Greedy algorithms for high-dimensional convex nonlinear and non-symmetric problems

V. Ehrlacher
Joint work with E. Cancès et T. Lelièvre

Financial support from Michelin is acknowledged.

CERMICS, Ecole des Ponts ParisTech & MicMac project-team, INRIA.

Outline

1. Convex nonlinear problems

2. Non-symmetric problems

Outline

1 Convex nonlinear problems

2 Non-symmetric problems

Goal

We want to approximate

$$u = \operatorname*{argmin}_{v \in V} \mathcal{E}(v)$$

where V is a Hilbert space of functions depending on d variables $x_1, \cdots, x_d$ (where d is typically very large) and $\mathcal{E} : V \mapsto \mathbb{R}$ whose unique global minimizer is u.

Separated variable representation

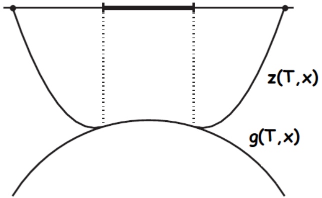
We consider an approach proposed by:

- Ladevèze *et al.* to do time-space variable separation
- Chinesta *et al.* to solve high-dimensional Fokker-Planck equations in the context of kinetic models for polymers
- Nouy *et al* in the context of UQ.

The solution is represented as linear combinations of tensor products of small-dimensional functions to avoid the curse of dimensionality:

$$
\begin{aligned}
u(x_1, \ldots, x_d) &= \sum_{k \geq 1} r_k^1(x_1) r_k^2(x_2) \ldots r_k^d(x_d) \\
&= \sum_{k \geq 1} \left(r_k^1 \otimes r_k^2 \ldots \otimes r_k^d \right) (x_1, x_2, \ldots, x_d).
\end{aligned}
$$

Prototypical problem: uncertainty quantification on an obstacle problem



T: Random vector modelizing some uncertain parameters in the model.
x: Position
x and T respectively take values in $\mathcal{X}$ and $\mathcal{T}$.

$z(T,x)$: Height of the rope
$g(T,x)$: Height of the obstacle
$f(T,x)$: Stresses applied to the rope

Notation: $L^2_T(\mathcal{T}, H_x) = \left\{ v : \mathcal{T} \to H_x \mid \mathbb{E}\left[\|v(T)\|^2_{H_x} \right] < +\infty \right\}$

$g \in L^2_T(\mathcal{T}, H^1_0(\mathcal{X}))$, $f \in L^2_T(\mathcal{T}, L^2(\mathcal{X}))$.

Prototypical problem: uncertainty quantification on an obstacle problem

With an obstacle:
Find $z \in L^2(\mathcal{T}, H^1_0(\mathcal{X}))$ such that

$$\left\{ \begin{array}{c} -\Delta_x z \geq f \;\; \text{sur } \mathcal{T} \times \mathcal{X}, \\ z \geq g \;\; \text{sur } \mathcal{T} \times \mathcal{X}, \\ (\Delta_x z + f)(z - g) = 0 \;\; \text{sur } \mathcal{T} \times \mathcal{X}, \\ z(T, x) = 0 \quad \forall (T, x) \in \mathcal{T} \times \partial \mathcal{X}. \end{array} \right.$$

Equivalent formulation:

$$\mathcal{K} = \left\{ v \in L^2(\mathcal{T}, H^1_0(\mathcal{X})) \mid v(T, x) \geq g(T, x) \; \forall (T, x) \in \mathcal{T} \times \mathcal{X} \right\}$$

$$z = \underset{v \in \mathcal{K}}{argmin} \, \mathcal{J}(v)$$

with $\mathcal{J}(v) = \mathbb{E}\left[\frac{1}{2} \int_{\mathcal{X}} |\nabla_x v(T, x)|^2 \, dx - \int_{\mathcal{X}} f(T, x) v(T, x) \, dx \right]$

Penalized formulation

Problem: $\mathcal{K}$ is not a Hilbert space!!

We introduce a series of approached problems: $\rho > 0$ (large)

$$z_\rho = \text{argmin}_{v \in L^2_T(\mathcal{T}, H^1_0(\mathcal{X}))} \mathcal{J}_\rho(v)$$

where

$$\boxed{\mathcal{J}_\rho(v) = \mathbb{E}\left[\frac{1}{2}\int_{\mathcal{X}} |\nabla_x v(T,x)|^2\,dx - \int_{\mathcal{X}} f(T,x)v(T,x)\,dx + \frac{\rho}{2}\int_{\mathcal{X}} [g(T,x) - v(T,x)]_+^2\,dx\right]}$$

$$z_\rho \underset{\rho \to +\infty}{\longrightarrow} z$$

Our aim is to approximate $z_\rho = u$ for a given value of ρ with the greedy algorithm.
Let us denote $V = L^2_T(\mathcal{T}, H^1_0(\mathcal{X}))$, $\mathcal{E} = \mathcal{J}_\rho$.

The penalized problem can be rewritten under a more general form:

$$\boxed{u = \underset{v \in V}{\text{argmin}}\, \mathcal{E}(v)}$$

Definition of the greedy algorithm

The idea is to look iteratively for the best tensor product.

$(t,x) \in \mathcal{T} \times \mathcal{X}$ $\qquad$ $u(t,x) = \sum_{k\geq 1} r_k(t) s_k(x).$

$u \to V$, $\quad r_k \to V_t$, $\quad s_k \to V_x$

where the solution u is the unique global minimizer of the functional $\mathcal{E} : V \to \mathbb{R}$.

Greedy algorithm: We define recursively

$$(r_n, s_n) \in \underset{(r,s)\in V_t \times V_x}{\text{argmin}} \; \mathcal{E}\left(\sum_{k=1}^{n-1} r_k \otimes s_k + r \otimes s \right) \tag{1}$$

Let us denote $u_n = \sum_{k=1}^{n} r_k \otimes s_k$.
Question: Does u_n converge towards u?

Previous results: the linear case

Assumptions: $\Sigma = \{r \otimes s, r \in V_t, s \in V_x\}$

(A1) $\text{Vect}\,\Sigma \subset V$ is dense.

(A2) $\mathcal{E}(v) = \|v - u\|_V^2$.

- **The Singular Value Decomposition Case:** $\|r \otimes s\|_V = \|r\|_{V_t} \|s\|_{V_x}$
  Strong convergence!
  Orthogonality relations: $\langle r_n, r_{n'} \rangle_{V_t} = \langle s_n, s_{n'} \rangle_{V_x} = 0$, $\forall n \neq n'$.
  Optimal (unique) decomposition: at iteration n, $u_n = \sum_{k=1}^n r_k \otimes s_k$ is the minimizer of $\|u - \sum_{k=1}^n \phi_k \otimes \psi_k\|_V^2$ over all possible $(\phi_k, \psi_k)_{1 \leq k \leq n} \in (V_t \times V_x)^n$.
- **The linear case:** $\|r \otimes s\|_V \neq \|r\|_{V_t} \|s\|_{V_x}$
  No more othogonality relations nor optimal decomposition, but still strong convergence! (Even rates of convergence...) ([De Vore, Temlyakov, 1996] [Le Bris, TL, Maday, 2009])

The nonlinear case

([Cancès, VE, Lelièvre, 2010])

Assumptions:

(H1) Vect $\Sigma \subset V$ is dense.

(H2) $\mathcal{E}$ is α-convex.

$$\forall v, w \in V,\ \mathcal{E}(v) \geq \mathcal{E}(w) + \langle \mathcal{E}'(w), v-w \rangle_V + \frac{\alpha}{2}\|v-w\|_V^2.$$

(H3) For any sequence $r_n \otimes s_n$ in Σ which is bounded in V, there exists a subsequence which weakly converges in V to an element of Σ.

(H4) $\mathcal{E}$ is differentiable, and its gradient is Lipschitz on bounded sets of V.

$$\forall K \text{ bdd } \subset V, \exists L_K > 0, \forall v, w \in K,\ \|\mathcal{E}'(v) - \mathcal{E}'(w)\|_V \leq L_K \|v-w\|_V.$$

Remark: In the obstacle problem, $V_t = L^2_T(\mathcal{T}, \mathbb{R})$ and $V_x = H^1_0(\mathcal{X})$ is a good choice that satisfies all these assumptions with $V = L^2_T(\mathcal{T}, H^1_0(\mathcal{X}))$.

Convergence results

$$(r_n, s_n) \in \underset{(r,s)\in V_t \times V_x}{\operatorname{argmin}} \; \mathcal{E}\left(u_{n-1} + r \otimes s\right) \tag{2}$$

Theorem

Under assumptions (H1), (H2), (H3) et (H4), the iterations are well-defined ((r_n, s_n) exists and is non-zero iff $u_{n-1} \neq u$). Moreover, the sequence $(u_n)_{n\in\mathbb{N}^*}$ strongly converges in V towards u.

Theorem

In the finite dimensional case, the convergence is exponentially fast:
$\exists C > 0, \sigma \in (0,1)$,

$$\mathcal{E}(u_n) - \mathcal{E}(u) \leq C\sigma^n,$$

$$\|u - u_n\|_V^2 \leq C\sigma^n.$$

These results also hold in the case of more than two variables!

Case of a minimum local

We add a supplementary assumption:

(H5) There exists $\beta, \gamma > 0$ such that for all $(r, s) \in V_t \times V_x$,

$$\beta \|r\|_{V_t} \|s\|_{V_x} \leq \|r \otimes s\|_V \leq \gamma \|r\|_{V_t} \|s\|_{V_x}.$$

It is automatically satisfied in the finite-dimensional setting.

Theorem

Under assumptions (H1), (H2), (H3), (H4) and (H5), if the pair (r_n, s_n) is only defined as a **local minimizer** *such that $\mathcal{E}\left(u_{n-1} + r_n \otimes s_n\right) < \mathcal{E}(u_{n-1})$, then the sequence $(u_n)_{n \in \mathbb{N}^*}$ still strongly converges towards u in V. Besides, the exponential rate of convergence still holds in the finite-dimensional setting.*

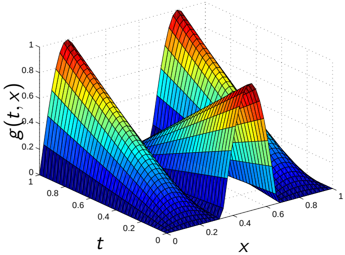
! Only in the case of two variables!

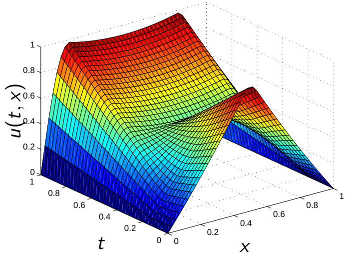
This last result is important since, in practice, only local minima can be computed.

Assumptions (H1), (H2), (H3), (H4) and (H5) are satisfied for the penalized formulation of the obstacle problem.
$\mathcal{X} = \mathcal{T} = (0,1)$. T uniform law of probability on $(0,1)$.

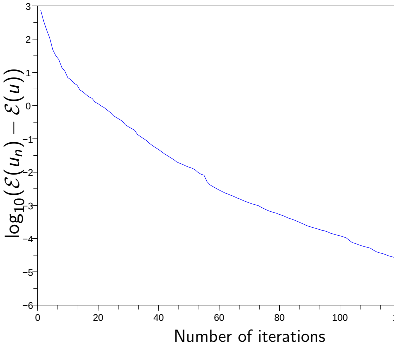
$$f(t,x) = -1 \text{ and } g(t,x) = t[sin(3\pi x)]_+ + (t-1)[sin(3\pi x)]_-.$$

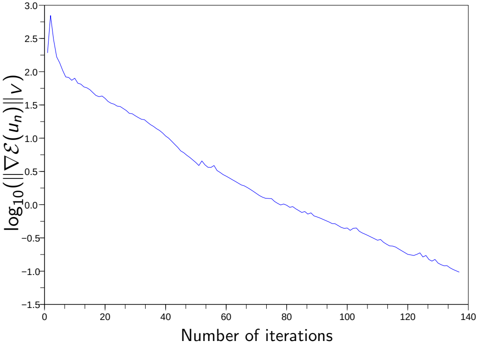
$\rho = 2500$





Rate of convergence



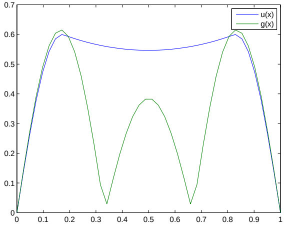


Open problems and future research

On greedy algorithms in general

- What happens in the case when $\mathcal{E}$ is not convex?
- What happens if u cannot be defined as the minimum of a functional? ([Nouy, Le Maitre, 2009]) Good numerical results on the Burgers equation

On the obstacle problem



- Problem of the penalized approach: we approximate z_ρ and not z! Idea: combine the greedy algorithm with other algorithms which converge towards the real solution of the obstacle problem: augmented lagrangian method
- Improve the rate of convergence by using alternative version of the algorithm: Orthogonal greedy, Minimum residual or Minimax ([Nouy 2010])

Outline

1. Convex nonlinear problems

2. Non-symmetric problems

The symmetric setting on a particular case

$$\left\{ \begin{array}{l} \text{find } u \in L^2((-1,1), H^1_{\mathrm{per}}(-1,1)) \text{ such that} \\ -\Delta_x u(t,x) + cu(t,x) = f(t,x) \quad \forall (t,x) \in (-1,1) \times \mathbb{R}, \end{array} \right.$$

$$\text{with } f \in L^2((-1,1), L^2_{\mathrm{per}}(-1,1)) \text{ and } c > 0.$$

The Euler-Lagrange equations associated to the first iterations of the greedy algorithm are the following:

$$\text{find } (r,s) \in L^2(-1,1) \times H^1_{\mathrm{per}}(-1,1) \text{ such that}$$

$$\left\{ \begin{array}{rcl} \left(\int_{-1}^1 |r(t)|^2 \, dt \right) (-\Delta_x s(x) + cs(x)) & = & \int_{-1}^1 f(t,x) r(t) \, dt, \\ \left(\int_{-1}^1 |\nabla_x s(x)|^2 + c|s(x)|^2 \, dx \right) r(t) & = & \int_{-1}^1 f(t,x) s(x) \, dx. \end{array} \right.$$

They have a non-zero solution if and only if $f \neq 0$!

The non-symmetric setting

$$\left\{ \begin{array}{l} \text{find } u \in L^2((-1,1), H^1_{\text{per}}(-1,1)) \text{ such that} \\ -\Delta_x u(t,x) + b\nabla_x u(t,x) + cu(t,x) = f(t,x) \quad \forall (t,x) \in (-1,1) \times \mathbb{R}, \end{array} \right.$$

$$\text{with } f \in L^2((-1,1), L^2_{\text{per}}(-1,1)), \quad c > 0 \text{ and } b \in \mathbb{R}.$$

The Euler-Lagrange equations associated to the first iterations of the greedy algorithm are the following:

$$\text{find } (r,s) \in L^2(-1,1) \times H^1_{\text{per}}(-1,1) \text{ such that}$$

$$\left\{ \begin{array}{rcl} \left(\int_{-1}^1 |r(t)|^2\, dt\right) (-\Delta_x s(x) + b\nabla_x s(x) + cs(x)) & = & \int_{-1}^1 f(t,x) r(t)\, dt, \\ \left(\int_{-1}^1 |\nabla_x s(x)|^2 + c|s(x)|^2\, dx\right) r(t) & = & \int_{-1}^1 f(t,x) s(x)\, dx. \end{array} \right.$$

$$\int_{-1}^1 s(x) \nabla_x s(x)\, dx = 0.$$

What is happening then?

There is a whole bunch of functions f's such that the only solution to the Euler-Lagrange equations is necessarily $r \otimes s = 0$ (even for b as small as I want!)

Typically, $f(t,x) = \phi(x-t)$ with $\phi \in L^2_{\rm per}(-1,1)$ an even or an odd function.

Let us reason by contradiction and assume that there exits a solution (r,s) such that $r \otimes s \neq 0$. I can choose the couple (r,s) such that

$$\int_{-1}^{1} |r(t)|^2 \, dt = \int_{-1}^{1} |\nabla_x s(x)|^2 + |s(x)|^2 \, dx = \lambda > 0.$$

$$\left\{ \begin{array}{rcl} -\Delta_x s(x) + b\nabla_x s(x) + cs(x) & = & \frac{1}{\lambda} \int_{-1}^{1} f(t,x) r(t) \, dt, \\ r(t) & = & \frac{1}{\lambda} \int_{-1}^{1} f(t,x) s(x) \, dx. \end{array} \right.$$

What is happening then?

Plugging the second equation into the first one yields

$$-\Delta_x s(x) + b\nabla_x s(x) + cs(x) = \frac{1}{\lambda^2}\int_{-1}^{1}\left(\int_{-1}^{1} f(t,x)f(t,x')\,dt\right) s(x')\,dx'. \tag{3}$$

In the case when ϕ is an even function (typically $f(t,x) = \cos(2\pi(x-t))$),

$$\begin{aligned} g(x,x') &= \int_{-1}^{1} f(t,x)f(t,x')\,dt = \int_{-1}^{1} \phi(x-t)\phi(x'-t)\,dt, \\ &= \int_{-1}^{1} \phi(x-t)\phi(t-x')\,dt = \int_{-1-x'}^{1-x'} \phi(x-x'-u)\phi(u)\,du, \\ &= \int_{-1}^{1} \phi(x-x'-u)\phi(u)\,du, \\ &= \phi * \phi(x-x') = h(x-x'). \end{aligned}$$

What is happening then?

Fourier transform of (3) yields that for all $k \in \pi\mathbb{Z}$,

$$|k|^2\widehat{s}(k) + ikb\widehat{s}(k) + c\widehat{s}(k) = \frac{1}{\lambda^2}\left(\widehat{\phi}(k)\right)^2\widehat{s}(k). \tag{4}$$

In the case when ϕ is an even function, $\widehat{\phi}(k) \in \mathbb{R}$ and this yields that $\widehat{s}(k) = 0$ for all $k \in \pi\mathbb{Z}$.

Morality: There are cases when Galerkin PGD will not work (even with update) since the existence of a non-zero solution to the Euler-Lagrange equations is not always satisfied!

Need for other algorithms such as X-Greedy or Minimax...

References

- A. Ammar, B. Mokdad, F. Chinesta, and R. Keunings. A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. J. Non-Newtonian Fluid Mech., 139:153-176, 2006.
- E. Cancès, V. Ehrlacher, TL. Convergence of a greedy algorithm for high-dimensional convex nonlinear problems, http://hal.archives-ouvertes.fr/hal-00469622
- C. Le Bris, TL and Y. Maday. Results and questions on a nonlinear approximation approach for solving high-dimensional partial differential equations, Constructive Approximation 30(3):621-651, 2009.
- A. Nouy, Recent Developments in Spectral Stochastic Methods for the Numerical Solution of Stochastic Partial Differential Equations, Archives of Computational Methods in Engineering 16:251-285, 2009.
- A. Nouy, A priori model reduction through Proper Generalized Decomposition for solving time-dependent partial differential equations, *preprint*, 2010.
- A. Nouy and O. Le Maitre. Generalized spectral decomposition for stochastic nonlinear problems, J. Comput. Phys. 228(1):202-235, 2009.
- V.N. Temlyakov. Greedy approximation. Acta Numerica, 17:235-409, 2008.

Thank you for your attention!