FISTA: from PEP or Lyapunov Function

Zhenghao Xu

Zhejiang University

July 18, 2021

Table of Contents

Table of Contents

FISTA

- Original paper [Beck and Teboulle, 2009]
- PEP analysis [Kim and Fessler, 2018]

FISTA method

Replace the gradient step in Nesterov's FGM with proximal step [Beck and Teboulle, 2009]

FISTA with constant stepsize

$$x_k = P_L(y_k) \ (\textit{proximal step})$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

$$y_{k+1} = x_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(x_k - x_{k-1})$$

Definitions

Objective function: $F(x) = f(x) + \phi(x)$.

- $f(x) \in \mathcal{F}_{0,L}$
- $\phi(x) \in \mathcal{F}_{0,\infty}$

Definitions

Objective function: $F(x) = f(x) + \phi(x)$.

- $f(x) \in \mathcal{F}_{0,L}$
- $\phi(x) \in \mathcal{F}_{0,\infty}$

Proximal step:

$$\begin{aligned} \mathrm{p}_L(y) &= \arg\min_x \left\{ f(y) + \langle x - y, \nabla f(y) \rangle + \frac{L}{2} \|x - y\|^2 + \phi(x) \right\} \\ &= \arg\min_x \left\{ \frac{L}{2} \left\| x - \left(y - \frac{1}{L} \nabla f(y) \right) \right\|^2 + \phi(x) \right\} \end{aligned}$$

Definitions

Objective function: $F(x) = f(x) + \phi(x)$.

- $f(x) \in \mathcal{F}_{0,L}$
- $\phi(x) \in \mathcal{F}_{0,\infty}$

Proximal step:

$$\begin{aligned} \mathrm{p}_L(y) &= \arg\min_x \left\{ f(y) + \langle x - y, \nabla f(y) \rangle + \frac{L}{2} \|x - y\|^2 + \phi(x) \right\} \\ &= \arg\min_x \left\{ \frac{L}{2} \left\| x - \left(y - \frac{1}{L} \nabla f(y) \right) \right\|^2 + \phi(x) \right\} \end{aligned}$$

Composite gradient mapping:

$$\tilde{\nabla}_L F(x) = -L(\mathrm{p}_L(x) - x)$$

Definitions

Composite gradient mapping:

$$\tilde{\nabla}_L F(x) = -L(\mathrm{p}_L(x) - x)$$

Key (in)equalities:

$$\tilde{\nabla}_L F(x) = \nabla f(x) + \phi'(\mathrm{p}_L(x))$$

$$\left\| F'(\mathrm{p}_L(x)) \right\| \leq 2 \left\| \tilde{\nabla}_L F(\mathrm{p}_L(x)) \right\|$$

Algorithm PGM

Input: $f \in \mathcal{F}_L(\mathbb{R}^d)$, $\boldsymbol{x}_0 \in \mathbb{R}^d$.

For $i = 0, \ldots, N-1$

$\quad \boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{x}_i)$

$$F(x_N) - F(x_*) \leq \frac{LR^2}{2N}$$

$$\min_i \left\|\tilde{\nabla}_L F(x_i)\right\| = \left\|\tilde{\nabla}_L F(x_N)\right\| \leq \frac{2LR}{\sqrt{(N-1)(N+2)}}$$

Methods with Known Bounds

Algorithm FPGM (FISTA)

Input: $f \in \mathcal{F}_L(\mathbb{R}^d),\ \boldsymbol{x}_0 \in \mathbb{R}^d,\ \boldsymbol{y}_0 = \boldsymbol{x}_0,\ t_0 = 1.$

For $i = 0, \dots, N-1$

$$\begin{aligned} \boldsymbol{x}_{i+1} &= \mathbf{p}_L(\boldsymbol{y}_i) \\ t_{i+1} &= \frac{1 + \sqrt{1 + 4t_i^2}}{2} \\ \boldsymbol{y}_{i+1} &= \boldsymbol{x}_{i+1} + \frac{t_i - 1}{t_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) \end{aligned} \tag{2.11}$$

$$F(x_N) - F(x_*) \leq \frac{LR^2}{2t_{N-1}^2} = \frac{2LR^2}{(N+1)^2}$$

for any $t_i^2 = \sum_{j=0}^{i} t_j$, $t_i \geq \frac{i+2}{2}$.

Methods with Known Bounds

To optimize gradient norm:

Algorithm FPGM-m

Input: $f \in \mathcal{F}_L(\mathbb{R}^d)$, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\boldsymbol{y}_0 = \boldsymbol{x}_0$, $t_0 = 1$.

For $i = 0, \ldots, N-1$

$$\boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i)$$

$$t_{i+1} = \frac{1 + \sqrt{1 + 4t_i^2}}{2}, \quad i \le m-1$$

$$\boldsymbol{y}_{i+1} = \begin{cases} \boldsymbol{x}_{i+1} + \frac{t_i - 1}{t_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i), & i \le m-1, \\ \boldsymbol{x}_{i+1}, & \text{otherwise.} \end{cases}$$

$$\min_i \left\| \tilde{\nabla}_L F(x_i) \right\| = \left\| \tilde{\nabla}_L F(x_N) \right\| \le \frac{2LR}{(m+1)\sqrt{N-m+1}}$$

for any $t_i^2 = \sum_{j=0}^{i} t_j$, $t_i \ge \frac{i+2}{2}$. N is given a priori.
$O(1/N^{\frac{3}{2}})$ with $m = \lfloor \frac{2N}{3} \rfloor$.

Methods with Known Bounds

FPGM-σ: Replace $\mathrm{p}_L(\cdot)$ in FPGM with $\mathrm{p}_{L/\sigma^2}(\cdot)$.

$$F(x_N) - F(x_*) \le \frac{LR^2}{2t_{N-1}^2} = \frac{2LR^2}{\sigma^2 N^2}$$

$$\min_i \left\| \tilde{\nabla}_{L/\sigma^2} F(x_i) \right\| \le \frac{2\sqrt{3}}{\sigma} \sqrt{\frac{1+\sigma}{1-\sigma}} \frac{LR}{N^{\frac{3}{2}}}$$

for any $t_i^2 = \sum_{j=0}^{i} t_j$, $t_i \ge \frac{i+2}{2}$. N is is not needed a priori.

Table of Contents

Algorithm Class FSFOM

Input: $f \in \mathcal{F}_L(\mathbb{R}^d)$, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\boldsymbol{y}_0 = \boldsymbol{x}_0$.

For $i = 0, \dots, N-1$

$$\boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i) = \boldsymbol{y}_i - \frac{1}{L}\tilde{\nabla}_L F(\boldsymbol{y}_i)$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i + \sum_{k=0}^{i} h_{i+1,k}\,(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k) = \boldsymbol{y}_i - \frac{1}{L}\sum_{k=0}^{i} h_{i+1,k}\,\tilde{\nabla}_L F(\boldsymbol{y}_k).$$

General First Order PG Method

Algorithm Class FSFOM

Input: $f \in \mathcal{F}_L(\mathbb{R}^d),\ \boldsymbol{x}_0 \in \mathbb{R}^d,\ \boldsymbol{y}_0 = \boldsymbol{x}_0.$

For $i = 0, \ldots, N-1$

$$\boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i) = \boldsymbol{y}_i - \frac{1}{L}\tilde{\nabla}_L F(\boldsymbol{y}_i)$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i + \sum_{k=0}^{i} h_{i+1,k}\,(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k) = \boldsymbol{y}_i - \frac{1}{L}\sum_{k=0}^{i} h_{i+1,k}\,\tilde{\nabla}_L F(\boldsymbol{y}_k).$$

Step coefficients for FPGM (FISTA):

$$h_{i+1,k} = \begin{cases} \frac{1}{t_{i+1}}\left(t_k - \sum_{j=k+1}^{i} h_{j,k}\right), & k = 0, \ldots, i-1 \\ 1 + \frac{t_i - 1}{t_{i+1}}, & k = i. \end{cases}$$

PEP Formulation

Algorithm Class FSFOM

Input: $f \in \mathcal{F}_L(\mathbb{R}^d)$, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\boldsymbol{y}_0 = \boldsymbol{x}_0$.

For $i = 0, \ldots, N-1$

$$\boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i) = \boldsymbol{y}_i - \frac{1}{L}\tilde{\nabla}_L F(\boldsymbol{y}_i)$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i + \sum_{k=0}^{i} h_{i+1,k}\,(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k) = \boldsymbol{y}_i - \frac{1}{L}\sum_{k=0}^{i} h_{i+1,k}\,\tilde{\nabla}_L F(\boldsymbol{y}_k).$$

Corresponding PEP:

$$\mathcal{B}_{\mathrm{P}}(\boldsymbol{h}, N, d, L, R) := \max_{\substack{F \in \mathcal{F}_L(\mathbb{R}^d), \\ \boldsymbol{x}_0, \cdots, \boldsymbol{x}_N \in \mathbb{R}^d,\ \boldsymbol{x}_* \in X_*(F) \\ \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{N-1} \in \mathbb{R}^d}} F(\boldsymbol{x}_N) - F(\boldsymbol{x}_*) \tag{P}$$

$$\text{s.t. } \boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i), \quad i = 0, \ldots, N-1, \quad ||\boldsymbol{x}_0 - \boldsymbol{x}_*|| \leq R,$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i + \sum_{k=0}^{i} h_{i+1,k}(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k), \quad i = 0, \ldots, N-2.$$

PEP Simplification

Relaxation inequality:

$$\frac{L}{2}||\mathbf{p}_L(\boldsymbol{y}) - \boldsymbol{y}||^2 - L\left\langle \mathbf{p}_L(\boldsymbol{x}) - \boldsymbol{x},\ \mathbf{p}_L(\boldsymbol{y}) - \boldsymbol{y} \right\rangle$$
$$\leq F(\mathbf{p}_L(\boldsymbol{x})) - F(\mathbf{p}_L(\boldsymbol{y})) + L\left\langle \mathbf{p}_L(\boldsymbol{y}) - \boldsymbol{y},\ \boldsymbol{x} - \boldsymbol{y} \right\rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d \quad (3.2)$$

PEP Simplification

Relaxation inequality:

$$\frac{L}{2}||\mathbf{p}_L(\boldsymbol{y}) - \boldsymbol{y}||^2 - L\left\langle \mathbf{p}_L(\boldsymbol{x}) - \boldsymbol{x},\ \mathbf{p}_L(\boldsymbol{y}) - \boldsymbol{y}\right\rangle \\ \leq F(\mathbf{p}_L(\boldsymbol{x})) - F(\mathbf{p}_L(\boldsymbol{y})) + L\left\langle \mathbf{p}_L(\boldsymbol{y}) - \boldsymbol{y},\ \boldsymbol{x} - \boldsymbol{y}\right\rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d \quad (3.2)$$

Relaxed PEP:

$$\begin{aligned} \mathcal{B}_{\mathrm{P1}}(\boldsymbol{h}, N, d, L, R) := \max_{\substack{\boldsymbol{G}\in\mathbb{R}^{N\times d},\\ \boldsymbol{\delta}\in\mathbb{R}^N}} \ & LR^2\delta_{N-1} \\ \text{s.t.}\ \ & \mathrm{Tr}\left\{\boldsymbol{G}^\top \check{\boldsymbol{A}}_{i-1,i}(\boldsymbol{h})\boldsymbol{G}\right\} \leq \delta_{i-1} - \delta_i, \quad i = 1, \ldots, N-1, \\ & \mathrm{Tr}\left\{\boldsymbol{G}^\top \check{\boldsymbol{D}}_i(\boldsymbol{h})\boldsymbol{G} + \boldsymbol{\nu}\boldsymbol{u}_i^\top \boldsymbol{G}\right\} \leq -\delta_i, \quad i = 0, \ldots, N-1, \end{aligned} \qquad (\mathrm{P1})$$

$$\begin{cases} \boldsymbol{g}_i := -\frac{1}{||\boldsymbol{y}_0 - \boldsymbol{x}_*||}(\mathbf{p}_L(\boldsymbol{y}_i) - \boldsymbol{y}_i) = \frac{1}{L||\boldsymbol{y}_0 - \boldsymbol{x}_*||}\tilde{\nabla}_L F(\boldsymbol{y}_i), \\ \delta_i := \frac{1}{L||\boldsymbol{y}_0 - \boldsymbol{x}_*||^2}(F(\mathbf{p}_L(\boldsymbol{y}_i)) - F(\boldsymbol{x}_*)), \end{cases}$$

$$\begin{cases} \check{\boldsymbol{A}}_{i-1,i}(\boldsymbol{h}) := \frac{1}{2}\boldsymbol{u}_i\boldsymbol{u}_i^\top - \frac{1}{2}\boldsymbol{u}_{i-1}\boldsymbol{u}_i^\top - \frac{1}{2}\boldsymbol{u}_i\boldsymbol{u}_{i-1}^\top + \frac{1}{2}\sum_{k=0}^{i-1} h_{i,k}(\boldsymbol{u}_i\boldsymbol{u}_k^\top + \boldsymbol{u}_k\boldsymbol{u}_i^\top), \\ \check{\boldsymbol{D}}_i(\boldsymbol{h}) := \frac{1}{2}\boldsymbol{u}_i\boldsymbol{u}_i^\top + \frac{1}{2}\sum_{j=1}^{i}\sum_{k=0}^{j-1} h_{j,k}(\boldsymbol{u}_i\boldsymbol{u}_k^\top + \boldsymbol{u}_k\boldsymbol{u}_i^\top), \end{cases} \qquad (3.5)$$

Dual PEP

Dual problem:

$$\mathcal{B}_{\mathrm{D}}(\boldsymbol{h}, N, L, R) := \min_{\substack{(\boldsymbol{\lambda},\boldsymbol{\tau})\in\Lambda,\\ \gamma\in\mathbb{R}}} \left\{ \frac{1}{2}LR^2\gamma \ : \ \begin{pmatrix} \boldsymbol{S}(\boldsymbol{h},\boldsymbol{\lambda},\boldsymbol{\tau}) & \frac{1}{2}\boldsymbol{\tau} \\ \frac{1}{2}\boldsymbol{\tau}^\top & \frac{1}{2}\gamma \end{pmatrix} \succeq 0 \right\}, \tag{D}$$

$$\Lambda := \left\{ (\boldsymbol{\lambda},\boldsymbol{\tau}) \in \mathbb{R}_+^{2N-1} \ : \ \begin{array}{l} \tau_0 = \lambda_1, \quad \lambda_{N-1} + \tau_{N-1} = 1, \\ \lambda_i - \lambda_{i+1} + \tau_i = 0, \ i = 1, \ldots, N-2, \end{array} \right\}, \tag{3.6}$$

$$\boldsymbol{S}(\boldsymbol{h},\boldsymbol{\lambda},\boldsymbol{\tau}) := \sum_{i=1}^{N-1} \lambda_i \check{\boldsymbol{A}}_{i-1,i}(\boldsymbol{h}) + \sum_{i=0}^{N-1} \tau_i \check{\boldsymbol{D}}_i(\boldsymbol{h}). \tag{3.7}$$

Lemma

Lemma 2 *For the following step coefficients:*

$$h_{i+1,k} = \begin{cases} \frac{t_{i+1}}{T_{i+1}} \left(t_k - \sum_{j=k+1}^{i} h_{j,k} \right), & k = 0, \ldots, i-1, \\ 1 + \frac{(t_i - 1)t_{i+1}}{T_{i+1}}, & k = i, \end{cases} \tag{3.8}$$

the choice of variables:

$$\lambda_i = \frac{T_{i-1}}{T_{N-1}}, \quad i = 1, \ldots, N-1, \quad \tau_i = \frac{t_i}{T_{N-1}}, \quad i = 0, \ldots, N-1, \quad \gamma = \frac{1}{T_{N-1}}, \tag{3.9}$$

is a feasible point of (D) *for any choice of* t_i *such that*

$$t_0 = 1, \quad t_i > 0, \quad \text{and} \quad t_i^2 \leq T_i := \sum_{l=0}^{i} t_l. \tag{3.10}$$

FISTA falls into this category with $T_i = t_i^2$.

General FPGM(FISTA)

From lemma we get generalized FPGM(FISTA):

Algorithm GFPGM

Input: $f \in \mathcal{F}_L(\mathbb{R}^d)$, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\boldsymbol{y}_0 = \boldsymbol{x}_0$, $t_0 = T_0 = 1$.

For $i = 0, \ldots, N-1$

$$\boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i)$$

Choose t_{i+1} s.t. $t_{i+1} > 0$ and $t_{i+1}^2 \le T_{i+1} := \sum_{l=0}^{i+1} t_l$

$$\boldsymbol{y}_{i+1} = \boldsymbol{x}_{i+1} + \frac{(T_i - t_i)t_{i+1}}{t_i T_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) + \frac{(t_i^2 - T_i)t_{i+1}}{t_i T_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{y}_i)$$

Produced sequence $\{x_i\}$ are identical (GFPGM and FSFOM with step sizes in lemma 2).

General FPGM(FISTA)

From lemma we get generalized FPGM(FISTA):

Algorithm GFPGM

Input: $f \in \mathcal{F}_L(\mathbb{R}^d)$, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\boldsymbol{y}_0 = \boldsymbol{x}_0$, $t_0 = T_0 = 1$.

For $i = 0, \ldots, N-1$

$$\boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i)$$

Choose t_{i+1} s.t. $t_{i+1} > 0$ and $t_{i+1}^2 \leq T_{i+1} := \sum_{l=0}^{i+1} t_l$

$$\boldsymbol{y}_{i+1} = \boldsymbol{x}_{i+1} + \frac{(T_i - t_i)t_{i+1}}{t_i T_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) + \frac{(t_i^2 - T_i)t_{i+1}}{t_i T_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{y}_i)$$

Produced sequence $\{x_i\}$ are identical (GFPGM and FSFOM with step sizes in lemma 2). Convergence rate from dual objective:

Theorem 3 *Let $F : \mathbb{R}^d \to \mathbb{R}$ be in $\mathcal{F}_L(\mathbb{R}^d)$ and let $\boldsymbol{x}_0, \cdots, \boldsymbol{x}_N \in \mathbb{R}^d$ be generated by GFPGM. Then for $N \geq 1$,*

$$F(\boldsymbol{x}_N) - F(\boldsymbol{x}_*) \leq \frac{LR^2}{2T_{N-1}}. \tag{3.11}$$

General FPGM(FISTA)

Select $t_i = \frac{i+a}{a}$ satisfying conditions:

Corollary 1 *Let* $F : \mathbb{R}^d \to \mathbb{R}$ *be in* $\mathcal{F}_L(\mathbb{R}^d)$ *and let* $\boldsymbol{x}_0, \cdots, \boldsymbol{x}_N \in \mathbb{R}^d$ *be generated by GFPGM with* $t_i = \frac{i+a}{a}$ *(FPGM-a) for any* $a \geq 2$*. Then for* $N \geq 1$*,*

$$F(\boldsymbol{x}_N) - F(\boldsymbol{x}_*) \leq \frac{aLR^2}{N(N+2a-1)}. \tag{3.13}$$

Conjecture on Optimality

Conjecture 1

FISTA is optimal. That is, the feasible point in Lemma 2 with $t_i^2 = T_i$ that corresponds to FISTA is a global minimizer of the dual problem. (Might require further derivations with KKT conditions.)

PEP in Composite Gradient Mapping Form

Above argument is for *cost function* form of PEP.

$$\mathcal{B}_{\mathrm{P}}(\boldsymbol{h}, N, d, L, R) := \max_{\substack{F \in \mathcal{F}_L(\mathbb{R}^d), \\ \boldsymbol{x}_0, \cdots, \boldsymbol{x}_N \in \mathbb{R}^d,\ \boldsymbol{x}_* \in X_*(F) \\ \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{N-1} \in \mathbb{R}^d}} F(\boldsymbol{x}_N) - F(\boldsymbol{x}_*) \tag{P}$$

$$\text{s.t. } \boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i), \quad i = 0, \ldots, N-1, \quad ||\boldsymbol{x}_0 - \boldsymbol{x}_*|| \le R,$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i + \sum_{k=0}^{i} h_{i+1,k}(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k), \quad i = 0, \ldots, N-2.$$

PEP in Composite Gradient Mapping Form

Above argument is for *cost function* form of PEP.

$$\mathcal{B}_{\mathrm{P}}(\boldsymbol{h}, N, d, L, R) := \max_{\substack{F \in \mathcal{F}_L(\mathbb{R}^d), \\ \boldsymbol{x}_0, \cdots, \boldsymbol{x}_N \in \mathbb{R}^d,\ \boldsymbol{x}_* \in X_*(F) \\ \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{N-1} \in \mathbb{R}^d}} F(\boldsymbol{x}_N) - F(\boldsymbol{x}_*) \tag{P}$$

$$\text{s.t. } \boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i), \quad i = 0, \ldots, N-1, \quad ||\boldsymbol{x}_0 - \boldsymbol{x}_*|| \leq R,$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i + \sum_{k=0}^{i} h_{i+1,k}(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k), \quad i = 0, \ldots, N-2.$$

Similarly for *composite gradient mapping* form of PEP.

$$\mathcal{B}_{\mathrm{P}'}(\boldsymbol{h}, N, d, L, R) := \max_{\substack{F \in \mathcal{F}_L(\mathbb{R}^d), \\ \boldsymbol{x}_0, \cdots, \boldsymbol{x}_N \in \mathbb{R}^d,\ \boldsymbol{x}_* \in X_*(F), \\ \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{N-1} \in \mathbb{R}^d}} \min_{\boldsymbol{x} \in \Omega_N} ||L\left(\mathbf{p}_L(\boldsymbol{x}) - \boldsymbol{x}\right)||^2$$

$$\text{s.t. } \boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i), \quad i = 0, \ldots, N-1, \quad ||\boldsymbol{x}_0 - \boldsymbol{x}_*|| \leq R, \tag{P$'$}$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i + \sum_{k=0}^{i} h_{i+1,k}(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k), \quad i = 0, \ldots, N-2.$$

PEP in Composite Gradient Mapping Form

$$
\mathcal{B}_{\mathrm{P}'}(\boldsymbol{h}, N, d, L, R) := \max_{\substack{F\in\mathcal{F}_L(\mathbb{R}^d),\\ \boldsymbol{x}_0,\cdots,\boldsymbol{x}_N\in\mathbb{R}^d,\ \boldsymbol{x}_*\in X_*(F),\\ \boldsymbol{y}_0,\cdots,\boldsymbol{y}_{N-1}\in\mathbb{R}^d}} \min_{\boldsymbol{x}\in\Omega_N} ||L\left(\mathbf{p}_L(\boldsymbol{x}) - \boldsymbol{x}\right)||^2
$$

$$
\text{s.t.}\ \ \boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i), \quad i = 0,\dots,N-1, \quad ||\boldsymbol{x}_0 - \boldsymbol{x}_*|| \le R, \tag{P$'$}
$$

$$
\boldsymbol{y}_{i+1} = \boldsymbol{y}_i + \sum_{k=0}^{i} h_{i+1,k}(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k), \quad i = 0,\dots,N-2.
$$

$\Omega_N = \{y_0, \dots, y_{N-1}, x_N\}$.

- if $\Omega_N = \{y_0, \dots, y_{N-1}\}$, then the bound would be worse.
- if $\Omega_N = \{x_0, \dots, x_{N-1}\}$, then no proven result yet.
- if min is not taken among history composite gradient mapping, then the bound would be $O(1/N)$ (worse).

PEP Derivation

After relaxation and dualization, we get the dual PEP:

$$\mathcal{B}_{\mathrm{D}'}(\boldsymbol{h}, N, L, R) := \min_{\substack{(\boldsymbol{\lambda}, \boldsymbol{\tau}, \eta, \boldsymbol{\beta}) \in \Lambda', \\ \gamma \in \mathbb{R}}} \left\{ \frac{1}{2} L^2 R^2 \gamma \ : \ \begin{pmatrix} \boldsymbol{S}'(\boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \eta, \boldsymbol{\beta}) & \frac{1}{2}[\boldsymbol{\tau}^\top, 0]^\top \\ \frac{1}{2}[\boldsymbol{\tau}^\top, 0] & \frac{1}{2}\gamma \end{pmatrix} \succeq 0 \right\} \tag{D$'$}$$

$$\Lambda' := \left\{ (\boldsymbol{\lambda}, \boldsymbol{\tau}, \eta, \boldsymbol{\beta}) \in \mathbb{R}_+^{3N+1} \ : \ \begin{array}{l} \tau_0 = \lambda_1, \quad \lambda_{N-1} + \tau_{N-1} = \eta, \quad \sum_{i=0}^{N} \beta_i = 1, \\ \lambda_i - \lambda_{i+1} + \tau_i = 0, \ i = 1, \ldots, N-2 \end{array} \right\}, \tag{4.4}$$

$$\boldsymbol{S}'(\boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \eta, \boldsymbol{\beta}) := \sum_{i=1}^{N-1} \lambda_i \bar{\boldsymbol{A}}_{i-1,i}(\boldsymbol{h}) + \sum_{i=0}^{N-1} \tau_i \bar{\boldsymbol{D}}_i(\boldsymbol{h}) + \frac{1}{2} \eta \bar{\boldsymbol{u}}_N \bar{\boldsymbol{u}}_N^\top - \sum_{i=0}^{N} \beta_i \bar{\boldsymbol{u}}_i \bar{\boldsymbol{u}}_i^\top. \tag{4.5}$$

Feasible Solution and Bound

Lemma 3 *For the step coefficients $\{h_{i+1,k}\}$ in (3.8), the choice of variables*

$$\lambda_i = T_{i-1}\tau_0,\quad i=1,\dots,N-1,\quad \tau_i = \begin{cases} \left(\frac{1}{2}\left(\sum_{k=0}^{N-1}\left(T_k - t_k^2\right) + T_{N-1}\right)\right)^{-1}, & i=0,\\ t_i\tau_0, & i=1,\dots,N-1, \end{cases} \tag{4.6}$$

$$\eta = T_{N-1}\tau_0,\quad \beta_i = \begin{cases} \frac{1}{2}\left(T_i - t_i^2\right)\tau_0, & i=0,\dots,N-1,\\ \frac{1}{2}T_{N-1}\tau_0, & i=N, \end{cases}\quad \gamma = \tau_0. \tag{4.7}$$

is a feasible point of (D′) *for any choice of t_i and T_i satisfying* (3.10).

Theorem 4 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be in $\mathcal{F}_L(\mathbb{R}^d)$ and let $\boldsymbol{x}_0,\cdots,\boldsymbol{x}_N,\boldsymbol{y}_0,\cdots,\boldsymbol{y}_{N-1} \in \mathbb{R}^d$ be generated by GFPGM. Then for $N \geq 1$,*

$$\min_{i\in\{0,\dots,N\}} ||\tilde{\nabla}_L F(\boldsymbol{x}_i)|| \leq \min_{\boldsymbol{x}\in\Omega_N} ||\tilde{\nabla}_L F(\boldsymbol{x})|| \leq \frac{LR}{\sqrt{\sum_{k=0}^{N-1}\left(T_k - t_k^2\right) + T_{N-1}}}. \tag{4.8}$$

Conjecture on Optimality

Choose

$$t_i = \begin{cases} 1, & i = 0, \\ \frac{1+\sqrt{1+4t_{i-1}^2}}{2}, & i = 1, \ldots, \left\lfloor \frac{N}{2} \right\rfloor - 1, \\ \frac{N-i+1}{2}, & i = \left\lfloor \frac{N}{2} \right\rfloor, \ldots, N-1, \end{cases} \tag{4.9}$$

Numerical result shows its optimality.

Conjecture 2

Above choice of t_i corresponds to the optimal methods (with respect to composite gradient mapping).

FPGM-OCG

$$t_i = \begin{cases} 1, & i = 0, \\ \frac{1+\sqrt{1+4t_{i-1}^2}}{2}, & i = 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor - 1, \\ \frac{N-i+1}{2}, & i = \left\lfloor \frac{N}{2} \right\rfloor, \dots, N-1, \end{cases} \tag{4.9}$$

Corresponding method (N specified in advance):

Algorithm FPGM-OCG (GFPGM with t_i in (4.9))

Input: $f \in C_L^{1,1}(\mathbb{R}^d)$ convex, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\boldsymbol{y}_0 = \boldsymbol{x}_0$, $t_0 = T_0 = 1$.

For $i = 0, \dots, N-1$

$$\boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i)$$

$$t_{i+1} = \begin{cases} \frac{1+\sqrt{1+4t_i^2}}{2}, & i = 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor - 2, \\ \frac{N-i}{2}, & i = \left\lfloor \frac{N}{2} \right\rfloor - 1, \dots, N-2, \end{cases}$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{x}_{i+1} + \frac{(T_i - t_i)t_{i+1}}{t_i T_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) + \frac{(t_i^2 - T_i)t_{i+1}}{t_i T_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{y}_i), \quad i < N-1$$

FPGM-OCG Bound

Algorithm FPGM-OCG (GFPGM with t_i in (4.9))

Input: $f \in C_L^{1,1}(\mathbb{R}^d)$ convex, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\boldsymbol{y}_0 = \boldsymbol{x}_0$, $t_0 = T_0 = 1$.

For $i = 0, \ldots, N-1$

$$\boldsymbol{x}_{i+1} = \mathbf{p}_L(\boldsymbol{y}_i)$$

$$t_{i+1} = \begin{cases} \frac{1+\sqrt{1+4t_i^2}}{2}, & i = 1, \ldots, \lfloor \frac{N}{2} \rfloor - 2, \\ \frac{N-i}{2}, & i = \lfloor \frac{N}{2} \rfloor - 1, \ldots, N-2, \end{cases}$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{x}_{i+1} + \frac{(T_i - t_i)t_{i+1}}{t_i T_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) + \frac{(t_i^2 - T_i)t_{i+1}}{t_i T_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{y}_i), \quad i < N-1$$

Theorem 5 *Let $F : \mathbb{R}^d \to \mathbb{R}$ be in $\mathcal{F}_L(\mathbb{R}^d)$ and let $\boldsymbol{x}_0, \cdots, \boldsymbol{x}_N, \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{N-1} \in \mathbb{R}^d$ be generated by FPGM-OCG. Then for $N \geq 1$,*

$$F(\boldsymbol{x}_N) - F(\boldsymbol{x}_*) \leq \frac{4L||\boldsymbol{x}_0 - \boldsymbol{x}_*||^2}{N(N+4)}, \tag{4.11}$$

and for $N \geq 3$,

$$\min_{i \in \{0,\ldots,N\}} ||\tilde{\nabla}_L F(\boldsymbol{x}_i)|| \leq \min_{\boldsymbol{x} \in \Omega_N} ||\tilde{\nabla}_L F(\boldsymbol{x})|| \leq \frac{2\sqrt{6}LR}{N\sqrt{N-2}}. \tag{4.12}$$

FPGM-a Bound

Another choice of t_i preventing pre-selection of N:

Corollary 2 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be in $\mathcal{F}_L(\mathbb{R}^d)$ and let $\boldsymbol{x}_0, \cdots, \boldsymbol{x}_N, \boldsymbol{y}_0, \cdots, \boldsymbol{y}_{N-1} \in \mathbb{R}^d$ be generated by GFPGM with $t_i = \frac{i+a}{a}$ (FPGM-a) for any $a \geq 2$. Then for $N \geq 1$, we have the following bound on the (smallest) composite gradient mapping:*

$$\min_{i \in \{0,\ldots,N\}} ||\tilde{\nabla}_L F(\boldsymbol{x}_i)|| \leq \min_{\boldsymbol{x} \in \Omega_N} ||\tilde{\nabla}_L F(\boldsymbol{x})|| \leq \frac{a\sqrt{6}LR}{\sqrt{N((a-2)N^2 + 3(a^2 - a + 1)N + (3a^2 + 2a - 1))}}. \tag{4.14}$$

Theoretic Results

Algorithm	Asymptotic worst-case bound		Require selecting N in advance
	Cost function $(\times LR^2)$	Proximal gradient $(\times LR)$	
PGM	$\frac{1}{2}N^{-1}$	$2N^{-1}$	No
FPGM	$\mathbf{2N^{-2}}$	$2N^{-1}$	No
FPGM-σ $(0 < \sigma < 1)$	$\frac{2}{\sigma^2}N^{-2}$	$\frac{2\sqrt{3}}{\sigma^2}\sqrt{\frac{1+\sigma}{1-\sigma}}N^{-\frac{3}{2}}$	No
FPGM-$(\sigma{=}0.78)$	$3.3N^{-2}$	$16.2N^{-\frac{3}{2}}$	
FPGM-$\left(m{=}\left\lfloor\frac{2N}{3}\right\rfloor\right)$	$4.5N^{-2}$	$5.2N^{-\frac{3}{2}}$	Yes
FPGM-OCG	$4N^{-2}$	$\mathbf{4.9N^{-\frac{3}{2}}}$	Yes
FPGM-a $(a > 2)$	aN^{-2}	$\frac{a\sqrt{6}}{\sqrt{a-2}}N^{-\frac{3}{2}}$	No
FPGM-$(a{=}4)$	$4N^{-2}$	$6.9N^{-\frac{3}{2}}$	

Table 1 Asymptotic worst-case bounds on the cost function $F(\boldsymbol{x}_N) - F(\boldsymbol{x}_*)$ and the norm of the composite gradient mapping $\min_{\boldsymbol{x}\in\Omega_N} ||\tilde{\nabla}_L F(\boldsymbol{x})||$ of PGM, FPGM, FPGM-σ, FPGM-m, FPGM-OCG, and FPGM-a. (The cost function bound for FPGM-m in the table corresponds to the bound for FPGM after m iterations because a tight bound for the final Nth iteration is unknown. The bound on $\min_{i\in\{0,\dots,N\}} ||\tilde{\nabla}_{L/\sigma^2} F(\boldsymbol{y}_i)||$ is used for FPGM-σ.)

Numerical Results

Cost function:

N	PGM	FPGM	FPGM -(σ=0.78)	FPGM -$\left(m=\left\lfloor\frac{2N}{3}\right\rfloor\right)$	FPGM -OCG	FPGM -(a=4)
1	4.00	4.00	2.43	4.00	4.00	4.00
2	8.00	8.00	4.87	8.00	8.00	8.00
4	16.00	19.35	11.77	17.13	17.60	17.23
10	40.00	79.07	48.11	56.47	59.25	55.88
20	80.00	261.66	159.19	163.75	170.10	159.17
30	120.00	546.51	332.49	321.56	331.97	312.03
40	160.00	932.89	567.57	502.37	544.55	514.73
47	188.00	1263.58	768.76	675.68	723.06	686.33
50	200.00	1420.45	864.20	752.90	807.66	767.37
Empi. $O(\cdot)$	$N^{-1.00}$	$N^{-1.89}$	$N^{-1.89}$	$N^{-1.75}$	$N^{-1.79}$	$N^{-1.80}$
Known $O(\cdot)$	N^{-1}	N^{-2}	N^{-2}	N^{-2}	N^{-2}	N^{-2}

Table 2 Tight worst-case bounds on the cost function $LR^2/(F(\boldsymbol{x}_N) - F(\boldsymbol{x}_*))$ of PGM, FPGM, FPGM-(σ=0.78), FPGM-$\left(m=\left\lfloor\frac{2N}{3}\right\rfloor\right)$, FPGM-OCG, and FPGM-($a$=4). We computed empirical rates by assuming that the bounds follow the form bN^c with constants b and c, and then by estimating c from points $N = 47, 50$. Note that the corresponding empirical rates are underestimated due to the simplified exponential model.

Min composite gradient mapping:

N	PGM	FPGM	FPGM -(σ=0.78)	FPGM -$\left(m=\left\lfloor\frac{2N}{3}\right\rfloor\right)$	FPGM -OCG	FPGM -(a=4)
1	1.84	1.84	1.18	1.84	1.84	1.84
2	2.83	2.83	1.78	2.83	2.83	2.83
4	4.81	5.65	3.50	5.09	5.21	5.12
10	10.80	13.24	8.74	14.91	15.60	14.76
20	20.78	27.19	18.83	39.70	39.61	29.21
30	30.78	43.49	30.82	64.45	64.40	47.14
40	40.78	61.76	44.39	92.82	91.99	67.82
47	47.77	75.60	54.73	113.92	113.41	83.67
50	50.77	81.78	59.35	123.54	123.17	90.78
Empi. $O(\cdot)$	$N^{-0.98}$	$N^{-1.27}$	$N^{-1.31}$	$N^{-1.31}$	$N^{-1.33}$	$N^{-1.32}$
Known $O(\cdot)$	N^{-1}	N^{-1}	$N^{-\frac{3}{2}}$	$N^{-\frac{3}{2}}$	$N^{-\frac{3}{2}}$	$N^{-\frac{3}{2}}$

Table 3 Tight worst-case bounds on the norm of the composite gradient mapping $LR/\left(\min_{\boldsymbol{x}\in\Omega_N}||\tilde{\nabla}_L F(\boldsymbol{x})||\right)$ of PGM, FPGM, FPGM-(σ=0.78), FPGM-$\left(m=\left\lfloor\frac{2N}{3}\right\rfloor\right)$, FPGM-OCG, and FPGM-($a$=4). Empirical rates were computed as described in Table 2. (The bound for FPGM-σ uses $\min_{\boldsymbol{x}\in\Omega_N}||\tilde{\nabla}_{L/\sigma^2}F(\boldsymbol{x})||$.)

Final composite gradient mapping:

N	PGM	FPGM	FPGM -(σ=0.78)	FPGM -$\left(m=\left\lfloor \frac{2N}{3} \right\rfloor\right)$	FPGM -OCG	FPGM -(a=4)
1	1.84	1.84	1.18	1.84	1.84	1.84
2	2.83	2.83	1.78	2.83	2.83	2.83
4	4.81	5.65	3.50	5.09	5.21	5.12
10	10.80	12.68	8.41	14.91	15.60	14.76
20	20.78	22.02	14.26	39.65	39.10	25.96
30	30.78	31.26	20.12	64.40	63.40	34.21
40	40.78	40.46	25.97	92.78	90.16	42.39
47	47.77	46.89	30.06	113.92	110.12	48.13
50	50.77	49.65	31.81	123.53	118.99	50.59
Empi. $O(\cdot)$	$N^{-0.98}$	$N^{-0.92}$	$N^{-0.92}$	$N^{-1.31}$	$N^{-1.25}$	$N^{-0.81}$
Known $O(\cdot)$	N^{-1}	N^{-1}	N^{-1}	$N^{-\frac{3}{2}}$	N^{-1}	N^{-1}

Table 4 Tight worst-case bounds on the norm of the final composite gradient mapping $LR/||\tilde{\nabla}_L F(\boldsymbol{x}_N)||$ of PGM, FPGM, FPGM-(σ=0.78), FPGM-$\left(m=\left\lfloor \frac{2N}{3} \right\rfloor\right)$, FPGM-OCG, and FPGM-($a$=4). Empirical rates were computed as described in Table 2. (The bound for FPGM-σ uses $||\tilde{\nabla}_{L/\sigma^2} F(\boldsymbol{x}_N)||$.)

Final subgradient bound:

N	PGM	FPGM	FPGM -(σ=0.78)	FPGM -$\left(m=\left\lfloor\frac{2N}{3}\right\rfloor\right)$	FPGM -OCG	FPGM -(a=4)
1	1.00	1.00	0.61	1.00	1.00	1.00
2	2.00	2.00	1.22	2.00	2.00	2.00
4	4.00	4.83	2.94	4.28	4.40	4.31
10	10.00	7.60	4.67	14.12	14.81	12.10
20	20.00	12.58	7.67	38.29	36.65	16.85
30	30.00	17.63	10.74	62.71	60.40	21.61
40	40.00	22.67	13.80	91.00	86.62	26.47
47	47.00	26.20	15.94	112.01	106.21	29.91
50	50.00	27.71	16.86	121.53	114.93	31.39
Empi. $O(\cdot)$	$N^{-1.00}$	$N^{-0.91}$	$N^{-0.91}$	$N^{-1.32}$	$N^{-1.27}$	$N^{-0.78}$
Known $O(\cdot)$	N^{-1}	N^{-1}	N^{-1}	$N^{-\frac{3}{2}}$	N^{-1}	N^{-1}

Table 5 Tight worst-case bounds on the subgradient norm $LR/||F'(\boldsymbol{x}_N)||$ of PGM, FPGM, FPGM-(σ=0.78), FPGM-$\left(m=\left\lfloor\frac{2N}{3}\right\rfloor\right)$, FPGM-OCG, and FPGM-($a$=4), where $F'(\boldsymbol{x}) \in \partial F(\boldsymbol{x})$ is a subgradient. Empirical rates were computed as described in Table 2.

Table of Contents

Lyapunov Analysis

- $\sqrt{2}$ acceleration [Park et al., 2021]
- Acceleration methods monograph [d'Aspremont et al., 2021]

FGM

FGM-form1

$$y_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$$
$$x_{k+1} = y_{k+1} + \frac{\theta_k - 1}{\theta_{k+1}}(y_{k+1} - y_k)$$

with $x_0 = y_0 = z_0, \quad \theta_{k+1}^2 - \theta_{k+1} = (\leq)\theta_k^2.$

FGM-form2

$$y_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$$
$$z_{k+1} = z_k - \frac{\theta_k}{L}\nabla f(x_k)$$
$$x_{k+1} = (1 - \frac{1}{\theta_{k+1}})y_{k+1} + \frac{1}{\theta_{k+1}}z_{k+1}$$

OGM

OGM-form1

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

$$x_{k+1} = y_{k+1} + \frac{\theta_k - 1}{\theta_{k+1}}(y_{k+1} - y_k) + \frac{\theta_k}{\theta_{k+1}}(y_{k+1} - x_k)$$

OGM-form2

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

$$z_{k+1} = z_k - \frac{2\theta_k}{L}\nabla f(x_k)$$

$$x_{k+1} = (1 - \frac{1}{\theta_{k+1}})y_{k+1} + \frac{1}{\theta_{k+1}}z_{k+1}$$

OGM-form3 (last step modification)

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

$$x_{k+1} = y_{k+1} + \frac{\theta_k - 1}{\theta_{k+1}}(y_{k+1} - y_k) + \frac{\theta_k}{\theta_{k+1}}(y_{k+1} - x_k)$$

$$\tilde{x}_{k+1} = y_{k+1} + \frac{\theta_k - 1}{\phi_{k+1}}(y_{k+1} - y_k) + \frac{\theta_k}{\phi_{k+1}}(y_{k+1} - x_k)$$

with $x_0 = y_0 = z_0, \quad \phi_{k+1}^2 - \phi_{k+1} = 2\theta_k^2.$

Lyapunov Function Selection

FGM: $U_k = \theta_{k-1}^2(f(y_k) - f_*) + \frac{L}{2}\left\|z_k - x_*\right\|^2$, $U_k \leq U_{k-1}$

Lyapunov Function Selection

FGM: $U_k = \theta_{k-1}^2(f(y_k) - f_*) + \frac{L}{2}\left\|z_k - x_*\right\|^2,\ U_k \le U_{k-1}$

$$f(y_k) - f_* \le \frac{U_0}{\theta_{k-1}^2} = \frac{LR^2}{2\theta_{k-1}^2} \le \frac{2LR^2}{(k+1)^2}.$$

Lyapunov Function Selection

OGM (primary sequence):
$U_k = 2\theta_{k-1}^2(f(x_k) - f_* - \frac{1}{2L}\left\|\nabla f(x_k)\right\|^2) + \frac{L}{2}\left\|z_{k+1} - x_*\right\|^2,\ U_k \leq U_{k-1}$

Lyapunov Function Selection

OGM (primary sequence):
$U_k = 2\theta_{k-1}^2(f(x_k) - f_* - \frac{1}{2L}\|\nabla f(x_k)\|^2) + \frac{L}{2}\|z_{k+1} - x_*\|^2$, $U_k \leq U_{k-1}$

$$f(y_k) - f_* \leq \frac{LR^2}{4\theta_{k-1}^2} \leq \frac{LR^2}{(k+1)^2}.$$

Lyapunov Function Selection

OGM (secondary sequence):
$\tilde{U}_k = \phi_k^2(f(\tilde{x}_k) - f_*) + \frac{L}{2}\left\|z_k - x_* - \frac{1}{L}\phi_k \nabla f(\tilde{x}_k)\right\|^2$, $\tilde{U}_k \leq U_{k-1}$

OGM (secondary sequence):

$\tilde{U}_k = \phi_k^2(f(\tilde{x}_k) - f_*) + \frac{L}{2}\left\|z_k - x_* - \frac{1}{L}\phi_k \nabla f(\tilde{x}_k)\right\|^2$, $\tilde{U}_k \leq U_{k-1}$

$$f(y_k) - f_* \leq \frac{LR^2}{2\phi_k^2} \leq \frac{LR^2}{(k+1+1/\sqrt{2})^2}.$$

Strongly Convex Case

In strongly convex case, we have similar result that OGM methods is $\sqrt{2}$ times faster than FGM,

Linear Coupling

First do gradient step, then do mirror descent step, finally couple them together.[Zhu and Orecchia, 2017]

$$
\begin{aligned}
y_{k+1} &= x_k - L^{-1}Q^{-1}\nabla f(x_k) \\
z_{k+1} &= \underset{y\in\mathbb{R}^n}{\arg\min}\,\{V_{z_k}(y) + \langle \alpha_{k+1}\nabla f(x_k), y - x_k\rangle\} \\
x_{k+1} &= (1-\tau_{k+1})y_{k+1} + \tau_{k+1}z_{k+1}
\end{aligned} \tag{LC}
$$

Linear Coupling

Theorem 5 *Assume (A1), (A2), and (A3). Let the positive sequence* $\{\alpha_k\}_{k=1}^{\infty}$ *satisfy* $0 \leq \alpha_{k+1}^2 L - 2\alpha_{k+1} \leq \alpha_k^2 L$ *for* $k = 1, 2 \dots$ *and* $\alpha_1 = \frac{2}{L}$*. Let* $\tau_k = \frac{2}{\alpha_{k+1} L}$ *for* $k = 1, 2, \dots$*. The* y_k*-sequence of (LC) exhibits the rate*

$$f(y_k) - f_\star \leq \frac{2V_{x_0}(x_\star)}{L\alpha_k^2}$$

for $k = 1, 2, \dots$*.*

Theorem 6 *In the setup of Theorem 5, let* $0 \leq \tilde{\alpha}_{k+1}^2 L - \tilde{\alpha}_{k+1} \leq \frac{1}{2}\alpha_k^2 L$ *and* $\tilde{\alpha}_1 = \frac{1}{L}$*. Then the* $\tilde{x}_k$*-sequence, the secondary sequence with last-step modification, of the linear coupling method (LC) exhibits the rate*

$$f(\tilde{x}_k) - f_\star \leq \frac{V_{x_0}(x_\star)}{L\tilde{\alpha}_{k+1}^2}$$

for $k = 0, 1, \dots$

Unify FGM and OGM

Unified method. Equivalent to FGM when $t = \frac{1}{2}$, equivalent to OGM when $t = 1$.

$$
\begin{aligned}
y_{k+1} &= x_k - \frac{1}{L}\nabla f(x_k) \\
z_{k+1} &= z_k - \frac{2t\theta_k}{L}\nabla f(x_k) \\
x_{k+1} &= \left(1 - \frac{1}{\theta_{k+1}}\right) y_{k+1} + \frac{1}{\theta_{k+1}} z_{k+1}
\end{aligned}
$$

Corollary 5 *Assume (A1), (A2) and (A3). Let $0 < t \leq 1$. Then*

$$
f(y_k) - f_\star \leq \frac{L\,\|x_0 - x_\star\|^2}{4t\theta_{k-1}^2}
$$

for $k = 1, 2, \ldots$

How to Find Lyapunov Function

Start from functions that contains all the available information at iteration k. Then derive coefficients with numerical insights. [Taylor et al., 2018] [d'Aspremont et al., 2021]

Bibliography

Beck, A. and Teboulle, M. (2009).
A fast iterative shrinkage-thresholding algorithm for linear inverse problems.
SIAM J. Imaging Sci., 2:183–202.

d'Aspremont, A., Scieur, D., and Taylor, A. B. (2021).
Acceleration methods.
ArXiv, abs/2101.09545.

Kim, D. and Fessler, J. (2018).
Another look at the fast iterative shrinkage/thresholding algorithm (fista).
SIAM journal on optimization : a publication of the Society for Industrial and Applied Mathematics, 28 1:223–250.

Park, C., Park, J., and Ryu, E. K. (2021).
Factor-$\sqrt{2}$ acceleration of accelerated gradient methods.
arXiv: Optimization and Control.