Towards Performance Estimation Problems on Quadratic Functions

Nizar Bousselmi

Supervisors: François Glineur and Julien Hendrickx

Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM)
Université catholique de Louvain (UCLouvain)

Worst-case performance of a method on a class of functions

Common question in optimization :

Worst-case performance of an optimization method $\mathcal{M}$ on

$$\min_x f(x)$$

where $f \in \mathcal{F}$ has some properties (smoothness, convexity,...) ?

Worst-case performance of a method on a class of functions

Common question in optimization :

Worst-case performance of an optimization method $\mathcal{M}$ on

$$\min_x f(x)$$

where $f \in \mathcal{F}$ has some properties (smoothness, convexity,...) ?

Example :

Worst-case performance of $\overbrace{\textbf{gradient method}}^{\mathcal{M}}$ on $\overbrace{L\textbf{-smooth convex functions}}^{\mathcal{F}}$ (after N iterations) ?

$$f(x_N) - f^* \leq \frac{L}{2} \frac{1}{2N+1}.$$

Performance Estimation Problem (PEP)

Theoretical and practical framework to analyze **performance** of optimization methods on problem classes.

- *Performance of first-order methods...*Drori & Teboulle 2013
- *Convex interpolation and performance estimation...*Taylor 2017

Performance Estimation Problem (PEP)

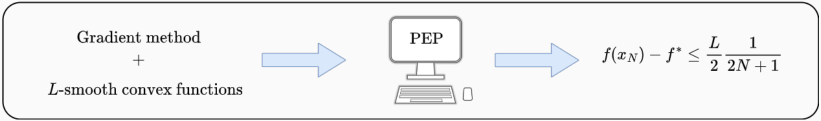
Theoretical and practical framework to analyze **performance** of optimization methods on problem classes.

- *Performance of first-order methods...* Drori & Teboulle 2013
- *Convex interpolation and performance estimation...* Taylor 2017

Theoretical and practical framework to analyze **performance** of optimization methods on problem classes.

- *Performance of first-order methods*...Drori & Teboulle 2013
- *Convex interpolation and performance estimation*...Taylor 2017

Example :



Performance Estimation Problem (PEP)

Theoretical and practical framework to analyze **performance** of optimization methods on problem classes.

- *Performance of first-order methods...* Drori & Teboulle 2013
- *Convex interpolation and performance estimation...* Taylor 2017

Our contribution :

Input = optimization method + problem class
Output = worst instance in problem class

Example:

- N steps of gradient method $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$
- L-smooth convex functions f

PEP

$$\begin{aligned} \max_{\text{points } x_k, x^*, \text{ function } f} \quad & f(x_N) - f(x^*) \\ \text{s.t.} \quad & f \; L\text{-smooth convex}, \\ & x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k), \\ & ||x^* - x_0|| \leq 1, \\ & \nabla f(x^*) = 0. \end{aligned}$$

INPUT = OPTIMIZATION METHOD + PROBLEM CLASS
OUTPUT = WORST INSTANCE IN PROBLEM CLASS

Example:

- N steps of gradient method $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$
- L-smooth convex functions f

PEP

$$
\begin{aligned}
\max_{\text{points } x_k, x^*, \text{ function } f} \quad & f(x_N) - f(x^*) \\
\text{s.t.} \quad & f \ L\text{-smooth convex}, \\
& x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k), \\
& ||x^* - x_0|| \leq 1, \\
& \nabla f(x^*) = 0.
\end{aligned}
$$

Input = optimization method + problem class
Output = worst instance in problem class

Example:

- N steps of gradient method $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$
- L-smooth convex functions f

PEP

$$\begin{aligned} \max_{\text{points } x_k, x^*, \text{ function } f} \quad & f(x_N) - f(x^*) \\ \text{s.t.} \quad & f \ L\text{-smooth convex}, \\ & x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k), \\ & ||x^* - x_0|| \leq 1, \\ & \nabla f(x^*) = 0. \end{aligned}$$

INPUT = OPTIMIZATION METHOD + PROBLEM CLASS
OUTPUT = WORST INSTANCE IN PROBLEM CLASS

Example:

- N steps of gradient method $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$
- L-smooth convex functions f

PEP

$$\max_{\text{points } x_k, x^*, \text{ function } f} \quad f(x_N) - f(x^*)$$

$$\begin{aligned} \text{s.t.} \quad & f \ L\text{-smooth convex}, \\ & x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k), \\ & ||x^* - x_0|| \leq 1, \\ & \nabla f(x^*) = 0. \end{aligned}$$

Input = optimization method + problem class
Output = worst instance in problem class

Example :

- N steps of gradient method $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$
- L-smooth convex functions f

PEP

$$\begin{aligned} \max_{\text{points } x_k, x^*, \text{ function } f} \quad & f(x_N) - f(x^*) \\ \text{s.t.} \quad & f \ L\text{-smooth convex}, \\ & x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k), \\ & \|x^* - x_0\| \leq 1, \\ & \nabla f(x^*) = 0. \end{aligned}$$

Output : $f(x_N) - f^* \leq \frac{L}{2}\frac{1}{2N+1}$ and worst function f achieving it.

PEP as finite-dimensional problem

f infinite-dimensional but algorithm only sees x_k, $f(x_k)$ and $\nabla f(x_k)$...

PEP

$$
\begin{aligned}
\max_{\text{points } x_k, x^*, \text{function } f} \quad & f(x_N) - f(x^*) \\
\text{s.t.} \quad & f \ L\text{-smooth convex}, \\
& x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k), \\
& ||x^* - x_0|| \leq 1, \\
& \nabla f(x^*) = 0.
\end{aligned}
$$

PEP as finite-dimensional problem

f infinite-dimensional but algorithm only sees x_k, $f(x_k)$ and $\nabla f(x_k)$...

PEP

$$\max_{\text{points } x_k, x^*, f_k, f^*, g_k, g^*} \quad f_N - f^*$$

$$\begin{aligned} \text{s.t.} \quad & \exists f \text{ } L\text{-smooth convex} : \quad f(x_k) = f_k, \quad \nabla f(x_k) = g_k, \\ & \qquad\qquad\qquad\qquad\qquad f(x^*) = f^*, \quad \nabla f(x^*) = g^*, \\ & x_{k+1} = x_k - \frac{1}{L} g_k \\ & ||x^* - x_0|| \leq 1, \\ & g^* = 0. \end{aligned}$$

PEP as finite-dimensional problem

f infinite-dimensional but algorithm only sees x_k, $f(x_k)$ and $\nabla f(x_k)$...

PEP

$$\begin{aligned}
\max_{\text{points } x_k, x^*, f_k, f^*, g_k, g^*} \quad & f_N - f^* \\
\text{s.t.} \quad & \exists f \ L\text{-smooth convex}: \quad f(x_k) = f_k, \ \ \nabla f(x_k) = g_k, \\
& \qquad\qquad\qquad\qquad\quad f(x^*) = f^*, \ \ \nabla f(x^*) = g^*, \\
& x_{k+1} = x_k - \frac{1}{L} g_k \\
& ||x^* - x_0|| \leq 1, \\
& g^* = 0.
\end{aligned}$$

Interpolation condition to reformulate.

PEP as finite-dimensional problem

f infinite-dimensional but algorithm only sees x_k, $f(x_k)$ and $\nabla f(x_k)$...

PEP

$$\begin{aligned}
\max_{\text{points } x_k, x^*, f_k, f^*, g_k, g^*} \quad & f_N - f^* \\
\text{s.t.} \quad & \exists f \ L\text{-smooth convex}: \quad f(x_k) = f_k, \quad \nabla f(x_k) = g_k, \\
& \qquad\qquad f(x^*) = f^*, \quad \nabla f(x^*) = g^*, \\
& x_{k+1} = x_k - \frac{1}{L} g_k \\
& ||x^* - x_0|| \leq 1, \\
& g^* = 0.
\end{aligned}$$

Interpolation condition to reformulate.

At the end, convex **semidefinite problem** efficiently solvable !

Currently formulable PEP

Interpolation conditions for L-smooth convex functions

Given x_k, g_k and f_k,

$\exists$ L-smooth convex f such that $\begin{cases} f(x_k) & = f_k \;\; \forall k, \\ \nabla f(x_k) & = g_k \;\; \forall k, \end{cases}$ if and only if

$$f_j \geq f_k + g_k^T(x_j - x_k) + \frac{1}{2L}||g_j - g_k||^2 \quad \forall j, k.$$

Interpolation conditions for L-smooth convex functions

Given x_k, g_k and f_k,

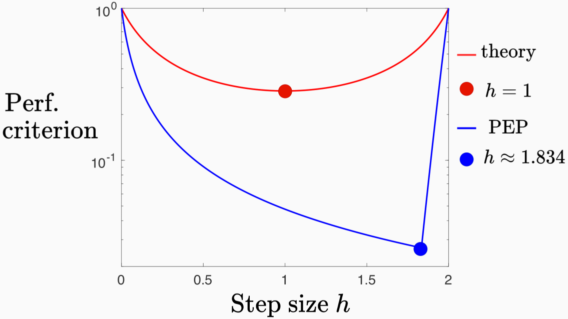
$\exists$ L-smooth convex f such that $\begin{cases} f(x_k) & = f_k \;\; \forall k, \\ \nabla f(x_k) & = g_k \;\; \forall k, \end{cases}$ if and only if

$$f_j \geq f_k + g_k^T(x_j - x_k) + \frac{1}{2L}||g_j - g_k||^2 \quad \forall j, k.$$

Remark : Interpolation conditions (and PEP formulation) for numerous function classes : **non-smooth, L-smooth, convex, μ-strongly convex,...**

Exploitation of PEP

Performance of 10 gradient method steps



Remarks :

- theory suggests a step size of 1 and PEP of ≈ 1.834;
- PEP provides tight results.

Extension of PEP : problems involving matrices

We would like to analyze the worst performance of methods on problems involving matrices:

- $\min_x \frac{1}{2}x^T Q x$
- $\min_x g(Ax)$
- ...

Extension of PEP : problems involving matrices

We would like to analyze the worst performance of methods on problems involving matrices:

- $\min_x \frac{1}{2}x^T Qx$
- $\min_x g(Ax)$
- ...

Desired interpolation conditions

Given x_k and y_k,

$\exists A$ such that $y_k = Ax_k \quad \forall k$ if and only if

(interpolation conditions).

Extension of PEP : problems involving matrices

We would like to analyze the worst performance of methods on problems involving matrices:

- $\min_x \frac{1}{2}x^TQx$
- $\min_x g(Ax)$
- ...

Desired interpolation conditions

Given x_k and y_k,

$\exists A$ such that $y_k = Ax_k \ \ \forall k$ if and only if

(interpolation conditions).

$$f(x) = \frac{1}{2}x^TQx \Rightarrow g_k = \nabla f(x_k) = Qx_k$$

$$f(x) = g(Ax) \Rightarrow g_k = \nabla f(x_k) = A^T\nabla g(Ax_k)$$

Extension of PEP : problems involving matrices

We would like to analyze the worst performance of methods on problems involving matrices:

- $\min_x \frac{1}{2} x^T Q x$
- $\min_x g(Ax)$
- ...

Desired interpolation conditions

Given x_k and y_k,

$\exists A$ such that $y_k = Ax_k \quad \forall k$ if and only if

(interpolation conditions).

$$f(x) = \frac{1}{2} x^T Q x \Rightarrow g_k = \nabla f(x_k) = Q x_k$$

$$f(x) = g(Ax) \Rightarrow g_k = \nabla f(x_k) = A^T \nabla g(Ax_k)$$

Gram matrix technique

Reminder : PEP is reformulated as an SDP.

Variables :

- Function values f_k;
- Scalar products $x_j^T x_k$, $x_j^T g_k$ and $g_j^T g_k$.

Gram matrix technique

Reminder : PEP is reformulated as an SDP.

Variables :

- Function values f_k;
- Scalar products $x_j^T x_k$, $x_j^T g_k$ and $g_j^T g_k$.

$$G = \begin{pmatrix} x_1 & \cdots & x_N & g_1 & \cdots & g_N \end{pmatrix}^T \begin{pmatrix} x_1 & \cdots & x_N & g_1 & \cdots & g_N \end{pmatrix}$$

$$= \begin{pmatrix} x_1^T x_1 & \cdots & x_1^T x_N & x_1^T g_1 & \cdots & x_1^T g_N \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_N^T x_1 & \cdots & x_N^T x_N & x_N^T g_1 & \cdots & x_N^T g_N \\ g_1^T x_1 & \cdots & g_1^T x_N & g_1^T g_1 & \cdots & g_1^T g_N \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ g_N^T x_1 & \cdots & g_N^T x_N & g_N^T g_1 & \cdots & g_N^T g_N \end{pmatrix}$$

Gram matrix technique

Reminder : PEP is reformulated as an SDP.

Variables :

- Function values f_k;
- Scalar products $x_j^T x_k$, $x_j^T g_k$ and $g_j^T g_k$.

$$G = \begin{pmatrix} x_1 & \cdots & x_N & g_1 & \cdots & g_N \end{pmatrix}^T \begin{pmatrix} x_1 & \cdots & x_N & g_1 & \cdots & g_N \end{pmatrix}$$

$$= \begin{pmatrix} x_1^T x_1 & \cdots & x_1^T x_N & x_1^T g_1 & \cdots & x_1^T g_N \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_N^T x_1 & \cdots & x_N^T x_N & x_N^T g_1 & \cdots & x_N^T g_N \\ g_1^T x_1 & \cdots & g_1^T x_N & g_1^T g_1 & \cdots & g_1^T g_N \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ g_N^T x_1 & \cdots & g_N^T x_N & g_N^T g_1 & \cdots & g_N^T g_N \end{pmatrix}$$

We can formulate interpolation conditions as semidefinite constraint i.e. linear matrix inequality, on the (blocks of the) Gram matrix.

Obtaining the interpolation conditions

Inspecting Gram matrix of sequences y_k and x_k linked by a symmetrix matrix Q, i.e. $y_k = Qx_k$, with $0 \preceq Q \preceq LI$

$$
\begin{aligned}
G &= \begin{pmatrix} x_1 & \cdots & x_N & y_1 & \cdots & y_N \end{pmatrix}^T \begin{pmatrix} x_1 & \cdots & x_N & y_1 & \cdots & y_N \end{pmatrix} \\
&= \begin{pmatrix} x_1 & \cdots & x_N & Qx_1 & \cdots & Qx_N \end{pmatrix}^T \begin{pmatrix} x_1 & \cdots & x_N & Qx_1 & \cdots & Qx_N \end{pmatrix} \\
&\triangleq \begin{pmatrix} X & QX \end{pmatrix}^T \begin{pmatrix} X & QX \end{pmatrix} = \begin{pmatrix} X^TX & X^TQX \\ X^TQX & X^TQ^2X \end{pmatrix} \triangleq \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}
\end{aligned}
$$

Inspecting Gram matrix of sequences y_k and x_k linked by a symmetrix matrix Q, i.e. $y_k = Qx_k$, with $0 \preceq Q \preceq LI$

$$\begin{aligned} G &= \begin{pmatrix} x_1 & \cdots & x_N & y_1 & \cdots & y_N \end{pmatrix}^T \begin{pmatrix} x_1 & \cdots & x_N & y_1 & \cdots & y_N \end{pmatrix} \\ &= \begin{pmatrix} x_1 & \cdots & x_N & Qx_1 & \cdots & Qx_N \end{pmatrix}^T \begin{pmatrix} x_1 & \cdots & x_N & Qx_1 & \cdots & Qx_N \end{pmatrix} \\ &\triangleq \begin{pmatrix} X & QX \end{pmatrix}^T \begin{pmatrix} X & QX \end{pmatrix} = \begin{pmatrix} X^TX & X^TQX \\ X^TQX & X^TQ^2X \end{pmatrix} \triangleq \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \end{aligned}$$

$$\Rightarrow \begin{cases} B = B^T, \\ B \succeq \frac{C}{L}. \end{cases}$$

Remark : Since G is a Gram matrix, G is symmetric and positive semidefinite.

Interpolation conditions for symmetric matrices

Let $G \triangleq \begin{pmatrix} X^TX & X^TY \\ Y^TX & Y^TY \end{pmatrix} \triangleq \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \succeq 0$ and $L \in \mathbb{R}$.

Theorem (Symmetric matrix with spectrum between 0 and L)

G can be written as $\begin{pmatrix} X^TX & X^TQX \\ X^TQX & X^TQ^2X \end{pmatrix}$ for a symmetric matrix Q with $0 \preceq Q \preceq LI$ if and only if

$$\begin{cases} B = B^T, \\ B \succeq \frac{C}{L}. \end{cases}$$

Interpolation conditions for symmetric matrices

Let $G \triangleq \begin{pmatrix} X^TX & X^TY \\ Y^TX & Y^TY \end{pmatrix} \triangleq \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \succeq 0$ and $\mu \leq L \in \mathbb{R}$.

Theorem (Symmetric matrix with spectrum between μ and L)

G can be written as $\begin{pmatrix} X^TX & X^TQX \\ X^TQX & X^TQ^2X \end{pmatrix}$ for a symmetric matrix Q with $\mu I \preceq Q \preceq LI$ if and only if

$$\begin{cases} B = B^T, \\ B \succeq \frac{\mu L}{\mu+L} A + \frac{1}{\mu+L} C. \end{cases}$$

Interpolation conditions for symmetric matrices

Let $G \triangleq \begin{pmatrix} X^TX & X^TY \\ Y^TX & Y^TY \end{pmatrix} \triangleq \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \succeq 0$ and $\mu \leq L \in \mathbb{R}$.

Theorem (Symmetric matrix with spectrum between μ and L)

G can be written as $\begin{pmatrix} X^TX & X^TQX \\ X^TQX & X^TQ^2X \end{pmatrix}$ for a symmetric matrix Q with $\mu I \preceq Q \preceq LI$ if and only if

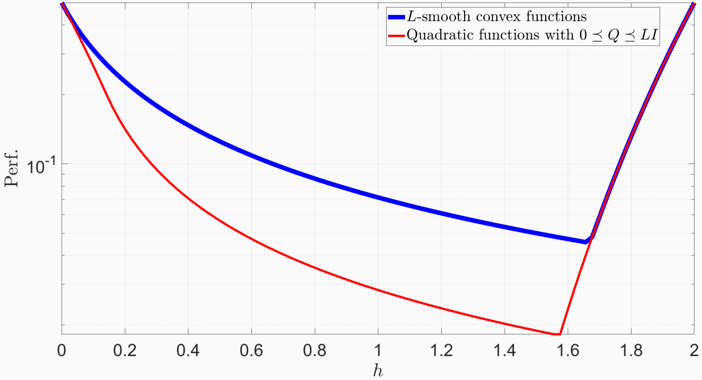
$$\begin{cases} B = B^T, \\ B \succeq \frac{\mu L}{\mu+L} A + \frac{1}{\mu+L} C. \end{cases}$$

Remark :

- We only consider homogeneous quadratic functions;
- Similar Theorem for **non-symmetric** matrix with bounded singular values.

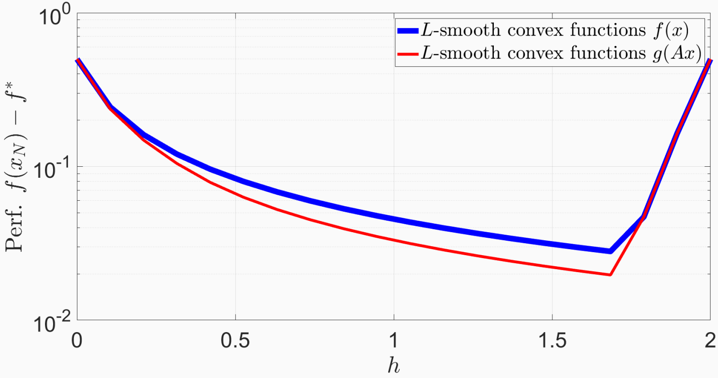
Performance of gradient method on quadratics

Worst-performance of gradient method on $\min_x \frac{1}{2}x^T Qx$.



Performance of gradient method on $g(Ax)$

Worst-performance of gradient method on $\min_x g(Ax)$.



PEP to analyze problems involving matrices

State of the art : PEP allows to obtain the worst-case performance of an optimization method on a class of functions.

Our contribution : Extending PEP to methods and classes involving matrices: $\frac{1}{2}x^TQX$, $g(Ax)$, . . .

Futur research : Analyzing more complex problems and identifying why gaps appear.

f is L-smooth when

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||.$$

First-order method of the form

$$x_N = x_0 - \sum_{i=0}^{N-1} h_{N,i}.$$

Case $\mu = L$

Let $G = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \succeq 0$ and $\mu = L \in \mathbb{R}$.

Theorem

G can be written as $\begin{pmatrix} X^T X & X^T Q X \\ X^T Q X & X^T Q^2 X \end{pmatrix}$ for a symmetric matrix Q with $LI \preceq Q \preceq LI$ if and only if

$$B = B^T,$$
$$C \preceq L^2 A.$$

Interpolation condition for L-smooth convex functions

f L-smooth convex if and only if

$$f(x) \geq f(y) + \nabla f^T(y)(x-y) + \frac{1}{2L}||\nabla f(x) - \nabla f(y)||^2 \quad \forall x, y$$

f L-smooth convex : $f(x_k) = f_k, \;\; \nabla f(x_k) = g_k$ if and only if

$$f_i \geq f_j + g_j^T(x_i - x_j) + \frac{1}{2L}||g_i - g_j||^2 \quad \forall i, j$$

SDP formulation

N steps of gradient method on L-smooth convex functions.

Matrix variable: $G = (g_0 \ \dots \ g_N \ x_0)^T (g_0 \ \dots \ g_N \ x_0) \in \mathbb{S}^{N+2}$
Parameters:

- $h_i = (0 \ \dots 0 \ \frac{-1}{L} \ 0 \ \dots \ 0 \ 1) \in \mathbb{R}^{N+2}$
- $u_i = (0 \dots 0 \ 1 \ 0 \ \dots 0) \in \mathbb{R}^{N+2}$
- $2A_{ij} = u_j(h_i - h_j)^T + (h_i - h_j)u_j^T + \frac{1}{L}(u_i - u_j)(u_i - u_j)^T$
- $A_R = u_{N+1}u_{N+1}^T$

$$
\begin{aligned}
\max_{G \in \mathbb{S}^{N+2}, f \in \mathbb{R}^{N+1}} \quad & f_N - f^* \\
\text{s.t.} \quad & f_j - f_i + \mathrm{Tr}(GA_{ij}) \leq 0, \quad \forall i, j \\
& \mathrm{Tr}(GA_{ij}) - R^2 \leq 0, \quad \forall i, j \\
& G \succeq 0.
\end{aligned}
$$