Parareal methods for optimal control II.

Julien Salomon

JLL & INRIA, ANGE project-team (Inria, Paris-Sorbonne Université, CNRS)

Ecole - Décomposition des domaines pour des problèmes de contrôle optimal, 7.9.22

Optimal control problems: *Minimize*

$$J(c) = F\left(u(T)\right) + \int_0^T G\left(u(t), c(t)\right) dt,$$

with the constraint $\dot{u} = f(u, c)$.

Context: We already have an optimal control solver.
Question: *How to parallelize it ?*

1. The Intermediate States Method
2. Quantum control problems
3. Various optimization solvers
4. Full efficiency ?
5. Conclusion

Outline

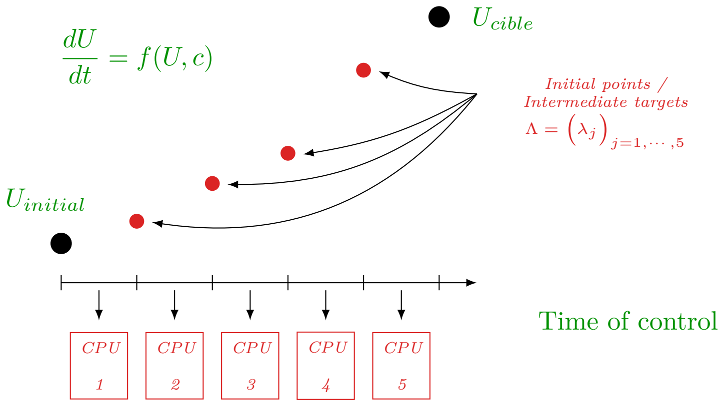
"Non-linear control" or "Bilinear control"

	Linear eq.	Non-linear eq.
"Linear" control	$\dot{y} = Ay + Bc$	$\dot{y} = f(y) + Bc$
Non-linear control	$\dot{y} = A(c)y$	$\dot{y} = f(y, c)$

- $y = y(t, x)$ state
- $c = c(t)$ or $c(t, x)$ control

The Intermediate States Method

Schematic description



Disclaimer : not a parareal algorithm.

And it follows:

- Independent sub-problems

$$J \to (J_j)_{j=1,\cdots,N},$$

- Need for an update formula for the intermediate states

$$\Lambda = (\lambda_j)_{j=1,\cdots,N}.$$

Algorithm:

Given c^k, Λ^k (intermediate targets) at step k:

1. *solve in parallel on $[T_j, T_{j+1}]$*

$$\max_{c_j} J_j(c_j) \to c_j^{k+1},$$

2. *define c^{k+1} as the concatenation of c_j^{k+1},*
3. *define Λ^{k+1} in a "relevant way" with c^{k+1}, so that the consistency lemma holds.*

Example of quantum control: Schrödinger Equation:

$$i\frac{\partial\psi(x,t)}{\partial t} = [-\frac{\hbar}{2m}\Delta + V(x) - \mu(x)c(t)]\psi(x,t)$$

Cost functional:

$$J(c) = 2\Re\langle\psi_{target}|\psi(.,T)\rangle - \int_0^T \alpha(t)c^2(t)dt$$

$$\left(= 2 - \|\psi_{target} - \psi(.,T)\|_{L^2}^2 - \int_0^T \alpha(t)c^2(t)dt\right)$$

...to be maximized.

$$J(c, \psi, \chi) = 2\Re\langle\psi_{target}|\psi(., T)\rangle - \int_0^T \alpha(t)c^2(t)dt$$

$$-2\Re\int_0^T \langle\chi(., t)|\partial_t + iH - \mu c(t)|\psi(., t)\rangle dt.$$

Optimality system:

$$\nabla_\chi J \rightarrow \begin{cases} i\frac{\partial}{\partial t}\psi(x,t) = (H - c(t)\mu(x))\psi(x,t) \\ \\ \psi(x,t=0) = \psi_0(x) \end{cases}$$

$$\nabla_\psi J \rightarrow \begin{cases} i\frac{\partial}{\partial t}\chi(x,t) = (H - c(t)\mu(x))\chi(x,t) \\ \\ \chi(x,t=T) = \psi_{target}(x) \end{cases}$$

$$\nabla_c J \rightarrow \quad \alpha(t)c(t) = -\Im < \chi(.,t)|\mu|\psi(.,t) >$$

Parallelization setting:

Define $\lambda_0 = \psi_0$, $\lambda_N = \psi_{target}$, $c_j = c_{|[T_j, T_{j+1}]}$ and $\beta_j = \frac{T}{T_{j+1} - T_j}$.

$$J_{\|}(c, \Lambda) = \sum_{j=0}^{N-1} \beta_j J_j(c_j, \lambda_j, \lambda_{j+1})$$

where J_j are the *parareal cost functionals*:

$$J_j(c_j, \lambda_j, \lambda_{j+1}) = \|\psi_j(T_{j+1}^-) - \lambda_{j+1}\|_{L^2}^2 + \int_{T_j}^{T_{j+1}} \alpha_j'(t) c_j(t)^2 dt,$$

$$\alpha_j'(t) = \frac{\alpha(t)}{\beta_j}, \ \psi_j(T_j^+) = \lambda_j.$$

Theorem: Given c, with the previous notations, let us define $\Lambda^c = (\lambda_j^c)_{j=1,\dots,N-1}$ by:

$$\lambda_j^c = (1-\gamma_j)\psi(T_j) + \gamma_j \chi(T_j),$$

where $\gamma_j = \frac{T_j}{T}$.
Then :

$$\Lambda^c = \mathrm{argmin}_\Lambda \left(J_{||}(c,\Lambda)\right).$$

Moreover we have:

$$J_{||}(c,\Lambda^c) = J(c).$$

Y. Maday, J. Salomon, G. Turinici, *SIAM J. Num. Anal., 45 (6) (2007)*

Convergence?

$$
\begin{aligned}
J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k}) \quad = \quad & J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k+1}) \\
& J_{\|}(c^{k},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k})
\end{aligned}
$$

Convergence?

$$\begin{aligned} J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k}) &= J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k+1}) \\ & \quad J_{\|}(c^{k},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k}) \\ &\geq J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k+1}). \end{aligned}$$

Convergence?

$$\begin{aligned} J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k}) &= J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k+1}) \\ & \quad J_{\|}(c^{k},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k}) \\ &\geq J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^{k},\Lambda^{k+1}). \end{aligned}$$

$\Rightarrow$ The proof of convergence is reduced to the one of the optimization solver.

Another example: Nuclear Magnetic Resonance

Aim : control spin using Magnetic fields.

Applications :

- Medical imaging
- Quantum computing
- Porous media identification
- ...

Toy model : Bloch equations

$$\begin{aligned} i\partial_t U(t) &= [H_0 + \sum_{\ell}^{L} \omega_\ell(t) H_\ell] U(t) \\ U(t=0) &= U_0. \end{aligned}$$

Optimal control problem:

find $\Omega^\star(t) = (\omega_1^\star(t), \cdots, \omega_L^\star(t))$, *that solves*

$$\Omega^\star = \text{argmax}\,(J(\Omega)) = \text{argmax}\,(Re\langle U(T), U_{target}\rangle)\,.$$

K. Riahi, J. Salomon, D. Sugny, *Physical Review A (2016).*

What about the optimization solver ?

We can show that :

$$\nabla J(c)_{|[T_j,T_{j+1}]} = \frac{T_{j+1}-T_j}{T}\nabla J_j(c_{|[T_j,T_{j+1}]}).$$

FOR EVERY c !

$\Rightarrow$ the intermediate target method provides a **decomposition of the gradient** that enables parallelization.

Constant step gradient method

$\Rightarrow$ Full efficiency !

Algebraic identity:

$$J(c') - J(c) = \int_0^T (c(t) - c'(t)) \Big(\alpha(t) (c(t) + c'(t)) + 2\Im \langle \chi(.,t) | \mu | \psi'(.,t) \rangle \Big) dt$$

Algebraic identity:

$$J(c') - J(c) = \int_0^T (c(t) - c'(t))\Big(\alpha(t)(c(t) + c'(t))$$

$$+ 2\Im\langle\chi(.,t)|\mu|\psi'(.,t)\rangle\Big)dt$$

$$\longrightarrow c^{k+1} = -\frac{1}{\alpha(t)}\Im\langle\chi^k(.,t)|\mu|\psi^{k+1}(.,t)\rangle$$

Algebraic identity:

$$J(c') - J(c) = \int_0^T (c(t) - c'(t))\Big(\alpha(t)(c(t) + c'(t)) + 2\Im\langle\chi(.,t)|\mu|\psi'(.,t)\rangle\Big)dt$$

$$\Rightarrow J(c^{k+1}) - J(c^k) = \int_0^T \alpha(t)\big(c^{k+1}(t) - c^k(t)\big)^2 dt \geq 0$$

Convergence?

$$\begin{aligned} J_{||}(c^{k+1},\Lambda^{k+1}) - J_{||}(c^k,\Lambda^k) &= J_{||}(c^{k+1},\Lambda^{k+1}) - J_{||}(c^k,\Lambda^{k+1}) \\ &\quad J_{||}(c^k,\Lambda^{k+1}) - J_{||}(c^k,\Lambda^k) \\ &\geq J_{||}(c^{k+1},\Lambda^{k+1}) - J_{||}(c^k,\Lambda^{k+1}) \end{aligned}$$

Convergence?

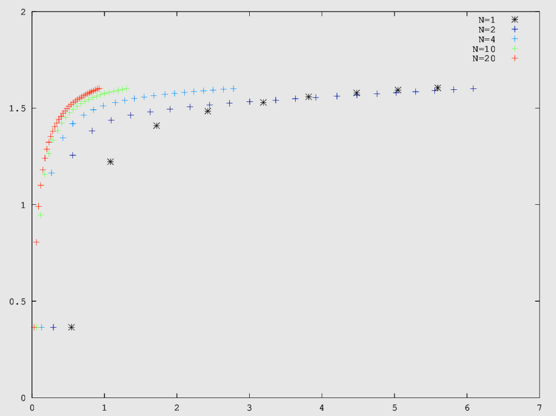
$$\begin{aligned} J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^k,\Lambda^k) &= J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^k,\Lambda^{k+1}) \\ & \quad J_{\|}(c^k,\Lambda^{k+1}) - J_{\|}(c^k,\Lambda^k) \\ &\geq J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^k,\Lambda^{k+1}) \\ &\geq \|c^{k+1} - c^k\|_\alpha^2. \end{aligned}$$

Convergence?

$$\begin{aligned} J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^k,\Lambda^k) &= J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^k,\Lambda^{k+1}) \\ &\quad J_{\|}(c^k,\Lambda^{k+1}) - J_{\|}(c^k,\Lambda^k) \\ &\geq J_{\|}(c^{k+1},\Lambda^{k+1}) - J_{\|}(c^k,\Lambda^{k+1}) \\ &\geq \|c^{k+1}-c^k\|_\alpha^2. \end{aligned}$$

$\Rightarrow$ Monotonicity preserved by the parallelization.

Monotonic algorithm



N	$\frac{Time_{seq}}{N \cdot Time_{\parallel}}$
1	-
2	41.7%
4	45.8%
10	39.4%
20	27.2%

Newton

Here, our parallelization method not only improves the Newton convergence makes it possible .

N	$N \cdot Time_{\|\|}$
1	-
2	-
4	33.722
10	3.2544
20	0.72559

Outline

Full efficiency ?

The optimization is achieved in parallel, but $\psi(t)$ and $\chi(t)$ seem to require solving on $[0, T]$ (full propagation) ?

Full efficiency ?

The optimization is achieved in parallel, but $\psi(t)$ and $\chi(t)$ seem to require solving on $[0, T]$ (full propagation) ?

NO !!! $\rightarrow$ only $\psi(t_j)$ and $\chi(t_j)$ are required.

$\Rightarrow$ For low dimensional systems, the propagators $t_j \rightarrow t_{j+1}$ can be computed in parallel, when computing the gradient !

For large systems:

- Use the parareal algorithm to achieve full propagations, see

  "Parareal in time intermediate targets methods for optimal control problem",
  Y. Maday, J. Salomon, K. Riahi, Proc. of " Control and Optimization of PDEs ",
  (Birkhäuser, Basel)
- Use model reduction...

Conclusion

- Generic approach with respect to the solver,
- Definition of the intermediate states depends on the problem,
- Limit preserving parallelization strategy,
- Full efficiency obtained in some cases.

Merci !