Time-parallelization of sequential data assimilation problems

Felix KWOK, Sebastián RIFFO*, Julien SALOMON

* GÉOAZUR, CNRS & Université Côte d'Azur

DD26, December 7-11, 2020

Outline

Luenberger observer

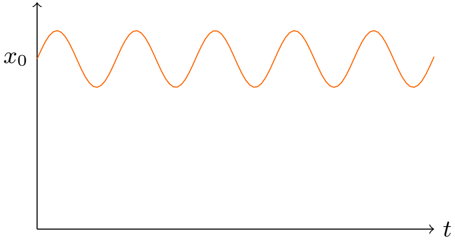
Time-parallelization setting

Parareal algorithm

Diamond strategy (Parareal case)

A textbook example

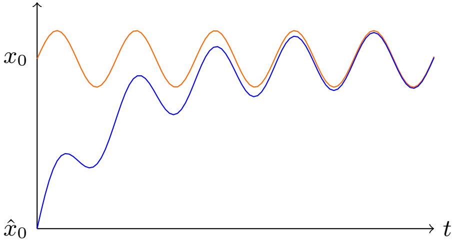
$$(1) \quad \begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ x(0) = x_0 \text{ unknown}, \\ y(t) = Cx(t). \end{cases}$$



Luenberger observer

(1) $$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ x(0) = x_0 \text{ unknown}, \\ y(t) = Cx(t). \end{cases}$$

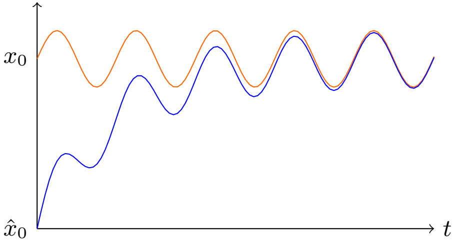
(2) $$\begin{cases} \dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) \\ \qquad + L\left[y(t) - \hat{y}(t)\right], \\ \hat{x}(0) = \hat{x}_0, \\ \hat{y}(t) = C\hat{x}(t). \end{cases}$$



Luenberger observer

(1) $$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), \\ x(0) = x_0 \text{ unknown}, \\ y(t) = Cx(t). \end{cases}$$

(2) $$\begin{cases} \dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) \\ \qquad\qquad + L\left[y(t) - \hat{y}(t)\right], \\ \hat{x}(0) = \hat{x}_0, \\ \hat{y}(t) = C\hat{x}(t). \end{cases}$$



Is there a *convenient* way to choose the *observer gain* L ?

Note that

$$(2) \iff \begin{cases} \dot{\hat{x}}(t) = (A - LC)\hat{x}(t) + (Bu(t) + Ly(t)), \\ \hat{x}(0) = \hat{x}_0. \end{cases}$$

and then $x(t) - \hat{x}(t) = \mathrm{e}^{(A-LC)t}\,(x(0) - \hat{x}(0))$

Note that

$$(2) \iff \begin{cases} \dot{\hat{x}}(t) = (A - LC)\hat{x}(t) + (Bu(t) + Ly(t)), \\ \hat{x}(0) = \hat{x}_0. \end{cases}$$

and then $x(t) - \hat{x}(t) = \mathrm{e}^{(A-LC)t}\,(x(0) - \hat{x}(0))$

Theorem ▶ Identity observer Theorem [Luenberger]

Given a completely *observable* system (1), an identity observer of the form (2) can be constructed, and the coefficients of the characteristic polynomial of the observer can be selected arbitrarily.

Note that

$$(2) \iff \begin{cases} \dot{\hat{x}}(t) = (A - LC)\hat{x}(t) + (Bu(t) + Ly(t)), \\ \hat{x}(0) = \hat{x}_0. \end{cases}$$

and then $x(t) - \hat{x}(t) = \mathrm{e}^{(A-LC)t}\,(x(0) - \hat{x}(0))$

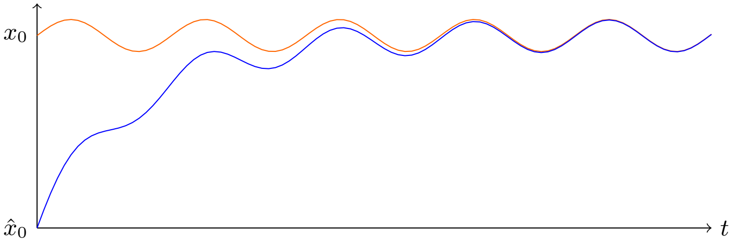
PROPOSITION 1.1

We assume System (1) is observable and the eigenvalues of $A - LC$ are negative and simple. Then, we have

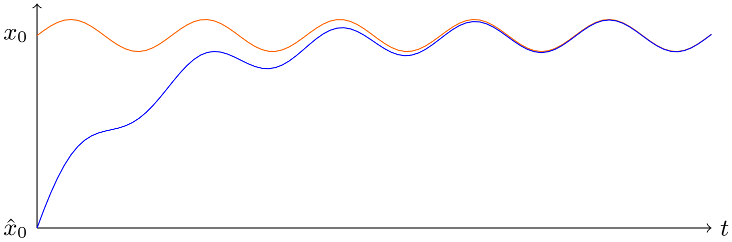
$$\left\| \mathrm{e}^{(A-LC)t} \right\| \leq \gamma\, \mathrm{e}^{-\mu t}$$

with $\mu := \min_{\nu \in \sigma(A-LC)} |\nu|$ and $\gamma := \mathrm{cond}(V) = \left\| V^{-1} \right\| \|V\|$, where V is the matrix whose rows are the eigenvectors of $A - LC$ and $\|\cdot\|$ represents the induced 2-norm of a matrix.

Diamond strategy



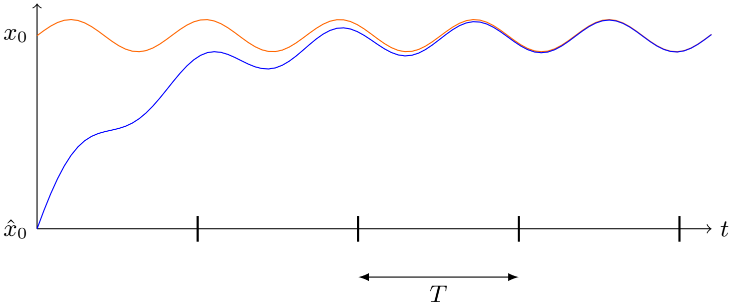
Diamond strategy



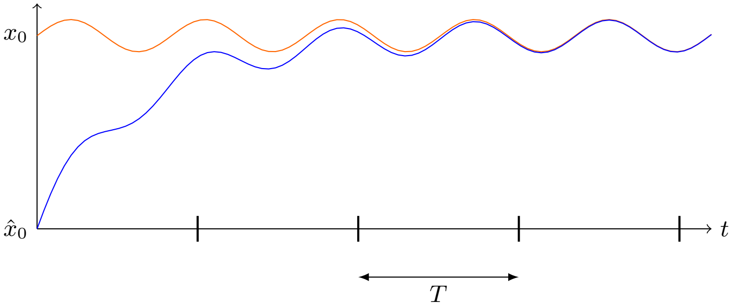
Luenberger observer

Time-parallel method

Diamond strategy

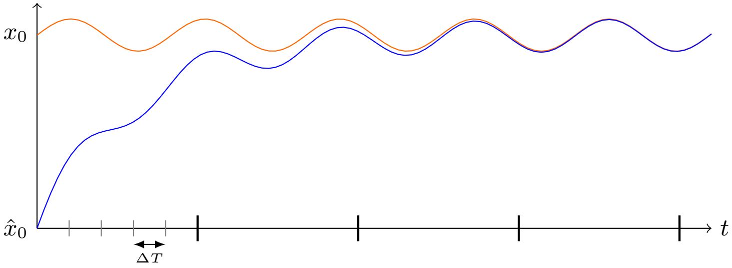


- Divide the time interval into *windows* W_ℓ of a given length $T > 0$.

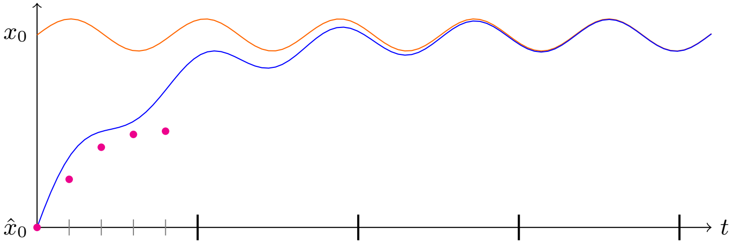


- Divide the time interval into *windows* W_ℓ of a given length $T > 0$.
- Solve Equation (2) on each window, in a sequential order, using a time-parallel algorithm.

Diamond strategy (time-parallelization)

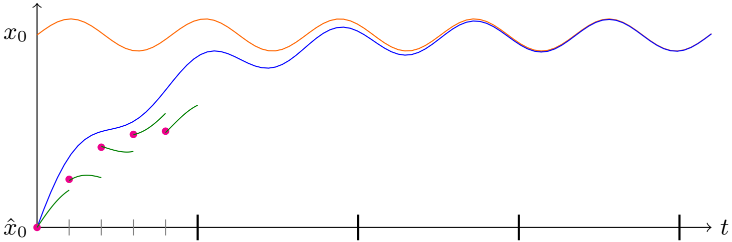


- Decompose W_ℓ into N *subintervals* of length ΔT.



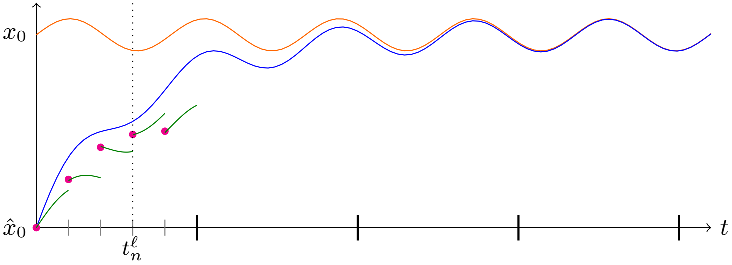
- Parallelizing in time requires the introduction of initial conditions $\hat{X}^h_{\ell,n}$.

Diamond strategy (time-parallelization)



- Parallelizing in time requires the introduction of initial conditions $\hat{X}^h_{\ell,n}$.
- We then construct a parallel version $\hat{x}_\parallel(t)$ of Equation (2) in each subinterval.

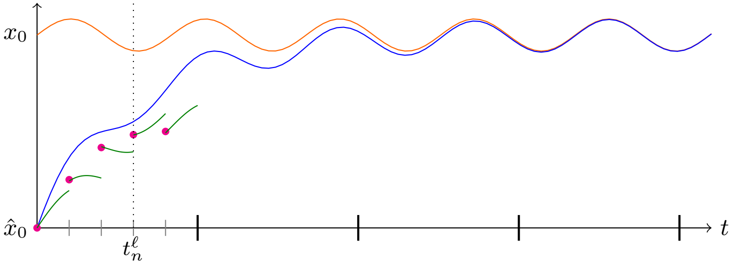
Diamond strategy (time-parallelization)



- Parallelizing in time requires the introduction of initial conditions $\hat{X}^h_{\ell,n}$.
- We then construct a parallel version $\hat{x}_\parallel(t)$ of Equation (2) in each subinterval.
- Imposing initial conditions induces discontinuities (jumps) at t_n^ℓ:

$$J^h_{\ell,n} := \hat{X}^h_{\ell,n} - \hat{x}_\parallel({t_n^\ell}^-).$$

Diamond strategy (time-parallelization)

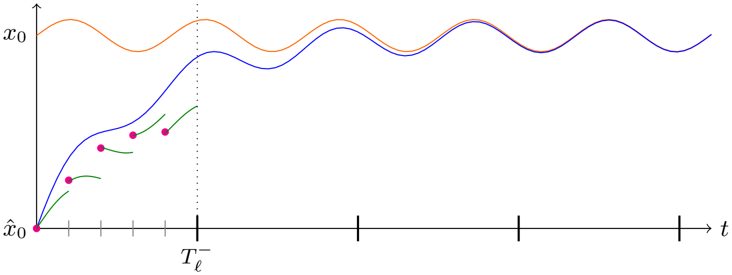


- Parallelizing in time requires the introduction of initial conditions $\hat{X}^h_{\ell,n}$.
- We then construct a parallel version $\hat{x}_{\|}(t)$ of Equation (2) in each subinterval.
- Imposing initial conditions induces discontinuities (jumps) at t_n^ℓ:

$$J^h_{\ell,n} := \hat{X}^h_{\ell,n} - \hat{x}_{\|}({t_n^\ell}^-).$$

Next step : define a suitable stopping criterion !

Diamond strategy (stopping criterion)



Lemma

Under the assumptions of Proposition 1.1, we have

$$\left\|(x - \hat{x}_\parallel)(T_\ell{}^-)\right\| \leq \gamma \left(\|x_0 - \hat{x}_0\| + \sum_{i=1}^{\ell} \mathrm{e}^{-\mu(i-1)T} \cdot \gamma \sum_{n=1}^{N-1} \mathrm{e}^{\mu n T} \left\| J_{i,n}^h \right\| \right) \mathrm{e}^{-\mu \ell T}$$

Diamond strategy (stopping criterion)

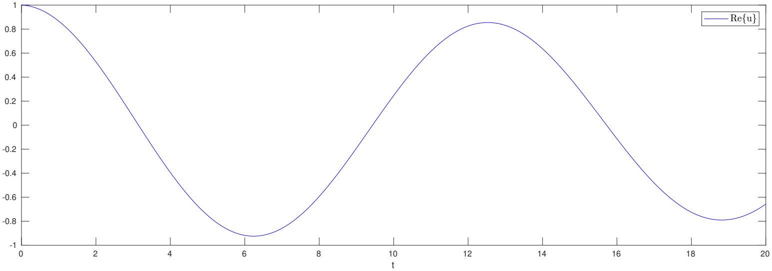
Proposition 2.1 ▶ A posteriori estimate

Let us assume that h is obtained from the stopping criterion in W_ℓ

$$2\gamma \sum_{n=1}^{N-1} \mathrm{e}^{\mu n \Delta T} \left\| J_{\ell,n}^{h} \right\| \leq \widetilde{\gamma} \frac{\mathrm{e}^{-\mu(\ell-1)T}}{2^{\ell-1}}$$

where $\widetilde{\gamma}$ is an arbitrary parameter. Then, the rate of convergence of $\hat{x}_{\|}(t)$ to $x(t)$ is bounded by μ, i.e.

$$\left\| (x - \hat{x}_{\|})(T_\ell{}^{-}) \right\| \leq \gamma \left(\| x(0) - \hat{x}(0) \| + \widetilde{\gamma} \right) \mathrm{e}^{-\mu \ell T}.$$

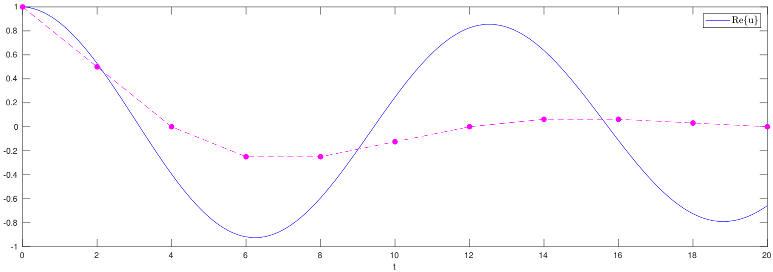


Dahlquist equation $\dot{u}(t) = -\frac{\mathrm{i}}{2}u$ in $[0, 20]$

To solve the problem

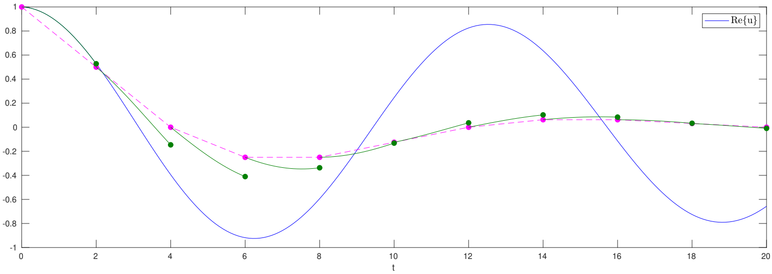
$$\begin{cases} \dot{u}(t) = f(u(t)), \quad t \in [0, T] \\ u(0) = u_0 \end{cases}$$

we decompose the time interval on N subintervals, denoted by (t_{n-1}, t_n).



- Impose arbitrary values on the subintervals by using the coarse solver $\mathcal{G}$:

$$U_0^0 = u_0,\ U_n^0 = \mathcal{G}(t_n, t_{n-1}, U_{n-1}^0).$$

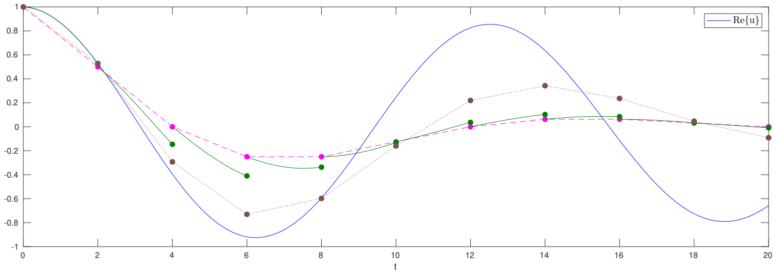


- Impose arbitrary values on the subintervals by using the coarse solver $\mathcal{G}$:

$$U_0^0 = u_0, \; U_n^0 = \mathcal{G}(t_n, t_{n-1}, U_{n-1}^0).$$

- Using the fine solver $\mathcal{F}$, solve in parallel

$$\begin{cases} \dot{u}(t) = f(u(t)), \quad t \in [t_{n-1}, t_n] \\ u(t_{n-1}) = U_{n-1}^0. \end{cases}$$



- Impose arbitrary values on the subintervals by using the coarse solver $\mathcal{G}$:

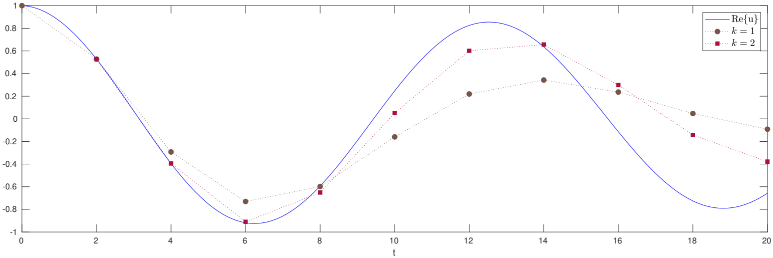
$$U_0^0 = u_0,\ U_n^0 = \mathcal{G}(t_n, t_{n-1}, U_{n-1}^0).$$

- Using the fine solver $\mathcal{F}$, solve in parallel

$$\begin{cases} \dot{u}(t) = f(u(t)), & t \in [t_{n-1}, t_n] \\ u(t_{n-1}) = U_{n-1}^0. \end{cases}$$

- Smooth the discontinuities previously introduced by defining

$$U_n^1 := \mathcal{F}(t_n, t_{n-1}, U_{n-1}^0) + \mathcal{G}(t_n, t_{n-1}, U_{n-1}^1) - \mathcal{G}(t_n, t_{n-1}, U_{n-1}^0).$$

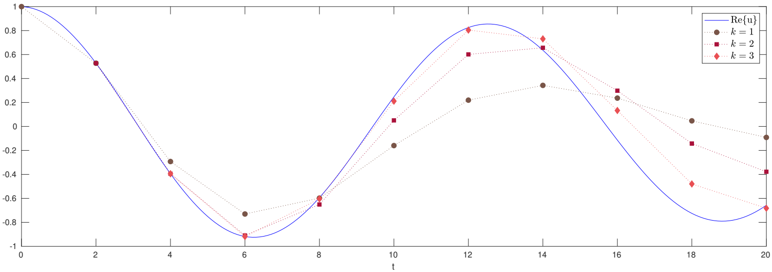


At iteration k:

- compute $\{\mathcal{F}(t_n, t_{n-1}, U_{n-1}^{k-1})\}_{n=1}^{N}$ in parallel.
- Update the sequence

$$U_n^k := \mathcal{F}(t_n, t_{n-1}, U_{n-1}^{k-1}) + \mathcal{G}(t_n, t_{n-1}, U_{n-1}^{k}) - \mathcal{G}(t_n, t_{n-1}, U_{n-1}^{k-1})$$

by computing $\{\mathcal{G}(t_n, t_{n-1}, U_{n-1}^{k})\}_{n=1}^{N}$ sequentially.



At iteration k:

- compute $\{\mathcal{F}(t_n, t_{n-1}, U_{n-1}^{k-1})\}_{n=1}^{N}$ in parallel.
- Update the sequence

$$U_n^k := \mathcal{F}(t_n, t_{n-1}, U_{n-1}^{k-1}) + \mathcal{G}(t_n, t_{n-1}, U_{n-1}^{k}) - \mathcal{G}(t_n, t_{n-1}, U_{n-1}^{k-1})$$

by computing $\{\mathcal{G}(t_n, t_{n-1}, U_{n-1}^{k})\}_{n=1}^{N}$ sequentially.

What about its convergence ?

Theorem ▶ Convergence of Parareal [Gander and Hairer]

(...) at iteration k of the Parareal algorithm, we have the bound

$$\left\|u(t_n) - U_n^k\right\| \leq \frac{C_3}{C_1} \frac{\left(C_1 \Delta T^{p+1}\right)^{k+1}}{k!} (1 + C_2 \Delta T)^{n-(k+1)} \prod_{j=0}^{k} (n - j).$$

Theorem ▶ Convergence of Parareal [Gander and Hairer]

(...) at iteration k of the Parareal algorithm, we have the bound

$$\left\|u(t_n) - U_n^k\right\| \leq \frac{C_3}{C_1} \frac{\left(C_1 \Delta T^{p+1}\right)^{k+1}}{k!} (1 + C_2 \Delta T)^{n-(k+1)} \prod_{j=0}^{k} (n-j).$$

- Superlinear rate of convergence.

Theorem ▶ Convergence of Parareal [Gander and Hairer]

(...) at iteration k of the Parareal algorithm, we have the bound

$$\left\| u(t_n) - U_n^k \right\| \leq \frac{C_3}{C_1} \frac{\left(C_1 \Delta T^{p+1}\right)^{k+1}}{k!} (1 + C_2 \Delta T)^{n-(k+1)} \prod_{j=0}^{k} (n-j).$$

- Superlinear rate of convergence.
- Among other assumptions, $\mathcal{F}(t_n, t_{n-1}, U_{n-1}^k)$ is the exact solution on (t_{n-1}, t_n), and $\mathcal{G}$ must satisfy

$$\left\| \mathcal{G}(t + \Delta T, t, x) - \mathcal{G}(t + \Delta T, t, y) \right\| \leq (1 + C_2 \Delta T) \left\| x - y \right\|.$$

Theorem ▶ Convergence of Parareal [Gander and Hairer]

(...) at iteration k of the Parareal algorithm, we have the bound

$$\left\|u(t_n) - U_n^k\right\| \leq \frac{C_3}{C_1}\frac{\left(C_1\Delta T^{p+1}\right)^{k+1}}{k!}(1 + C_2\Delta T)^{n-(k+1)}\prod_{j=0}^{k}(n-j).$$

- Superlinear rate of convergence.
- Among other assumptions, $\mathcal{F}(t_n, t_{n-1}, U_{n-1}^k)$ is the exact solution on (t_{n-1}, t_n), and $\mathcal{G}$ must satisfy

$$\|\mathcal{G}(t+\Delta T, t, x) - \mathcal{G}(t+\Delta T, t, y)\| \leq (1 + C_2\Delta T)\,\|x - y\|\,.$$

- The result is well suited for non-decaying problems.

THEOREM ▶ CONVERGENCE OF PARAREAL [GANDER AND HAIRER]

(...) at iteration k of the Parareal algorithm, we have the bound

$$\left\|u(t_n) - U_n^k\right\| \leq \frac{C_3}{C_1}\frac{\left(C_1 \Delta T^{p+1}\right)^{k+1}}{k!}\left(1 + C_2 \Delta T\right)^{n-(k+1)} \prod_{j=0}^{k}(n-j).$$

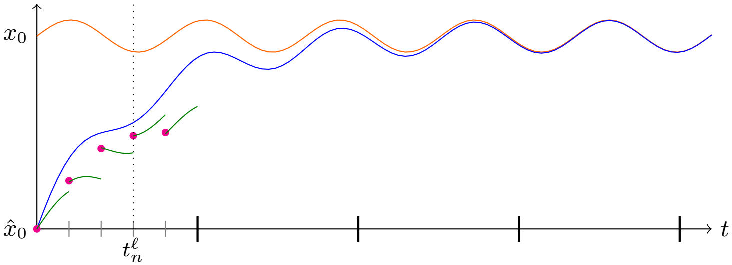
THEOREM 3.1 ▶ CONVERGENCE OF PARAREAL FOR DECAYING PROBLEMS

(...) We also assume that $\mathcal{F}$ and $\mathcal{G}$ are Lipschitz with respect to the initial conditions:

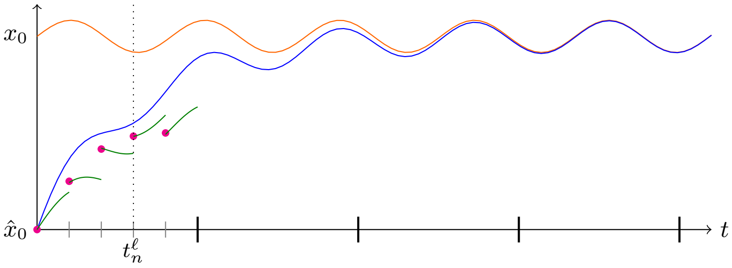
$$\max\left\{\|\mathcal{F}(t_n, t_{n-1}, y) - \mathcal{F}(t_n, t_{n-1}, z)\|, \|\mathcal{G}(t_n, t_{n-1}, y) - \mathcal{G}(t_n, t_{n-1}, z)\|\right\} \leq \varepsilon \|y - z\|,$$

for a constant $\varepsilon \in (0,1)$. Then, after k iterations of the Parareal algorithm, we have

$$\left\|U_n^k - u(t_n)\right\| \leq \begin{cases} 0 & n \leq k \\ B_n^k := \alpha\beta^k \sum_{i=0}^{n-k-1} \binom{k+i}{k}\varepsilon^i & n > k. \end{cases}$$



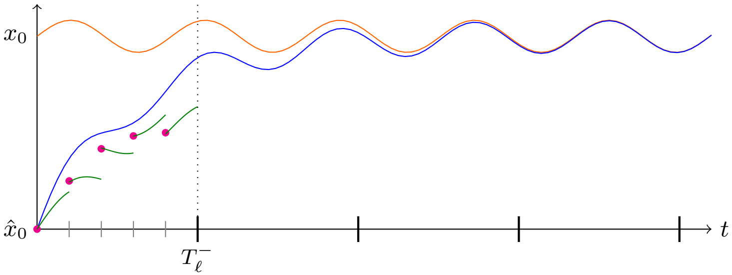
- The Luenberger observer $\hat{x}(t)$ is a decaying problem (Proposition 1.1).



- The Luenberger observer $\hat{x}(t)$ is a decaying problem (Proposition 1.1).
- The number of parareal iterations $\{k_\ell\}_\ell$ can be determined from
  (a) Proposition 2.1 (*a posteriori* estimate)

$$2\gamma \sum_{n=1}^{N-1} \mathrm{e}^{\mu n \Delta T} \left\| \hat{X}_{\ell,n}^{k_\ell} - \hat{x}_{\|}(t_n^\ell) \right\| \leq \widetilde{\gamma} \frac{\mathrm{e}^{-\mu(\ell-1)T}}{2^{\ell-1}}, \quad \widetilde{\gamma} \text{ arbitrary.}$$

Diamond strategy (Parareal case)



- The Luenberger observer $\hat{x}(t)$ is a decaying problem (Proposition 1.1).
- The number of parareal iterations $\{k_\ell\}_\ell$ can be determined from
  (a) Proposition 2.1 (*a posteriori* estimate)

$$2\gamma \sum_{n=1}^{N-1} \mathrm{e}^{\mu n \Delta T} \left\| \hat{X}_{\ell,n}^{k_\ell} - \hat{x}_{\|}(t_n^\ell) \right\| \leq \widetilde{\gamma} \frac{\mathrm{e}^{-\mu(\ell-1)T}}{2^{\ell-1}}, \quad \widetilde{\gamma} \text{ arbitrary.}$$

  (b) Theorem 1.3 (*a priori* bound)

$$\left\| \hat{X}_{\ell,n}^{k_\ell} - \hat{x}_{\|}(t_n^\ell) \right\| \leq B_n^k(\alpha, \beta, \varepsilon)$$

Diamond strategy (Parareal case)

Theorem 4.1

We keep the assumptions of Proposition 1.1 and Theorem 1.3. For a window W_ℓ and $\widetilde{\gamma} > 0$, we define

$$k_\ell = \begin{cases} \min S_\ell & S_\ell \neq \emptyset \\ k_{\ell-1} & S_\ell = \emptyset \end{cases}$$

where

$$S_\ell = \left\{ k \in \mathbb{N}^*, k \leq N-1 :\, 2\gamma \sum_{n=1}^{N-1} \mathrm{e}^{\mu n \Delta T} B_n^k(\alpha, \beta, \varepsilon) \leq \widetilde{\gamma} \frac{\mathrm{e}^{-\mu(\ell-1)T}}{2^{\ell-1}} \right\}.$$

Suppose that we apply the *Diamond strategy* using k_ℓ iterations of the Parareal algorithm. Then, the stopping criterion is satisfied.

A TEXTBOOK EXAMPLE

Consider the second order system

$$\begin{cases} \ddot{z} + 2\eta\omega\dot{z} + \omega^2 z = 5 + e^{-100t}\sin(\frac{3}{4}t), \\ z(0),\, \dot{z}(0) \text{ unknown}, \end{cases}$$

with $\dot{z}(t)$ as output.

A textbook example

Consider the second order system (in matricial form)

$$\begin{cases} \dot{x} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2\eta\omega \end{bmatrix} x + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t), \\ x(0) = x_0 \text{ unknown}, \\ y = \begin{pmatrix} 0 & 1 \end{pmatrix} x, \end{cases}$$

with $x = (z \;\; \dot{z})^\top$ and $u(t) = 5 + e^{-100t} \sin(\frac{3}{4}t)$.

Consider the second order system (in matricial form)

$$\begin{cases} \dot{x} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2\eta\omega \end{bmatrix} x + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t), \\ x(0) = x_0 \text{ unknown}, \\ y = \begin{pmatrix} 0 & 1 \end{pmatrix} x, \end{cases}$$

with $x = (z \;\; \dot{z})^\top$ and $u(t) = 5 + e^{-100t} \sin(\frac{3}{4}t)$.

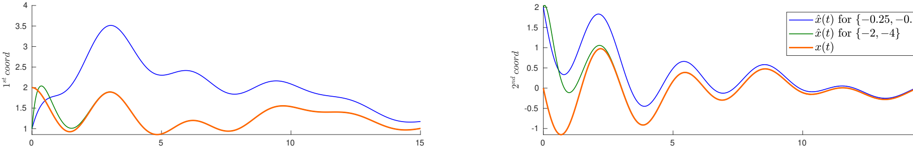
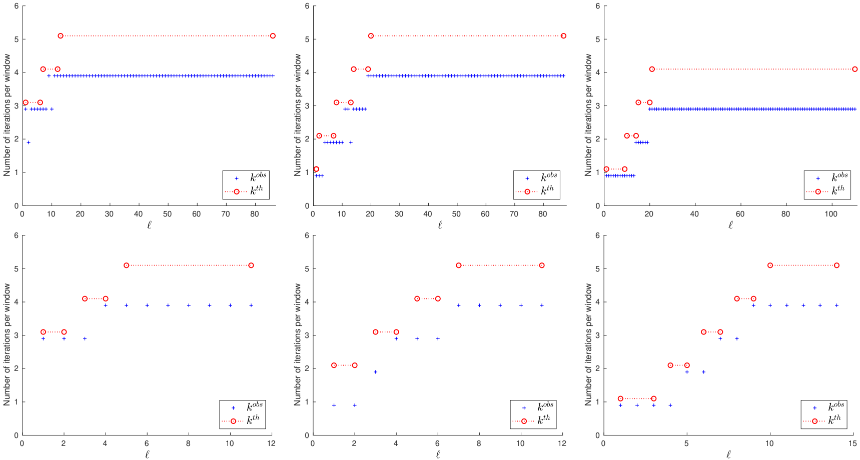


Figure: $\eta = 0.1$, $\omega = 2$



(a) $\widetilde{\gamma} = 10^{-3}$ (b) $\widetilde{\gamma} = 1$. (c) $\widetilde{\gamma} = 10^{3}$.

Figure: Comparison between k^{th} and k^{obs}, for $N = 16$ and $\delta t = \frac{\Delta T}{2^5}$. The eigenvalues of $A - LC$ are $\{-0.25, -0.5\}$ (top) and $\{-2, -4\}$ (bottom).

We define the efficiency of the algorithm as

$$E = \frac{\tau_s}{N\tau_p}$$

where τ_s, τ_p are the CPU times required to reach a given tolerance Tol by using a sequential and parallel solver, respectively; and N represents the number of available processors.

Complexity analysis (Parareal case)

Theorem

The efficiency of the *Diamond Strategy* satisfies

$$E \leq \frac{\ell^* \tau_{\Delta T}^{\mathcal{F}}}{\tau_{\Delta T}^{\mathcal{F}} + N \tau_{\Delta T}^{\mathcal{G}}} \left(\sum_{\ell=0}^{\ell_{\|}^* - 1} k_\ell \right)^{-1},$$

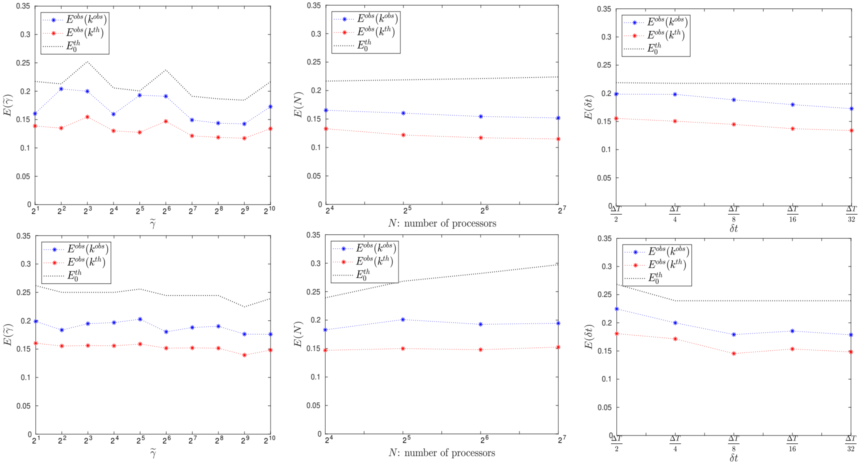
where $\tau_{\Delta T}^{\mathcal{F}}$, $\tau_{\Delta T}^{\mathcal{G}}$ corresponds to the computational time associated to one solving of (2) on a subinterval of length ΔT, for $\mathcal{F}$ and $\mathcal{G}$ respectively; and ℓ^*, $\ell_{\|}$ denotes the number of windows required to reach a given tolerance Tol by using a sequential or parallel solver.

Theorem

The efficiency of the *Diamond Strategy* satisfies

$$E \leq E_0^{th} = \ell^* \left(\sum_{\ell=0}^{\ell_{\parallel}^* - 1} k_\ell \right)^{-1},$$

where $\tau_{\Delta T}^{\mathcal{F}}$, $\tau_{\Delta T}^{\mathcal{G}}$ corresponds to the computational time associated to one solving of (2) on a subinterval of length ΔT, for $\mathcal{F}$ and $\mathcal{G}$ respectively; and ℓ^*, $\ell_{\parallel}$ denotes the number of windows required to reach a given tolerance Tol by using a sequential or parallel solver.



(a) $E(\widetilde{\gamma})$, for $N = 16$ and $\delta t = \frac{\Delta T}{2^5}$.

(b) $E(N)$, for $\delta t = \frac{\Delta T}{2^5}$ and $\widetilde{\gamma} = 2^{10}$.

(c) $E(\delta t)$, for $N = 16$ and $\widetilde{\gamma} = 2^{10}$.

Figure: Comparison between $E^{obs}(k^{obs})$, $E^{obs}(k^{th})$ and E_0^{th}. The eigenvalues of $A - LC$ are $\{-0.25, -0.5\}$ (top) and $\{-2, -4\}$ (bottom).

Perspectives

- Extension to nonlinear observers, Kalman filters.
- Use of other time-parallelization algorithms (e.g. ParaExp).
- Aplication to space-time problems.

Thank you for your attention !