Paraopt algorithm & Runge-Kutta methods

Norbert D. Tognon
djahou-norbert.tognon@inria.fr
Supervisor : Julien Salomon
INRIA Paris & Sorbonne University (France)

Joint work with Félix Kwok, Laval University (Canada)

21st Copper Mountain Conference On Multigrid Methods

16-20 April 2023, Copper Mountain (USA)

Table of contents

Optimal control problem

- Let us consider the following Cauchy problem on $[0, T]$,

$$\dot{y}(t) - f\left(y(t)\right) = u(t),$$
$$y(0) = y_i.$$

- $y, u \in \mathbb{R}^r$.
- y_i: the initial state,
- y_{tg}: the target state.

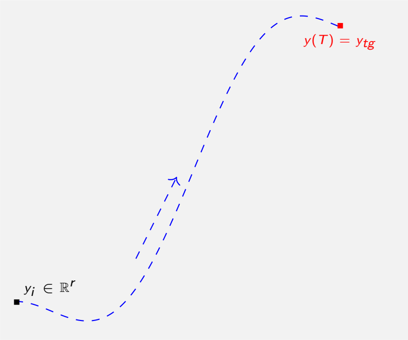
Assumption

We assume that this equation is controllable, .i.e, the application $u \longmapsto y(T)$ is subjective.

Optimal control problem

Find the optimal control u such that

$$y(T) = y_{tg}.$$



Optimal control problem

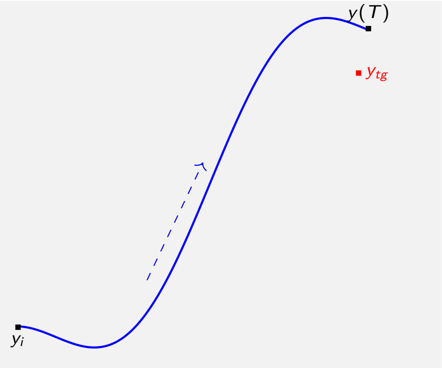
- The optimal control problem becomes the following optimization problem.

Minimization problem

$$\min_u \mathfrak{J}(u) := \frac{1}{2}\|y(T) - y_{tg}\|^2 + \frac{\alpha}{2}\int_0^T \|u\|^2(t)dt,$$

subject to

$$\dot{y}(t) - f(y(t)) = u(t), t \in [0, T]$$
$$y(0) = y_i.$$



Optimality system

- Using an adjoint variable λ, the Lagrange operator becomes

$$\mathfrak{L}(u, y, \lambda) = \mathcal{J}(u) - \int_0^T (\dot{y} - f(y) - u)^T \lambda dt.$$

- Taking $\nabla \mathfrak{L} = 0$, we get the optimality system

$$\begin{cases} \dot{y} - f(y) = u \\ y(0) = y_i, \end{cases} \qquad \begin{cases} \dot{\lambda} + [f'(y)]^T \lambda = 0 \\ \lambda(T) = y(T) - y_{tg}, \end{cases}$$

$$\alpha u = -\lambda.$$

Reduced optimality system

$$\begin{cases} \dot{y}(t) = f(y(t)) - \frac{1}{\alpha}\lambda(t), \\ \dot{\lambda}(t) = -[f'(y(t))]^T \lambda(t), \end{cases} \tag{1}$$

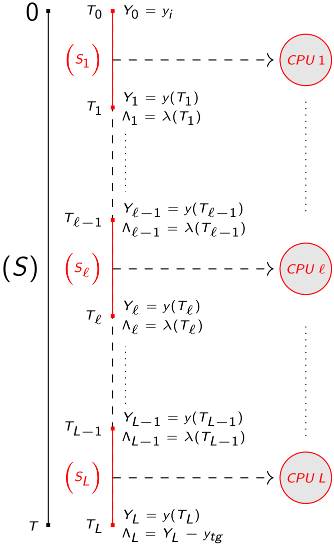
with the initial and final conditions $y(0) = y_i$ and $\lambda(T) = y(T) - y_{tg}$ respectively.

Multiple shooting

- Partition of $[0, T]$ into $[T_{\ell-1}, T_\ell]$, $T_0 = 0, T_\ell = \ell\Delta T$ and $\ell = 1, \ldots, L$.
- Subproblem notation on $[T_{\ell-1}, T_\ell]$

$$(\mathcal{S}_\ell) : \begin{cases} \dot{y}_\ell = f(y_\ell) - \frac{1}{\alpha}\lambda_\ell \\ \dot{\lambda}_\ell = -\left[f'(y_\ell)\right]^T \lambda_\ell, \end{cases}$$

with initial and final conditions $y_\ell(T_{\ell-1}) = Y_{\ell-1}$ and $\lambda_\ell(T_\ell) = \Lambda_\ell$ respectively.



Multiple shooting

- Introduce the solution operators to solve the subproblems.

$$\begin{aligned} y(T_\ell) =& P(Y_{\ell-1}, \Lambda_\ell) \\ \lambda(T_{\ell-1}) =& Q(Y_{\ell-1}, \Lambda_\ell). \end{aligned}$$

- The solutions must match at interfaces, which lead to the equations

$$\left.\begin{array}{rcl} Y_0 - y_i &=& 0 \\ Y_1 - P(Y_0, \Lambda_1) &=& 0 \\ Y_2 - P(Y_1, \Lambda_2) &=& 0 \\ & \vdots & \\ Y_L - P(Y_{L-1}, \Lambda_L) &=& 0 \end{array}\right| \begin{array}{rcl} && \\ \Lambda_1 - Q(Y_1, \Lambda_2) &=& 0 \\ \Lambda_2 - Q(Y_2, \Lambda_3) &=& 0 \\ & \vdots & \\ \Lambda_L - Y_L + y_{tg} &=& 0. \end{array}$$

Nonlinear system

- Collecting the unknowns in the vector X we obtain the nonlinear system

$$\mathcal{F}(X) := \begin{pmatrix} Y_0 - y_i \\ Y_1 - P(Y_0, \Lambda_1) \\ Y_2 - P(Y_1, \Lambda_2) \\ \vdots \\ Y_L - P(Y_{L-1}, \Lambda_L) \\ \Lambda_1 - Q(Y_1, \Lambda_2) \\ \Lambda_2 - Q(Y_2, \Lambda_3) \\ \vdots \\ \Lambda_L - Y_L + y_{tg} \end{pmatrix} = 0, \quad X := \begin{pmatrix} Y_0 \\ \vdots \\ \vdots \\ Y_L \\ \hline \Lambda_1 \\ \vdots \\ \vdots \\ \Lambda_L \end{pmatrix}. \tag{2}$$

Newton Method

Newton method

$$\mathcal{F}'\left(X^k\right)\left(X^{k+1}-X^k\right)=-\mathcal{F}\left(X^k\right),$$

where

$$\mathcal{F}'(X) = \left(\begin{array}{ccccc|cccc}
I & & & & & & & & \\
-P_y(Y_0,\Lambda_1) & I & & & & -P_\lambda(Y_0,\Lambda_1) & & & \\
 & -P_y(Y_1,\Lambda_2) & I & & & & P_\lambda(Y_1,\Lambda_2) & & \\
 & & \ddots & \ddots & & & & \ddots & \\
 & & & -P_y(Y_{L-1},\Lambda_L) & I & & & & -P_\lambda(Y_{L-1},\Lambda_L) \\
\hline
 & -Q_y(Y_1,\Lambda_2) & & & & I & -Q_\lambda(Y_1,\Lambda_2) & & \\
 & & \ddots & & & & \ddots & \ddots & \\
 & & & -Q_y(Y_{L-1},\Lambda_L) & & & & I & -Q_\lambda(Y_{L-1},\Lambda_L) \\
 & & & & -I & & & & I
\end{array}\right).$$

Jacobian approximation and Parareal Idea

- Like in Parareal the remaining expensive fine grid $P(Y_{\ell-1}, \Lambda_\ell)$ and $Q(Y_{\ell-1}, \Lambda_\ell)$ can now all be performed in parallel.
- Derivative Parareal idea [Gander & Hairer 2014]:

$$\begin{aligned} P_y(Y^k_{\ell-1}, \Lambda^k_\ell)(Y^{k+1}_{\ell-1} - Y^k_{\ell-1}) &\approx P^G_y(Y^k_{\ell-1}, \Lambda^k_\ell)(Y^{k+1}_{\ell-1} - Y^k_{\ell-1}), \\ P_\lambda(Y^k_{\ell-1}, \Lambda^k_\ell)(\Lambda^{k+1}_\ell - \Lambda^k_\ell) &\approx P^G_\lambda(Y^k_{\ell-1}, \Lambda^k_\ell)(\Lambda^{k+1}_\ell - \Lambda^k_\ell), \\ Q_\lambda(Y^k_{\ell-1}, \Lambda^k_\ell)(\Lambda^{k+1}_\ell - \Lambda^k_\ell) &\approx Q^G_\lambda(Y^k_{\ell-1}, \Lambda^k_\ell)(\Lambda^{k+1}_\ell - \Lambda^k_\ell), \\ Q_y(Y^k_{\ell-1}, \Lambda^k_\ell)(Y^{k+1}_{\ell-1} - Y^k_{\ell-1}) &\approx Q^G_y(Y^k_{\ell-1}, \Lambda^k_\ell)(Y^{k+1}_{\ell-1} - Y^k_{\ell-1}). \end{aligned}$$

$\longrightarrow$ ***Derivative Parareal***: *M. Gander and E. Hairer, Analysis for parareal algorithms applied to Hamiltonian differential equations, 2014.*

Paraopt algorithm

- Application to the Jacobian $\mathcal{F}'$ by solving linear subproblems in parallel using a coarse solver.
- $\mathcal{J}^G$ is the coarse approximation of $\mathcal{F}'$.

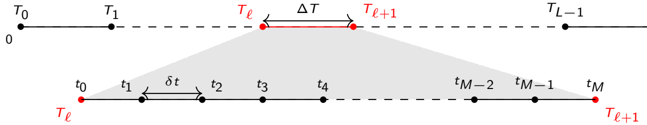
Paraopt algorithm

$$\mathcal{J}^G\left(X^k\right)\left(X^{k+1}-X^k\right)=-\mathcal{F}\left(X^k\right).$$

$\longrightarrow$ ***ParaOpt*** *: M. Gander, F. Kwok, J. Salomon, ParaOpt: A Parareal algorithm for optimal control systems, 2020.*

Discrete cost functional

- Let $M_0, M \in \mathbb{N}$, $\delta t = T/M_0$, $\Delta T = M\delta t$, $M_0 = ML$ and $t_n = n\delta t, n = 0, \dots, M_0$.



- We consider a quadrature formula (d, c). Discrete cost functional :

$$\mathfrak{J}_{\delta t}(u) = \frac{1}{2}\|y_{M_0} - y_{tg}\|^2 + \frac{\alpha}{2}\delta t \sum_{n=0}^{M_0-1} \sum_{j=1}^{s} d_j \|u_{n,j}\|^2, \tag{3}$$

with $y_{M_0} \approx y(T)$ and $u_{n,j} = u(t_n + c_j \delta t)$.

- Runge-Kutta method (A, b^T, c) for the linear dynamic

$$g_i = f\left(y_n + \delta t \sum_{j=1}^{s} a_{i,j} g_j\right) + u_{n,i}, \quad i = 1, \dots, s,$$

$$y_{n+1} = y_n + \delta t \sum_{j=1}^{s} b_j g_j, \; y_n \approx y(t_n).$$

Discrete constraint

- Matrix notation
$$y_{n+1} = y_n + \delta t(b_1 I, \dots, b_s I)(g_1^T, \dots, g_s^T)^T.$$
- We consider the linear dynamic given by
$$\dot{y}(t) - \mathcal{L}y(t) = u(t).$$
- The stage approximations g_i satisfy
$$\begin{pmatrix} g_1 \\ \vdots \\ g_s \end{pmatrix} = (I - \delta t A \otimes \mathcal{L})^{-1} \begin{pmatrix} \mathcal{L}y_n + u_{n,1} \\ \vdots \\ \mathcal{L}y_n + u_{n,s} \end{pmatrix}.$$
- Setting $(W_1, W_2, \dots, W_s) = (b_1 I, \dots, b_s I)(I - \delta t A \otimes \mathcal{L})^{-1}$
$$y_{n+1} = y_n + \delta t(W_1, W_2, \dots, W_s) \begin{pmatrix} \mathcal{L}y_n + u_{n,1} \\ \vdots \\ \mathcal{L}y_n + u_{n,s} \end{pmatrix}.$$
- The discrete constraint:
$$y_{n+1} = (I + \delta t W \mathcal{L})\, y_n + \delta t \sum_{j=1}^{s} W_j u_{n,j}, \quad W = \sum_{i=1}^{s} W_i. \tag{4}$$

Discrete optimality system

- The optimality system

$$\begin{aligned} y_0 &= y_i \\ y_{n+1} &= (I + \delta t W\mathcal{L}) y_n + \delta t \sum_{j=1}^{s} W_j u_{n,j} \\ \lambda_n &= (I + \delta t W\mathcal{L})^T \lambda_{n+1} \\ \lambda_{M_0} &= y_{M_0} - y_{tg} \\ \alpha d_j u_{n,j} &= -W_j^T \lambda_{n+1}. \end{aligned}$$

- Reduced optimality system

$$\begin{aligned} y_{n+1} &= (I + \delta t W\mathcal{L}) y_n - \frac{\delta t}{\alpha} \left(\sum_{j=1}^{s} \frac{1}{d_j} W_j W_j^T \right) \lambda_{n+1} \\ \lambda_n &= (I + \delta t W\mathcal{L})^T \lambda_{n+1}, \end{aligned}$$

with initial and final conditions $y_0 = y_i$ and $\lambda_{M_0} = y_{M_0} - y_{tg}$ respectively.

Discrete optimality system

- Eliminating the interior unknowns (y_n) and (λ_n) on each subinterval $[T_\ell, T_{\ell+1}]$

$$
\left.\begin{array}{rcl}
Y_0 - y_i & = & 0 \\
Y_1 - \mathcal{S}_{\delta t} Y_0 + \frac{1}{\alpha}\mathcal{R}_{\delta t}\Lambda_1 & = & 0 \\
Y_2 - \mathcal{S}_{\delta t} Y_1 + \frac{1}{\alpha}\mathcal{R}_{\delta t}\Lambda_2 & = & 0 \\
 & \vdots & \\
Y_L - \mathcal{S}_{\delta t} Y_{L-1} + \frac{1}{\alpha}\mathcal{R}_{\delta t}\Lambda_L & = & 0
\end{array}\right|
\begin{array}{rl}
 & \\
\Lambda_1 - \mathcal{S}_{\delta t}^T \Lambda_2 & = 0 \\
 & \vdots \\
\Lambda_{L-1} - \mathcal{S}_{\delta t}^T \Lambda_L & = 0 \\
\Lambda_L - Y_L + y_{tg} & = 0
\end{array} \quad ,
$$

where

$$
\mathcal{S}_{\delta t} := (I + \delta t W \mathcal{L})^M,
$$

$$
\mathcal{R}_{\delta t} := \delta t \sum_{n=0}^{M-1} (I + \delta t W \mathcal{L})^n \left(\sum_{j=1}^{s} \frac{1}{d_j} W_j W_j^T \right) [(I + \delta t W \mathcal{L})^T]^n.
$$

Discrete optimality system

- The function $\mathcal{F}_{\delta t}(X) := \mathcal{M}_{\delta t} X - \mathbf{f} = 0$;

$$
\mathcal{M}_{\delta t} = \left(\begin{array}{cccc|cccc}
I & & & & 0 & & & \\
-\mathcal{S}_{\delta t} & \ddots & & & \mathcal{R}_{\delta t}/\alpha & \ddots & & \\
 & \ddots & \ddots & & & \ddots & & 0 \\
 & & -\mathcal{S}_{\delta t} & I & & & & \mathcal{R}_{\delta t}/\alpha \\
\hline
 & & & & I & -\mathcal{S}_{\delta t}^T & & \\
 & & & & & \ddots & \ddots & \\
 & & & & & & \ddots & -\mathcal{S}_{\delta t}^T \\
 & & & -I & & & & I
\end{array}\right), \quad \mathbf{f} = \begin{pmatrix} y_i \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ -y_{tg} \end{pmatrix}.
$$

Discrete formulation

- We introduce the coarse time step $\Delta t = \Delta T/N$ with

$$\delta t \leq \Delta t \leq \Delta T.$$

- The Paraopt algorithm becomes the following iteration

$$\mathcal{M}_{\Delta t}\left(X^{k+1} - X^k\right) = -\left(\mathcal{M}_{\delta t} X^k - \mathbf{f}\right), \tag{5}$$

  or

$$X^{k+1} = \mathcal{M}_{\Delta t}^{-1}(\mathcal{M}_{\Delta t} - \mathcal{M}_{\delta t})X^k + \mathcal{M}_{\Delta t}^{-1}\mathbf{f}.$$

- How do behave the convergence factor of the iteration matrix $\mathcal{M}_{\Delta t}^{-1}(\mathcal{M}_{\Delta t} - \mathcal{M}_{\delta t})$?

Convergence analysis

- The Dahlquist problem case where $\mathcal{L} \in \mathbb{R}^-$.
- Backward Euler method: $(A, b^T, c) = (1, 1, 1)$.
- Quadrature formula: backward Euler method $(d, c) = (1, 1)$.

Theorem (Gander et al)

Let $\Delta T, \Delta t, \delta t$ and α be fixed. Then for all $\mathcal{L} < 0$, the spectral radius of $\mathcal{M}_{\Delta t}^{-1}(\mathcal{M}_{\Delta t} - \mathcal{M}_{\delta t})$ satisfies

$$\max_{\mathcal{L}<0} \rho(\mathcal{L}) \leq \frac{0.79\Delta t}{\alpha + \sqrt{\alpha \Delta t}} + 0.3.$$

Thus, if $\alpha > 0.4544\Delta t$, then the linear Paraopt algorithm converges.

Convergence analysis

Definition

- Let $\mathcal{E}_i, i \in \{0,1\}$ be the finite sequence set

$$\mathcal{E}_i := \{Y = (Y_\ell)_{\ell=i,\dots,L} : Y_\ell \in \mathbb{R}^r \text{ and } \|Y\|^2_{\Delta T} = \Delta T \sum_{\ell=i}^{L} \|Y_\ell\|^2 < \infty\}.$$

- For $X = \begin{pmatrix} Y \\ \Lambda \end{pmatrix} \in \mathcal{E}_0 \times \mathcal{E}_1$,

$$\|X\|^2_* := \|Y\|^2_{\Delta T} + \alpha^{-2}\|\Lambda\|^2_{\Delta T} = \Delta T \left(\sum_{\ell=0}^{L} \|Y_\ell\|^2 + \alpha^{-2} \sum_{\ell=1}^{L} \|\Lambda_\ell\|^2 \right).$$

- The induced matrix norm

$$\|\mathcal{M}_{\delta t}\|_* = \inf\{\kappa;\ \|\mathcal{M}_{\delta t} X\|_* \le \kappa \|X\|_*, X \in \mathcal{E}_0 \times \mathcal{E}_1\}.$$

Stability condition

Assumption 1

We assume that the Runge-Kutta method satisfies the stability condition

$$\|I + \Delta t W \mathcal{L}\| < 1.$$

- Let $\{\nu_j, j = 1, \dots, r\}$ be the spectrum of $\mathcal{L}$ and F the stability function of the Runge-Kutta method.
- The assumption 1 means that

$$\{\Delta t.\nu_j, j = 1, \dots, r\} \subset \{z \in \mathbb{C}; |F(z)| < 1\}.$$

Convergence results

Lemma (Kwok, Salomon, T)

Let the integers p, q be given, $k = \min\{p, q\}$ and the assumption 1 holds. We assume that the Runge-Kutta method and the quadrature formula are of order p and q respectively. Then there exist $c_{\mathcal{S}} > 0$ and $c_{\mathcal{R}} > 0$ independent on δt and Δt such that

$$\|\mathcal{S}_{\Delta t} - \mathcal{S}_{\delta t}\| \leq c_{\mathcal{S}}(\Delta t - \delta t)\Delta t^{p-1} \text{ and } \|\mathcal{R}_{\Delta t} - \mathcal{R}_{\delta t}\| \leq c_{\mathcal{R}}(\Delta t - \delta t)\Delta t^{k-1}.$$

Theorem (Kwok, Salomon, T)

Let the integers p, q be given, $k = \min\{p, q\}$ and the assumption 1 holds. We assume that the Runge-Kutta method and the quadrature formula are of order p and q respectively. Then there exists $c_{\mathcal{M}} > 0$ independent on δt and Δt such that

$$\|\mathcal{M}_{\Delta t} - \mathcal{M}_{\delta t}\|_* \leq c_{\mathcal{M}}(\Delta t - \delta t)\Delta t^{k-1}.$$

Convergence results

Theorem (Kwok, Salomon, T)

Let us assume that the assumption 1 holds. Then there exists $c_{\mathcal{M}^{-1}} > 0$ independent on Δt such that

$$\|\mathcal{M}_{\Delta t}^{-1}\|_* \leq \frac{c_{\mathcal{M}^{-1}}}{\Delta T}(1+\alpha^{-1}).$$

- $\rho \leq \|\mathcal{M}_{\Delta t}^{-1}\left(\mathcal{M}_{\Delta t} - \mathcal{M}_{\delta t}\right)\|_*$.

Spectral radius

$$\rho \leq c_0(\Delta t - \delta t)\Delta t^{k-1},$$

$c_0 = \frac{1}{\Delta T} c_{\mathcal{M}} c_{\mathcal{M}^{-1}}(1+\alpha^{-1})$.

Numerical results

- We consider the heat equation in one dimension

$$\begin{aligned} \partial_t y(x,t) - \partial_x^2 y(x,t) =& u(x,t), \;\; 0 \le t \le T \\ y(x,0) =& y_0(x), \;\; 0 \le x \le 1, \\ y(0,t) =& y(1,t) = 0. \end{aligned}$$

- A semi-discretization in space of this equation gives

$$\begin{aligned} \partial_t y(t) =& \mathcal{L} y(t) + u(t), t \in [0,T] \\ y(0) =& y_0 \end{aligned}$$

where $y = (y_n)_{n=1,\ldots,r}$ and $u = (u_n)_{n=1,\ldots,r}$ and

$$\mathcal{L} = -\frac{1}{\delta x^2} \mathrm{tridiag}(-1,2,-1).$$

Test 1

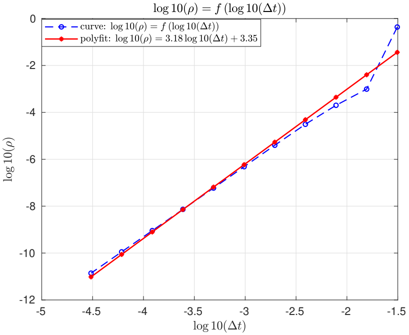
- The spectral radius ρ of the iteration matrix $\mathcal{M}_{\Delta t}^{-1}(\mathcal{M}_{\Delta t} - \mathcal{M}_{\delta t})$ satisfies

$$\rho \leq c_0 \Delta t^k.$$

- Singly Diagonal Implicit Runge-Kutta (SDIRK) of order 3, $\gamma = \frac{3-\sqrt{3}}{6}$

$$A = \begin{pmatrix} \gamma & 0 \\ 1-2\gamma & \gamma \end{pmatrix} \quad b = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \quad c = \begin{pmatrix} \gamma \\ 1-\gamma \end{pmatrix}$$

- The stability condition is satisfied for $\delta t, \Delta t < \frac{1}{2}\left(3 + 2\sqrt{3}\right)\delta x^2$.
- Problem parameters: $T = 10$, $\alpha = 1$.
- Discretization parameters: $L = 20$, $r = 10$, $\delta x = 0.1$, $\Delta T = 0.5$, $\delta t = \Delta T/2^{16}$, $\Delta t = \Delta T/2^n$, $n = 4, \ldots, 14$.
- We plot the log ρ on y-axis and log Δt on x-axis.



Test 2

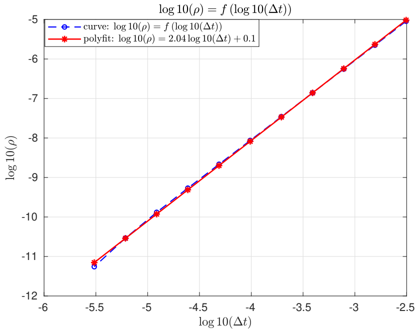
- The spectral radius ρ of the iteration matrix $\mathcal{M}_{\Delta t}^{-1}(\mathcal{M}_{\Delta t} - \mathcal{M}_{\delta t})$ satisfies

$$\rho \leq c_0 \Delta t^k.$$

- Heun method of order 2,

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad b = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \quad c = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

- The stability condition is satisfied for $\delta t, \Delta t < \frac{\delta x^2}{2}$.
- Optimal control problem parameters: $T = 10$, $\alpha = 1$.
- Discretization parameters: $L = 20$, $r = 4$, $\delta x = 0.25$, $\Delta T = 0.5$, $\delta t = \Delta T/(10 \times 2^{16})$, $\Delta t = \Delta T/(10 \times 2^n)$, $n = 4, \dots, 14$.
- We plot the $\log \rho$ on y-axis and $\log \Delta t$ on x-axis.



Thanks

Thank you!

References

M. Gander, S. Vandewalle
Analysis of the parareal time-parallel time-integration method,
SIAM, 2007.

E Hairer, M. Gander,
Analysis for parareal algorithms applied to Hamiltonian differential equations,
Computational and Applied Mathematics, 2014.

M. Gander, J. Salomon, F. Kwok.
Paraopt: A Parareal algorithm for optimality system,
SIAM, 2020.