Performance Estimation Approach

Zhenghao Xu

Zhejiang University

July 11, 2021

Table of Contents

Table of Contents

Framework

Problem

$$\min_{x \in \mathbb{R}^d} \quad f(x) \tag{M}$$

Function $f : \mathbb{R}^d \to \mathbb{R}$ satisfies $f \in \mathcal{F}_{0,L}$:

- convex: $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$
- smooth: $\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$

Framework

Algorithm GM

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

$$x_{i+1} = x_i - \frac{h}{L} \nabla f(x_i)$$

For $h = 1$, an usual convergence bound holds:

$$f(x_N) - f(x_*) \leq \frac{L \left\| x_0 - x_* \right\|^2}{2N} \tag{1}$$

PEP for GM

Performance estimation problem formulation:

$$
\begin{aligned}
\max \quad & f(x_N) - f(x_*) \\
\text{s.t.} \quad & f \in C_L^{1,1}(\mathbb{R}^d),\ f \text{ is convex}, \\
& x_{i+1} = \mathcal{A}(x_0, \ldots, x_i; f(x_0), \ldots, f(x_i); f'(x_0), \ldots, f'(x_i)),\ i = 0, \ldots, N-1, \\
& x_* \in X_*(f), \|x_* - x_0\| \le R, \\
& x_0, \ldots, x_N, x_* \in \mathbb{R}^d.
\end{aligned}
\qquad \text{(P)}
$$

PEP for GM (discretization)

Discretize (relax) the functional constraint:

$$
\begin{aligned}
& f \in \mathcal{F}_{0,L} \\
\rightarrow & \frac{1}{2L}\left\|\nabla f(x_i) - \nabla f(x_j)\right\|^2 \le f(x_i) - f(x_j) - \langle \nabla f(x_j), x_i - x_j \rangle
\end{aligned}
$$

Define new variables and remove the gradient ties:

$$
\begin{aligned}
\delta_i &\triangleq \frac{1}{L\left\|x_* - x_0\right\|^2}(f(x_i) - f(x_*)) \\
g_i &\triangleq \frac{1}{L\left\|x_* - x_0\right\|^2}\nabla f(x_i)
\end{aligned}
$$

PEP for GM (discretized)

$$\begin{aligned}
\max_{\substack{x_0,\dots,x_N,x_*\in\mathbb{R}^d,\\ g_0,\dots,g_N\in\mathbb{R}^d,\\ \delta_0,\dots,\delta_N\in\mathbb{R}}} \quad & L\|x_*-x_0\|^2\delta_N \\
\text{s.t.} \quad & x_{i+1}=x_i-h\|x_*-x_0\|g_i, \quad i=0,\dots,N-1, \\
& \frac{1}{2}\|g_i-g_j\|^2 \le \delta_i-\delta_j-\frac{\langle g_j, x_i-x_j\rangle}{\|x_*-x_0\|}, \quad i,j=0,\dots,N,*, \\
& \|x_*-x_0\|\le R.
\end{aligned}$$

PEP for GM (simplification)

Let ν be the unit direction vector of $x_* - x_0$ and eliminate equality constraints:

$$
\begin{aligned}
\max_{g_i \in \mathbb{R}^d, \delta_i \in \mathbb{R}} \quad & LR^2 \delta_N \\
\text{s.t.} \quad & \frac{1}{2}\|g_i - g_j\|^2 \le \delta_i - \delta_j - \langle g_j, \sum_{t=i+1}^{j} h g_{t-1} \rangle, \quad i < j = 0, \dots, N, \\
& \frac{1}{2}\|g_i - g_j\|^2 \le \delta_i - \delta_j + \langle g_j, \sum_{t=j+1}^{i} h g_{t-1} \rangle, \quad j < i = 0, \dots, N, \\
& \frac{1}{2}\|g_i\|^2 \le \delta_i, \quad i = 0, \dots, N, \\
& \frac{1}{2}\|g_i\|^2 \le -\delta_i - \langle g_i, \nu + \sum_{t=1}^{i} h g_{t-1} \rangle, \quad i = 0, \dots, N.
\end{aligned}
$$

PEP for GM (matrix form)

Let $u_i = e_{i+1}$, $G = (g_0, \ldots, g_N)^T$ and define auxiliary matrices:

$$
\begin{aligned}
A_{i,j} &= \frac{1}{2}(u_i - u_j)(u_i - u_j)^T + \frac{1}{2}\sum_{t=i+1}^{j} h(u_j u_{t-1}^T + u_{t-1} u_j^T),\\
B_{i,j} &= \frac{1}{2}(u_i - u_j)(u_i - u_j)^T - \frac{1}{2}\sum_{t=j+1}^{i} h(u_j u_{t-1}^T + u_{t-1} u_j^T),\\
C_i &= \frac{1}{2}u_i u_i^T,\\
D_i &= \frac{1}{2}u_i u_i^T + \frac{1}{2}\sum_{t=1}^{i} h(u_i u_{t-1}^T + u_{t-1} u_i^T),
\end{aligned}
\tag{3.5}
$$

then we have:

$$
\begin{aligned}
\max_{G \in \mathbb{R}^{(N+1)\times d}, \delta \in \mathbb{R}^{N+1}} \quad & LR^2 \delta_N \\
\text{s.t.} \quad & \operatorname{tr}(G^T A_{i,j} G) \le \delta_i - \delta_j, \quad i < j = 0, \ldots, N,\\
& \operatorname{tr}(G^T B_{i,j} G) \le \delta_i - \delta_j, \quad j < i = 0, \ldots, N,\\
& \operatorname{tr}(G^T C_i G) \le \delta_i, \quad i = 0, \ldots, N,\\
& \operatorname{tr}(G^T D_i G + \nu u_i^T G) \le -\delta_i, \quad i = 0, \ldots, N.
\end{aligned}
\tag{G}
$$

PEP for GM (relaxation)

Drop some constraints:

$$\begin{aligned} \max_{G\in\mathbb{R}^{(N+1)\times d},\delta\in\mathbb{R}^{N+1}} \quad & LR^2\delta_N \\ \text{s.t.} \quad & \operatorname{tr}(G^T A_{i-1,i} G) \le \delta_{i-1} - \delta_i, \quad i = 1,\dots,N, \\ & \operatorname{tr}(G^T D_i G + \nu u_i^T G) \le -\delta_i, \quad i = 0,\dots,N. \end{aligned} \tag{G$'$}$$

Corresponding dual problem:

$$\min_{\lambda\in\mathbb{R}^N, t\in\mathbb{R}} \{\frac{1}{2}LR^2 t : \ \lambda \in \Lambda, \ S(\lambda,t) \succeq 0\}, \tag{DG$'$}$$

$$S(\lambda,t) = \begin{pmatrix} (1-h)S_0 + hS_1 & q \\ q^T & t \end{pmatrix}, \qquad S_0 = \begin{pmatrix} 2\lambda_1 & -\lambda_1 & & & & \\ -\lambda_1 & 2\lambda_2 & -\lambda_2 & & & \\ & -\lambda_2 & 2\lambda_3 & -\lambda_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\lambda_{N-1} & 2\lambda_N & -\lambda_N \\ & & & & -\lambda_N & 1 \end{pmatrix} \qquad S_1 = \begin{pmatrix} 2\lambda_1 & \lambda_2-\lambda_1 & \dots & \lambda_N-\lambda_{N-1} & 1-\lambda_N \\ \lambda_2-\lambda_1 & 2\lambda_2 & & \lambda_N-\lambda_{N-1} & 1-\lambda_N \\ \vdots & & \ddots & & \vdots \\ \lambda_N-\lambda_{N-1} & \lambda_N-\lambda_{N-1} & & 2\lambda_N & 1-\lambda_N \\ 1-\lambda_N & 1-\lambda_N & \dots & 1-\lambda_N & 1 \end{pmatrix}.$$

GM Bound

Let $\lambda_i = \frac{i}{2N+1-i}$, $i = 1, \ldots, N$, $t = \frac{1}{2Nh+1}$, then $\lambda = (\lambda_1, \ldots, \lambda_N) \in \Lambda$, $S(\lambda, t) \succeq 0$ is a feasible solution to (DG')

Theorem 3.1. *Let $f \in C_L^{1,1}(\mathbb{R}^d)$ and let $x_0, \ldots, x_N \in \mathbb{R}^d$ be generated by Algorithm GM with $0 < h \le 1$. Then*

$$f(x_N) - f(x_*) \le \frac{LR^2}{4Nh+2}. \tag{3.11}$$

Theorem 3.2. *Let $L > 0$, $N \in \mathbb{N}$ and $d \in \mathbb{N}$. Then for every $h > 0$ there exists a convex function $\varphi \in C_L^{1,1}(\mathbb{R}^d)$ and a point $x_0 \in \mathbb{R}^d$ such that after N iterations, Algorithm GM reaches an approximate solution x_N with following absolute inaccuracy*

$$\varphi(x_N) - \varphi^* = \frac{LR^2}{2} \max\left(\frac{1}{2Nh+1}, (1-h)^{2N}\right).$$

Lower bounds:

$$\varphi_1(x) = \begin{cases} \frac{1}{2Nh+1}\|x\| - \frac{1}{2(2Nh+1)^2} & \|x\| \ge \frac{1}{2Nh+1} \\ \frac{1}{2}\|x\|^2 & \|x\| < \frac{1}{2Nh+1} \end{cases}. \qquad \varphi_2(x) = \frac{1}{2}\|x\|^2$$

GM Bound Conjecture

For $0 < h \leq 1$, the exact upper bound is proved. For $0 < h < 2$ we only have conjecture:

Conjecture 3.1. *Suppose the sequence $x_0, \ldots, x_N$ is generated by the gradient method GM with $0 < h < 2$, then*

$$f(x_N) - f(x_*) \leq \frac{LR^2}{2} \max\left(\frac{1}{2Nh+1}, (1-h)^{2N}\right).$$

This is checked numerically with error less than 10^{-7} [Taylor et al., 2017], but still not proved yet.

Nesterov's FGM method:

Algorithm FGM

0. Input: $f \in C_L^{1,1}(\mathbb{R}^d)$, $x_0 \in \mathbb{R}^d$,
1. $y_1 \leftarrow x_0$, $t_1 \leftarrow 1$,
2. For $i = 1, \dots, N$ compute:
   (a) $x_i \leftarrow y_i - \frac{1}{L} f'(y_i)$,
   (b) $t_{i+1} \leftarrow \frac{1+\sqrt{1+4t_i^2}}{2}$,
   (c) $y_{i+1} \leftarrow x_i + \frac{t_i - 1}{t_{i+1}}(x_i - x_{i-1})$.

main sequence: $\{x_i\}$; auxiliary sequence $\{y_i\}$

Another form fits in FO scheme:

Algorithm FGM′

0. Input: $f \in C_L^{1,1}(\mathbb{R}^d)$, $x_0 \in \mathbb{R}^d$,
1. $y_1 \leftarrow x_0$,
2. For $i = 1, \dots, N-1$ compute:
   (a) $y_{i+1} \leftarrow y_i - \frac{1}{L} \sum_{k=1}^{i} h_k^{(i+1)} f'(y_k)$,
3. $x_N \leftarrow y_N - \frac{1}{L} f'(y_N)$,

with

$$h_k^{(i+1)} = \begin{cases} \frac{t_i - 1}{t_{i+1}} h_k^{(i)} & k+2 \le i, \\ \frac{t_i - 1}{t_{i+1}} (h_{i-1}^{(i)} - 1) & k = i-1, \\ 1 + \frac{t_i - 1}{t_{i+1}} & k = i \end{cases} \qquad (i = 1, \dots, N-1,\ k = 1, \dots, i). \tag{4.2}$$

and

$$t_i = \begin{cases} 1 & i = 1 \\ \frac{1 + \sqrt{1 + 4t_{i-1}^2}}{2} & i > 1. \end{cases}$$

These two methods produces the same sequences. (Prop 4.1)

Relaxed PEP:

$$\max_{G\in\mathbb{R}^{(N+1)\times d},\delta_i\in\mathbb{R}} \quad LR^2\delta_N$$
$$\text{s.t.} \quad \mathrm{tr}(G^T\tilde{A}_{i-1,i}G)\le \delta_{i-1}-\delta_i, \quad i=1,\dots,N, \qquad (\mathrm{Q}')$$
$$\mathrm{tr}(G^T\tilde{D}_iG+\nu u_i^TG)\le -\delta_i, \quad i=0,\dots,N.$$

$$\tilde{A}_{i,j}=\frac{1}{2}(u_i-u_j)(u_i-u_j)^T+\frac{1}{2}\sum_{t=i+1}^{j}\sum_{k=0}^{t-1}h_k^{(t)}(u_ju_k^T+u_ku_j^T),$$

$$\tilde{D}_i=\frac{1}{2}u_iu_i^T+\frac{1}{2}\sum_{t=1}^{i}\sum_{k=0}^{t-1}h_k^{(t)}(u_iu_k^T+u_ku_i^T)$$

Corresponding dual with constraints:

$$\min_{\lambda,\tau,t} \ \frac{1}{2}t$$
$$\text{s.t.} \quad \begin{pmatrix}\sum_{i=1}^N\lambda_i\tilde{A}_{i-1,i}+\sum_{i=0}^N\tau_i\tilde{D}_i & \frac{1}{2}\tau\\ \frac{1}{2}\tau^T & \frac{1}{2}t\end{pmatrix}\succeq 0, \qquad (\mathrm{DQ}')$$
$$(\lambda,\tau)\in\tilde{\Lambda},$$

$$\tilde{\Lambda}=\{(\lambda,\tau)\in\mathbb{R}_+^N\times\mathbb{R}_+^{N+1}:\tau_0=\lambda_1,\ \lambda_i-\lambda_{i+1}+\tau_i=0,\ i=1,\dots,N-1,\ \lambda_N+\tau_N=1\}.$$

FGM Bound

Prior results (only proved for main sequence $\{x_i\}$):

$$f(x_N) - f^* \leq \frac{2L\|x_0 - x_*\|^2}{(N+1)^2}, \quad \forall x_* \in X_*(f).$$

$$f(x_N) - f^* \geq \frac{3L\|x_0 - x_*\|^2}{32(N+1)^2}, \quad \forall x_* \in X_*(f).$$

Numerical evaluation leads to conjecture:

Conjecture 4.1. *Let $x_0, x_1, \ldots$ and $y_1, y_2, \ldots$ be the main and auxiliary sequences defined by FGM (respectively), then $\{f(x_i)\}$ and $\{f(y_i)\}$ converge to the optimal value of the problem with the same rate of convergence.*

New OGM Method

With PEP, we can do minimax optimization to find step size(s) h which achieves the best worst-case convergence rate.

$$\min_{h,\lambda,\tau,t} \left\{ \frac{1}{2}t : \begin{pmatrix} \sum_{i=1}^{N} \lambda_i \tilde{A}_{i-1,i}(h) + \sum_{i=0}^{N} \tau_i \tilde{D}_i(h) & \frac{1}{2}\tau \\ \frac{1}{2}\tau^T & \frac{1}{2}t \end{pmatrix} \succeq 0, (\lambda,\tau) \in \tilde{\Lambda} \right\}, \qquad \text{(BIL)}$$

$$\min_{r,\lambda,\tau,t} \left\{ \frac{1}{2}t : \begin{pmatrix} S(r,\lambda,\tau) & \frac{1}{2}\tau \\ \frac{1}{2}\tau^T & \frac{1}{2}t \end{pmatrix} \succeq 0,\ (\lambda,\tau) \in \tilde{\Lambda} \right\}, \qquad \text{(LIN)}$$

where

$$S(r,\lambda,\tau) = \frac{1}{2}\sum_{i=1}^{N} \lambda_i (u_{i-1}-u_i)(u_{i-1}-u_i)^T + \frac{1}{2}\sum_{i=0}^{N} \tau_i u_i u_i^T + \frac{1}{2}\sum_{i=1}^{N}\sum_{k=0}^{i-1} r_{i,k}(u_i u_k^T + u_k u_i^T).$$

Theorem 5.1. *Suppose $(r^*,\lambda^*,\tau^*,t^*)$ is an optimal solution for (LIN), then (h,λ^*,τ^*,t^*) is an optimal solution for (BIL), where $h = (h_k^{(i)})_{0\le k<i\le N}$ is defined by the following recursive rule*

$$h_k^{(i)} = \begin{cases} \frac{\tau_i^* \sum_{t=k+1}^{i-1} h_k^{(t)} - r_{i,k}^*}{\lambda_i^*} & \lambda_i^* \neq 0 \\ 0 & \lambda_i^* = 0 \end{cases}, \quad i = 1,\ldots,N,\ k = 0,\ldots,i-1. \qquad (5.2)$$

Numerical OGM

Solve the SDP problem numerically gives numerical step sizes.

$$
\begin{aligned}
x_1 &\leftarrow x_0 + \tfrac{1.6180}{L} f'(x_0) \\
x_2 &\leftarrow x_1 + \tfrac{0.1741}{L} f'(x_0) + \tfrac{2.0194}{L} f'(x_1) \\
x_3 &\leftarrow x_2 + \tfrac{0.0756}{L} f'(x_0) + \tfrac{0.4425}{L} f'(x_1) + \tfrac{2.2317}{L} f'(x_2) \\
x_4 &\leftarrow x_3 + \tfrac{0.0401}{L} f'(x_0) + \tfrac{0.2350}{L} f'(x_1) + \tfrac{0.6541}{L} f'(x_2) + \tfrac{2.3656}{L} f'(x_3) \\
x_5 &\leftarrow x_4 + \tfrac{0.0178}{L} f'(x_0) + \tfrac{0.1040}{L} f'(x_1) + \tfrac{0.2894}{L} f'(x_2) + \tfrac{0.6043}{L} f'(x_3) + \tfrac{2.0778}{L} f'(x_4)
\end{aligned}
$$

Figure 4: A first-order algorithm with optimal step-sizes for $N = 5$.

OGM Performance

Numerical result shows that OGM is 2 times faster than FGM.

N	Heavy Ball	FGM, main	FGM, auxiliary	Known bound on FGM (4.1)
1	$LR^2/6.00$	$LR^2/6.00$	$LR^2/2.00$	$LR^2/2.0=2LR^2/(1+1)^2$
2	$LR^2/7.99$	$LR^2/10.00$	$LR^2/6.00$	$LR^2/4.5=2LR^2/(2+1)^2$
3	$LR^2/9.00$	$LR^2/15.13$	$LR^2/11.13$	$LR^2/8.0=2LR^2/(3+1)^2$
4	$LR^2/12.35$	$LR^2/21.35$	$LR^2/17.35$	$LR^2/12.5=2LR^2/(4+1)^2$
5	$LR^2/16.41$	$LR^2/28.66$	$LR^2/24.66$	$LR^2/18.0=2LR^2/(5+1)^2$
10	$LR^2/39.63$	$LR^2/81.07$	$LR^2/77.07$	$LR^2/60.5=2LR^2/(10+1)^2$
20	$LR^2/89.45$	$LR^2/263.65$	$LR^2/259.65$	$LR^2/220.5=2LR^2/(20+1)^2$
40	$LR^2/188.99$	$LR^2/934.89$	$LR^2/930.89$	$LR^2/840.5=2LR^2/(40+1)^2$
80	$LR^2/387.91$	$LR^2/3490.22$	$LR^2/3486.22$	$LR^2/3280.5=2LR^2/(80+1)^2$
160	$LR^2/785.68$	$LR^2/13427.43$	$LR^2/13423.43$	$LR^2/12960.5=2LR^2/(160+1)^2$
500	$LR^2/2476.11$	$LR^2/127224.44$	$LR^2/127220.32$	$LR^2/125500.5=2LR^2/(500+1)^2$
1000	$LR^2/4962.01$	$LR^2/504796.99$	$LR^2/504798.28$	$LR^2/501000.5=2LR^2/(1000+1)^2$

N	val(LIN)
1	$LR^2/8.00$
2	$LR^2/16.16$
3	$LR^2/26.53$
4	$LR^2/39.09$
5	$LR^2/53.80$
10	$LR^2/159.07$
20	$LR^2/525.09$
40	$LR^2/1869.22$
80	$LR^2/6983.13$
160	$LR^2/26864.04$
500	$LR^2/254482.61$
1000	$LR^2/1009628.17$

DT's defect & Further Works

Some problems remain unsolved in [Drori and Teboulle, 2014]:

- conjecture on GM for $h > 1$ (unsolved? numerical evidence from [Taylor et al., 2017])
- conjecture on FGM's auxiliary sequence (solved by [Kim and Fessler, 2017])
- no analytical OGM (given by [Kim and Fessler, 2016], two times faster confirmed)
- too much relaxation (tightened by [Taylor et al., 2017])
- limited to $\mathcal{F}_{0,L}$ and function value criteria (general $\mathcal{F}_{\mu,L}$ cases and other criteria included in [Taylor et al., 2017])

Table of Contents

FGM Recap

Here we specify $\{y_i\}$ as the main sequence and $\{x_i\}$ as auxiliary sequence.

Algorithm Class FO

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

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

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \frac{1}{L}\sum_{k=0}^{i} h_{i+1,k} f'(\boldsymbol{x}_k). \tag{1.1}$$

Algorithm FGM1

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

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

$$\boldsymbol{y}_{i+1} = \boldsymbol{x}_i - \frac{1}{L} f'(\boldsymbol{x}_i)$$

$$t_{i+1} = \frac{1 + \sqrt{1 + 4t_i^2}}{2}$$

$$\boldsymbol{x}_{i+1} = \boldsymbol{y}_{i+1} + \frac{t_i - 1}{t_{i+1}}(\boldsymbol{y}_{i+1} - \boldsymbol{y}_i). \tag{3.1}$$

$$\bar{h}_{i+1,k} = \begin{cases} \frac{t_i-1}{t_{i+1}}\bar{h}_{i,k}, & k = 0, \dots, i-2, \\ \frac{t_i-1}{t_{i+1}}(\bar{h}_{i,i-1} - 1), & k = i-1, \\ 1 + \frac{t_i-1}{t_{i+1}}, & k = i, \end{cases} \tag{3.3}$$

PEP Recap

PEP for FO:

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

$$\text{s.t.} \;\; \boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \frac{1}{L} \sum_{k=0}^{i} h_{i+1,k} f'(\boldsymbol{x}_k), \quad i = 0, \ldots, N-1,$$

$$||\boldsymbol{x}_0 - \boldsymbol{x}_*|| \leq R.$$

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

$$\Lambda = \left\{ (\boldsymbol{\lambda}, \boldsymbol{\tau}) \in \mathbb{R}_+^N \times \mathbb{R}_+^{N+1} \colon \begin{array}{l} \tau_0 = \lambda_1, \;\; \lambda_N + \tau_N = 1 \\ \lambda_i - \lambda_{i+1} + \tau_i = 0, \;\; i = 1, \ldots, N-1 \end{array} \right\}, \tag{4.3}$$

FGM Bound

Lemma (lemma 1 [Kim and Fessler, 2016])

Let $\lambda_i = \frac{t_{i-1}^2}{t_N^2}$, $\tau_i = \frac{t_i}{t_N^2}$, $\gamma = \frac{1}{t_N^2}$, and $\lambda = (\lambda_1, \ldots, \lambda_N)$, $\tau = (\tau_0, \ldots, \tau_N)$, then (λ, τ, γ) is a feasible point of problem (D) for h given by FGM.

Theorem 1 *Let $f \colon \mathbb{R}^d \to \mathbb{R}$ be convex and $C_L^{1,1}$, and let $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots \in \mathbb{R}^d$ be generated by FGM1 or FGM2. Then for $n \geq 1$,*

$$f(\boldsymbol{x}_n) - f(\boldsymbol{x}_*) \leq \frac{L||\boldsymbol{x}_0 - \boldsymbol{x}_*||^2}{2t_n^2} \leq \frac{2L||\boldsymbol{x}_0 - \boldsymbol{x}_*||^2}{(n+2)^2}, \quad \forall \boldsymbol{x}_* \in X_*(f). \tag{5.5}$$

Same constant 2 with classical result.

Following DT's work, to find the optimized steps h can be reduced from finding solution of the primal problem to the dual problem.

$$\hat{\boldsymbol{h}} \triangleq \underset{\boldsymbol{h} \in \mathbb{R}^{N(N+1)/2}}{\arg\min} \; \mathcal{B}_{\mathrm{D}}(\boldsymbol{h}, N, L, R), \qquad \text{(HD)}$$

Analytical OGM

Following DT's work, to find the optimized steps h can be reduced from finding solution of the primal problem to the dual problem.

$$\hat{\boldsymbol{h}} \triangleq \mathop{\arg\min}_{\boldsymbol{h} \in \mathbb{R}^{N(N+1)/2}} \mathcal{B}_{\mathrm{D}}(\boldsymbol{h}, N, L, R), \qquad \text{(HD)}$$

Convert to new variables:

$$r_{i,k} = \lambda_i h_{i,k} + \tau_i \sum_{j=k+1}^{i} h_{j,k} \qquad (6.1)$$

Analytical OGM (RD)

$$\hat{\boldsymbol{r}} \triangleq \underset{\boldsymbol{r}\in\mathbb{R}^{N(N+1)/2}}{\arg\min}\ \breve{\mathcal{B}}_{\mathrm{D}}(\boldsymbol{r}, N, L, R), \tag{RD}$$

where

$$\breve{\mathcal{B}}_{\mathrm{D}}(\boldsymbol{r}, N, L, R) \triangleq \min_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^N,\\ \boldsymbol{\tau}\in\mathbb{R}^{N+1},\\ \gamma\in\mathbb{R}}} \left\{ \frac{1}{2}LR^2\gamma \colon \begin{pmatrix} \breve{\mathbf{S}}(\boldsymbol{r}, \boldsymbol{\lambda}, \boldsymbol{\tau}) & \frac{1}{2}\boldsymbol{\tau} \\ \frac{1}{2}\boldsymbol{\tau}^\top & \frac{1}{2}\gamma \end{pmatrix} \succeq 0,\ (\boldsymbol{\lambda}, \boldsymbol{\tau}) \in \Lambda \right\},$$

$$\begin{aligned} \breve{\mathbf{S}}(\boldsymbol{r}, \boldsymbol{\lambda}, \boldsymbol{\tau}) \triangleq & \frac{1}{2}\sum_{i=1}^{N} \lambda_i (\boldsymbol{u}_{i-1} - \boldsymbol{u}_i)(\boldsymbol{u}_{i-1} - \boldsymbol{u}_i)^\top + \frac{1}{2}\sum_{i=0}^{N} \tau_i \boldsymbol{u}_i \boldsymbol{u}_i^\top \\ & + \frac{1}{2}\sum_{i=1}^{N}\sum_{k=0}^{i-1} r_{i,k} \left(\boldsymbol{u}_i \boldsymbol{u}_k^\top + \boldsymbol{u}_k \boldsymbol{u}_i^\top \right). \end{aligned} \tag{6.2}$$

Utilize the fact that $S_{N,N} = \frac{1}{2}$, one can split the last variable in w and optimize it first, thus reducing the problem to:

$$\hat{\boldsymbol{r}} = \underset{\boldsymbol{r} \in \mathbb{R}^{N(N+1)/2}}{\arg\min} \ \breve{\mathcal{B}}_{\mathrm{D1}}(\boldsymbol{r}, N, L, R), \tag{RD1}$$

where

$$\breve{\mathcal{B}}_{\mathrm{D1}}(\boldsymbol{r}, N, L, R) \triangleq \min_{\substack{\boldsymbol{\lambda} \in \mathbb{R}^N, \\ \boldsymbol{\tau} \in \mathbb{R}^{N+1}, \\ \gamma \in \mathbb{R}}} \left\{ \frac{1}{2} L R^2 \gamma : \begin{pmatrix} \breve{\boldsymbol{Q}} - 2\breve{\boldsymbol{q}}\breve{\boldsymbol{q}}^\top & \frac{1}{2}(\breve{\boldsymbol{\tau}} - 2\breve{\boldsymbol{q}}\tau_N) \\ \frac{1}{2}(\breve{\boldsymbol{\tau}} - 2\breve{\boldsymbol{q}}\tau_N)^\top & \frac{1}{2}(\gamma - \tau_N^2) \end{pmatrix} \succeq 0, \ (\boldsymbol{\lambda}, \boldsymbol{\tau}) \in \Lambda \right\},$$

$$\begin{aligned} \breve{\boldsymbol{Q}}(\boldsymbol{r}, \boldsymbol{\lambda}, \boldsymbol{\tau}) &= \frac{1}{2} \sum_{i=1}^{N-1} \lambda_i (\breve{\boldsymbol{u}}_{i-1} - \breve{\boldsymbol{u}}_i)(\breve{\boldsymbol{u}}_{i-1} - \breve{\boldsymbol{u}}_i)^\top + \frac{1}{2} \lambda_N \breve{\boldsymbol{u}}_{N-1} \breve{\boldsymbol{u}}_{N-1}^\top \\ &\quad + \frac{1}{2} \sum_{i=0}^{N-1} \tau_i \breve{\boldsymbol{u}}_i \breve{\boldsymbol{u}}_i^\top + \frac{1}{2} \sum_{i=1}^{N-1} \sum_{k=0}^{i-1} r_{i,k} \left(\breve{\boldsymbol{u}}_i \breve{\boldsymbol{u}}_k^\top + \breve{\boldsymbol{u}}_k \breve{\boldsymbol{u}}_i^\top \right), \end{aligned} \tag{6.7}$$

$$\breve{\boldsymbol{q}}(\boldsymbol{r}, \boldsymbol{\lambda}, \boldsymbol{\tau}) = \frac{1}{2} \sum_{k=0}^{N-2} r_{N,k} \breve{\boldsymbol{u}}_k, + \frac{1}{2}(r_{N,N-1} - \lambda_N) \breve{\boldsymbol{u}}_{N-1} \tag{6.8}$$

Lemma 2 *A feasible point of both (RD) and (RD1) is* $(\hat{\boldsymbol{r}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\tau}}, \hat{\gamma})$*, where*

$$\hat{r}_{i,k} = \begin{cases} \frac{4\theta_i\theta_k}{\theta_N^2}, & i = 2, \ldots, N-1,\ k = 0, \ldots, i-2, \\ \frac{4\theta_i\theta_{i-1}}{\theta_N^2} + \frac{2\theta_{i-1}^2}{\theta_N^2}, & i = 1, \ldots, N-1,\ k = i-1, \\ \frac{2\theta_k}{\theta_N}, & i = N,\ k = 0, \ldots, i-2, \\ \frac{2\theta_{N-1}}{\theta_N} + \frac{2\theta_{N-1}^2}{\theta_N^2}, & i = N,\ k = i-1, \end{cases} \tag{6.9}$$

$$\hat{\lambda}_i = \frac{2\theta_{i-1}^2}{\theta_N^2}, \quad i = 1, \ldots, N, \tag{6.10}$$

$$\hat{\tau}_i = \begin{cases} \frac{2\theta_i}{\theta_N^2}, & i = 0, \ldots, N-1, \\ 1 - \frac{2\theta_{N-1}^2}{\theta_N^2} = \frac{1}{\theta_N}, & i = N, \end{cases} \tag{6.11}$$

$$\hat{\gamma} = \frac{1}{\theta_N^2}, \tag{6.12}$$

for

$$\theta_i = \begin{cases} 1, & i = 0, \\ \frac{1+\sqrt{1+4\theta_{i-1}^2}}{2}, & i = 1, \ldots, N-1, \\ \frac{1+\sqrt{1+8\theta_{i-1}^2}}{2} & i = N. \end{cases} \tag{6.13}$$

Analytical OGM

Lemma (lemma 3 [Kim and Fessler, 2016])

The choice of $(r, \lambda, \tau, \gamma)$ *given by lemma 2 is optimal solution to both (RD) and (RD1) as KKT conditions hold.*

Lemma (lemma 4 [Kim and Fessler, 2016])

The choice of h *given* θ *in lemma 2, which is*

$$h_{i+1,k} = \begin{cases} \frac{1}{\theta_{i+1}}\left(2\theta_k - \sum_{j=k+1}^{i} h_{j,k}\right), & k = 0, \ldots, i-1, \\ 1 + \frac{2\theta_i - 1}{\theta_{i+1}}, & k = i, \end{cases}$$

is an optimal solution of (HD).

OGM Bound

With given feasible $\gamma = \frac{1}{\theta_N^2}$, we have a bound on OGM:

Theorem 2 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex and $C_L^{1,1}$ and let $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_N \in \mathbb{R}^d$ be generated by Algorithm FO with $\hat{\boldsymbol{h}}$ (6.16) for a given $N \geq 1$. Then*

$$f(\boldsymbol{x}_N) - f(\boldsymbol{x}_*) \leq \frac{L||\boldsymbol{x}_0 - \boldsymbol{x}_*||^2}{2\theta_N^2} \leq \frac{L||\boldsymbol{x}_0 - \boldsymbol{x}_*||^2}{(N+1)(N+1+\sqrt{2})}, \quad \forall \boldsymbol{x}_* \in X_*(f). \tag{6.17}$$

This bound has constant 1. Two times faster than FGM confirmed!

OGM Lower Bound

Similar to GM, OGM is also lower bounded by the case of piecewise affine-quadratic function.

Theorem 3 *For the following convex functions in $C_L^{1,1}(\mathbb{R}^d)$ for all $d \geq 1$:*

$$\phi(\boldsymbol{x}) = \begin{cases} \frac{LR}{\theta_N^2}||\boldsymbol{x}|| - \frac{LR^2}{2\theta_N^4}, & \text{if } ||\boldsymbol{x}|| \geq \frac{R}{\theta_N^2}, \\ \frac{L}{2}||\boldsymbol{x}||^2, & \text{otherwise}, \end{cases} \tag{8.1}$$

both OGM1 and OGM2 exactly achieve the smallest upper bound in (6.17), i.e.,

$$\phi(\boldsymbol{x}_N) - \phi(\boldsymbol{x}_*) = \frac{L||\boldsymbol{x}_0 - \boldsymbol{x}_*||^2}{2\theta_N^2}.$$

This shows that OGM upper bound is tight.

Collect the pieces we get the practical OGM method:

Algorithm OGM1

Input: $f \in C_L^{1,1}(\mathbb{R}^d)$ convex, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\boldsymbol{y}_0 = \boldsymbol{x}_0$, $\theta_0 = 1$.

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

$$\boldsymbol{y}_{i+1} = \boldsymbol{x}_i - \frac{1}{L} f'(\boldsymbol{x}_i)$$

$$\theta_{i+1} = \begin{cases} \frac{1+\sqrt{1+4\theta_i^2}}{2}, & i \leq N-2 \\ \frac{1+\sqrt{1+8\theta_i^2}}{2}, & i = N-1 \end{cases}$$

$$\boldsymbol{x}_{i+1} = \boldsymbol{y}_{i+1} + \frac{\theta_i - 1}{\theta_{i+1}}(\boldsymbol{y}_{i+1} - \boldsymbol{y}_i) + \frac{\theta_i}{\theta_{i+1}}(\boldsymbol{y}_{i+1} - \boldsymbol{x}_i)$$

Table of Contents

Convergence Analysis for Auxiliary Sequence

Recall the conjecture from [Drori and Teboulle, 2014]

Conjecture 4.1. *Let $x_0, x_1, \dots$ and $y_1, y_2, \dots$ be the main and auxiliary sequences defined by FGM (respectively), then $\{f(x_i)\}$ and $\{f(y_i)\}$ converge to the optimal value of the problem with the same rate of convergence.*

In [Kim and Fessler, 2017], this is proved by following theorem:

Theorem 4.1 *Let $f \in \mathcal{F}_L(\mathbb{R}^d)$ and let $\boldsymbol{y}_0, \dots, \boldsymbol{y}_N \in \mathbb{R}^d$ be generated by OGM1 and OGM2. Then for $1 \leq i \leq N$, the primary sequence for OGM satisfies:*

$$f(\boldsymbol{y}_i) - f(\boldsymbol{x}_*) \leq \frac{LR^2}{4t_{i-1}^2} \leq \frac{LR^2}{(i+1)^2}. \tag{20}$$

Table of Contents

PEP Discretization Recap

In DT and KF's work, inequality

$$\frac{1}{2L}\|\nabla f(x_i) - \nabla f(x_j)\|^2 \leq f(x_i) - f(x_j) - \langle \nabla f(x_j), x_i - x_j \rangle$$

is used to discretize the convex smooth function f, converting the infinite dimensional constraint into finite dimensional constraint.

But this strategy may fail if

- function f is not smooth
- function f is strongly convex

We need more general characterization for $\mathcal{F}_{\mu,L}$.

Convex Interpolation

Definition ($\mathcal{F}_{\mu,L}$-interpolation)

Let I be an index set. Consider set of triples $S = \{(x_i, g_i, f_i)\}_{i \in I}$ where $x_i, g_i \in \mathbb{R}^d$, $f_i \in \mathbb{R}$. Then S is $\mathcal{F}_{\mu,L}$-interpolable if and only if there exists a function $f \in \mathcal{F}_{\mu,L}(\mathbb{R}^d)$ such that $g_i \in \partial f(x_i)$ and $f(x_i) = f_i$ for all $i \in I$.

Theorem (theorem 1 [Taylor et al., 2017])

Set $\{(x_i, g_i, f_i)\}$ is $\mathcal{F}_{0,\infty}$-interpolable if and only if

$$f_i \geq f_j + \langle g_j, x_i - x_j \rangle \quad \forall i, j \in I.$$

Smooth Strongly Convex Interpolation

Use Legendre-Fenchel conjugation's properties and strongly convex conditions, we have following steps to check interpolability:

1. Reformulate $\mathcal{F}_{\mu,L}$ interpolation into $\mathcal{F}_{0,L-\mu}$ interpolation using minimal curvature subtraction.
2. Write $\mathcal{F}_{0,L-\mu}$ interpolation into $\mathcal{F}_{1/(L-\mu),\infty}$ interpolation using Legendre-Fenchel conjugation.
3. Transform $\mathcal{F}_{1/(L-\mu),\infty}$ interpolation into $\mathcal{F}_{0,\infty}$ interpolation using minimal curvature subtraction.

Smooth Strongly Convex Interpolation

Theorem (theorem 4 [Taylor et al., 2017])

Set $\{(x_i, g_i, f_i)\}$ *is* $\mathcal{F}_{\mu,L}$*-interpolable if and only if for any* $i, j \in I$

$$f_i - f_j - \langle g_j, x_i - x_j \rangle \geq \frac{1}{2(1-\mu/L)}\Big(\frac{1}{L}\|g_i - g_j\|^2 + \mu\|x_i - x_j\|^2 - 2\frac{\mu}{L}\langle g_j - g_i, x_j - x_i\rangle\Big).$$

If $\mu = 0$, the inequality just reduces to

$$\frac{1}{2L}\|\nabla f(x_i) - \nabla f(x_j)\|^2 \leq f(x_i) - f(x_j) - \langle \nabla f(x_j), x_i - x_j \rangle,$$

which means DT's discretization does not cause relaxation.

Convex Formulation for PEP

Original PEP:

$$w^f(\mathcal{F}, R, \mathcal{M}, N, \mathcal{P}) = \sup_{\{x_i, g_i, f_i\}_{i \in I}} \mathcal{P}\left(\{x_i, g_i, f_i\}_{i \in I}\right), \tag{f-PEP}$$

$$\text{such that there exists } f \in \mathcal{F} \text{ such that } \mathcal{O}_f(x_i) = \{f_i, g_i\} \ \forall i \in I,$$
$$g_* = 0,$$
$$x_1, \ldots, x_N \text{ is generated from } x_0 \text{ by method } \mathcal{M} \text{ with } \{f_i, g_i\}_{i \in \{0, \ldots, N-1\}},$$
$$\|x_0 - x_*\|_2 \leq R.$$

$\Longleftrightarrow$ Discrete PEP:

$$w^{(d)}_{\mu,L}(R, \mathcal{M}, N, \mathcal{P}) = \sup_{\{x_i, g_i, f_i\}_{i \in I} \in \left(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}\right)^{N+2}} \mathcal{P}\left(\{x_i, g_i, f_i\}_{i \in I}\right), \tag{d-PEP}$$

$$\text{such that } \{x_i, g_i, f_i\}_{i \in I} \text{ is } \mathcal{F}_{\mu,L}\text{-interpolable},$$
$$x_1, \ldots, x_N \text{ is generated from } x_0 \text{ by method } \mathcal{M} \text{ with (2)},$$
$$\{x_*, g_*, f_*\} = \{0^d, 0^d, 0\} \text{ and } \|x_0 - x_*\|_2 \leq R.$$

Convex Formulation for PEP

With definitions:

$$G = \{G_{i,j}\}_{0\le i,j\le N} \text{ with } \begin{cases} G_{i,j} = g_i^\top g_j & \text{for any } 0 \le i,j \le N, \\ G_{N+1,j} = x_0^\top g_j & \text{for any } 0 \le j \le N, \\ G_{i,N+1} = g_i^\top x_0 & \text{for any } 0 \le i \le N, \\ G_{N+1,N+1} = x_0^\top x_0 \end{cases}$$

$$h_i^\top = [-h_{i,0} \;\; -h_{i,1} \;\; \dots \;\; -h_{i,i-1} \;\; 0 \;\; \dots \;\; 0 \;\; 1], \quad h_*^\top = [0 \;\; \dots \;\; 0],$$

$$\begin{aligned} 2A_{ij} &= \frac{L}{L-\mu}\left(u_j(h_i-h_j)^\top + (h_i-h_j)u_j^\top\right) + \frac{1}{L-\mu}(u_i-u_j)(u_i-u_j)^\top \\ &\quad + \frac{\mu}{L-\mu}\left(u_i(h_j-h_i)^\top + (h_j-h_i)u_i^\top\right) + \frac{L\mu}{L-\mu}(h_i-h_j)(h_i-h_j)^\top, \quad \text{for all } i,j \in I, \\ A_R &= u_{N+1}u_{N+1}^\top. \end{aligned}$$

Discrete PEP $\overset{thrm\ 5}{\Longleftrightarrow}$ Semidefinite PEP:

$$\begin{aligned} w_{\mu,L}^{sdp}(R,H,N,b,C) = \sup_{G\in\mathbb{S}^{N+2},\, f\in\mathbb{R}^{N+1}} \;& b^\top f + \operatorname{Tr}(CG) \\ \textit{such that } & f_j - f_i + \operatorname{Tr}(GA_{ij}) \le 0, \qquad i,j \in I, \\ & \operatorname{Tr}(GA_R) - R^2 \le 0, \\ & G \succeq 0, \end{aligned} \tag{sdp-PEP}$$

Theorem 5 [Taylor et al., 2017] guarantees equivalence.

Convex Formulation for PEP

Semidefinite PEP $\overset{thrm\ 6}{\Longleftrightarrow}$ Dual PEP:

$$\begin{aligned} \inf_{\lambda_{ij},\tau} \ \tau R^2 \ \text{ such that } \ \tau A_R - C + \sum_{i,j\in I} \lambda_{ij} A_{ij} &\succeq 0, && \text{(d-sdp-PEP)} \\ b - \sum_{i,j\in I} \lambda_{ij}(u_j - u_i) &= 0, \\ \lambda_{ij} &\geq 0, \qquad i,j \in I, \\ \tau &\geq 0, \end{aligned}$$

Theorem 6 [Taylor et al., 2017] guarantees zero duality gap under assumption that $h_{i,i-1} \neq 0$.

Numerical Performance

N	h_{opt}	Conjecture 1	Value computed in [10]	Rel. error	Value from (sdp-PEP)	Rel. error
1	1.5000	$LR^2/8.00$	$LR^2/8.00$	0.00	$LR^2/8.00$	7e-09
2	1.6058	$LR^2/14.85$	$LR^2/14.54$	2e-02	$LR^2/14.85$	5e-09
5	1.7471	$LR^2/36.94$	$LR^2/32.57$	1e-01	$LR^2/36.94$	1e-08
10	1.8341	$LR^2/75.36$	$LR^2/59.80$	3e-01	$LR^2/75.36$	3e-08
20	1.8971	$LR^2/153.77$	$LR^2/109.58$	4e-01	$LR^2/153.77$	6e-08
30	1.9238	$LR^2/232.85$	$LR^2/156.23$	5e-01	$LR^2/232.85$	7e-08
40	1.9388	$LR^2/312.21$	$LR^2/201.10$	6e-01	$LR^2/312.21$	3e-08
50	1.9486	$LR^2/391.72$	$LR^2/244.70$	6e-01	$LR^2/391.72$	1e-07
100	1.9705	$LR^2/790.22$	$LR^2/451.72$	7e-01	$LR^2/790.22$	1e-07

Table 1 Gradient Method with $\mu = 0$, worst-case computed with relaxation from [10] and worst-case obtained by exact formulation (sdp-PEP) for the criterion $f(x_N) - f^*$. Error is measured relatively to the conjectured result. Results obtained with MOSEK [18].

Conjectures

The duality gap is zero, which leads to tons of conjectures inspired by numerical results.

Any feasible solution to (sdp-PEP) can lead to a lower bound, and such examples can be constructed explicitly through (smooth) convex interpolation; any feasible solution to (d-sdp-PEP) can derive a proof for upper bound.

Conjectures

Conjecture 2 Any sequence of iterates $\{x_i\}$ generated by the gradient method GM with constant normalized step sizes $0 \le h \le 2$ on a smooth strongly convex function $f \in \mathcal{F}_{\mu,L}(\mathbb{R}^d)$ satisfies

$$f(x_N) - f_* \le \frac{LR^2}{2} \max\left(\frac{\kappa}{(\kappa-1) + (1-\kappa h)^{-2N}}, (1-h)^{2N}\right).$$

Conjecture 3 Any sequence of iterates $\{x_i\}$ generated by the gradient method GM with constant normalized step sizes $0 \le h \le 2$ on a smooth strongly convex function $f \in \mathcal{F}_{\mu,L}(\mathbb{R}^d)$ satisfies

$$\|\nabla f(x_N)\|_2 \le LR \max\left(\frac{\kappa}{(\kappa-1) + (1-\kappa h)^{-N}}, |1-h|^{N}\right).$$

Conjecture 4 Any (primary) sequence of iterates $\{y_i\}$ generated by the fast gradient methd FGM (resp. optimized gradient method OGM) on a smooth convex function $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ satisfies

$$f(y_N) - f_* \le f_{1,\tau_1}(y_{1,N}) = \frac{LR^2}{2} \frac{1}{2\sum_{k=0}^{N-2} h_{N-1,k} + 3},$$

where $y_{1,N}$ is the final (primary) iterate computed by FGM (resp. OGM) applied to f_{1,τ_1} starting from $x_0 = R$, and quantities $h_{N-1,k}$ are the fixed coefficients of the last step of FGM (resp. OGM).

Conjecture 5 Any (secondary) sequence of iterates $\{x_i\}$ generated by the fast gradient methd FGM (resp. optimized gradient method OGM) on a smooth convex function $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ satisfies

$$f(x_N) - f_* \le f_{1,\tau_2}(x_{1,N}) = \frac{LR^2}{2} \frac{1}{2\sum_{k=0}^{N-1} h_{N,k} + 1},$$

where $x_{1,N}$ is the final (secondary) iterate computed by FGM (resp. OGM) applied to f_{1,τ_2} starting from $x_0 = R$, and quantities $h_{N,k}$ are the fixed coefficients of the last step of FGM (resp. OGM).

Bibliography

Drori, Y. and Teboulle, M. (2014).
Performance of first-order methods for smooth convex minimization: a novel approach.
Mathematical Programming, 145:451–482.

Kim, D. and Fessler, J. (2016).
Optimized first-order methods for smooth convex minimization.
Mathematical Programming, 159:81–107.

Kim, D. and Fessler, J. (2017).
On the convergence analysis of the optimized gradient method.
Journal of Optimization Theory and Applications, 172:187–205.

Taylor, A. B., Hendrickx, J., and Glineur, F. (2017).
Smooth strongly convex interpolation and exact worst-case performance of first-order methods.
Mathematical Programming, 161:307–345.