Augmented Lagrangian for infinite-dimensional equality constrained optimization

Romain DUJOL

Facultés Universitaires Notre-Dame de la Paix (Namur, Belgium)
Department of Mathematics / Numerical Analysis Unit

Student Day, Toulouse
Tuesday, September 9th 2008

Outline

Algorithm

Context and references

Infinite-dimensional optimization with equality constraints

$$\begin{cases} \min\ f(x) \\ \quad \text{s.t. } c(x) = 0 \end{cases} \qquad \begin{array}{l} f : X \to \mathbf{R},\ X \text{ hilbertian} \\ c : X \to Y,\ Y \text{ hilbertian} \end{array}$$

Augmented Lagrangian algorithm

A. R. Conn, N. I. Gould, A. Sartenaer, Ph. L. Toint,
Convergence properties of an augmented Lagrangian for optimization with a combination of general equality and linear constraints,
SIAM Journal on Optimization, **6**(3):674-703 (1996)

$$\begin{cases} \min\ f(x) \\ \quad \text{s.t. } c(x) = 0 \text{ and } Ax \leq b \end{cases}$$

(handles "disaggregated constraint formulation")
$\Rightarrow$ **LANCELOT software** (bound constraints)

Finite-dimensional algorithm

$$\min_{c(x)=0} f(x) \qquad x \in X,\ c : X \to Y \qquad \dim X < \infty,\ \dim Y < \infty$$

Conn, Gould, Sartenaer, Toint (simplified version)

$\Phi(x, \lambda, \mu) \triangleq f(x) + \lambda \cdot c(x) + \|c(x)\|^2/(2\mu)$ *Augmented Lagrangian*

0. **Initialization:** $x_0 \in X$, $\lambda_0 \in Y$, $\mu_0 \in]0, 1[$
1. **Inner iteration:** Find $x_k \in X$ (approximate) solution of $\min_x \Phi(x, \lambda_k, \mu_k) \rightsquigarrow \|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\|_X \leq \omega_k$
2. **Convergence test:** IF $\|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\|_X \leq \omega_*$ and $\|c(x_k)\|_Y \leq \eta_*$ THEN **STOP**
3. **Update** λ_k, μ_k, ω_k and η_k:
   - IF $\|c(x_k)\|_Y \leq \eta_k$ THEN $\lambda_{k+1} \leftarrow \lambda_k + c(x_k)/\mu_k$
   - IF $\|c(x_k)\|_Y > \eta_k$ THEN $\mu_{k+1} \leftarrow \tau \mu_k$
4. $k \leftarrow k + 1$. **Go back to 1**

Infinite-dimensional algorithm: ALINF

$$\min_{c(x)=0} f(x) \qquad x \in X,\ c : X \to Y \qquad X \text{ hilbertian},\ Y \text{ hilbertian}$$

E. W. Sachs, A. Sartenaer (2005, in revision)

$\Phi(x, \lambda, \mu) \triangleq f(x) + \langle \lambda, c(x) \rangle_X + \|c(x)\|_X^2/(2\mu)$

0. **Initialization:** $x_0 \in X$, $\lambda_0 \in Y$, $\mu_0 \in]0, 1[$
1. **Inner iteration:** Find $x_k \in X$ (approximate) solution of $\min_x \Phi(x, \lambda_k, \mu_k) \rightsquigarrow \|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\|_X \leq \omega_k$
2. **Convergence test:** IF $\|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\|_X \leq \omega_*$ and $\|c(x_k)\|_Y \leq \eta_*$ THEN **STOP**
3. **Update** λ_k, μ_k, ω_k and η_k $(0 < \gamma_1 < 1 < \gamma_2)$:
   - IF $\|c(x_k)\|_Y \leq \gamma_1 \eta_k$ THEN $\lambda_{k+1} \leftarrow \lambda_k + c(x_k)/\mu_k$
   - IF $\|c(x_k)\|_Y > \gamma_2 \eta_k$ THEN $\mu_{k+1} \leftarrow \tau \mu_k$
   - ELSE, do either λ or μ update
4. $k \leftarrow k + 1$. **Go back to 1**

Discretized infinite-dimensional algorithm: ALDISCR

$$\min_{c_n(x^n)=0} f_n(x^n) \qquad x^n \in X_n,\ c_n : X_n \to Y_n \quad X_n \text{ hilbertian},\ Y_n \text{ hilbertian}$$

E. W. Sachs, A. Sartenaer (2005, in revision)

$$\Phi_n(x^n, \lambda^n, \mu) \triangleq f_n(x^n) + \langle \lambda^n , c_n(x^n) \rangle_X + \|c_n(x^n)\|_X^2/(2\mu)$$

0. **Initialization:** $x_0 \in X_{n_0}$, $\lambda_0 \in Y_{n_0}$, $\mu_0 \in]0, 1[$
1. **Inner iteration:** Find $x_k \in X_{n_k}$ (approximate) solution of $\min_x \Phi_{n_k}(x, \lambda_k, \mu_k) \rightsquigarrow \|\nabla_x \Phi_{n_k}(x_k, \lambda_k, \mu_k)\|_X \leq \omega_k$
2. **Convergence test:** IF $\|\nabla_x \Phi(x_k, \lambda_k, \mu_k)\|_X \leq \omega_*$ and $\|c(x_k)\|_Y \leq \eta_*$ THEN **STOP**
3. **Update** λ_k, μ_k, ω_k and η_k:
   - IF $\|c_{n_k}(x_k)\|_Y \leq \eta_k$ THEN $\lambda_{k+1} \leftarrow \lambda_k + c_{n_k}(x_k)/\mu_k$
   - IF $\|c_{n_k}(x_k)\|_Y > \eta_k$ THEN $\mu_{k+1} \leftarrow \tau \mu_k$

   Zone $[\gamma_1 \eta_k, \gamma_2 \eta_k]$ *"filled" by discretization error* $c_{n_k}(x_k) - c(x_k)$
4. **Update** n_k. $k \leftarrow k+1$. **Go back to 1**

A few more words

Main idea

- At each iteration, solve Lagrangian subproblems
  $\Rightarrow$ *Usual finite-dimensional augmented Lagrangian iterations*
- Level is updated (increased) after each iteration

Level update ("Update n_k")

- Straightforward: $n_{k+1} \leftarrow n_k + 1$
- Refined: n_{k+1} is such that :
  $\|c_n(x_k) - c(x_k)\|_Y < \min\{\alpha\eta_{k+1},\ \mu_{k+1}\omega_{k+1}\}$ and
  $\|\nabla_x \Phi_n(x_k,\ \lambda_k,\ \mu_k) - \nabla_x \Phi(x_k,\ \lambda_k,\ \mu_k)\|_X \leq \omega_{k+1}/2$

If used too soon, refined update used may be counter-productive.

$\Rightarrow$ *Adaptive update between both strategies ?*

Convergence results : assumptions

1. f, c, f_n and c_n are C^2 Fréchet-differentiable for every $n \in \mathbf{N}$.
2. There exists $n^* \in \mathbf{N}$ such that :
   $\forall x \in X_{n^*}, \forall n \geq n^*, \|c_n(x) - c(x)\|_Y \xrightarrow[n\to\infty]{} 0$
   $\forall x \in X_{n^*}, \forall \lambda \in Y_{n^*}, \forall \mu > 0, \forall n \geq n^*,$
   $\|\nabla_x \Phi_n(x, \lambda, \mu) - \nabla_x \Phi(x, \lambda, \mu)\|_X \xrightarrow[n\to\infty]{} 0$
3. Refined level update is used at least for as subsequence of iterations.
4. The iterates $(x_k)_k$ lie in a compact set.
5. For every x_*, limit point of $(x_k)_k$, $c'(x_*)$ is surjective.
6. f'' and c'' are Lipschitz-continuous at x_*
7. If λ_* is the Lagrange mutliplier associated to x_*, we assume $\begin{pmatrix} \nabla_{xx}\ell(x_*, \lambda_*) & c'(x_*)^* \\ c'(x_*) & 0 \end{pmatrix}$ to have a continous inverse where $\ell = f + \langle \lambda, c \rangle$ (This implies assumption 5).

Convergence results (Sachs, Sartenaer)

Theorem (Convergence)

Under assumptions 1–5, if $(x_k)_k$ converges to x_* with $c(x_*) = 0$,

- *x_* is a first-order stationary point*
  the Lagrange multiplier is $\lambda_* = -c'(x)^+ \nabla_x f(x_*)$
- *$\lambda_k \to \lambda_*$, $c(x_k) \to 0$ and $\nabla_x \Phi(x_k, \lambda_k, \mu_k) \to \nabla_x \ell(x^*, \lambda^*)$.*

Theorem (Convergence rate)

Under assumptions 1–7,

- *$(\mu_k)_k$ is bounded away from 0 and $\mu_{min} = \min_k \mu_k \in]0, 1[$.*
- *$(x_k)_k$ and $(\lambda_k)_k$ converge at least R-linearly with R-factor at most $\mu_{min}^{\beta_\eta}$.*

Current results and implementation

Fields of applications

Many various fields of applications like

- PDE solving (Lax-Milgram theorem),
- Optimal control on ODE and PDE with state and/or mixed constraints
- Shape optimization
- ...

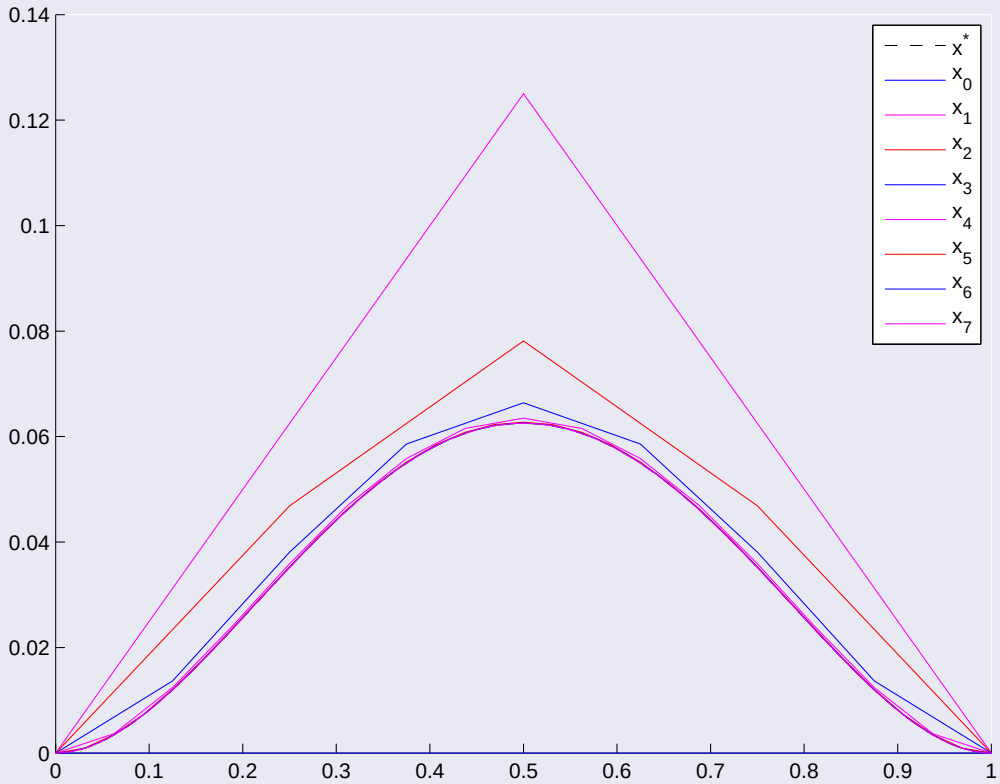
Remark: Many applications involves functional spaces.

As we need Hilbert spaces, Sobolev spaces are the usual setting. One will usually use $\mathbf{L^2}$- **or** $\mathbf{H^1}$- **spaces** and perform mesh-based discretizations.

Example : One-dimensional Poisson's equation

$$\begin{cases} -\Delta u = g \quad \text{a.e. on } [0,1] \\ u(0) = u(1) = 0 \end{cases} \quad \underset{\text{Lax-Milgram}}{\Longleftrightarrow} \quad \min_u \frac{1}{2}\int_0^1 \dot{u}^2 - \int_0^1 g \cdot u$$

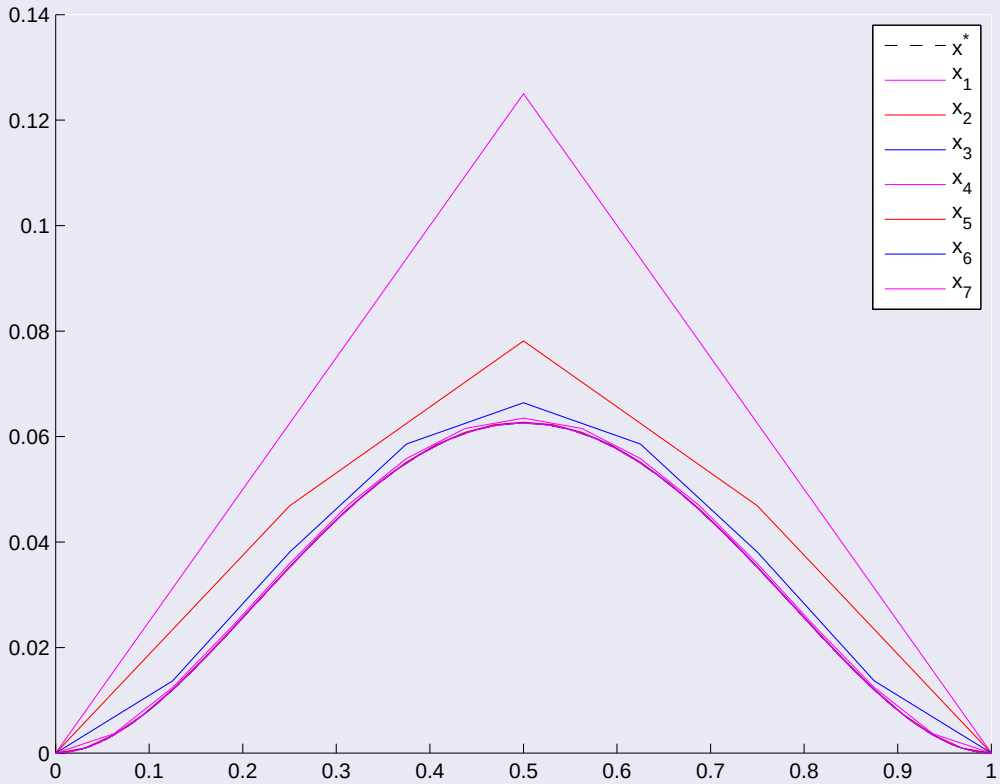
- No constraints
- $X = \mathrm{H}_0^1([0,1], \mathbf{R})$
- $\langle u, v \rangle_X = \int_0^1 \dot{u}\dot{v}$
- Nested subdivisions: $\Sigma_n = \{i/2^n\}_{0 \leq i \leq 2^n}$
- Straightfoward level update
- Exact gradient



Example : One-dimensional Poisson's equation

$$\begin{cases} -\Delta u = g \quad \text{a.e. on } [0,1] \\ u(0) = u(1) = 0 \end{cases} \quad \text{Lax-Milgram} \iff \quad \min_u \frac{1}{2}\int_0^1 \dot{u}^2 - \int_0^1 g \cdot u$$

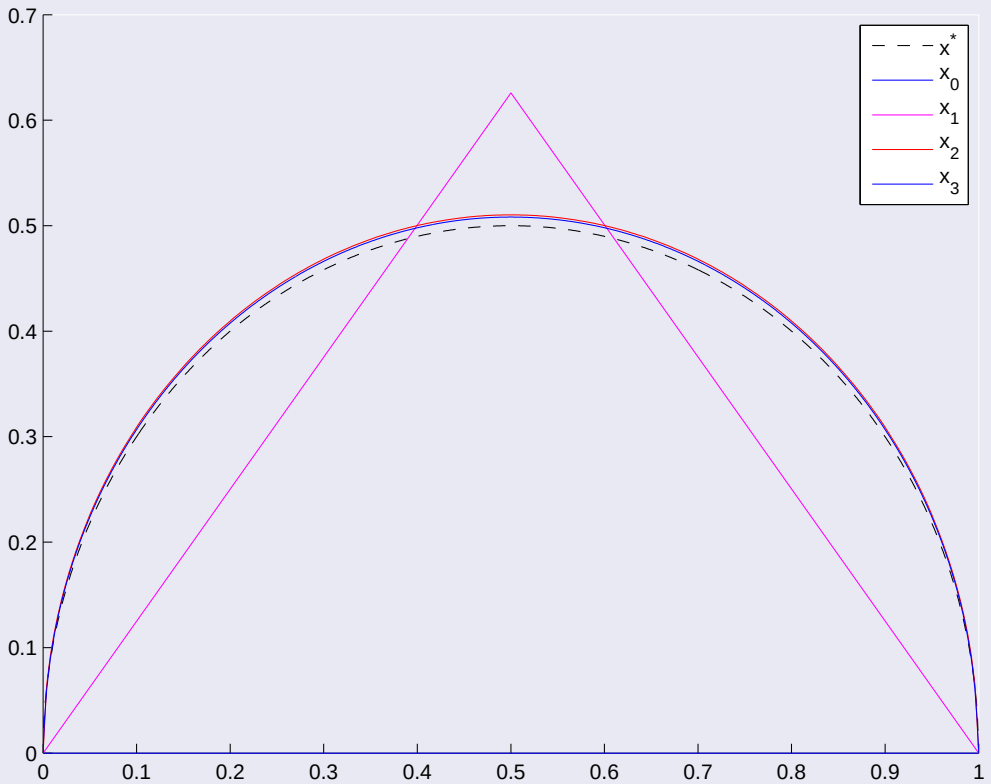
- No constraints
- $X = \mathrm{H}_0^1([0,1], \mathbf{R})$
- $\langle u, v \rangle_X = \int_0^1 \dot{u}\dot{v}$
- Nested subdivisions: $\Sigma_n = \{i/2^n\}_{0 \leq i \leq 2^n}$
- Straightfoward level update
- Finite-difference



Example : Modified Dido (or isoperimetric) problem

$$\max_u \int_0^1 u(t)\,dt \quad \text{s.t.} \quad \int_0^1 \sqrt{1+\dot{u}(t)^2}\,dt = \frac{\pi}{2}$$

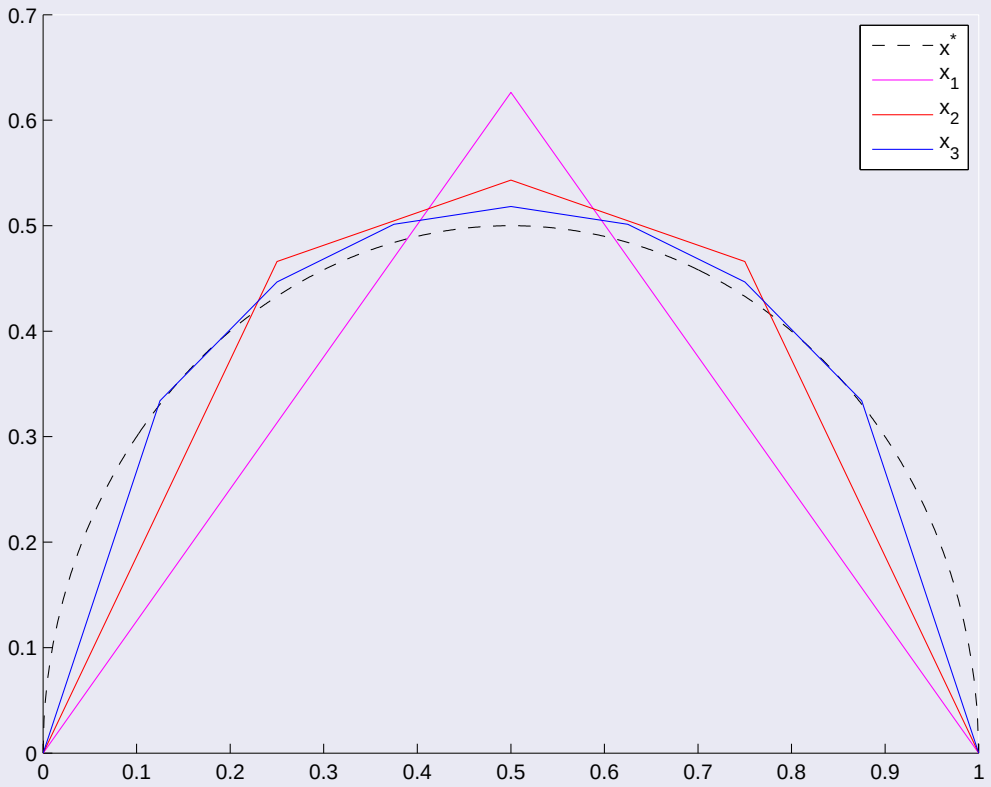
- Nonlinear constraints
- $X = \mathrm{H}_0^1([0,1], \mathbf{R})$
- $\langle u\,, v\rangle_X = \int_0^1 \dot{u}\dot{v}$
- Nested subdivisions: $\Sigma_n = \{i/2^n\}_{0\leq i\leq 2^n}$
- Straightfoward level update
- Exact gradient



Example : Modified Dido (or isoperimetric) problem

$$\max_u \int_0^1 u(t)\,dt \quad \text{s.t.} \quad \int_0^1 \sqrt{1+\dot{u}(t)^2}\,dt = \frac{\pi}{2}$$

- Nonlinear constraints
- $X = \mathrm{H}_0^1([0,1], \mathbf{R})$
- $\langle u\,, v\rangle_X = \int_0^1 \dot{u}\dot{v}$
- Nested subdivisions: $\Sigma_n = \{i/2^n\}_{0\leq i\leq 2^n}$
- Straightfoward level update
- Finite-difference



Implementation

Inner iteration solver used

- Trust-region algorithm
- Steihaug-Toint method for step computation
- BFGS Hessian update

Inner product and norm computation

- Inner product and norm on X_n induced by X:

$$\langle \cdot, \cdot \rangle_{X_n} = \langle \cdot, \cdot \rangle_X \qquad \| \cdot \|_{X_n} = \| \cdot \|_X$$

- Computation from the representation of discretized elements

Example: $X = \mathrm{L}^2(\Omega, \mathbf{R})$ with $\langle u, v \rangle_X = \int_\Omega u \cdot v$

X_n: set of piecewise-constant functions on a subdivision of Ω

$$\langle u, v \rangle_{X_n} = \sum_i h_i u(t_i) v(t_i)$$

"Infinite-dimensional gap" : part one (conceptually)

Structure gap

- Implementation presente uses the definition of inner product of X_n (induced by X).
- Most solvers designed for our inner iterations fully use the canonical Euclidian structure of finite-dimensional spaces.

Should we replace directly X structure by the Euclidian one ?

- Structures are equivalent when considering each iteration separately.
- As levels go higher, this equivalence is gradually lost.
  $\Rightarrow$ Replacing directly $\|\cdot\|_X$ with $\|\cdot\|_2$ in algorithm yields too strong conditions
  $\Rightarrow$ Counter-productive and costly.

$\Rightarrow$ **Must keep structure of X for coherence through levels**

"Infinite-dimensional gap" : part two (practically)

Straightforward view: adapt inner solver to new structure

- Adapt solver to a general inner product
  Alters subproblem definition, step and gradient computation
- Changing inner product affects gradient and adjoint definition

Problem: requires lot of theoretical computations from user (gradient computation, namely)

Reverse view: adapt algorithm to Euclidian structure

- Use inner solver "as is" with adaptation of stopping condition
- The structure of X should be kept only when necesary (for outer iteration)

Advantage: enable use of many finite-dimensional techniques
Less computation required: gradient computed in the classical way

Results

Good results still hold with the reverse view, since it is mainly a rewriting.

So far, so good...

- Pre-validation with finite-dimensional optimization problems
- Algorithm converges numerically to required solution

... but some issues still remain

- Infinite-dimensional framework requires a high pre-analysis. (Estimation of discretization errors on c and $\nabla\Phi$)
- Refined level update to be revised (Poisson : $n_0 = 1, n_1 = 11$)

Conclusion and prospects

Conclusion

Conclusion

- **Infinite**-dimensional optimization with equality constraints
- The well-known **augmented Lagrangian** algorithm can be extended with only a few adjustments.
  The result algorithm is then discretized.
- "Infinite-dimensional gap": infinite-dimensioanl structure is required to keep coherence throughout levels
  Efficient solution: only use infinite-dimensional structure when necessary
- Various fields of applications: PDEs, optimal control...
- Preliminary relevant problems have been solved (quadratic criterion, linear criterion with nonlinear constraints).

Prospects

Prospects

- Thorough investigation on level update rule: when use the refined rule ?
- Development of an ODE optimal control layer (in progress)
  Layer converting optimal control definition to optimization definition readable by `ALDISCR`
- Integration with recursive model framework ?
- Gradient is costly and finite-difference less efficient
  $\Rightarrow$ DFO ? Automatic differentiation ?

Merci pour votre attention !

Q- and R-convergence

Definition (Q-linear convergence)

A converging sequence $(x_n)_n$ is said to converge Q-linearly if

$$\exists\, Q > 0,\ \exists\, N > 0,\ \forall\, n > N,\ \|x_{n+1} - x_n\| \leq Q\|x_n - x_*\|$$

where $x_* = \lim_{n \to +\infty} x_k$.

Definition (R-linear convergence)

A converging sequence $(x_n)_n$ is said to converge R-linearly if it is dominated by a sequence $(r_n)_n$ converging Q-linearly to 0:

$$\|x_n - x_*\| \leq r_n \quad \text{and} \quad (r_n)_n \to 0 \ \text{ Q-linearly}$$

where $x_* = \lim_{n \to +\infty} x_k$.

▸ Back to talk