Convex optimization: first-order methods and slightly beyond

Juan PEYPOUQUET
Universidad Técnica Federico Santa María

Workshop on Variational and Stochastic Analysis

Santiago, March 15, 2017

Content

- Basic first-order descent methods
- Nesterov's acceleration
  - Main properties
  - Dynamic interpretation: Damped Inertial Gradient System
  - Limitations
- A first-order variant bearing second-order information in time and space

Content

- Basic first-order descent methods
- Nesterov's acceleration
  - Main properties
  - Dynamic interpretation: Damped Inertial Gradient System
  - Limitations
- A first-order variant bearing second-order information in time and space

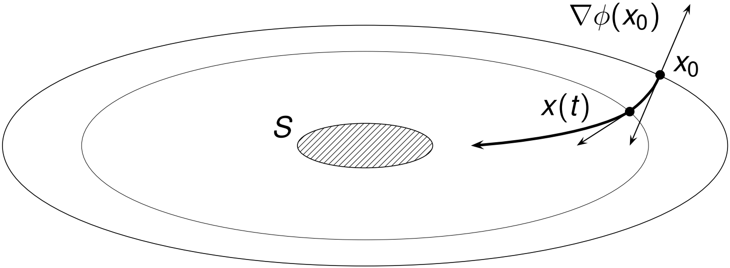
Content

- Basic first-order descent methods
- Nesterov's acceleration
  - Main properties
  - Dynamic interpretation: Damped Inertial Gradient System
  - Limitations
- A first-order variant bearing second-order information in time and space

Basic Descent Methods

Basic (first-order) descent methods

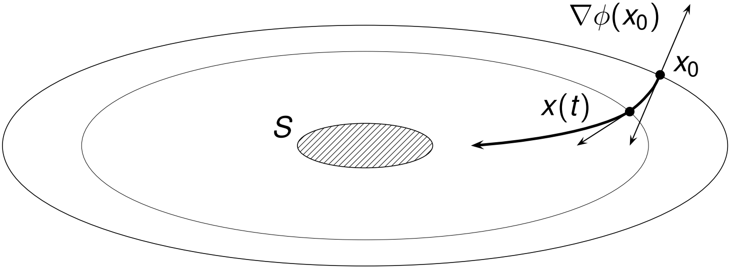
Steepest descent dynamics: $\dot{x}(t) = -\nabla\phi(x(t))$, $x(0) = x_0$



$$\frac{d}{dt}\phi(x(t)) = \langle \nabla\phi(x(t)), \dot{x}(t)\rangle = -\|\nabla\phi(x(t))\|^2 = -\|\dot{x}(t)\|^2$$

Basic (first-order) descent methods

Steepest descent dynamics: $\dot{x}(t) = -\nabla\phi(x(t))$, $x(0) = x_0$



$$\frac{d}{dt}\phi(x(t)) = \langle \nabla\phi(x(t)), \dot{x}(t)\rangle = -\|\nabla\phi(x(t))\|^2 = -\|\dot{x}(t)\|^2$$

Basic (first-order) descent methods

Explicit discretization → gradient method (Cauchy 1847):

$$\frac{x_{k+1} - x_k}{\lambda} = -\nabla\phi(x_k) \quad \Longleftrightarrow \quad x_{k+1} = x_k - \lambda\nabla\phi(x_k).$$

Implicit discretization → proximal method (Martinet 1970):

$$\frac{z_{k+1} - z_k}{\lambda} = -\nabla\phi(z_{k+1}) \quad \Longleftrightarrow \quad z_{k+1} + \lambda\nabla\phi(z_{k+1}) = z_k$$

$$\Longleftrightarrow \quad z_{k+1} = \text{Argmin}\left\{\phi(\zeta) + \frac{1}{2\lambda}\|\zeta - z_k\|^2\right\}.$$

Basic (first-order) descent methods

Explicit discretization $\rightarrow$ gradient method (Cauchy 1847):

$$\frac{x_{k+1}-x_k}{\lambda} = -\nabla\phi(x_k) \quad \Longleftrightarrow \quad x_{k+1} = x_k - \lambda\nabla\phi(x_k).$$

Implicit discretization $\rightarrow$ proximal method (Martinet 1970):

$$\frac{z_{k+1}-z_k}{\lambda} = -\nabla\phi(z_{k+1}) \quad \Longleftrightarrow \quad z_{k+1} + \lambda\nabla\phi(z_{k+1}) = z_k$$

$$\Longleftrightarrow \quad z_{k+1} = \operatorname{Argmin}\left\{\phi(\zeta) + \frac{1}{2\lambda}\|\zeta - z_k\|^2\right\}.$$

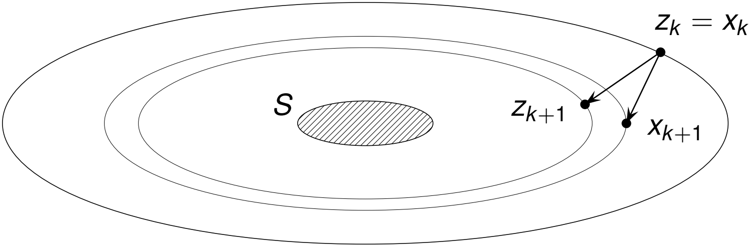
Basic (first-order) descent methods

Gradient

$x_{k+1} = x_k - \lambda \nabla \phi(x_k)$

Proximal

$z_{k+1} + \lambda \nabla \phi(z_{k+1}) = z_k$



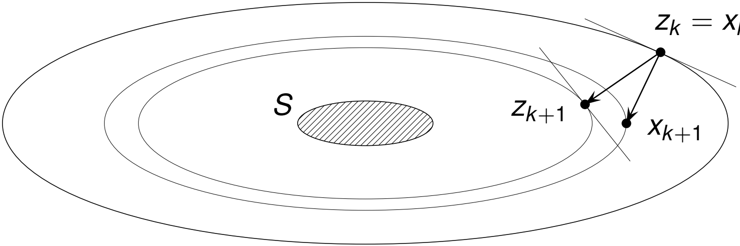
Basic (first-order) descent methods

Gradient

$x_{k+1} = x_k - \lambda \nabla \phi(x_k)$

Proximal

$z_{k+1} + \lambda \nabla \phi(z_{k+1}) = z_k$



Pros and cons

Gradient method

+ Lower computational cost per iteration (explicit formula), easy implementation
– Convergence depends strongly on the regularity of the function (typically $\phi \in \mathcal{C}^{1,1}$) and on the step sizes

Proximal point algorithm

+ More stability, convergence certificate for a larger class of functions ($\nabla\phi \to \partial\phi$), independent of the step size
– Higher computational cost per iteration (implicit formula), often requires inexact computation

Pros and cons

Gradient method

+ Lower computational cost per iteration (explicit formula), easy implementation
– Convergence depends strongly on the regularity of the function (typically $\phi \in \mathcal{C}^{1,1}$) and on the step sizes

Proximal point algorithm

+ More stability, convergence certificate for a larger class of functions ($\nabla\phi \to \partial\phi$), independent of the step size
– Higher computational cost per iteration (implicit formula), often requires inexact computation

Combining smooth and nonsmooth functions

Problem

$$\min\{\Phi(x) := F(x) + G(x) : x \in H\},$$

where F is not smooth but G is.

Forward-Backward Method ($x_k \to x_{k+\frac{1}{2}} \to x_{k+1}$)

$$x_{k+1} + \lambda \partial F(x_{k+1}) \ni x_{k+\frac{1}{2}} = x_k - \lambda \nabla G(x_k)$$

$$x_{k+1} = \text{Prox}_{\lambda F} \circ \text{Grad}_{\lambda G}(x_k)$$

Combining smooth and nonsmooth functions

Problem

$$\min\{\Phi(x) := F(x) + G(x) : x \in H\},$$

where F is not smooth but G is.

Forward-Backward Method $(x_k \to x_{k+\frac{1}{2}} \to x_{k+1})$

$$x_{k+1} + \lambda \partial F(x_{k+1}) \ni x_{k+\frac{1}{2}} = x_k - \lambda \nabla G(x_k)$$

$$x_{k+1} = \mathrm{Prox}_{\lambda F} \circ \mathrm{Grad}_{\lambda G}(x_k)$$

Problem

$$\min\{\Phi(x) := F(x) + G(x) : x \in H\},$$

where F is not smooth but G is.

Forward-Backward Method $(x_k \to x_{k+\frac{1}{2}} \to x_{k+1})$

$$x_{k+1} + \lambda \partial F(x_{k+1}) \ni x_{k+\frac{1}{2}} = x_k - \lambda \nabla G(x_k)$$

$$x_{k+1} = \text{Prox}_{\lambda F} \circ \text{Grad}_{\lambda G}(x_k)$$

Combining smooth and nonsmooth functions

Gradient projection:
Goldstein 1964, Levitin-Polyak 1966, with $F = \delta_C$

General setting:
Lions-Mercier 1979, Passty 1979

Iterative Shrinkage-Thresholding Algorithm (ISTA):
Daubechies-Defrise-DeMol 2004, Combettes-Wajs 2005, for "$\ell^1 + \ell^2$" minimization

$$\Phi(x) = F(x) + G(x) = \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|^2$$

Combining smooth and nonsmooth functions

Gradient projection:
Goldstein 1964, Levitin-Polyak 1966, with $F = \delta_C$

General setting:
Lions-Mercier 1979, Passty 1979

Iterative Shrinkage-Thresholding Algorithm (ISTA):
Daubechies-Defrise-DeMol 2004, Combettes-Wajs 2005, for "$\ell^1 + \ell^2$" minimization

$$\Phi(x) = F(x) + G(x) = \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|^2$$

Combining smooth and nonsmooth functions

Gradient projection:
Goldstein 1964, Levitin-Polyak 1966, with $F = \delta_C$

General setting:
Lions-Mercier 1979, Passty 1979

Iterative Shrinkage-Thresholding Algorithm (ISTA):
Daubechies-Defrise-DeMol 2004, Combettes-Wajs 2005, for "$\ell^1 + \ell^2$" minimization

$$\Phi(x) = F(x) + G(x) = \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|^2$$

Theorem

Let $\Phi = F + G$, where G is closed and convex, and F is convex with ∇F L-Lipschitz. Assume Φ has minimizers, and let (x_k) be obtained by the FB method with $\lambda \leq 1/L$. Then

- *As $k \to \infty$, (x_k) converges* to a minimizer of Φ; and*
- *$\Phi(x_k) - \min \Phi = \mathcal{O}(k^{-1})$: More precisely,*

$$\Phi(x_k) - \min \Phi \leq \frac{dist(x_0, S)^2}{2\lambda k}.$$

Convergence of the forward-backward method

Theorem

Let $\Phi = F + G$, where G is closed and convex, and F is convex with ∇F L-Lipschitz. Assume Φ has minimizers, and let (x_k) be obtained by the FB method with $\lambda \leq 1/L$. Then

- *As $k \to \infty$, (x_k) converges* to a minimizer of Φ; and*
- *$\Phi(x_k) - \min \Phi = \mathcal{O}(k^{-1})$: More precisely,*

$$\Phi(x_k) - \min \Phi \leq \frac{dist(x_0, S)^2}{2\lambda k}.$$

Convergence of the forward-backward method

Theorem

Let $\Phi = F + G$, where G is closed and convex, and F is convex with ∇F L-Lipschitz. Assume Φ has minimizers, and let (x_k) be obtained by the FB method with $\lambda \leq 1/L$. Then

- *As $k \to \infty$, (x_k) converges* to a minimizer of Φ; and*
- $\Phi(x_k) - \min \Phi = \mathcal{O}(k^{-1})$*: More precisely,*

$$\Phi(x_k) - \min \Phi \leq \frac{dist(x_0, S)^2}{2\lambda k}.$$

Convergence of ISTA

Let $\Phi : \mathbb{R}^N \to \mathbb{R}$ be defined by

$$\Phi(x) = \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|^2.$$

Recently,

- Local linear convergence results; and
- implicit complexity bounds

have been established.

Convergence of ISTA

Let $\Phi : \mathbb{R}^N \to \mathbb{R}$ be defined by

$$\Phi(x) = \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|^2.$$

Recently,

- Local linear convergence results; and
- implicit complexity bounds

have been established.

Convergence of ISTA

Let $\Phi : \mathbb{R}^N \to \mathbb{R}$ be defined by

$$\Phi(x) = \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|^2.$$

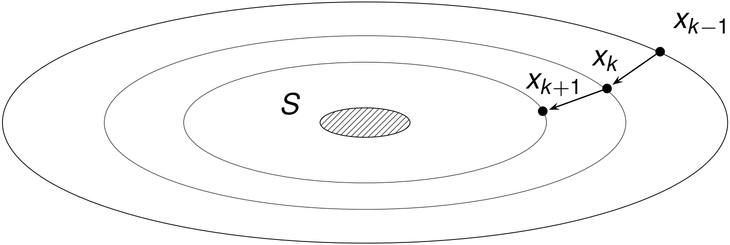
Theorem (Bolte-Nguyen-P.-Suter 2015)

Let (x_k) be obtained by the FB method with step size λ. Then, there exist $x^* \in S$ and explicit constants $c, d > 0$ such that

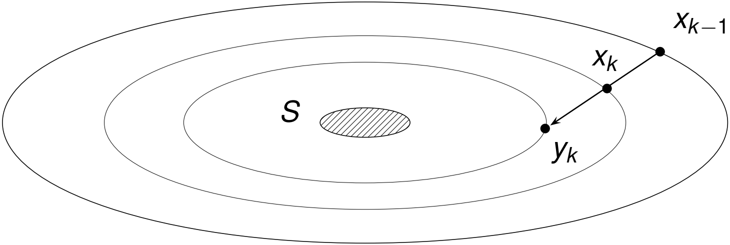
$$c\|x_k - x^*\|^2 \leq \Phi(x_k) - \min \Phi \leq \frac{\Phi(x_0) - \min \Phi}{(1 + d\lambda)^{2k}}.$$

NESTEROV'S ACCELERATION

The main idea is the following: Instead of doing

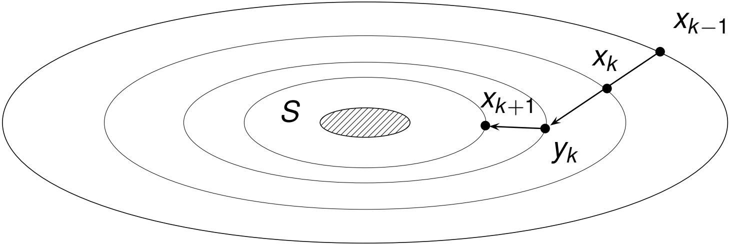


Better try



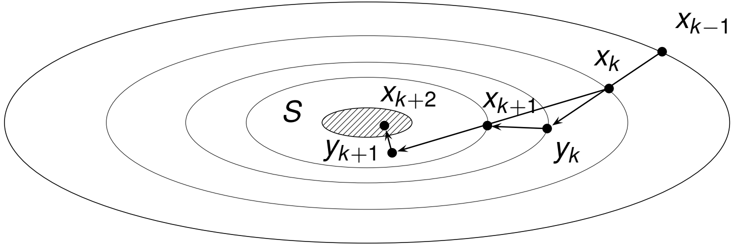
Acceleration

Better try



Acceleration

Better try



Some remarks

- Convergence and its rate are sensitive to the choice of y_k
- This simple procedure (Nesterov 1983) can take the theoretical rate of worst-case convergence for the values from the typical $\mathcal{O}(1/k)$ down to $\mathcal{O}(1/k^2)$
- No convergence proof for the iterates x_k
- Current common practice is

$$y_k = x_k + \left(1 - \frac{3}{k}\right)(x_k - x_{k-1})$$

Keynote example in image processing: FISTA (Beck-Teboulle 2009)

Some remarks

- Convergence and its rate are sensitive to the choice of y_k
- This simple procedure (Nesterov 1983) can take the theoretical rate of worst-case convergence for the values from the typical $\mathcal{O}(1/k)$ down to $\mathcal{O}(1/k^2)$
- No convergence proof for the iterates x_k
- Current common practice is

$$y_k = x_k + \left(1 - \frac{3}{k}\right)(x_k - x_{k-1})$$

Keynote example in image processing: FISTA (Beck-Teboulle 2009)

Some remarks

- Convergence and its rate are sensitive to the choice of y_k
- This simple procedure (Nesterov 1983) can take the theoretical rate of worst-case convergence for the values from the typical $\mathcal{O}(1/k)$ down to $\mathcal{O}(1/k^2)$
- No convergence proof for the iterates x_k
- Current common practice is

$$y_k = x_k + \left(1 - \frac{3}{k}\right)(x_k - x_{k-1})$$

Keynote example in image processing: FISTA (Beck-Teboulle 2009)

Some remarks

- Convergence and its rate are sensitive to the choice of y_k
- This simple procedure (Nesterov 1983) can take the theoretical rate of worst-case convergence for the values from the typical $\mathcal{O}(1/k)$ down to $\mathcal{O}(1/k^2)$
- No convergence proof for the iterates x_k
- Current common practice is

$$y_k = x_k + \left(1 - \frac{3}{k}\right)(x_k - x_{k-1})$$

Keynote example in image processing: FISTA (Beck-Teboulle 2009)

PROPERTIES OF THE ACCELERATED FORWARD-BACKWARD METHOD

Basic properties

Recall that

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\ x_{k+1} = \mathrm{Prox}_{\lambda F} \circ \mathrm{Grad}_{\lambda G}(y_k) \end{cases}$$

Theorem (Attouch-Chbani-P.-Redont 2015)

If $\alpha > 0$, then

- $\lim_{k \to +\infty} \Phi(x_k) = \inf(\Phi)$; *and*
- *every weak limit point of x_k, as $k \to +\infty$, minimizes Φ.*

Recall that

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\ x_{k+1} = \mathrm{Prox}_{\lambda F} \circ \mathrm{Grad}_{\lambda G}(y_k) \end{cases}$$

Theorem (Attouch-Chbani-P.-Redont 2015)

If $\alpha > 0$, then

- $\lim_{k\to+\infty} \Phi(x_k) = \inf(\Phi)$*; and*
- *every weak limit point of x_k, as $k \to +\infty$, minimizes Φ.*

Further properties

Theorem (Nesterov 1983, Beck-Teboulle 2009, Su-Boyd-Candès 2014, ACPR)

If $\alpha \geq 3$ and Φ has minimizers, then

$$\Phi(x_k) - \min \Phi = \mathcal{O}(1/k^2) \qquad \textit{and} \qquad \|x_k - x_{k-1}\| = \mathcal{O}(1/k).$$

Theorem (Nesterov 1983)

There is $\Phi : \mathbb{R}^N \to \mathbb{R}$ such that

$$\Phi(x_k) - \min \Phi \geq \frac{3dist(x_0, S)^2}{32(k+1)^2}$$

as long as $2k \leq N - 1$.

Further properties

Theorem (Nesterov 1983, Beck-Teboulle 2009, Su-Boyd-Candès 2014, ACPR)

If $\alpha \geq 3$ and Φ has minimizers, then

$$\Phi(x_k) - \min \Phi = \mathcal{O}(1/k^2) \qquad \textit{and} \qquad \|x_k - x_{k-1}\| = \mathcal{O}(1/k).$$

Theorem (Nesterov 1983)

There is $\Phi : \mathbb{R}^N \to \mathbb{R}$ such that

$$\Phi(x_k) - \min \Phi \geq \frac{3 dist(x_0, S)^2}{32(k+1)^2}$$

as long as $2k \leq N - 1$.

Convergence

Theorem (Chambolle-Dossal 2015, ACPR)

If $\alpha > 3$ and Φ has minimizers, then:

- *x_k converges weakly, as $k \to +\infty$, to a minimizer of Φ.*
- *Strong convergence holds if Φ is even, uniformly convex, or if Argmin(Φ) has nonempty interior.*

Finer convergence rates

Theorem (Attouch-P. 2016)

If $\alpha > 3$ and Φ has minimizers, then:

- $\|x_k - x_{k-1}\| = o(1/k)$; *and*
- $\Phi(x_k) - \min \Phi = o(1/k^2)$.

Corollary

Nesterov's optimality bound cannot hold for all k.

Finer convergence rates

Theorem (Attouch-P. 2016)

If $\alpha > 3$ and Φ has minimizers, then:

- $\|x_k - x_{k-1}\| = o(1/k)$; *and*
- $\Phi(x_k) - \min \Phi = o(1/k^2)$.

Corollary

Nesterov's optimality bound cannot hold for all k.

Stochastic coordinate descent

Work in Progress

Instead of computing full gradients, we use partial gradients with respect to some randomly chosen variables at each iteration. Under suitable assumptions, a similar analysis can be carried out:

- *Expected gap decreases at* $o(1/k^2)$.
- *We conjecture convergence (of iterates) with probability 1.*

Dynamic interpretation (SBC)

A finite-difference discretization of the Damped Inertial Gradient System

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \partial F(x(t)) + \nabla G(x(t)) \ni 0 (DIGS)$$

yields

$$\begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\ x_{k+1} &= \mathrm{Prox}_{\lambda F} \circ \mathrm{Grad}_{\lambda G}(y_k) \end{cases}$$

Basic properties

Theorem (ACPR)

If $\alpha > 0$, then

- $\lim_{t\to+\infty} \Phi(x(t)) = \inf(\Phi) \in \mathbb{R} \cup \{-\infty\}$.
- *Every weak limit point of $x(t)$, as $t \to \infty$, minimizes Φ.*
- *Either Φ has minimizers and all trajectories are bounded, or it does not and all trajectories diverge to $+\infty$ in norm.*
- *If Φ is bounded from below, then $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0$.*

Convergence rate

Theorem (SBC)

If $\alpha \geq 3$ and Φ has minimizers, every solution satisfies

$$\Phi(x(t)) - \min(\Phi) \leq C/t^2$$

for all $t \geq t_0$, where C depends on α and the initial data.

Theorem (ACPR)

For each $p > 2$, there is Φ such that every solution satisfies

$$\Phi(x(t)) - \min(\Phi) = C_p/t^p$$

for some C_p and all $t \geq t_0$.

Convergence rate

Theorem (SBC)

If $\alpha \geq 3$ and Φ has minimizers, every solution satisfies

$$\Phi(x(t)) - \min(\Phi) \leq C/t^2$$

for all $t \geq t_0$, where C depends on α and the initial data.

Theorem (ACPR)

For each $p > 2$, there is Φ such that every solution satisfies

$$\Phi(x(t)) - \min(\Phi) = C_p/t^p$$

for some C_p and all $t \geq t_0$.

Strongly convex case

If Φ is strongly convex, convergence is arbitrarily fast, as α grows.

Theorem (ACPR)

Let Φ be strongly convex and let x^* be its unique minimizer. Every solution satisfies

$$D\|x(t) - x^*\|^2 \leq \Phi(x(t)) - \min(\Phi) \leq Ct^{-\frac{2}{3}\alpha}$$

for all $t \geq t_0$, where C and D depend on α, the strong convexity parameter and the initial data.

Convergence

Theorem (ACPR)

If $\alpha > 3$ and Φ has minimizers, then

- *$x(t)$ converges weakly, as $t \to +\infty$, to a minimizer of Φ.*
- *Convergence is strong if either Φ is uniformly convex, int(Argmin(Φ)) $\neq \emptyset$, or Φ is even.*

Theorem (ACPR, May 2016)

If $\alpha > 3$ and Φ has minimizers, then

$$\|\dot{x}(t)\| = o(1/t) \qquad \text{and} \qquad \Phi(x(t)) - \min(\Phi) = o(1/t^2).$$

Convergence

Theorem (ACPR)

If $\alpha > 3$ and Φ has minimizers, then

- *$x(t)$ converges weakly, as $t \to +\infty$, to a minimizer of Φ.*
- *Convergence is strong if either Φ is uniformly convex, $int(Argmin(\Phi)) \neq \emptyset$, or Φ is even.*

Theorem (ACPR, May 2016)

If $\alpha > 3$ and Φ has minimizers, then

$$\|\dot{x}(t)\| = o(1/t) \qquad \text{and} \qquad \Phi(x(t)) - \min(\Phi) = o(1/t^2).$$

LIMITATIONS

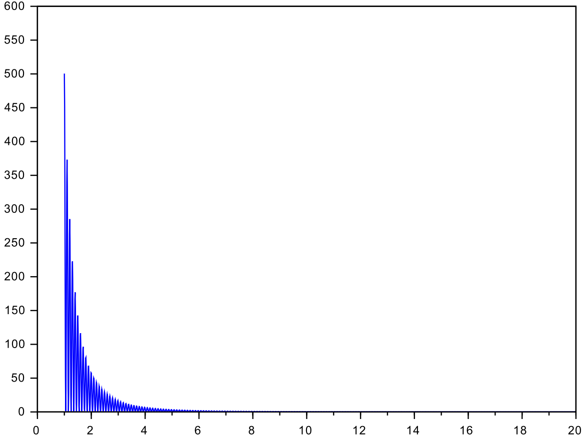
A simple example

We consider the function $\Phi(x_1, x_2) = \frac{1}{2}(x_1^2 + 1000x_2^2)$. We show the behavior of a solution to

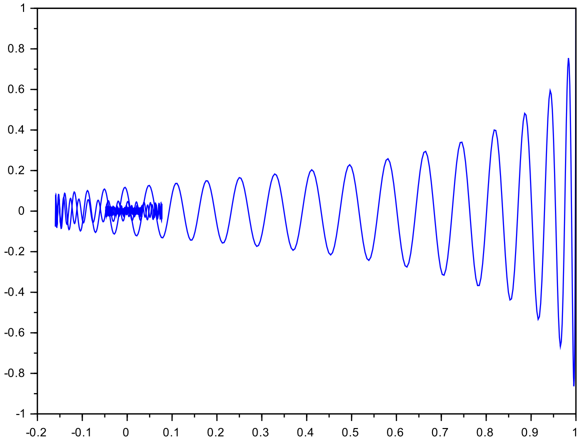
$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = 0$$

on the interval $[1, 20]$ with $\alpha = 3.1$.

Function values



Trajectory



CAN WE DO BETTER?

Idea: Newton / Levenberg-Marquardt

Pros:

- Is fast.
- Compensates the effect of ill-conditioning.

Cons:

- Requires higher regularity (to compute and invert the Hessian).
- Is costly to implement.

Idea: Newton / Levenberg-Marquardt

Pros:

- Is fast.
- Compensates the effect of ill-conditioning.

Cons:

- Requires higher regularity (to compute and invert the Hessian).
- Is costly to implement.

Newton Inertial Gradient System

$$(NIGS) \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0.$$

Seems much more complicated, but

Proposition (Attouch-P.-Redont 2016)

(NIGS) is equivalent to

$$\begin{cases} \dot{x}(t) + \beta\nabla\Phi(x(t)) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)x(t) + \frac{1}{\beta}y(t) & = & 0 \\ \dot{y}(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right)x(t) + \frac{1}{\beta}y(t) & = & 0. \end{cases}$$

Newton Inertial Gradient System

$$(NIGS) \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0.$$

Seems much more complicated, but

Proposition (Attouch-P.-Redont 2016)

(NIGS) is equivalent to

$$\begin{cases} \dot{x}(t) + \beta\nabla\Phi(x(t)) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)x(t) + \frac{1}{\beta}y(t) & = & 0 \\ \dot{y}(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right)x(t) + \frac{1}{\beta}y(t) & = & 0. \end{cases}$$

Newton Inertial Gradient System

$$(NIGS) \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0.$$

Seems much more complicated, but

Proposition (Attouch-P.-Redont 2016)

(NIGS) is equivalent to

$$\begin{cases} \dot{x}(t) + \beta\nabla\Phi(x(t)) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)x(t) + \frac{1}{\beta}y(t) &= 0 \\ \dot{y}(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right)x(t) + \frac{1}{\beta}y(t) &= 0. \end{cases}$$

Nonsmooth functions

Using variable $Z = (x, y)$, this is

$$\dot{Z}(t) + \nabla\mathcal{G}(Z(t)) + D(t, Z(t)) \ni 0,$$

where $\mathcal{G}(Z) = \beta\Phi(x)$ and D is a regular linear perturbation.

So, we can consider

$$(NIGS') \qquad \dot{Z}(t) + \partial\mathcal{G}(Z(t)) + D(t, Z(t)) \ni 0,$$

whose maximal solutions exist, even for nondifferentiable Φ.

Nonsmooth functions

Using variable $Z = (x, y)$, this is

$$\dot{Z}(t) + \nabla\mathcal{G}(Z(t)) + D(t, Z(t)) \ni 0,$$

where $\mathcal{G}(Z) = \beta\Phi(x)$ and D is a regular linear perturbation.

So, we can consider

$$(NIGS') \qquad \dot{Z}(t) + \partial\mathcal{G}(Z(t)) + D(t, Z(t)) \ni 0,$$

whose maximal solutions exist, even for nondifferentiable Φ.

Convergence results

Theorem (APR)

Let Φ be closed and convex, and let $\beta > 0$.

- *All the conclusions obtained for the solutions of (DIGS) are also true for the solutions of (NIGS').*
- *But also* $\lim_{t\to\infty} \|\nabla\Phi(x(t))\| = 0$.
- *If $\nabla\Phi$ is locally Lipschitz-continuous, then* $\lim_{t\to\infty} \|\ddot{x}(t)\| = 0$.

Convergence results

Theorem (APR)

Let Φ be closed and convex, and let $\beta > 0$.

- *All the conclusions obtained for the solutions of (DIGS) are also true for the solutions of (NIGS').*
- *But also* $\lim_{t\to\infty} \|\nabla\Phi(x(t))\| = 0$.
- *If $\nabla\Phi$ is locally Lipschitz-continuous, then* $\lim_{t\to\infty} \|\ddot{x}(t)\| = 0$.

Convergence results

Theorem (APR)

Let Φ be closed and convex, and let $\beta > 0$.

- *All the conclusions obtained for the solutions of (DIGS) are also true for the solutions of (NIGS').*
- *But also* $\lim_{t\to\infty} \|\nabla\Phi(x(t))\| = 0$.
- *If $\nabla\Phi$ is locally Lipschitz-continuous, then* $\lim_{t\to\infty} \|\ddot{x}(t)\| = 0$.

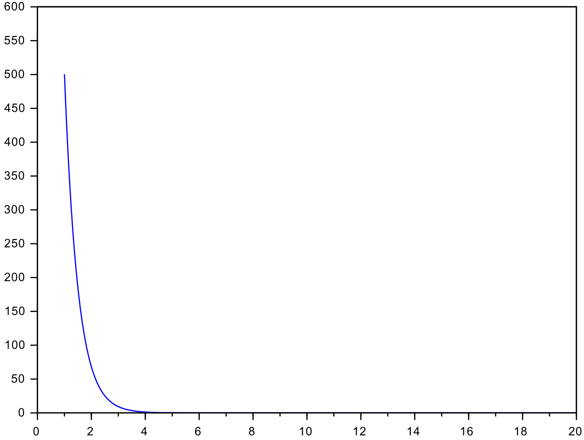
The same simple example

We consider the function $\Phi(x_1, x_2) = \frac{1}{2}(x_1^2 + 1000x_2^2)$. We show the behavior of a solution to

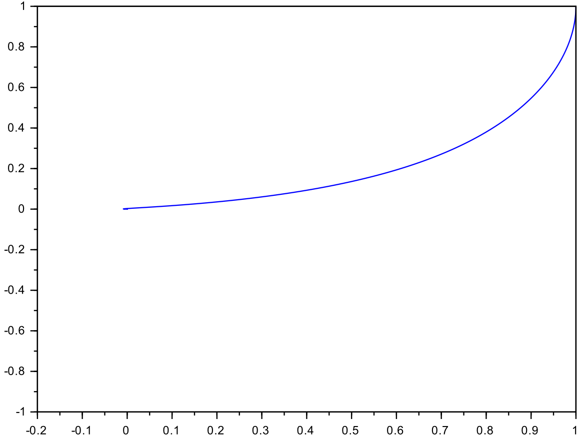
$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0$$

on the interval $[1, 20]$ with $\alpha = 3.1$ and $\beta = 1$.

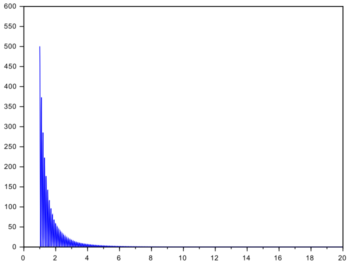
Function values

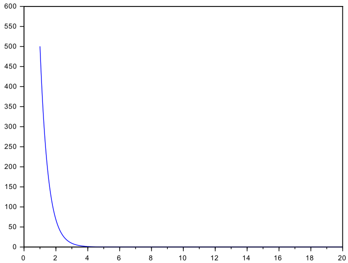


Trajectory

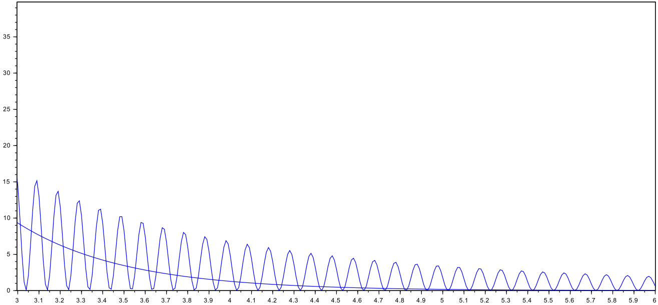


DIGS vs NIGS: Function values

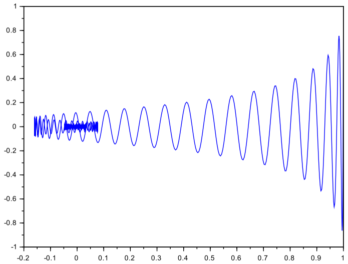


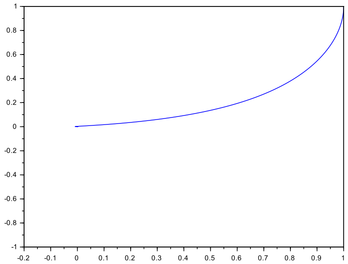


DIGS vs NIGS: Function values



DIGS vs NIGS: Trajectory





Algorithmic implementation

Several discretizations are possible, giving different iterative algorithms.

Work in Progress

An appropriate discretization defines an algorithm with the same convergence properties as the continuous-time system (NDIGS').