Principled Analyses of First-Order Methods with Inexact Proximal Operators

Mathieu Barré, Adrien Taylor and Francis Bach

EUROPT 2021 - July, 8

What is this work about?

- Practical tool for worst-case analyses of algorithm with **inexact proximal computations**.

What is this work about?

- Practical tool for worst-case analyses of algorithm with **inexact proximal computations**.
- Built on Performance Estimation Problem (PEP) (Drori & Teboulle 2014) (Taylor, Hendrickx & Glineur 2017).

What is this work about?

- Practical tool for worst-case analyses of algorithm with **inexact proximal computations**.
- Built on Performance Estimation Problem (PEP) (Drori & Teboulle 2014) (Taylor, Hendrickx & Glineur 2017).
- Implemented in PESTO Toolbox (everything is reproducible) /github.com/AdrienTaylor/Performance-Estimation-Toolbox.

Proximal computations

Computing a proximal step corresponds to

$$\operatorname{prox}_{\lambda h}(z) = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \underbrace{\lambda h(x) + \frac{1}{2}\|x - z\|^2}_{\Phi(x)}. \tag{Prox}$$

Proximal computations

Computing a proximal step corresponds to

$$\mathrm{prox}_{\lambda h}(z) = \underset{x \in \mathbb{R}^d}{\mathrm{argmin}} \underbrace{\lambda h(x) + \tfrac{1}{2}\|x - z\|^2}_{\Phi(x)}. \tag{Prox}$$

- Introduced in optimization by Martinet and Rockafellar.

Proximal computations

Computing a proximal step corresponds to

$$\operatorname{prox}_{\lambda h}(z) = \operatorname*{argmin}_{x \in \mathbb{R}^d} \underbrace{\lambda h(x) + \tfrac{1}{2}\|x - z\|^2}_{\Phi(x)} . \qquad \text{(Prox)}$$

- Introduced in optimization by Martinet and Rockafellar.
- Base primitive for many optimization algorithms

Proximal computations

Computing a proximal step corresponds to

$$\mathrm{prox}_{\lambda h}(z) = \underset{x \in \mathbb{R}^d}{\mathrm{argmin}} \underbrace{\lambda h(x) + \frac{1}{2}\|x - z\|^2}_{\Phi(x)}. \qquad \text{(Prox)}$$

- Introduced in optimization by Martinet and Rockafellar.
- Base primitive for many optimization algorithms (e.g. splitting methods, augmented Lagrangian, HPE, Catalyst, high order tensor methods).

Proximal computations

Computing a proximal step corresponds to

$$\operatorname{prox}_{\lambda h}(z) = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \underbrace{\lambda h(x) + \tfrac{1}{2}\|x - z\|^2}_{\Phi(x)}. \qquad \text{(Prox)}$$

- Introduced in optimization by Martinet and Rockafellar.
- Base primitive for many optimization algorithms (e.g. splitting methods, augmented Lagrangian, HPE, Catalyst, high order tensor methods).
- Exact solutions to (Prox) known in some cases (e.g. $\|\cdot\|_p$, indicator functions, ... see e.g. `http://proximity-operator.net`).

Proximal computations

Computing a proximal step corresponds to

$$\operatorname{prox}_{\lambda h}(z) = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \underbrace{\lambda h(x) + \frac{1}{2}\|x - z\|^2}_{\Phi(x)}. \tag{Prox}$$

- Introduced in optimization by Martinet and Rockafellar.
- Base primitive for many optimization algorithms (e.g. splitting methods, augmented Lagrangian, HPE, Catalyst, high order tensor methods).
- Exact solutions to (Prox) known in some cases (e.g. $\|\cdot\|_p$, indicator functions, ... see e.g. `http://proximity-operator.net`).
- In many situations, no closed formula, (Prox) has to be approximated.

PEPs on a simple example

We wish to solve problem

$$\min_{x \in \mathbb{R}^d} h(x),$$

with $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

PEPs on a simple example

We wish to solve problem

$$\min_{x \in \mathbb{R}^d} h(x),$$

with $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

Using e.g.

proximal point algorithm:

$$x_{k+1} = x_k - \lambda \; \partial h(x_{k+1})$$

PEPs on a simple example

We wish to solve problem

$$\min_{x \in \mathbb{R}^d} h(x),$$

with $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

Using e.g.

proximal point algorithm:

$$x_{k+1} = x_k - \lambda \, \partial h(x_{k+1})$$

Interested in worst-case bounds of the form

$$h(x_N) - h(x_*) \leq C(N) \|x_0 - x_*\|^2,$$

$x_* \in \underset{x}{\operatorname{argmin}} \; h(x)$

PEPs on a simple example

We wish to solve problem

$$\min_{x \in \mathbb{R}^d} h(x),$$

with $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

Using e.g.

proximal point algorithm:

$$x_{k+1} = x_k - \lambda \, \partial h(x_{k+1})$$

Interested in worst-case bounds of the form

$$h(x_N) - h(x_*) \leq C(N) \|x_0 - x_*\|^2,$$

$x_* \in \operatorname{argmin}_x h(x)$

Smallest value for $C(N)$?

PEPs on a simple example

We wish to solve problem

$$\min_{x \in \mathbb{R}^d} h(x),$$

with $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

Using e.g.
relatively inexact proximal point algorithm:

$$x_{k+1} = x_k - \lambda \ \partial h(x_{k+1})$$

Interested in worst-case bounds of the form

$$h(x_N) - h(x_*) \leq C(N) \|x_0 - x_*\|^2,$$

$x_* \in \underset{x}{\operatorname{argmin}} \ h(x)$

Smallest value for $C(N)$?

PEPs on a simple example

We wish to solve problem

$$\min_{x \in \mathbb{R}^d} h(x),$$

with $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

Using e.g.
relatively inexact proximal point algorithm:

$$x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k),$$

Interested in worst-case bounds of the form

$$h(x_N) - h(x_*) \leq C(N)\|x_0 - x_*\|^2,$$

$x_* \in \underset{x}{\operatorname{argmin}}\ h(x)$

Smallest value for $C(N)$?

PEPs on a simple example

We wish to solve problem

$$\min_{x \in \mathbb{R}^d} h(x),$$

with $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

Using e.g.
relatively inexact proximal point algorithm:

$$x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k),$$

with $\|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2} \|x_{k+1} - x_k\|^2$.

Interested in worst-case bounds of the form

$$h(x_N) - h(x_*) \leq C(N) \|x_0 - x_*\|^2,$$

$x_* \in \underset{x}{\operatorname{argmin}}\ h(x)$

Smallest value for $C(N)$?

PEPs on a simple example

Find smallest $C(N)$ such that

$$h(x_N) - h(x_*) \leq C(N)\|x_0 - x_*\|^2,$$

for any execution of the method on all $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

PEPs on a simple example

Find smallest $C(N)$ such that

$$h(x_N) - h(x_*) \leq C(N)\|x_0 - x_*\|^2,$$

for any execution of the method on all $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

$$C(N) = \quad \text{maximize} \quad \frac{h(x_N) - h(x_*)}{\|x_0 - x_*\|^2}$$

PEPs on a simple example

Find smallest $C(N)$ such that

$$h(x_N) - h(x_*) \leq C(N)\|x_0 - x_*\|^2,$$

for any execution of the method on all $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

$$\begin{aligned} C(N) = \quad & \text{maximize} \quad && \tfrac{h(x_N) - h(x_*)}{\|x_0 - x_*\|^2} \\ & \text{subject to} \quad && d \in \mathbb{N}^*, h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\ & && x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \end{aligned}$$

PEPs on a simple example

Find smallest $C(N)$ such that

$$h(x_N) - h(x_*) \le C(N)\|x_0 - x_*\|^2,$$

for any execution of the method on all $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

$$\begin{aligned} C(N) = \quad & \text{maximize} \quad \frac{h(x_N) - h(x_*)}{\|x_0 - x_*\|^2} \\ & \text{subject to} \quad d \in \mathbb{N}^*, h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\ & \qquad x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\ & \qquad x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \ldots, N-1, \\ & \qquad 0 \in \partial h(x_*), \end{aligned}$$

PEPs on a simple example

Find smallest $C(N)$ such that

$$h(x_N) - h(x_*) \leq C(N)\|x_0 - x_*\|^2,$$

for any execution of the method on all $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

$$\begin{aligned} C(N) = \quad & \text{maximize} && \tfrac{h(x_N)-h(x_*)}{\|x_0 - x_*\|^2} \\ & \text{subject to} && d \in \mathbb{N}^*, h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\ &&& x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\ &&& x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \ldots, N-1, \\ &&& 0 \in \partial h(x_*), \\ &&& \|e_k\|^2 \leq \tfrac{\sigma^2}{\lambda^2}\|x_{k+1} - x_k\|^2. \end{aligned}$$

PEPs on a simple example

Find smallest $C(N)$ such that

$$h(x_N) - h(x_*) \leq C(N)\|x_0 - x_*\|^2,$$

for any execution of the method on all $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

$$\begin{aligned} C(N) = \quad & \text{maximize} \quad && \tfrac{h(x_N)-h(x_*)}{\|x_0 - x_*\|^2} \\ & \text{subject to} \quad && d \in \mathbb{N}^*, h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\ &&& x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \dots, N-1, \\ &&& x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \dots, N-1, \\ &&& 0 \in \partial h(x_*), \\ &&& \|e_k\|^2 \leq \tfrac{\sigma^2}{\lambda^2}\|x_{k+1} - x_k\|^2. \end{aligned}$$

Problem : Infinite dimensional

PEPs on a simple example

Find smallest $C(N)$ such that

$$h(x_N) - h(x_*) \leq C(N)\|x_0 - x_*\|^2,$$

for any execution of the method on all $h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$.

$$\begin{aligned} C(N) = \quad & \text{maximize} \quad && \tfrac{h(x_N)-h(x_*)}{\|x_0 - x_*\|^2} \\ & \text{subject to} \quad && d \in \mathbb{N}^*, h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\ &&& x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\ &&& x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \ldots, N-1, \\ &&& 0 \in \partial h(x_*), \\ &&& \|e_k\|^2 \leq \tfrac{\sigma^2}{\lambda^2}\|x_{k+1} - x_k\|^2. \end{aligned}$$

Problem : Infinite dimensional

Can be reformulated as SDP

PEPs on a simple example

$$
\begin{aligned}
C(N) = \quad & \text{maximize} \quad \frac{h(x_N) - h(x_*)}{\|x_0 - x_*\|^2} \\
& \text{subject to} \quad d \in \mathbb{N}^*,\ h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\
& \qquad x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\
& \qquad x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \ldots, N-1, \\
& \qquad 0 \in \partial h(x_*) \\
& \qquad \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2} \|x_{k+1} - x_k\|^2.
\end{aligned}
$$

PEPs on a simple example

$$
\begin{aligned}
C(N) = \quad & \text{maximize} \quad \frac{h(x_N) - h(x_*)}{\|x_0 - x_*\|^2} \\
& \text{subject to} \quad d \in \mathbb{N}^*,\ h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\
& \qquad x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\
& \qquad x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \ldots, N-1, \\
& \qquad 0 \in \partial h(x_*) \\
& \qquad \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2} \|x_{k+1} - x_k\|^2.
\end{aligned}
$$

$$
C(N) = \quad \text{maximize} \quad \frac{h_N - h_*}{\|x_0 - x_*\|^2}
$$

PEPs on a simple example

$$C(N) = \quad \text{maximize} \quad \frac{h(x_N) - h(x_*)}{\|x_0 - x_*\|^2}$$

$$\text{subject to} \quad d \in \mathbb{N}^*, h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d),$$

$$x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1,$$

$$x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \ldots, N-1,$$

$$0 \in \partial h(x_*)$$

$$\|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2} \|x_{k+1} - x_k\|^2.$$

$$C(N) = \quad \text{maximize} \quad \frac{h_N - h_*}{\|x_0 - x_*\|^2}$$

$$\text{subject to} \quad d \in \mathbb{N}^*, \exists h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d) \text{ s.t. } h_i = h(x_i),$$

$$g_i \in \partial h(x_i), \text{ for } i \in \{*, 0, \ldots, N\},$$

$$x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1,$$

PEPs on a simple example

$$
\begin{aligned}
C(N) = \quad & \text{maximize} \quad \frac{h(x_N)-h(x_*)}{\|x_0-x_*\|^2} \\
& \text{subject to} \quad d \in \mathbb{N}^*, h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\
& \qquad x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\
& \qquad x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \ldots, N-1, \\
& \qquad 0 \in \partial h(x_*) \\
& \qquad \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2}\|x_{k+1}-x_k\|^2.
\end{aligned}
$$

$$
\begin{aligned}
C(N) = \quad & \text{maximize} \quad \frac{h_N-h_*}{\|x_0-x_*\|^2} \\
& \text{subject to} \quad d \in \mathbb{N}^*, \exists h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d) \text{ s.t. } h_i = h(x_i), \\
& \qquad g_i \in \partial h(x_i), \text{ for } i \in \{*, 0, \ldots, N\}, \\
& \qquad x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\
& \qquad x_{k+1} = x_k - \lambda(g_{k+1} + e_k) \text{ for } k = 0, \ldots, N-1, \\
& \qquad g_* = 0
\end{aligned}
$$

PEPs on a simple example

$$
\begin{aligned}
C(N) = \quad & \text{maximize} \quad \frac{h(x_N)-h(x_*)}{\|x_0-x_*\|^2} \\
& \text{subject to} \quad d \in \mathbb{N}^*, h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\
& \qquad x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\
& \qquad x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \ldots, N-1, \\
& \qquad 0 \in \partial h(x_*) \\
& \qquad \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2}\|x_{k+1} - x_k\|^2.
\end{aligned}
$$

$$
\begin{aligned}
C(N) = \quad & \text{maximize} \quad \frac{h_N-h_*}{\|x_0-x_*\|^2} \\
& \text{subject to} \quad d \in \mathbb{N}^*, \exists h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d) \text{ s.t. } h_i = h(x_i), \\
& \qquad g_i \in \partial h(x_i), \text{ for } i \in \{*, 0, \ldots, N\}, \\
& \qquad x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\
& \qquad x_{k+1} = x_k - \lambda(g_{k+1} + e_k) \text{ for } k = 0, \ldots, N-1, \\
& \qquad g_* = 0 \\
& \qquad \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2}\|x_{k+1} - x_k\|^2.
\end{aligned}
$$

PEPs on a simple example

$$\begin{aligned} C(N) = \quad & \text{maximize} \quad \frac{h(x_N) - h(x_*)}{\|x_0 - x_*\|^2} \\ & \text{subject to} \quad d \in \mathbb{N}^*, h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d), \\ & \qquad x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\ & \qquad x_{k+1} = x_k - \lambda(\partial h(x_{k+1}) + e_k) \text{ for } k = 0, \ldots, N-1, \\ & \qquad 0 \in \partial h(x_*) \\ & \qquad \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2} \|x_{k+1} - x_k\|^2. \end{aligned}$$

$$\begin{aligned} C(N) = \quad & \text{maximize} \quad \frac{h_N - h_*}{\|x_0 - x_*\|^2} \\ & \text{subject to} \quad d \in \mathbb{N}^*, \exists h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d) \text{ s.t. } h_i = h(x_i), \\ & \qquad g_i \in \partial h(x_i), \text{ for } i \in \{*, 0, \ldots, N\}, \\ & \qquad x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\ & \qquad x_{k+1} = x_k - \lambda(g_{k+1} + e_k) \text{ for } k = 0, \ldots, N-1, \\ & \qquad g_* = 0 \\ & \qquad \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2} \|x_{k+1} - x_k\|^2. \end{aligned}$$

Finite number of function evaluations and subgradients

$\Longrightarrow$ use interpolation (or extension) thms.

Convex interpolation

Given a set of triplets $S = \{(x_i, g_i, h_i)\}_i$

Convex interpolation

Given a set of triplets $S = \{(x_i, g_i, h_i)\}_i$

Convex interpolation:

$$\exists h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d) \text{ s.t. } h_i = h(x_i), g_i \in \partial h(x_i), \text{ for all } (x_i, g_i, h_i) \in S$$

equivalent to

Convex interpolation

Given a set of triplets $S = \{(x_i, g_i, h_i)\}_i$

Convex interpolation:

$$\exists h \in \mathcal{F}_{0,\infty}(\mathbb{R}^d) \text{ s.t. } h_i = h(x_i), g_i \in \partial h(x_i), \text{ for all } (x_i, g_i, h_i) \in S$$

equivalent to

$$h_j \geq h_i + \langle g_i, x_j - x_i \rangle, \text{ for all } (x_i, g_i, h_i), (x_j, g_j, h_j) \in S$$

PEPs on a simple example

$$C(N) = \quad \text{maximize} \quad \frac{h_N - h_*}{\|x_0 - x_*\|^2}$$

PEPs on a simple example

$$
\begin{aligned}
C(N) = \quad & \text{maximize} && \frac{h_N - h_*}{\|x_0 - x_*\|^2} \\
& \text{subject to} && d \in \mathbb{N}^*, \\
& && x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \ldots, N-1, \\
& && h_j \geq h_i + \langle g_i, x_j - x_i \rangle \text{ for } i, j \in \{*, 0, \ldots, N\},
\end{aligned}
$$

PEPs on a simple example

$$
\begin{aligned}
C(N) = \quad & \text{maximize} && \frac{h_N - h_*}{\|x_0 - x_*\|^2} \\
& \text{subject to} && d \in \mathbb{N}^*, \\
& && x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \dots, N-1, \\
& && h_j \geq h_i + \langle g_i, x_j - x_i \rangle \text{ for } i, j \in \{*, 0, \dots, N\}, \\
& && x_{k+1} = x_k - \lambda(g_{k+1} + e_k) \text{ for } k = 0, \dots, N-1, \\
& && g_* = 0, \\
& && \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2} \|x_{k+1} - x_k\|^2.
\end{aligned}
$$

PEPs on a simple example

$+$ rescaling argument:

$\frac{h_i}{\|x_0 - x_*\|^2}, \frac{g_i}{\|x_0 - x_*\|}, \frac{x_i}{\|x_0 - x_*\|}, \frac{e_i}{\|x_0 - x_*\|}$ are still feasibles

$\implies$ we can set $\|x_0 - x_*\| = 1$.

PEPs on a simple example

+ rescaling argument:

$\frac{h_i}{\|x_0 - x_*\|^2}, \frac{g_i}{\|x_0 - x_*\|}, \frac{x_i}{\|x_0 - x_*\|}, \frac{e_i}{\|x_0 - x_*\|}$ are still feasibles

$\implies$ we can set $\|x_0 - x_*\| = 1$.

$$
\begin{aligned}
C(N) = \quad & \text{maximize} && h_N - h_* \\
& \text{subject to} && d \in \mathbb{N}^*, \\
& && x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \dots, N-1, \\
& && h_j \geq h_i + \langle g_i, x_j - x_i \rangle \text{ for } i, j \in \{*, 0, \dots, N\}, \\
& && x_{k+1} = x_k - \lambda(g_{k+1} + e_k) \text{ for } k = 0, \dots, N-1, \\
& && g_* = 0, \\
& && \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2} \|x_{k+1} - x_k\|^2, \\
& && \|x_0 - x_*\|^2 = 1.
\end{aligned}
$$

PEPs on a simple example

+ rescaling argument:

$\frac{h_i}{\|x_0 - x_*\|^2}, \frac{g_i}{\|x_0 - x_*\|}, \frac{x_i}{\|x_0 - x_*\|}, \frac{e_i}{\|x_0 - x_*\|}$ are still feasibles

$\implies$ we can set $\|x_0 - x_*\| = 1$.

$$
\begin{aligned}
C(N) = \quad & \text{maximize} \quad && h_N - h_* \\
& \text{subject to} \quad && d \in \mathbb{N}^*, \\
& && x_0 \in \mathbb{R}^d,\ e_k \in \mathbb{R}^d \text{ for } k = 0, \dots, N-1, \\
& && h_j \geq h_i + \langle g_i, x_j - x_i \rangle \text{ for } i, j \in \{*, 0, \dots, N\}, \\
& && x_{k+1} = x_k - \lambda(g_{k+1} + e_k) \text{ for } k = 0, \dots, N-1, \\
& && g_* = 0, \\
& && \|e_k\|^2 \leq \frac{\sigma^2}{\lambda^2} \|x_{k+1} - x_k\|^2, \\
& && \|x_0 - x_*\|^2 = 1.
\end{aligned}
$$

Problem is linear in h_i's and in $\langle x, y \rangle$, with $x, y \in \{x_i\} \cup \{e_i\} \cup \{g_i\}$

$\implies$ SDP reformulation.

Numerical worst-case for simple relatively inexact PPA.

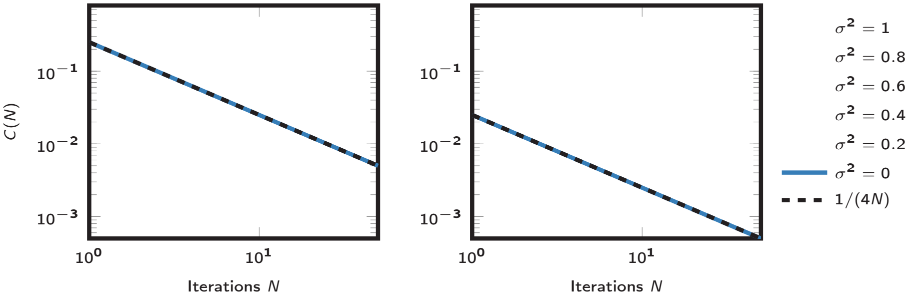

Figure: Worst-case guarantees on $C(N)$ bound on $[h(x_N) - h(x_*)]/\|x_0 - x_*\|^2$, as function of N. Left: $\lambda = 1$. Right: $\lambda = 10$.

Numerical worst-case for simple relatively inexact PPA.

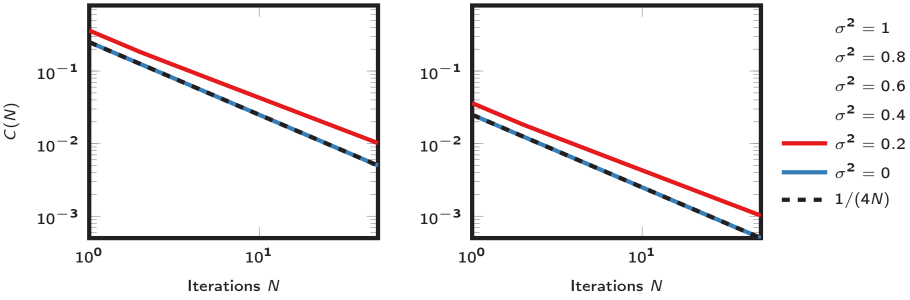

Figure: Worst-case guarantees on $C(N)$ bound on $[h(x_N) - h(x_*)]/\|x_0 - x_*\|^2$, as function of N. Left: $\lambda = 1$. Right: $\lambda = 10$.

Numerical worst-case for simple relatively inexact PPA.

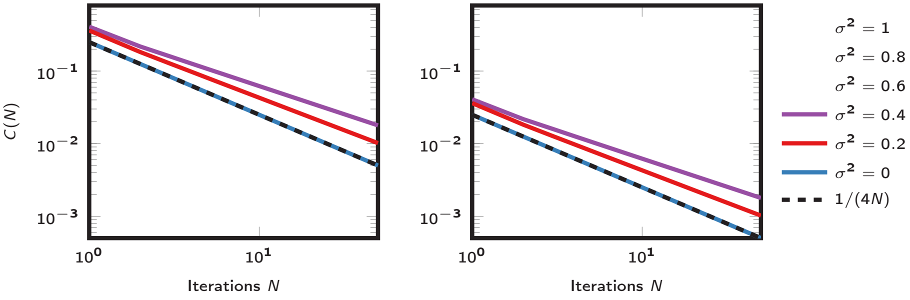

Figure: Worst-case guarantees on $C(N)$ bound on $[h(x_N) - h(x_*)]/\|x_0 - x_*\|^2$, as function of N. Left: $\lambda = 1$. Right: $\lambda = 10$.

Numerical worst-case for simple relatively inexact PPA.

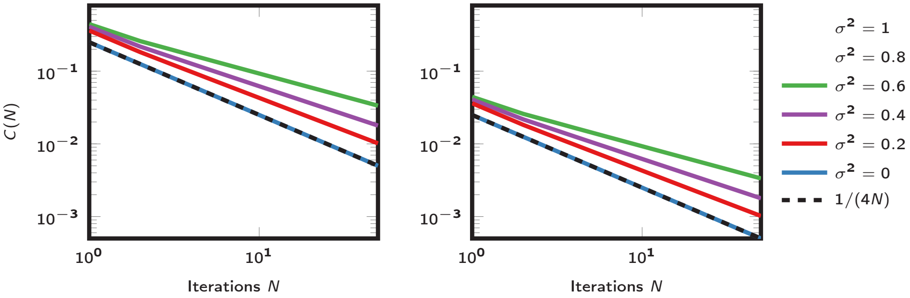

Figure: Worst-case guarantees on $C(N)$ bound on $[h(x_N) - h(x_*)]/\|x_0 - x_*\|^2$, as function of N. Left: $\lambda = 1$. Right: $\lambda = 10$.

Numerical worst-case for simple relatively inexact PPA.

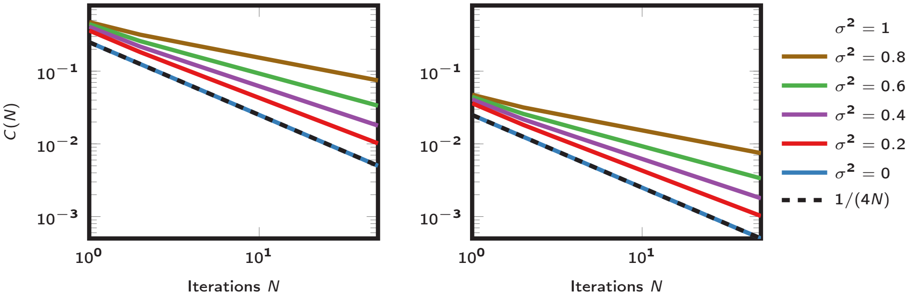

Figure: Worst-case guarantees on $C(N)$ bound on $[h(x_N) - h(x_*)]/\|x_0 - x_*\|^2$, as function of N. Left: $\lambda = 1$. Right: $\lambda = 10$.

Numerical worst-case for simple relatively inexact PPA.

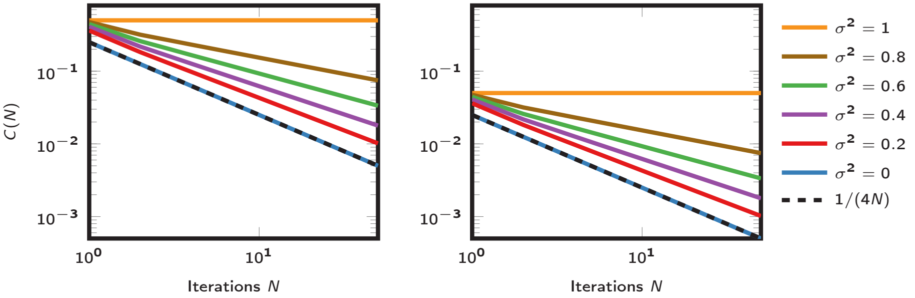

Figure: Worst-case guarantees on $C(N)$ bound on $[h(x_N) - h(x_*)]/\|x_0 - x_*\|^2$, as function of N. Left: $\lambda = 1$. Right: $\lambda = 10$.

Numerical worst-case for simple relatively inexact PPA.

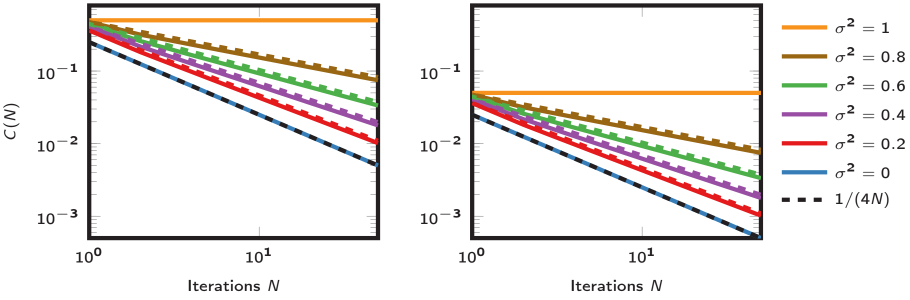

Figure: Worst-case guarantees on $C(N)$ bound on $[h(x_N) - h(x_*)]/\|x_0 - x_*\|^2$, as function of N. Left: $\lambda = 1$. Right: $\lambda = 10$.

Dashed curves: $N \to \frac{1+\sigma}{4\lambda N\sqrt{1-\sigma^2}}$ (conjecture).

SDP reformulation

If objective and constraints are **affine** in function values $h(\cdot)$ or $h^*(\cdot)$ and in $\langle x, y\rangle$ with x, y some iterates or subgradients

SDP reformulation

If objective and constraints are **affine** in function values $h(\cdot)$ or $h^*(\cdot)$ and in $\langle x, y \rangle$ with x, y some iterates or subgradients

$\implies$

problem can be reformulated as SDP

SDP reformulation

If objective and constraints are **affine** in function values $h(\cdot)$ or $h^*(\cdot)$ and in $\langle x, y \rangle$ with x, y some iterates or subgradients

$\Longrightarrow$

problem can be reformulated as SDP

- Other method.

SDP reformulation

If objective and constraints are **affine** in function values $h(\cdot)$ or $h^*(\cdot)$ and in $\langle x, y\rangle$ with x, y some iterates or subgradients

$\implies$

problem can be reformulated as SDP

- Other method.
- Other objective criteria.

SDP reformulation

If objective and constraints are **affine** in function values $h(\cdot)$ or $h^*(\cdot)$ and in $\langle x, y \rangle$ with x, y some iterates or subgradients

$\Longrightarrow$

problem can be reformulated as SDP

- Other method.
- Other objective criteria.
- Other inexactness criteria

SDP reformulation

If objective and constraints are **affine** in function values $h(\cdot)$ or $h^*(\cdot)$ and in $\langle x, y\rangle$ with x, y some iterates or subgradients

$\Longrightarrow$

problem can be reformulated as SDP

- Other method.
- Other objective criteria.
- Other inexactness criteria
  - $\frac{x_k - x_{k+1}}{\lambda} \in \partial_\varepsilon h(x_{k+1})$ with $\varepsilon \leq \dots$

SDP reformulation

If objective and constraints are **affine** in function values $h(\cdot)$ or $h^*(\cdot)$ and in $\langle x, y \rangle$ with x, y some iterates or subgradients

$\Longrightarrow$

problem can be reformulated as SDP

- Other method.
- Other objective criteria.
- Other inexactness criteria
  - $\frac{x_k - x_{k+1}}{\lambda} \in \partial_\varepsilon h(x_{k+1})$ with $\varepsilon \leq \dots$
  - Primal-dual gap of proximal problem (Prox) $\leq \dots$

SDP reformulation

If objective and constraints are **affine** in function values $h(\cdot)$ or $h^*(\cdot)$ and in $\langle x, y\rangle$ with x, y some iterates or subgradients

$\Longrightarrow$

problem can be reformulated as SDP

- Other method.
- Other objective criteria.
- Other inexactness criteria
  - $\frac{x_k - x_{k+1}}{\lambda} \in \partial_\varepsilon h(x_{k+1})$ with $\varepsilon \leq \dots$
  - Primal-dual gap of proximal problem (Prox) $\leq \dots$
  - See paper for review of inexactness criteria.

Numerical example I.

Inexact proximal-gradient $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} \quad \approx \operatorname{prox}_{\frac{(1-\sigma^2)g}{L}}\left(x_k - \frac{1-\sigma^2}{L}\nabla f(x_k)\right) \\ \end{cases}$$

Numerical example I.

Inexact proximal-gradient $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{(1-\sigma^2)g}{L}}\left(x_k - \frac{1-\sigma^2}{L}\nabla f(x_k)\right) \\ & \text{s.t } \left\| x_{k+1} - x_k + \frac{1-\sigma^2}{L}\nabla f(x_k) + \frac{1-\sigma^2}{L}\partial g(x_{k+1}) \right\|^2 \leq \sigma^2 \|x_{k+1} - x_k\|^2 \end{cases}$$

Numerical example I.

Inexact proximal-gradient $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{(1-\sigma^2)g}{L}}\left(x_k - \frac{1-\sigma^2}{L}\nabla f(x_k)\right) \\ & \text{s.t } \left\| x_{k+1} - x_k + \frac{1-\sigma^2}{L}\nabla f(x_k) + \frac{1-\sigma^2}{L}\partial g(x_{k+1})\right\|^2 \leq \sigma^2 \|x_{k+1} - x_k\|^2 \end{cases}$$

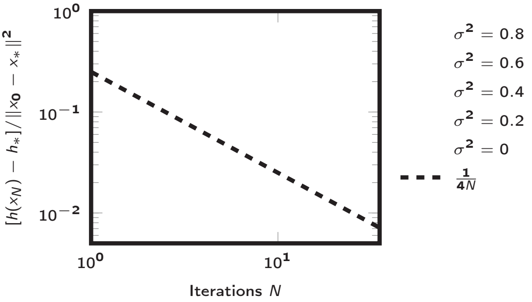

Figure: Worst-case guarantees on $(h(x_N) - h_*)/\|x_0 - x_*\|^2$ as function of N.

Numerical example I.

Inexact proximal-gradient $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{(1-\sigma^2)g}{L}}\left(x_k - \frac{1-\sigma^2}{L}\nabla f(x_k)\right) \\ & \text{s.t } \left\| x_{k+1} - x_k + \frac{1-\sigma^2}{L}\nabla f(x_k) + \frac{1-\sigma^2}{L}\partial g(x_{k+1}) \right\|^2 \leq \sigma^2 \|x_{k+1} - x_k\|^2 \end{cases}$$

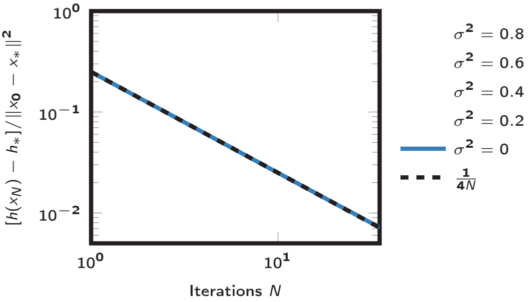

Figure: Worst-case guarantees on $(h(x_N) - h_*)/\|x_0 - x_*\|^2$ as function of N.

Numerical example I.

Inexact proximal-gradient $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{(1-\sigma^2)g}{L}}\left(x_k - \frac{1-\sigma^2}{L}\nabla f(x_k)\right) \\ & \text{s.t } \left\| x_{k+1} - x_k + \frac{1-\sigma^2}{L}\nabla f(x_k) + \frac{1-\sigma^2}{L}\partial g(x_{k+1}) \right\|^2 \leq \sigma^2 \|x_{k+1} - x_k\|^2 \end{cases}$$

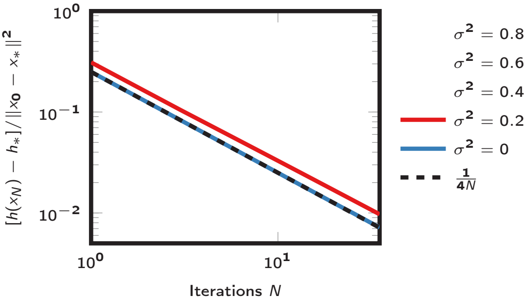

Figure: Worst-case guarantees on $(h(x_N) - h_*)/\|x_0 - x_*\|^2$ as function of N.

Numerical example I.

Inexact proximal-gradient $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \mathrm{prox}_{\frac{(1-\sigma^2)g}{L}}\left(x_k - \frac{1-\sigma^2}{L}\nabla f(x_k)\right) \\ & \text{s.t } \left\| x_{k+1} - x_k + \frac{1-\sigma^2}{L}\nabla f(x_k) + \frac{1-\sigma^2}{L}\partial g(x_{k+1}) \right\|^2 \leq \sigma^2 \|x_{k+1} - x_k\|^2 \end{cases}$$

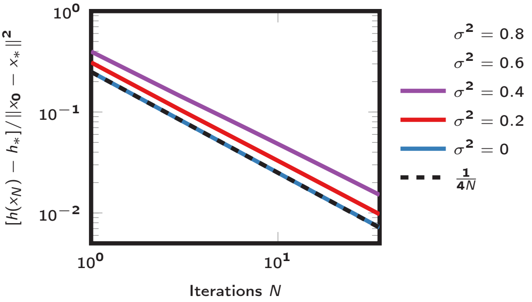

Figure: Worst-case guarantees on $(h(x_N) - h_*)/\|x_0 - x_*\|^2$ as function of N.

Numerical example I.

Inexact proximal-gradient $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{(1-\sigma^2)g}{L}}\left(x_k - \frac{1-\sigma^2}{L}\nabla f(x_k)\right) \\ & \text{s.t } \left\| x_{k+1} - x_k + \frac{1-\sigma^2}{L}\nabla f(x_k) + \frac{1-\sigma^2}{L}\partial g(x_{k+1}) \right\|^2 \leq \sigma^2 \|x_{k+1} - x_k\|^2 \end{cases}$$

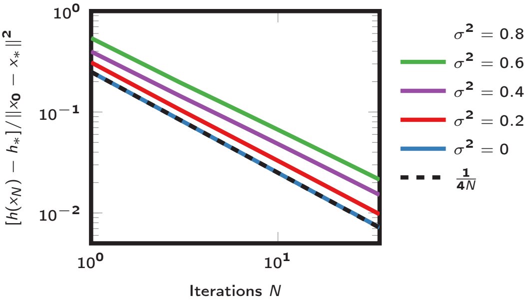

Figure: Worst-case guarantees on $(h(x_N) - h_*)/\|x_0 - x_*\|^2$ as function of N.

Numerical example I.

Inexact proximal-gradient $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{(1-\sigma^2)g}{L}}\left(x_k - \frac{1-\sigma^2}{L}\nabla f(x_k)\right) \\ & \text{s.t } \left\|x_{k+1} - x_k + \frac{1-\sigma^2}{L}\nabla f(x_k) + \frac{1-\sigma^2}{L}\partial g(x_{k+1})\right\|^2 \leq \sigma^2\|x_{k+1} - x_k\|^2 \end{cases}$$

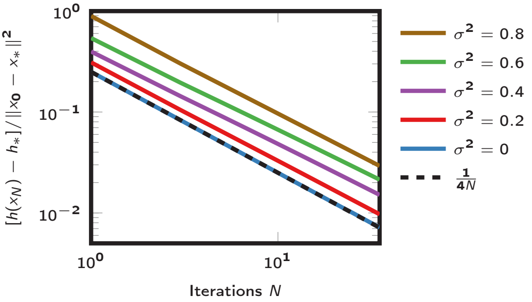

Figure: Worst-case guarantees on $(h(x_N) - h_*)/\|x_0 - x_*\|^2$ as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k)) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \\ & \end{cases}$$

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011),
$h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k)) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011),
$h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \mathrm{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k)) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\mathrm{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

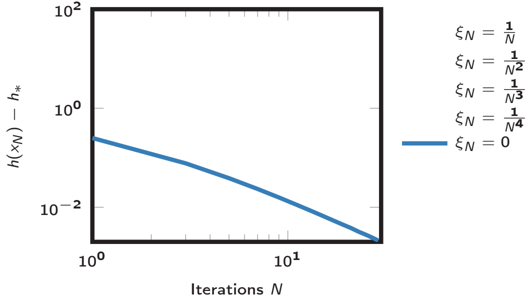

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \leq 1$ and $L = 1$, as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011),
$h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \mathrm{prox}_{\frac{g}{L}}\left(y_k - \frac{1}{L}\nabla f(y_k)\right) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\mathrm{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

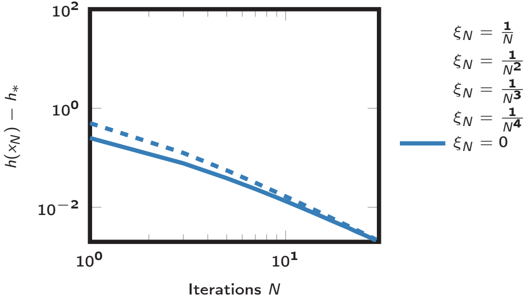

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \leq 1$ and $L = 1$, as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011),
$h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \mathrm{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k)) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\mathrm{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

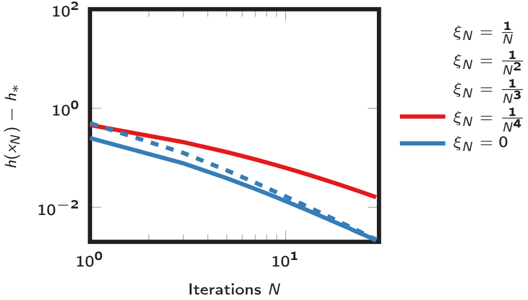

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \leq 1$ and $L = 1$, as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k)) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

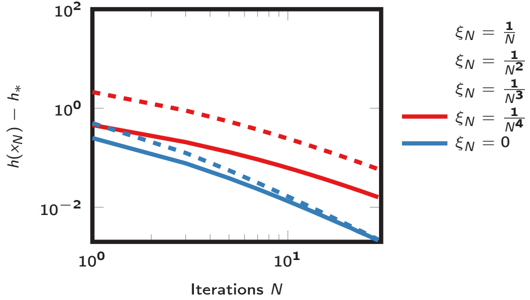

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \leq 1$ and $L = 1$, as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k)) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

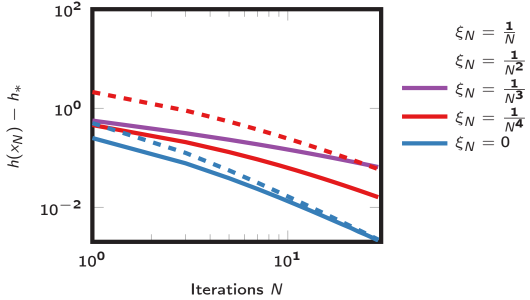

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \leq 1$ and $L = 1$, as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{g}{L}}(y_k - \tfrac{1}{L}\nabla f(y_k)) \\ y_{k+1} & = x_{k+1} + \tfrac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\operatorname{prox}_{\frac{g}{L}}(y_k - \tfrac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \tfrac{1}{L}g(x) + \tfrac{1}{2}\left\|x - y_k + \tfrac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

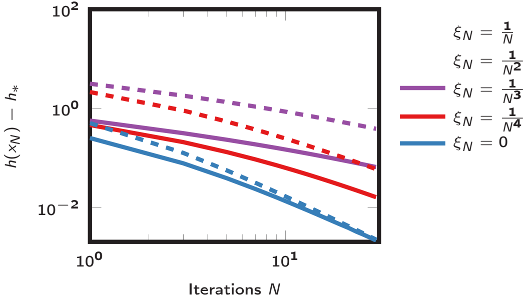

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \leq 1$ and $L = 1$, as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k)) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \le \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

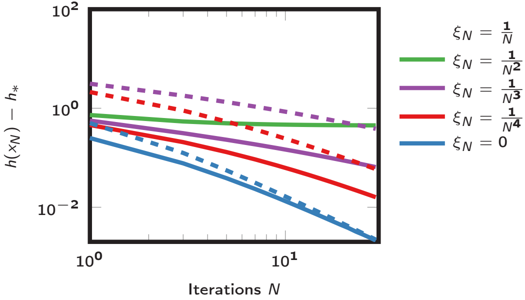

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \le 1$ and $L = 1$, as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{g}{L}}\left(y_k - \frac{1}{L}\nabla f(y_k)\right) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

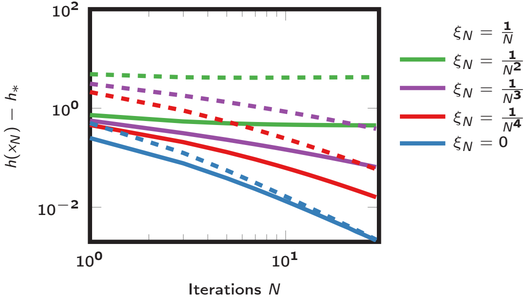

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \leq 1$ and $L = 1$, as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \text{prox}_{\frac{g}{L}}\left(y_k - \frac{1}{L}\nabla f(y_k)\right) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\text{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

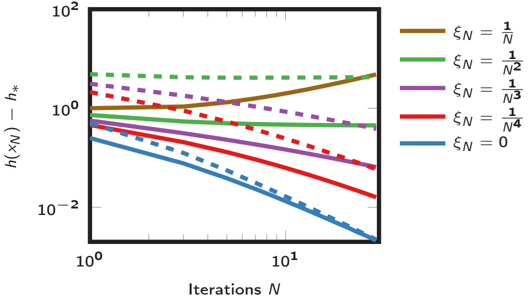

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \leq 1$ and $L = 1$, as function of N.

Numerical example II

Inexact accelerated proximal-gradient (Schmidt, Le Roux & Bach, 2011),
$h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$, $y_0 = x_0$

$$\begin{cases} x_{k+1} & \approx \operatorname{prox}_{\frac{g}{L}}\left(y_k - \frac{1}{L}\nabla f(y_k)\right) \\ y_{k+1} & = x_{k+1} + \frac{k}{k+3}(x_{k+1} - x_k) \\ & \text{s.t } \Phi_k(x_{k+1}) - \Phi_k(\operatorname{prox}_{\frac{g}{L}}(y_k - \frac{1}{L}\nabla f(y_k))) \leq \xi_{k+1}, \\ & \quad \text{with } \Phi_k(x) = \frac{1}{L}g(x) + \frac{1}{2}\left\|x - y_k + \frac{1}{L}\nabla f(y_k)\right\|^2. \end{cases}$$

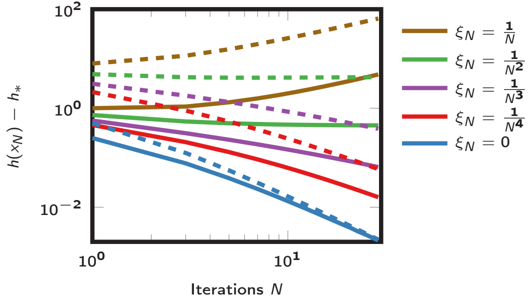

Figure: Worst-case guarantees on $h(x_N) - h_*$ with initial condition $\|x_0 - x_*\|^2 \leq 1$ and $L = 1$, as function of N.

Numerical example III

Relatively inexact Douglas-Rachford (Eckstein & Yao, 2018), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$,

$$\begin{cases} x_k & \approx \mathrm{prox}_{\lambda f}(z_k), \\ y_k & = \mathrm{prox}_{\lambda g}(x_k - \lambda \nabla f(x_k)), \\ z_{k+1} & = z_k + y_k - x_k \\ \end{cases}$$

Numerical example III

Relatively inexact Douglas-Rachford (Eckstein & Yao, 2018), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$,

$$\begin{cases} x_k & \approx \mathrm{prox}_{\lambda f}(z_k), \\ y_k & = \mathrm{prox}_{\lambda g}(x_k - \lambda \nabla f(x_k)), \\ z_{k+1} & = z_k + y_k - x_k \\ & \quad \text{s.t} \|x_k - z_k + \lambda \nabla f(x_k)\|^2 \leq \sigma^2 \|y_k - z_k + \lambda \nabla f(x_k)\|^2. \end{cases}$$

Numerical example III

Relatively inexact Douglas-Rachford (Eckstein & Yao, 2018), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$,

$$\begin{cases} x_k & \approx \mathrm{prox}_{\lambda f}(z_k), \\ y_k & = \mathrm{prox}_{\lambda g}(x_k - \lambda \nabla f(x_k)), \\ z_{k+1} & = z_k + y_k - x_k \\ & \quad \text{s.t} \|x_k - z_k + \lambda \nabla f(x_k)\|^2 \leq \sigma^2 \|y_k - z_k + \lambda \nabla f(x_k)\|^2. \end{cases}$$

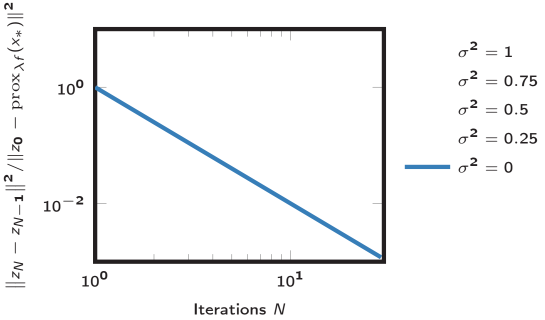

Figure: Worst-case guarantees on $\|z_N - z_{N-1}\|^2 / \|z_0 - \mathrm{prox}_{\lambda f}(x_*)\|^2$, as function of N.

Numerical example III

Relatively inexact Douglas-Rachford (Eckstein & Yao, 2018), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$,

$$\begin{cases} x_k & \approx \mathrm{prox}_{\lambda f}(z_k), \\ y_k & = \mathrm{prox}_{\lambda g}(x_k - \lambda \nabla f(x_k)), \\ z_{k+1} & = z_k + y_k - x_k \\ & \quad \text{s.t} \|x_k - z_k + \lambda \nabla f(x_k)\|^2 \leq \sigma^2 \|y_k - z_k + \lambda \nabla f(x_k)\|^2. \end{cases}$$

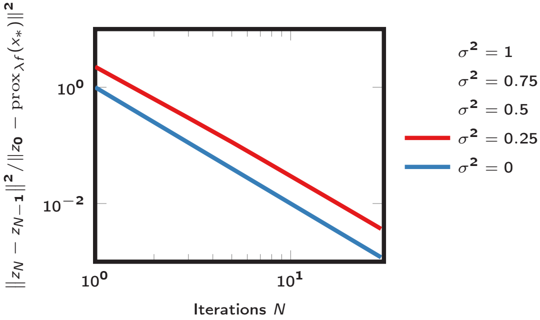

Figure: Worst-case guarantees on $\|z_N - z_{N-1}\|^2 / \|z_0 - \mathrm{prox}_{\lambda f}(x_*)\|^2$, as function of N.

Numerical example III

Relatively inexact Douglas-Rachford (Eckstein & Yao, 2018), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$,

$$\begin{cases} x_k & \approx \mathrm{prox}_{\lambda f}(z_k), \\ y_k & = \mathrm{prox}_{\lambda g}(x_k - \lambda \nabla f(x_k)), \\ z_{k+1} & = z_k + y_k - x_k \\ & \quad \text{s.t} \|x_k - z_k + \lambda \nabla f(x_k)\|^2 \leq \sigma^2 \|y_k - z_k + \lambda \nabla f(x_k)\|^2. \end{cases}$$

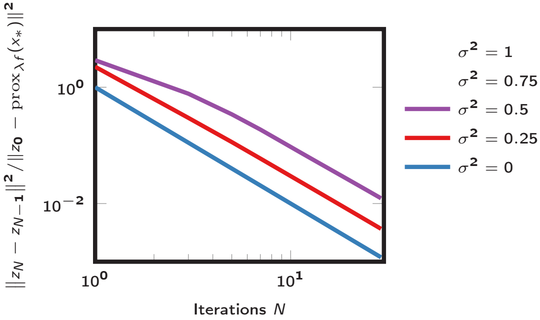

Figure: Worst-case guarantees on $\|z_N - z_{N-1}\|^2 / \|z_0 - \mathrm{prox}_{\lambda f}(x_*)\|^2$, as function of N.

Numerical example III

Relatively inexact Douglas-Rachford (Eckstein & Yao, 2018), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$,

$$\begin{cases} x_k & \approx \mathrm{prox}_{\lambda f}(z_k), \\ y_k & = \mathrm{prox}_{\lambda g}(x_k - \lambda \nabla f(x_k)), \\ z_{k+1} & = z_k + y_k - x_k \\ & \quad \text{s.t} \|x_k - z_k + \lambda \nabla f(x_k)\|^2 \leq \sigma^2 \|y_k - z_k + \lambda \nabla f(x_k)\|^2. \end{cases}$$

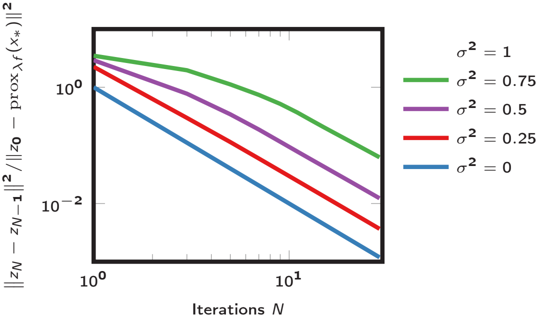

Figure: Worst-case guarantees on $\|z_N - z_{N-1}\|^2 / \|z_0 - \mathrm{prox}_{\lambda f}(x_*)\|^2$, as function of N.

Numerical example III

Relatively inexact Douglas-Rachford (Eckstein & Yao, 2018), $h = f + g$, with $f \in \mathcal{F}_{0,L}(\mathbb{R}^d)$ and $g \in \mathcal{F}_{0,\infty}(\mathbb{R}^d)$,

$$\begin{cases} x_k & \approx \mathrm{prox}_{\lambda f}(z_k), \\ y_k & = \mathrm{prox}_{\lambda g}(x_k - \lambda \nabla f(x_k)), \\ z_{k+1} & = z_k + y_k - x_k \\ & \quad \text{s.t} \|x_k - z_k + \lambda \nabla f(x_k)\|^2 \leq \sigma^2 \|y_k - z_k + \lambda \nabla f(x_k)\|^2. \end{cases}$$

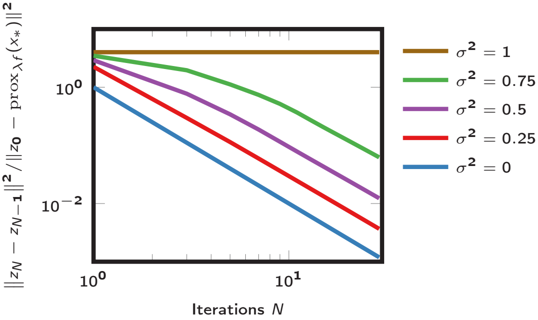

Figure: Worst-case guarantees on $\|z_N - z_{N-1}\|^2 / \|z_0 - \mathrm{prox}_{\lambda f}(x_*)\|^2$, as function of N.

Conclusion

- Practical worst-case analyses of “Outer” algorithms
  (whatever subroutines for prox approximations)

Conclusion

- Practical worst-case analyses of “Outer” algorithms (whatever subroutines for prox approximations)

- PEPs can provide proofs.

Conclusion

- Practical worst-case analyses of “Outer” algorithms (whatever subroutines for prox approximations)
- PEPs can provide proofs.
- PEPs allow designing methods with optimized worst-case behaviors. (see ORI-PPA in the paper)

Conclusion

- Practical worst-case analyses of "Outer" algorithms (whatever subroutines for prox approximations)
- PEPs can provide proofs.
- PEPs allow designing methods with optimized worst-case behaviors. (see ORI-PPA in the paper)
- Can search for Lyapunov arguments in PEP $\implies$ simpler proofs.

References I

M. Barré, A. Taylor, and F. Bach.
Principled analyses and design of first-order methods with inexact proximal operators.
arXiv preprint arXiv:2006.06041, 2020.

Y. Drori and M. Teboulle.
Performance of first-order methods for smooth convex minimization: a novel approach.
Mathematical Programming, 145(1-2):451–482, 2014.

J. Eckstein and W. Yao.
Relative-error approximate versions of Douglas–Rachford splitting and special cases of the ADMM.
Mathematical Programming, 170(2):417–444, 2018.

B. Martinet.
Détermination approchée d'un point fixe d'une application pseudo-contractante. cas de l'application prox.
Comptes rendus hebdomadaires des séances de l'Académie des sciences de Paris, 274:163–165, 1972.

R. T. Rockafellar.
Monotone operators and the proximal point algorithm.
SIAM journal on control and optimization, 14(5):877–898, 1976.

M. Schmidt, N. Le Roux, and F. Bach.
Convergence rates of inexact proximal-gradient methods for convex optimization.
In *Advances in neural information processing systems (NIPS)*, pages 1458–1466, 2011.

References II

A. B. Taylor, J. M. Hendrickx, and F. Glineur.
Smooth strongly convex interpolation and exact worst-case performance of first-order methods.
Mathematical Programming, 161(1-2):307–345, 2017.