Worst-case analyses for first-order optimization methods

Adrien Taylor[*], Baptiste Goujaud[†]

Current version: June 30, 2022

Foreword & Acknowledgements

Those notes were written for accompanying the TraDE-OPT workshop on algorithmic and continuous optimization. If you have any comment, remark, or if you found a typo/mistake, please don't hesitate to feedback the authors!

Funding. A. Taylor acknowledges support from the European Research Council (grant SEQUOIA 724063). This work was partly funded by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). The work of B. Goujaud is partially supported by ANR-19-CHIA-0002-01/chaire SCAI, and Hi!Paris.

[*]INRIA, SIERRA project-team, and D.I. Ecole normale supérieure, Paris, France. Email: adrien.taylor@inria.fr

[†]CMAP, École Polytechnique, Institut Polytechnique de Paris, France. Email: baptiste.goujaud@gmail.com

1 Introduction

This document provides a series of exercises for getting familiar with "performance estimation problems" and the use of semidefinite programming for analyzing the worst-case behaviors of first-order optimization methods. An informal introduction can be found in this blog post.

Notations.

2 Exercises

2.1 Warming up; modelling problems

take some stupid method and ask a series of questions about it...

Exercise 1 (Gradient method) *The goal of this exercise is to show that the smallest τ such that the inequality*

$$\|x_{k+1} - x_\star\|^2 \leqslant \tau \|x_k - x_\star\|^2;$$

is true for any ... and $x_{k+1} = x_k - \frac{\gamma_k}{L}\nabla f(x_k)$ can be framed as a convex optimization problem.

1. *show that it can be formulated as ...*
2. *show that the previous problem can be framed using a discrete version...*
3. *show that it is equivalent to XXX (remove denominator)*
4. *show that it is equivalent to the SDP XXXX*
5. *show that τ is a function of $\kappa = \frac{L}{\mu}$ only.*
6. *how should you adapt the SDP formulation to obtain bounds on*

   $$\|\nabla f(x_{k+1}\|^2 \leqslant \tau \|\nabla f(x_k)\|^2;$$

7. *how should you adapt the SDP formulation to obtain bounds on*

   $$f(x_{k+1}) - f(x_\star) \leqslant \tau (f(x_k) - f(x_\star));$$

8. *how should you adapt the SDP formulation to obtain bounds on*

   $$f(x_{k+1}) - f(x_\star) \leqslant \tau \|x_0 - x_\star\|^2;$$

9. *using duality show that ... (dual SDP)*
10. *numerical trials*

Exercise 2 (Acceleration) *Show that the smallest τ such that the inequality*

$$\ldots$$

1. *show that it can be formulated as ...*
2. *show that the previous problem can be framed using a discrete version...*
3. *show that it is equivalent to the SDP XXXX*
4. *numerical trials*

Exercise 3 (Acceleration) *Show that the smallest τ such that the inequality*

$$\ldots$$

1. *show that it can be formulated as ...*
2. *show that the previous problem can be framed using a discrete version...*
3. *show that it is equivalent to the SDP XXXX*
4. *numerical trials*

Exercise 4 (Fixed-point iterations) *Show that the smallest τ such that the inequality*

...

1. *show that it can be formulated as ...*
2. *show that the previous problem can be framed using a discrete version...*
3. *show that it is equivalent to the SDP XXXX*
4. *using duality show that ... (dual SDP)*
5. *numerical trials*

Exercise 5 (Stochastic gradient descent) *Show that the smallest τ such that the inequality*

...

1. *show that it can be formulated as ...*
2. *show that the previous problem can be framed using a discrete version...*
3. *show that it is equivalent to the SDP XXXX*
4. *using duality show that ... (dual SDP)*
5. *numerical trials*

Exercise 6 (Proximal point method) *Show that the smallest τ such that the inequality*

...

1. *show that it can be formulated as ...*
2. *show that the previous problem can be framed using a discrete version...*
3. *show that it is equivalent to the SDP XXXX*
4. *numerical trials*

2.2 Numerical worst-case analyses

All previous exercises: use the model for computing XXX...

here: give (1) class of problems (2) method (3) metrics (initial conditions/perf measure) and let the student do the job

Exercise 7 *Using PEPit or PESTO, compute ... ++ use documentation of PEPit ++ try to guess the numerics?*

1. *Gradient descent:*
2. *Acceleration:*
3. *Fixed-point: Halpern iteration*
4. *Proximal gradient*

2.2.1 Closed-form worst-case analyses

2.2.2 Easy-to-solve

Gradient method I

Gradient method II

2.2.3 Easy-to-guess

Gradient method

Mirror descent/Bregman gradient

2.2.4 More tricky

Douglas-Rachford

3 Background material and useful facts

3.1 Standard definitions

smoothness, strong convexity...

3.2 Interpolation/extension theorems

This section gathers useful elements allowing to answer certain questions ...

Theorem 1 *interpolation 1... (ccp)*

Theorem 2 *interpolation 2... (smooth str convex)*

Theorem 3 *interpolation 3... (smooth nonconvex)*