\tau_k}\|x - y_k\|^2$$

such that the following holds (choose $\overline{L}_k \geq \overline{L}_{k-1}$)

$$f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{\overline{L}_k}{2}\|x_{k+1} - y_k\|^2$$
$$f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle - \frac{\underline{L}_k}{2}\|x_k - y_k\|^2$$

and (for simplicity, set $\tau_k = \overline{L}_k^{-1}$)

$$\|x_k - y_k\|^2 \overset{!}{\leq} \frac{\overline{L}_{k-1} - \varepsilon}{\overline{L}_{k-1} + \underline{L}_k}\|x_k - x_{k-1}\|^2 \overset{\text{incr. in } \overline{L} > \varepsilon}{\leq} \frac{\overline{L}_k - \varepsilon}{\overline{L}_k + \underline{L}_k}\|x_k - x_{k-1}\|^2$$

Convex–Convace Inertial Backtracking (CoCaIn):

- Find $\underline{L}_k$ and γ_k such that

$$y_k = x_k + \gamma_k(x_k - x_{k-1})$$

satisfies

$$\begin{cases} f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle - \frac{\underline{L}_k}{2} \|x_k - y_k\|^2 \\ \|x_k - y_k\|^2 \leq \frac{\overline{L}_{k-1} - \varepsilon}{\overline{L}_{k-1} + \underline{L}_k} \|x_k - x_{k-1}\|^2 \,. \end{cases}$$

- Find $\overline{L}_k \geq \overline{L}_{k-1}$ such that

$$x_{k+1} = \operatorname*{argmin}_x \, \langle \nabla f(y_k), x - y_k \rangle + \frac{\overline{L}_k}{2} \|x - y_k\|^2$$

satisfies

$$f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{\overline{L}_k}{2} \|x_{k+1} - y_k\|^2 \,.$$

Convex–Convace Inertial Backtracking (CoCaIn):

- Find $\underline{L}_k$ and γ_k such that

$$y_k = x_k + \gamma_k (x_k - x_{k-1})$$

satisfies

$$\begin{cases} f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle - \frac{\underline{L}_k}{2} \|x_k - y_k\|^2 \\ \|x_k - y_k\|^2 \leq \frac{\overline{L}_{k-1} - \varepsilon}{\overline{L}_{k-1} + \underline{L}_k} \|x_k - x_{k-1}\|^2 , \quad \text{e.g. } \gamma = \sqrt{\frac{\overline{L}_{k-1} - \varepsilon}{\overline{L}_{k-1} + \underline{L}_k}} \,. \end{cases}$$

- Find $\overline{L}_k \geq \overline{L}_{k-1}$ such that

$$x_{k+1} = \operatorname*{argmin}_x \, \langle \nabla f(y_k), x - y_k \rangle + \frac{\overline{L}_k}{2} \|x - y_k\|^2$$

satisfies

$$f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{\overline{L}_k}{2} \|x_{k+1} - y_k\|^2 \,.$$

Convex–Convace Inertial Backtracking (CoCaIn):

- Find $\underline{L}_k$ and γ_k such that

$$y_k = x_k + \gamma_k(x_k - x_{k-1})$$

satisfies

$$\begin{cases} f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle - \dfrac{\underline{L}_k}{2} \|x_k - y_k\|^2 \\ \|x_k - y_k\|^2 \leq \dfrac{\overline{L}_{k-1} - \varepsilon}{\overline{L}_{k-1} + \underline{L}_k} \|x_k - x_{k-1}\|^2 \,. \end{cases}$$

- Find $\overline{L}_k \geq \overline{L}_{k-1}$ such that

$$x_{k+1} = \operatorname*{argmin}_x \, \langle \nabla f(y_k), x - y_k \rangle + \frac{\overline{L}_k}{2} \|x - y_k\|^2$$

satisfies

$$f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{\overline{L}_k}{2} \|x_{k+1} - y_k\|^2 \,.$$

Convex–Convace Inertial Backtracking (CoCaIn): $\min_x \underbrace{g(x)}_{\text{convex, non-smooth, simple}} + f(x)$

- Find $\underline{L}_k$ and γ_k such that

$$y_k = x_k + \gamma_k(x_k - x_{k-1})$$

satisfies

$$\begin{cases} f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle - \frac{\underline{L}_k}{2} \|x_k - y_k\|^2 \\ \|x_k - y_k\|^2 \leq \frac{\overline{L}_{k-1} - \varepsilon}{\overline{L}_{k-1} + \underline{L}_k} \|x_k - x_{k-1}\|^2 . \end{cases}$$

- Find $\overline{L}_k \geq \overline{L}_{k-1}$ such that

$$x_{k+1} = \operatorname*{argmin}_x g(x) + \langle \nabla f(y_k), x - y_k \rangle + \frac{\overline{L}_k}{2} \|x - y_k\|^2$$

satisfies

$$f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{\overline{L}_k}{2} \|x_{k+1} - y_k\|^2 .$$

Convex–Convace Inertial Backtracking (CoCaIn): $\min_x \underbrace{g(x)}_{\substack{\text{convex}\\\text{non-smooth}\\\text{simple}}} + f(x)$

- Find $\underline{L}_k$ and γ_k such that

$$y_k = x_k + \gamma_k(x_k - x_{k-1})$$

satisfies

$$\begin{cases} f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle - \frac{\underline{L}_k}{2} \|x_k - y_k\|^2 \\ \|x_k - y_k\|^2 \leq \frac{\overline{L}_{k-1} - \varepsilon}{\overline{L}_{k-1} + \underline{L}_k} \|x_k - x_{k-1}\|^2 . \end{cases}$$

- Find $\overline{L}_k \geq \overline{L}_{k-1}$ such that

$$x_{k+1} = \operatorname*{argmin}_x \, g(x) + \langle \nabla f(y_k), x - y_k \rangle + \frac{\overline{L}_k}{2} \|x - y_k\|^2$$

satisfies

$$f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \frac{\overline{L}_k}{2} \|x_{k+1} - y_k\|^2 .$$

Convex–Convace Inertial Backtracking (CoCaIn)

Convex–Convace Inertial Backtracking (CoCaIn): $\min_x \underbrace{g(x)}_{\substack{\text{convex}\\\text{non-smooth}\\\text{simple}}} + f(x)$

- Find $\underline{L}_k$ and γ_k such that

$$y_k = x_k + \gamma_k(x_k - x_{k-1})$$

satisfies

$$\begin{cases} f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle - \underline{L}_k D_h(x_k, y_k) \\ D_h(x_k, y_k) \leq \frac{\overline{L}_{k-1} - \varepsilon}{\overline{L}_{k-1} + \underline{L}_k} D_h(x_{k-1}, x_k) \,. \end{cases}$$

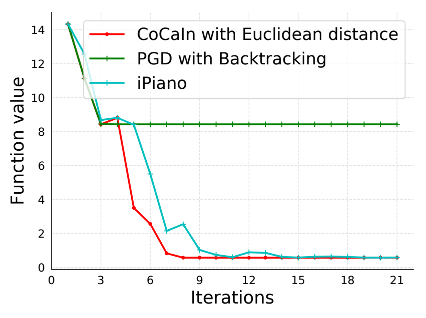
- Find $\overline{L}_k \geq \overline{L}_{k-1}$ such that

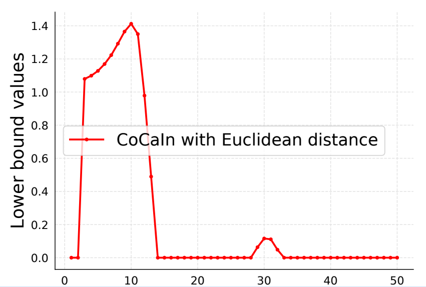
$$x_{k+1} = \underset{x}{\operatorname{argmin}}\; g(x) + \langle \nabla f(y_k), x - y_k \rangle + \overline{L}_k D_h(x, y_k)$$

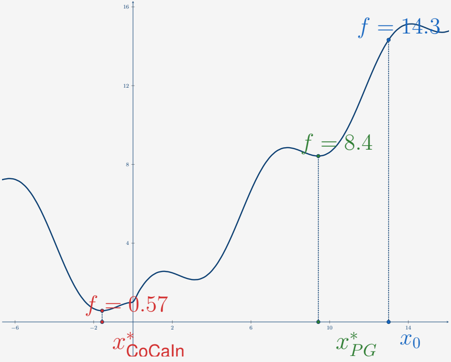
satisfies

$$f(x_{k+1}) \leq f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \overline{L}_k D_h(x_{k+1}, y_k) \,.$$

Simple Function: adapt to "local convexity"



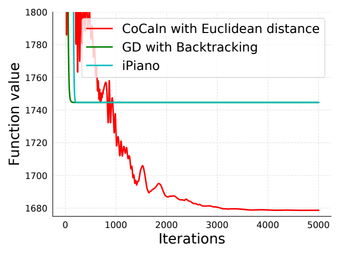


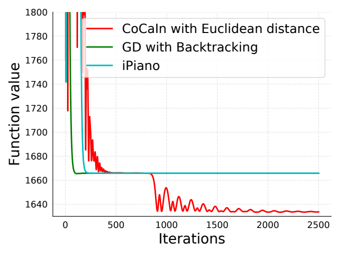


Matrix Factorization

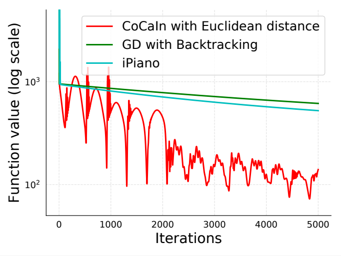
$$\min_{U,Z} \frac{1}{2}\|A - UZ\|_F^2 + \lambda F_1(U) + \lambda F_2(Z)$$

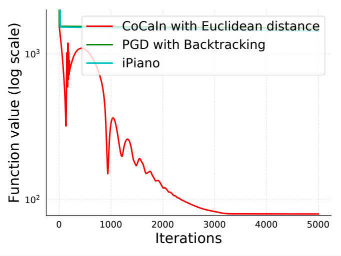
toy experiment





Medulloblastoma dataset [Brunet, Tamayo, Golub, Mesirov 2004]





L2 regularization

L1 regularization

CoCaIn Bregman Proximal Gradient Algorithm

- Gradient Descent only explores quadratic upper bounds.
- **Inertial algorithms** (Nesterov extrapolation) require a **lower bound**.
- Lower bound can "**detect local convexity**" to speed up.
- Presented an efficient **convex–concave backtracking strategy**.
- Same story in **non-smooth (additive composite setting)** and suitable class of **Bregman distances**.