Recent advances on fully well-balanced methods: high-order accuracy, hydrodynamic reconstruction and hybridization with machine learning

Victor Michel-Dansac[*], joint work with
Christophe Berthon[†], Solène Bulteau[‡], Emmanuel Franck[*],
Françoise Foucher[†§], Meissa M'Baye[‖], Laurent Navoret[*]

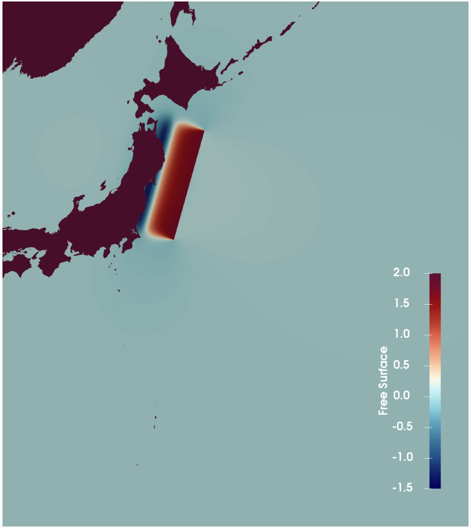
May 02, 2024

Séminaire Calcul Scientifique et Modélisation, Bordeaux

[*] Université de Strasbourg, CNRS, Inria, IRMA, France
[†] LMJL, Université de Nantes, France
[‡] ICAM La Roche-sur-Yon, France
[§] École Centrale de Nantes, France
[‖] LMDAN, Université Cheikh Anta Diop, Dakar, Senegal

Tsunami simulation: naive numerical method

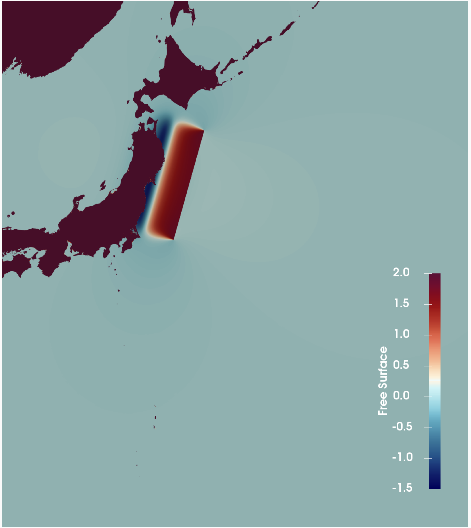
Tsunami initialization



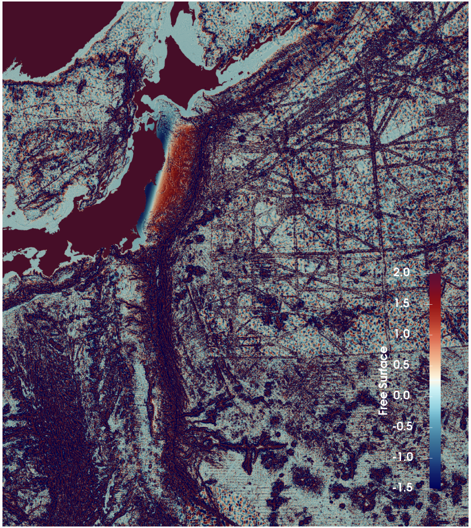
Simulation with a naive numerical method

Tsunami simulation: naive numerical method

Tsunami initialization



Simulation with a naive numerical method



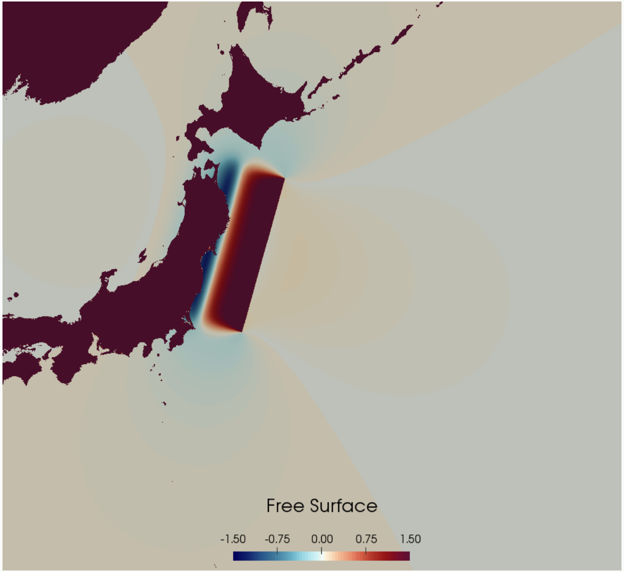
Tsunami simulation: failure

⇝ **The simulation is not usable!**

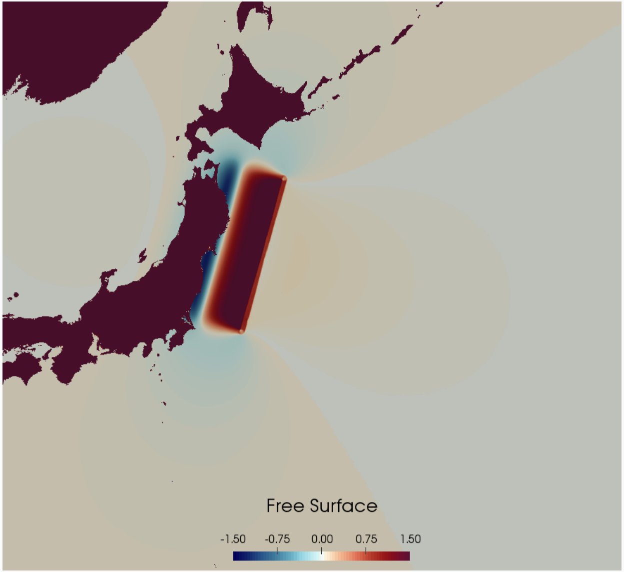
Indeed, the ocean at rest, far from the tsunami, started spontaneously producing waves.

This comes from the non-preservation of stationary solutions, hence the need to develop numerical methods that **preserve stationary solutions**: so-called **well-balanced** methods.

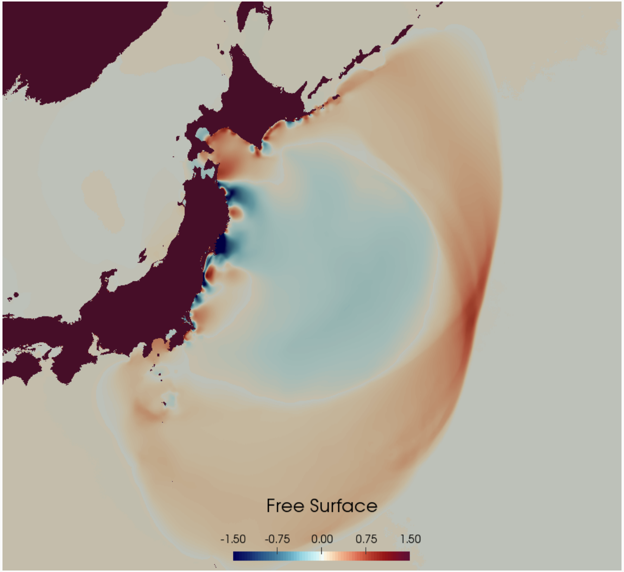
Tsunami simulation: well-balanced method



Tsunami simulation: well-balanced method



Tsunami simulation: well-balanced method



Generic systems of balance laws

We consider a generic **system of balance laws**:

$$\partial_t W + \partial_x F(W) = S(W, x), \quad x \in \mathbb{R}, \quad t > 0,$$

where:

- W is the vector of unknowns,
- F is the physical flux function,
- S is the source term.

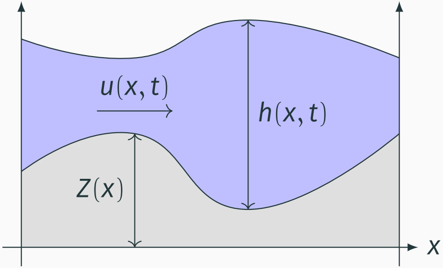
We assume that the homogeneous system is **hyperbolic**.

Example: the shallow water equations with topography.

The shallow water equations with topography

$$\begin{cases} \partial_t h + \partial_x q = 0, \\ \partial_t q + \partial_x \left(\dfrac{q^2}{h} + \dfrac{1}{2} g h^2 \right) = -gh\partial_x Z(x) \end{cases}$$

The equations are written under the form $\partial_t W + \partial_x F(W) = S(W)$.



- $h(x,t)$: water height
- $u(x,t)$: water velocity
- $q = hu$: water discharge
- $Z(x)$: known topography
- g: gravity constant

We pay particular attention to solutions of prime importance: the **steady solutions**.

Steady solutions and well-balanced schemes

Definition: steady solution

W is a steady solution of $\partial_t W + \partial_x F(W) = S(W,x)$ if, and only if, $\partial_t W = 0$, i.e. W satisfies the following ODE:

$$\partial_x F(W) = S(W,x).$$

Example: For the shallow water equations with topography, the ODE governing smooth steady solutions can be simplified.

Steady solutions and well-balanced schemes

Definition: steady solution

W is a steady solution of $\partial_t W + \partial_x F(W) = S(W, x)$ if, and only if, $\partial_t W = 0$, i.e. W satisfies the following ODE:

$$\partial_x F(W) = S(W, x).$$

Example: For the shallow water equations with topography, the ODE governing smooth steady solutions can be simplified.

Definition: well-balanced scheme

A numerical method approximating the solution of a balance law is called well-balanced if it exactly preserves the steady solutions.

Shallow water equations: steady solutions

Taking $\partial_t W = 0$ in the shallow water system yields

$$\begin{cases} \partial_x q = 0, \\ \partial_x \left(\dfrac{q^2}{h} + \dfrac{1}{2} g h^2 \right) = -g h \partial_x Z, \end{cases} \quad \xRightarrow[\text{solution}]{\text{smooth}} \quad \begin{cases} q = \text{cst} = q_0, \\ \partial_x \left(\dfrac{q_0^2}{2h^2} + g(h + Z) \right) = 0. \end{cases}$$

We summarize the second relation by introducing a function B such that, for a steady solution, $B(h, q_0, Z) = B_0$.

Shallow water equations: steady solutions

Taking $\partial_t W = 0$ in the shallow water system yields

$$\begin{cases} \partial_x q = 0, \\ \partial_x \left(\dfrac{q^2}{h} + \dfrac{1}{2} g h^2 \right) = -g h \partial_x Z, \end{cases} \quad \overset{\text{smooth}}{\underset{\text{solution}}{\Longrightarrow}} \quad \begin{cases} q = \text{cst} = q_0, \\ \partial_x \left(\dfrac{q_0^2}{2h^2} + g(h+Z) \right) = 0. \end{cases}$$

We summarize the second relation by introducing a function B such that, for a steady solution, $B(h, q_0, Z) = B_0$.

Two cases are distinguished:

- $\mathbf{q = 0} \rightsquigarrow$ lake at rest
  we get $B(h, q_0, Z) = h + Z = B_0$: linear equation in h
- $\mathbf{q \neq 0} \rightsquigarrow$ moving steady solution
  we get $B(h, q_0, Z) = \dfrac{q_0^2}{2h^2} + g(h+Z) = B_0$: nonlinear equation in h!

General objectives

Main objectives of this work: develop numerical schemes that

- are high-order accurate and fully well-balanced (WB) (e.g. preserving equilibria with nonzero velocity),
- can be applied to generic hyperbolic systems of balance laws,
- have a low computational cost, mostly ensured by avoiding costly inversions of nonlinear systems.

We present **three strategies**:

1. build a high-order WB extension of a given first-order WB scheme,
2. build a first-order WB scheme for the shallow water equations that uses an arbitrary consistent numerical flux function,
3. enhance a Discontinuous Galerkin scheme using a prior on the steady solution.

Motivation and general objectives

1/ Extending a first-order well-balanced scheme to high-order accuracy

Berthon, Bulteau, Foucher, M'Baye, M.-D., SIAM SISC, 2022

Issues of classical approaches

Well-balanced correction

2/ Making any consistent numerical flux fully WB for the shallow water equations

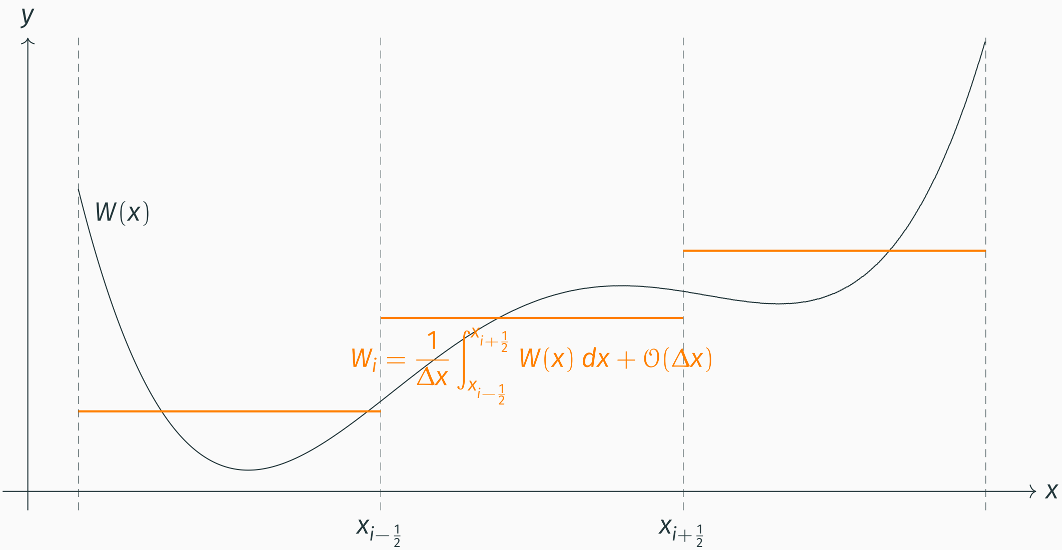
Berthon, M.-D., J. Numer. Math., 2024

3/ Enhancing Discontinuous Galerkin bases with a prior

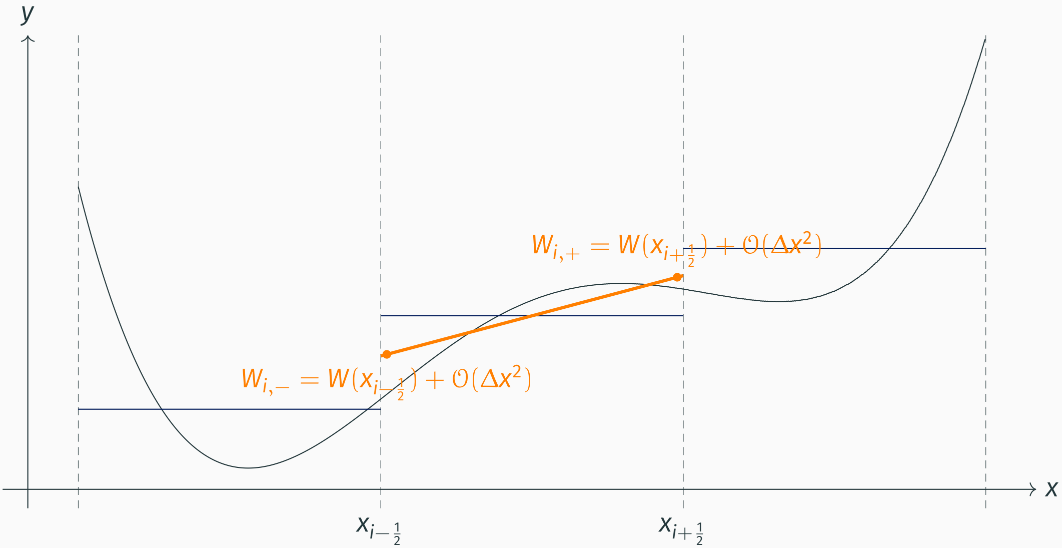
Franck, M.-D., Navoret, in revision, 2024

Conclusion and perspectives

First-order accuracy in space: visualized



Second-order accuracy in space: visualized



Second-order accuracy in space: the reconstruction

The naive approach[1] consists in adding the slope between cells $i+1$ and $i-1$ to the value in cell i, to get:

$$W_{i,\pm}^n = W_i^n \pm \frac{W_{i+1}^n - W_{i-1}^n}{4}.$$

[1] Usually, one uses a slope limiter to avoid spurious oscillations; it is omitted here for clarity, but the forthcoming discussion remains valid when a limiter is involved.

Second-order accuracy in space: the reconstruction

The naive approach[1] consists in adding the slope between cells $i+1$ and $i-1$ to the value in cell i, to get:

$$W_{i,\pm}^n = W_i^n \pm \frac{W_{i+1}^n - W_{i-1}^n}{4}.$$

If the first-order scheme is well-balanced, then the second-order scheme will also be well-balanced if

$$\text{steady solution for } (W_{i-1}^n, W_i^n, W_{i+1}^n) \implies \text{steady solution for } W_{i,\pm}^n.$$

Usual method: Reconstruct other variables, and deduce $W_{i,\pm}^n$ from these reconstructions.

[1] Usually, one uses a slope limiter to avoid spurious oscillations; it is omitted here for clarity, but the forthcoming discussion remains valid when a limiter is involved.

Simple case of the lake at rest: $\mathbf{q} = \mathbf{0}$

The steady solutions are given by $q = 0$ and $h + Z = \text{cst} =: H_0$: we reconstruct these quantities, as well as Z. For instance, we get

$$
\begin{aligned}
(h+Z)^n_{i,\pm} &= (h+Z)^n_i \pm \frac{1}{4}\Big((h+Z)^n_{i+1} - (h+Z)^n_{i-1}\Big), \\
&= H_0 \pm \frac{1}{4}(H_0 - H_0) \text{ for a steady solution,} \\
&= H_0.
\end{aligned}
$$

Application to the shallow water equations: $q = 0$

Simple case of the lake at rest: $\mathbf{q} = \mathbf{0}$

The steady solutions are given by $q = 0$ and $h + Z = \text{cst} =: H_0$: we reconstruct these quantities, as well as Z. For instance, we get

$$\begin{aligned}(h+Z)^n_{i,\pm} &= (h+Z)^n_i \pm \frac{1}{4}\big((h+Z)^n_{i+1} - (h+Z)^n_{i-1}\big), \\ &= H_0 \pm \frac{1}{4}(H_0 - H_0) \text{ for a steady solution,} \\ &= H_0.\end{aligned}$$

The reconstructed height $h^n_{i,\pm}$ is deduced as follows:

$$h^n_{i,\pm} = (h+Z)^n_{i,\pm} - Z^n_{i,\pm} \implies h^n_{i,\pm} + Z^n_{i,\pm} = H_0 \text{ for a steady solution}.$$

The same reasoning applies to the discharge q.

Therefore, the reconstruction coincides with the steady solution, and the second-order scheme preserves the steady states at rest.

Application to the shallow water equations: $q \neq 0$

General case of the moving steady solutions: $\mathbf{q} \neq \mathbf{0}$

The steady solutions are given by $q = \text{cst} \coloneqq q_0$ and $B(h, q_0, Z) = \text{cst} \eqqcolon B_0$: we reconstruct these quantities, as well as Z.

Thus, we get $q_{i,\pm}^n = q_0$ and $B_{i,\pm}^n = B_0$; recall that

$$B(h, q_0, Z) = \frac{q_0^2}{2h^2} + g(h + Z).$$

General case of the moving steady solutions: $\mathbf{q} \neq \mathbf{0}$

The steady solutions are given by $q = \text{cst} \coloneqq q_0$ and $B(h, q_0, Z) = \text{cst} \eqqcolon B_0$: we reconstruct these quantities, as well as Z.

Thus, we get $q_{i,\pm}^n = q_0$ and $B_{i,\pm}^n = B_0$; recall that

$$B(h, q_0, Z) = \frac{q_0^2}{2h^2} + g(h + Z).$$

The reconstructed height $h_{i,\pm}^n$ should be deduced as follows:

$$\frac{(q_{i,\pm}^n)^2}{2(h_{i,\pm}^n)^2} + g(h_{i,\pm}^n + Z_{i,\pm}^n) = B_{i,\pm}^n : \quad \text{this is a nonlinear equation in } h_{i,\pm}^n!$$

Therefore, for the reconstruction to coincide with the steady solution, two nonlinear equations must be solved in each cell and at each time step! This leads to a very heavy computational cost.

Motivation and general objectives

1/ Extending a first-order well-balanced scheme to high-order accuracy

Berthon, Bulteau, Foucher, M'Baye, M.-D., SIAM SISC, 2022

Issues of classical approaches

Well-balanced correction

2/ Making any consistent numerical flux fully WB for the shallow water equations

Berthon, M.-D., J. Numer. Math., 2024

3/ Enhancing Discontinuous Galerkin bases with a prior

Franck, M.-D., Navoret, in revision, 2024

Conclusion and perspectives

Main idea behind the well-balanced correction

A remark on accuracy and well-balancing

A well-balanced scheme is *exact* on steady solutions; therefore, it is *more accurate than any high-order scheme.* Well-balanced schemes are *formally of infinite order* on steady solutions.

Main idea behind the well-balanced correction

A remark on accuracy and well-balancing

A well-balanced scheme is *exact* on steady solutions; therefore, it is *more accurate than any high-order scheme*. Well-balanced schemes are *formally of infinite order* on steady solutions.

Main idea behind the well-balanced correction:

- if the solution is steady, use the first-order well-balanced scheme, since it is exact;
- otherwise, use a high-order scheme.

This idea is implemented by introducing a convex combination.

This convex combination concerns both the source term and the reconstruction; in this talk, **we focus on the reconstruction**.

Very easy well-balanced correction: the reconstruction

Recall that the naive reconstruction is given by

$$W_{i,-}^n = W_i^n - \frac{W_{i+1}^n - W_{i-1}^n}{4} \quad \text{and} \quad W_{i,+}^n = W_i^n + \frac{W_{i+1}^n - W_{i-1}^n}{4}.$$

$$\left.\begin{array}{r}\text{well-balanced first-order scheme, and}\\ \text{(steady solution)} \implies (W_{i,\pm}^n = W_i^n)\end{array}\right\} \implies \begin{array}{c}\text{steady solutions preserved by}\\ \text{the second-order scheme}\end{array}$$

How to modify the reconstruction close to a steady solution?

Very easy well-balanced correction: the reconstruction

Recall that the naive reconstruction is given by

$$W_{i,-}^n = W_i^n - \frac{W_{i+1}^n - W_{i-1}^n}{4} \qquad \text{and} \qquad W_{i,+}^n = W_i^n + \frac{W_{i+1}^n - W_{i-1}^n}{4}.$$

$$\left.\begin{array}{c}\text{well-balanced first-order scheme, and}\\ \text{(steady solution)} \implies (W_{i,\pm}^n = W_i^n)\end{array}\right\} \implies \begin{array}{c}\text{steady solutions preserved by}\\ \text{the second-order scheme}\end{array}$$

How to modify the reconstruction close to a steady solution?

We introduce the following modification of the reconstruction:

$$\overline{W}_{i,-}^n = W_i^n - \theta_{i-\frac{1}{2}}^n \frac{W_{i+1}^n - W_{i-1}^n}{4} \qquad \text{and} \qquad \overline{W}_{i,+}^n = W_i^n + \theta_{i+\frac{1}{2}}^n \frac{W_{i+1}^n - W_{i-1}^n}{4},$$

where $\theta_{i\pm\frac{1}{2}}^n$ is a steady solution indicator (defined on the next slide):

- $\theta_{i\pm\frac{1}{2}}^n \simeq 1$ far away from a steady solution;
- $\theta_{i\pm\frac{1}{2}}^n = 0$ when a steady solution is reached.

An expression of $\theta^n_{i+1/2}$

The convex combination parameter $\theta^n_{i+\frac{1}{2}}$ must therefore satisfy the following properties:

- vanish when the pair (W_i^n, W_{i+1}^n) defines a steady state;
- be an approximation of 1 up to $\mathcal{O}(\Delta x^2)$ otherwise.

An expression of $\theta^n_{i+1/2}$

The convex combination parameter $\theta^n_{i+\frac{1}{2}}$ must therefore satisfy the following properties:

- vanish when the pair (W_i^n, W_{i+1}^n) defines a steady state;
- be an approximation of 1 up to $\mathcal{O}(\Delta x^2)$ otherwise.

Define G such that

$$(W_i^n, W_{i+1}^n) \text{ is a steady solution} \quad \Longleftrightarrow \quad G(W_i^n, x_i) = G(W_{i+1}^n, x_{i+1}).$$

We propose the following expression, with C some scaling parameter:

$$\theta^n_{i+\frac{1}{2}} = \frac{\varepsilon^n_{i+\frac{1}{2}}}{\varepsilon^n_{i+\frac{1}{2}} + C\Delta x^2}, \qquad \text{with} \qquad \varepsilon^n_{i+\frac{1}{2}} = \frac{1}{\Delta x}\left\| G(W_{i+1}^n, x_{i+1}) - G(W_i^n, x_i) \right\|,$$

Extended definition of $\theta^n_{i+1/2}$

To handle a scheme of order δ and non-unique equilibria given by L functions[2] G_ℓ, we propose the following expression of $\theta^n_{i+\frac{1}{2}}$:

$$\theta^n_{i+\frac{1}{2}} = \frac{\varepsilon^n_{i+\frac{1}{2}}}{\varepsilon^n_{i+\frac{1}{2}} + C\Delta x^\delta}, \qquad \text{with} \qquad \varepsilon^n_{i+\frac{1}{2}} = \prod_{\ell=1}^{L} \frac{1}{\Delta x} \left\| G_\ell(W^n_{i+1}, x_{i+1}) - G_\ell(W^n_i, x_i) \right\|.$$

We have made two changes to the steady solution detector:

- the exponent in $\theta^n_{i+\frac{1}{2}}$,
- the expression of $\varepsilon^n_{i+\frac{1}{2}}$.

Next step: Build a first-order well-balanced scheme; we consider the example of the shallow water equations.

[2] For instance, for the Euler system with gravity, we get $L = 3$.

Motivation and general objectives

1/ Extending a first-order well-balanced scheme to high-order accuracy

Berthon, Bulteau, Foucher, M'Baye, M.-D., SIAM SISC, 2022

2/ Making any consistent numerical flux fully WB for the shallow water equations

Berthon, M.-D., J. Numer. Math., 2024

- The hydrodynamic reconstruction
- Suitable expression of $\mathcal{H}$
- Numerical experiments

3/ Enhancing Discontinuous Galerkin bases with a prior

Franck, M.-D., Navoret, in revision, 2024

Conclusion and perspectives

Finite volume scheme

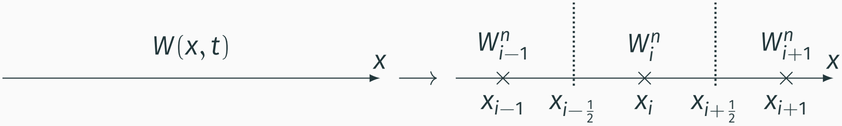
Recall the compact form of the shallow water equations:

$$\partial_t W + \partial_x F(W) = S(W).$$

We take a generic finite volume numerical scheme approximating the shallow water equations:

$$\frac{W_i^{n+1} - W_i^n}{\Delta t} + \frac{1}{\Delta x}\left[\mathcal{F}(W_i^n, W_{i+1}^n) - \mathcal{F}(W_{i-1}^n, W_i^n)\right] = \mathcal{S}(W_{i-1}^n, W_i^n, W_{i+1}^n),$$

with $\mathcal{F}$ a consistent numerical flux, i.e. $\mathcal{F}(W, W) = F(W)$, and $\mathcal{S}$ a consistent numerical source term.



Finite volume scheme

Recall the compact form of the shallow water equations:

$$\partial_t W + \partial_x F(W) = S(W).$$

We take a generic finite volume numerical scheme approximating the shallow water equations:

$$\frac{W_i^{n+1} - W_i^n}{\Delta t} + \frac{1}{\Delta x}\left[\mathcal{F}(W_i^n, W_{i+1}^n) - \mathcal{F}(W_{i-1}^n, W_i^n)\right] = \mathcal{S}(W_{i-1}^n, W_i^n, W_{i+1}^n),$$

with $\mathcal{F}$ a consistent numerical flux, i.e. $\mathcal{F}(W, W) = F(W)$, and $\mathcal{S}$ a consistent numerical source term.

Question: can we make this generic finite volume scheme well-balanced without changing the numerical flux?

Lake at rest preservation: the hydrostatic reconstruction

The **hydrostatic reconstruction** was introduced in E. Audusse et al., *SIAM J. Sci. Comput.* (2004), as a way to make it possible for any finite volume scheme to capture the **lake at rest** steady solution.

It relies on:

1. providing a relevant expression for $\mathcal{S}$,
2. evaluating the numerical flux at a specific reconstruction of W.

$$\frac{W_i^{n+1} - W_i^n}{\Delta t} + \frac{1}{\Delta x}\left[\mathcal{F}\left(W_i^n , W_{i+1}^n \right) - \mathcal{F}\left(W_{i-1}^n , W_i^n \right)\right] = \mathcal{S}_i^n$$

Lake at rest preservation: the hydrostatic reconstruction

The **hydrostatic reconstruction** was introduced in E. Audusse et al., *SIAM J. Sci. Comput.* (2004), as a way to make it possible for any finite volume scheme to capture the **lake at rest** steady solution.

It relies on:

1. providing a relevant expression for $\mathcal{S}$,
2. evaluating the numerical flux at a specific reconstruction of W.

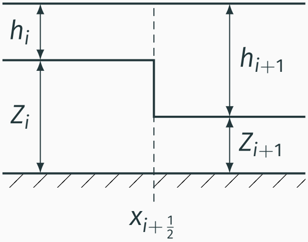
$$\frac{W_i^{n+1} - W_i^n}{\Delta t} + \frac{1}{\Delta x}\left[\mathcal{F}\left(W^n_{i+\frac{1}{2},-}, W^n_{i+\frac{1}{2},+}\right) - \mathcal{F}\left(W^n_{i-\frac{1}{2},-}, W^n_{i-\frac{1}{2},+}\right)\right] = \mathcal{S}_i^n$$

Lake at rest preservation: the hydrostatic reconstruction

The **hydrostatic reconstruction** was introduced in E. Audusse et al., *SIAM J. Sci. Comput.* (2004), as a way to make it possible for any finite volume scheme to capture the **lake at rest** steady solution.

It relies on:

1. providing a relevant expression for $\mathcal{S}$,
2. evaluating the numerical flux at a specific reconstruction of W.

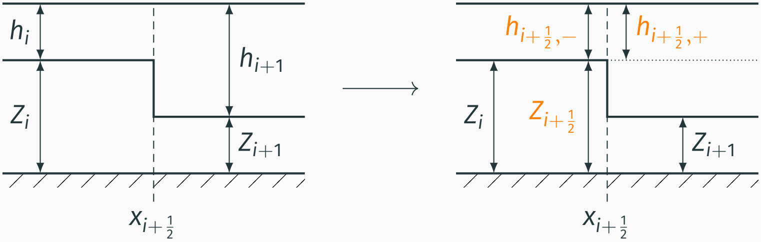


Lake at rest preservation: the hydrostatic reconstruction

The **hydrostatic reconstruction** was introduced in E. Audusse et al., *SIAM J. Sci. Comput.* (2004), as a way to make it possible for any finite volume scheme to capture the **lake at rest** steady solution.

It relies on:

1. providing a relevant expression for $\mathcal{S}$,
2. evaluating the numerical flux at a specific reconstruction of W.



Objectives

Main goal of this work: Provide a reconstruction able to capture the steady solutions with $q_0 = 0$ or $q_0 \neq 0$, **without solving nonlinear equations**.

The objectives of our **hydrodynamic reconstruction** include:

- making sure that the resulting scheme is consistent,
- ensuring the capture of steady solutions with $q_0 = 0$ or $q_0 \neq 0$,
- handling dry areas and transitions between wet and dry areas (not presented in this talk).

Motivation and general objectives

1/ Extending a first-order well-balanced scheme to high-order accuracy

Berthon, Bulteau, Foucher, M'Baye, M.-D., SIAM SISC, 2022

2/ Making any consistent numerical flux fully WB for the shallow water equations

Berthon, M.-D., J. Numer. Math., 2024

The hydrodynamic reconstruction

Suitable expression of $\mathcal{H}$

Numerical experiments

3/ Enhancing Discontinuous Galerkin bases with a prior

Franck, M.-D., Navoret, in revision, 2024

Conclusion and perspectives

Expression of the hydrodynamic reconstruction

Away from dry areas, the hydrostatic reconstruction reads:

$$h^n_{i+\frac{1}{2},-} = h^n_i + \left(Z_i - Z_{i+\frac{1}{2}}\right),$$

$$h^n_{i+\frac{1}{2},+} = h^n_{i+1} + \left(Z_{i+1} - Z_{i+\frac{1}{2}}\right).$$

Expression of the hydrodynamic reconstruction

Away from dry areas, the hydrodynamic reconstruction reads:

$$
\begin{aligned}
h^n_{i+\frac{1}{2},-} = h^n_i &+ \left(Z_i - Z_{i+\frac{1}{2}}\right) \\
&+ 2\mathrm{Fr}^2\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i\right) \mathcal{H}\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i, Z_{i+\frac{1}{2}} - Z_i\right), \\
h^n_{i+\frac{1}{2},+} = h^n_{i+1} &+ \left(Z_{i+1} - Z_{i+\frac{1}{2}}\right) \\
&+ 2\mathrm{Fr}^2\left(h^n_{i+1}, h^n_{i+\frac{1}{2}}, q^n_{i+1}\right) \mathcal{H}\left(h^n_{i+1}, h^n_{i+\frac{1}{2}}, q^n_{i+1}, Z_{i+\frac{1}{2}} - Z_{i+1}\right),
\end{aligned}
$$

with $\mathcal{H}$ a function of h_L, h_R, $\bar{q}$ and $\Delta Z \coloneqq Z_R - Z_L$, and with

$$
\mathrm{Fr}^2(h_L, h_R, \bar{q}) = \frac{\bar{q}^2(h_L + h_R)}{2g h_L^2 h_R^2}.
$$

Expression of the hydrodynamic reconstruction

Away from dry areas, the hydrodynamic reconstruction reads:

$$
\begin{aligned}
h^n_{i+\frac{1}{2},-} &= h^n_i + \left(Z_i - Z_{i+\frac{1}{2}}\right) \\
&\quad + 2\mathrm{Fr}^2\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i\right) \mathcal{H}\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i, Z_{i+\frac{1}{2}} - Z_i\right), \\
h^n_{i+\frac{1}{2},+} &= h^n_{i+1} + \left(Z_{i+1} - Z_{i+\frac{1}{2}}\right) \\
&\quad + 2\mathrm{Fr}^2\left(h^n_{i+1}, h^n_{i+\frac{1}{2}}, q^n_{i+1}\right) \mathcal{H}\left(h^n_{i+1}, h^n_{i+\frac{1}{2}}, q^n_{i+1}, Z_{i+\frac{1}{2}} - Z_{i+1}\right),
\end{aligned}
$$

with $\mathcal{H}$ a function of h_L, h_R, $\bar{q}$ and $\Delta Z \coloneqq Z_R - Z_L$, and with

$$
\mathrm{Fr}^2(h_L, h_R, \bar{q}) = \frac{\bar{q}^2(h_L + h_R)}{2g h_L^2 h_R^2}.
$$

The hydrodynamic reconstruction relies on deriving a suitable function $\mathcal{H}$.

For instance, for consistency, $\mathcal{H}$ should vanish when ΔZ does.

Characterization of interface steady relations

$$h^n_{i+\frac{1}{2},-} = h^n_i + \left(Z_i - Z_{i+\frac{1}{2}}\right) + 2\mathrm{Fr}^2\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i\right)\mathcal{H}\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i, Z_{i+\frac{1}{2}} - Z_i\right)$$

Define the interface state by

$$(h^n_{i+\frac{1}{2}}, Z_{i+\frac{1}{2}}) = \begin{cases} (h^n_i, Z_i) & \text{if } Z_i > Z_{i+1}, \\ (h^n_{i+1}, Z_{i+1}) & \text{otherwise.} \end{cases}$$

The relations $h^n_{i+\frac{1}{2},-} = h^n_{i+\frac{1}{2}} = h^n_{i+\frac{1}{2},+}$ have to hold for steady solutions.

Characterization of interface steady relations

$$h^n_{i+\frac{1}{2},-} = h^n_i + \left(Z_i - Z_{i+\frac{1}{2}}\right)$$
$$+ 2\mathrm{Fr}^2 \left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i\right) \mathcal{H}\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i, Z_{i+\frac{1}{2}} - Z_i\right)$$

Define the interface state by

$$(h^n_{i+\frac{1}{2}}, Z_{i+\frac{1}{2}}) = \begin{cases} (h^n_i, Z_i) & \text{if } Z_i > Z_{i+1}, \\ (h^n_{i+1}, Z_{i+1}) & \text{otherwise.} \end{cases}$$

The relations $h^n_{i+\frac{1}{2},-} = h^n_{i+\frac{1}{2}} = h^n_{i+\frac{1}{2},+}$ **have to hold for steady solutions.**

When the solution is steady, setting $\bar{q} = q_i = q_{i+1}$, we get:

$$B(h_i, \bar{q}, Z_i) = B(h_{i+1}, \bar{q}, Z_{i+1}).$$

Characterization of interface steady relations

$$h^n_{i+\frac{1}{2},-} = h^n_i + \left(Z_i - Z_{i+\frac{1}{2}}\right) + 2\mathrm{Fr}^2 \left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i\right) \mathcal{H}\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i, Z_{i+\frac{1}{2}} - Z_i\right)$$

Define the interface state by

$$(h^n_{i+\frac{1}{2}}, Z_{i+\frac{1}{2}}) = \begin{cases} (h^n_i, Z_i) & \text{if } Z_i > Z_{i+1}, \\ (h^n_{i+1}, Z_{i+1}) & \text{otherwise.} \end{cases}$$

The relations $h^n_{i+\frac{1}{2},-} = h^n_{i+\frac{1}{2}} = h^n_{i+\frac{1}{2},+}$ have to hold for steady solutions.

When the solution is steady, setting $\bar{q} = q_i = q_{i+1}$, we get:

$$B(h_i, \bar{q}, Z_i) = B(h_{i+\frac{1}{2}}, \bar{q}, Z_{i+\frac{1}{2}}) = B(h_{i+1}, \bar{q}, Z_{i+1}).$$

Characterization of interface steady relations

$$h^n_{i+\frac{1}{2},-} = h^n_i + \left(Z_i - Z_{i+\frac{1}{2}}\right) + 2\mathrm{Fr}^2\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i\right) \mathcal{H}\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i, Z_{i+\frac{1}{2}} - Z_i\right)$$

Define the interface state by

$$(h^n_{i+\frac{1}{2}}, Z_{i+\frac{1}{2}}) = \begin{cases} (h^n_i, Z_i) & \text{if } Z_i > Z_{i+1}, \\ (h^n_{i+1}, Z_{i+1}) & \text{otherwise.} \end{cases}$$

The relations $h^n_{i+\frac{1}{2},-} = h^n_{i+\frac{1}{2}} = h^n_{i+\frac{1}{2},+}$ have to hold for steady solutions.

When the solution is steady, setting $\bar{q} = q_i = q_{i+1}$, we get:

$$\frac{\bar{q}^2}{2h_i^2} + g(h_i + Z_i) = \frac{\bar{q}^2}{2h_{i+\frac{1}{2}}^2} + g(h_{i+\frac{1}{2}} + Z_{i+\frac{1}{2}}) = \frac{\bar{q}^2}{2h_{i+1}^2} + g(h_{i+1} + Z_{i+1}).$$

Well-balancing requirement on $\mathcal{H}$ – statement

Some algebraic manipulations allow us to write

$$\frac{\bar{q}^2}{2h_i^2} + g(h_i + Z_i) = \frac{\bar{q}^2}{2h_{i+\frac{1}{2}}^2} + g(h_{i+\frac{1}{2}} + Z_{i+\frac{1}{2}})$$

$$\Longleftrightarrow$$

$$Z_{i+\frac{1}{2}} - Z_i = -\left(h_{i+\frac{1}{2}} - h_i\right)\left(1 - \mathrm{Fr}^2\left(h_i, h_{i+\frac{1}{2}}, \bar{q}\right)\right),$$

which is nothing but the usual discrete characterization of smooth steady solutions.

Well-balancing requirement on $\mathcal{H}$ – statement

Some algebraic manipulations allow us to write

$$\frac{\bar{q}^2}{2h_i^2} + g(h_i + Z_i) = \frac{\bar{q}^2}{2h_{i+\frac12}^2} + g(h_{i+\frac12} + Z_{i+\frac12})$$

$$\Longleftrightarrow$$

$$Z_{i+\frac12} - Z_i = -\left(h_{i+\frac12} - h_i\right)\left(1 - \mathrm{Fr}^2\left(h_i, h_{i+\frac12}, \bar{q}\right)\right),$$

which is nothing but the usual discrete characterization of smooth steady solutions.

We claim that imposing the following property on $\mathcal{H}$ will be enough to preserve steady solutions:

$$\Delta Z = -(h_R - h_L)(1 - \mathrm{Fr}^2(h_L, h_R, \bar{q})) \implies \mathcal{H}(h_L, h_R, \bar{q}, \Delta Z) = \frac{h_R - h_L}{2}.$$

Indeed, assuming that the solution is steady, we obtain the following identities:

$$h^n_{i+\frac{1}{2},-} = h^n_i + \left(Z_i - Z_{i+\frac{1}{2}}\right) + 2\mathrm{Fr}^2\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i\right) \mathcal{H}\left(h^n_i, h^n_{i+\frac{1}{2}}, q^n_i, Z_{i+\frac{1}{2}} - Z_i\right),$$

Well-balancing requirement on $\mathcal{H}$ – proof

Indeed, assuming that the solution is steady, we obtain the following identities:

$$
\begin{aligned}
h_{i+\frac{1}{2},-}^n = h_i^n &+ \left(Z_i - Z_{i+\frac{1}{2}}\right) \\
&+ 2\mathrm{Fr}^2\left(h_i^n, h_{i+\frac{1}{2}}^n, q_i^n\right) \mathcal{H}\left(h_i^n, h_{i+\frac{1}{2}}^n, q_i^n, Z_{i+\frac{1}{2}} - Z_i\right), \\
h_{i+\frac{1}{2},-}^n = h_i^n &+ \left(Z_i - Z_{i+\frac{1}{2}}\right) \\
&+ \mathrm{Fr}^2\left(h_i^n, h_{i+\frac{1}{2}}^n, q_i^n\right) \left(h_{i+\frac{1}{2}}^n - h_i^n\right),
\end{aligned}
$$

Well-balancing requirement on $\mathcal{H}$ – proof

Indeed, assuming that the solution is steady, we obtain the following identities:

$$
\begin{aligned}
h^n_{i+\frac12,-} &= h^n_i + \left(Z_i - Z_{i+\frac12}\right) \\
&\quad + 2\mathrm{Fr}^2\left(h^n_i, h^n_{i+\frac12}, q^n_i\right) \mathcal{H}\left(h^n_i, h^n_{i+\frac12}, q^n_i, Z_{i+\frac12} - Z_i\right), \\
h^n_{i+\frac12,-} &= h^n_i + \left(Z_i - Z_{i+\frac12}\right) \\
&\quad + \mathrm{Fr}^2\left(h^n_i, h^n_{i+\frac12}, q^n_i\right) \left(h^n_{i+\frac12} - h^n_i\right), \\
h^n_{i+\frac12,-} &= h^n_i + \left(Z_i - Z_{i+\frac12}\right) \\
&\quad + \left(Z_{i+\frac12} - Z_i\right) + \left(h^n_{i+\frac12} - h^n_i\right),
\end{aligned}
$$

Well-balancing requirement on $\mathcal{H}$ – proof

Indeed, assuming that the solution is steady, we obtain the following identities:

$$h^n_{i+\frac12,-} = h^n_i + \left(Z_i - Z_{i+\frac12}\right) + 2\mathrm{Fr}^2\left(h^n_i, h^n_{i+\frac12}, q^n_i\right)\mathcal{H}\left(h^n_i, h^n_{i+\frac12}, q^n_i, Z_{i+\frac12} - Z_i\right),$$

$$h^n_{i+\frac12,-} = h^n_i + \left(Z_i - Z_{i+\frac12}\right) + \mathrm{Fr}^2\left(h^n_i, h^n_{i+\frac12}, q^n_i\right)\left(h^n_{i+\frac12} - h^n_i\right),$$

$$h^n_{i+\frac12,-} = h^n_i + \left(Z_i - Z_{i+\frac12}\right) + \left(Z_{i+\frac12} - Z_i\right) + \left(h^n_{i+\frac12} - h^n_i\right),$$

$$h^n_{i+\frac12,-} = h^n_{i+\frac12},$$

which proves that the scheme is well-balanced.

Summary and source term discretization

To summarize, for the reconstruction to be consistent and well-balanced, we require the **following two properties** on the bounded function $\mathcal{H}$:

1. $\mathcal{H}(h_L, h_R, \bar{q}, \Delta Z) \underset{\Delta Z \to 0}{=} \mathcal{O}(\Delta Z)$,
2. $\Delta Z = -(h_R - h_L)(1 - \mathrm{Fr}^2(h_L, h_R, \bar{q})) \implies \mathcal{H}(h_L, h_R, \bar{q}, \Delta Z) = \dfrac{h_R - h_L}{2}$.

Summary and source term discretization

To summarize, for the reconstruction to be consistent and well-balanced, we require the **following two properties** on the bounded function $\mathcal{H}$:

1. $\mathcal{H}(h_L, h_R, \bar{q}, \Delta Z) \underset{\Delta Z \to 0}{=} \mathcal{O}(\Delta Z)$,
2. $\Delta Z = -(h_R - h_L)(1 - \mathrm{Fr}^2(h_L, h_R, \bar{q})) \implies \mathcal{H}(h_L, h_R, \bar{q}, \Delta Z) = \dfrac{h_R - h_L}{2}$.

In addition, the whole scheme will also be consistent and well-balanced if the following **numerical source term** is used:

$$\Delta x (\mathcal{S}_q)_i^n = -g \frac{2 h^n_{i-\frac{1}{2},+} h^n_{i+\frac{1}{2},-}}{h^n_{i-\frac{1}{2},+} + h^n_{i+\frac{1}{2},-}} \left(Z_{i+\frac{1}{2}} - Z_{i-\frac{1}{2}} \right) + \frac{4g}{h^n_{i-\frac{1}{2},+} + h^n_{i+\frac{1}{2},-}} \mathcal{H}\left(h^n_{i-\frac{1}{2},+}, h^n_{i+\frac{1}{2},-}, q_i, Z_{i+\frac{1}{2}} - Z_{i-\frac{1}{2}} \right)^3.$$

The proof results from algebraic manipulations (not detailed here).

Summary and source term discretization

To summarize, for the reconstruction to be consistent and well-balanced, we require the **following two properties** on the bounded function $\mathcal{H}$:

1. $\mathcal{H}(h_L, h_R, \bar{q}, \Delta Z) \underset{\Delta Z \to 0}{=} \mathcal{O}(\Delta Z)$,
2. $\Delta Z = -(h_R - h_L)(1 - \mathrm{Fr}^2(h_L, h_R, \bar{q})) \implies \mathcal{H}(h_L, h_R, \bar{q}, \Delta Z) = \dfrac{h_R - h_L}{2}$.

In addition, the whole scheme will also be consistent and well-balanced if the following **numerical source term** is used:

$$\Delta x (\mathcal{S}_q)_i^n = -g \frac{2h^n_{i-\frac{1}{2},+} h^n_{i+\frac{1}{2},-}}{h^n_{i-\frac{1}{2},+} + h^n_{i+\frac{1}{2},-}} \left(Z_{i+\frac{1}{2}} - Z_{i-\frac{1}{2}} \right) + \frac{4g}{h^n_{i-\frac{1}{2},+} + h^n_{i+\frac{1}{2},-}} \mathcal{H}\left(h^n_{i-\frac{1}{2},+}, h^n_{i+\frac{1}{2},-}, q_i, Z_{i+\frac{1}{2}} - Z_{i-\frac{1}{2}} \right)^3 .$$

The proof results from algebraic manipulations (not detailed here).

Next step: obtain a suitable expression of $\mathcal{H}$.

Satisfying the well-balanced property

Recall that, when $\mathcal{H}$ is applied to a discrete steady solution, we need

$$\mathcal{H}(h_L, h_R, \bar{q}, \Delta Z) = \frac{h_R - h_L}{2}.$$

To obtain an expression of $\mathcal{H}$ satisfying this property, we need to understand how $(h_R - h_L)/2$ behaves for discrete steady solutions.

Satisfying the well-balanced property

Recall that, when $\mathcal{H}$ is applied to a discrete steady solution, we need

$$\mathcal{H}(h_L, h_R, \bar{q}, \Delta Z) = \frac{h_R - h_L}{2}.$$

To obtain an expression of $\mathcal{H}$ satisfying this property, we need to understand how $(h_R - h_L)/2$ behaves for discrete steady solutions.

We now seek a relation to characterize **the jump of h at the interface**, i.e. an expression of $(h_R - h_L)/2$ for steady solutions.

We assume that the solution is steady, and introduce notation

$$\bar{h} \coloneqq \frac{h_L + h_R}{2} \quad \text{and} \quad [h] \coloneqq \frac{h_R - h_L}{2},$$

so that h_L and h_R satisfy

$$h_L = \bar{h} - [h] \quad \text{and} \quad h_R = \bar{h} + [h].$$

The goal is now to rewrite the steady relation in terms of $\bar{h}$ and $[h]$.

A local relation to characterize steady solutions

Recall that the steady solutions are governed by

$$B(h, q_0, Z) = \frac{q_0^2}{2h^2} + g(h + Z) = B_0.$$

That is to say, at the interface between states W_L and W_R, the solution is locally steady if $q_L = q_R = \bar{q}$ and

$$B(h_L, \bar{q}, Z_L) = B(h_R, \bar{q}, Z_R) \iff \frac{\bar{q}^2}{2h_L^2} + g(h_L + Z_L) = \frac{\bar{q}^2}{2h_R^2} + g(h_R + Z_R).$$

We set out to rewrite the above relation using $\bar{h}$ and $[h]$ instead of h_L and h_R.

A nonlinear relation for the interface jump

$$\frac{\bar{q}^2}{2h_L^2} + g(h_L + Z_L) = \frac{\bar{q}^2}{2h_R^2} + g(h_R + Z_R)$$

$$\Longleftrightarrow$$

$$\frac{\bar{q}^2}{2(\bar{h} - [h])^2} + g(\bar{h} - [h] + Z_L) = \frac{\bar{q}^2}{2(\bar{h} + [h])^2} + g(\bar{h} + [h] + Z_R)$$

$$\Longleftrightarrow$$

$$\dots$$

$$\Longleftrightarrow$$

$$2[h]\left(g\left(\bar{h}^2 - [h]^2\right)^2 - \bar{q}^2\bar{h}\right) = -g(Z_R - Z_L)\left(\bar{h}^2 - [h]^2\right)^2.$$

“Quadratized” relation

$$2\mathcal{H}\left(g\left(\bar{h}^2-\mathcal{H}^2\right)^2-\bar{q}^2\bar{h}\right)=-g\Delta Z\left(\bar{h}^2-\mathcal{H}^2\right)^2 \qquad (*)$$

“Quadratized” relation

$$2\mathcal{H}\left(g\left(\bar{h}^2-\mathcal{H}^2\right)^2-\bar{q}^2\bar{h}\right)=-g\Delta Z\left(\bar{h}^2-\mathcal{H}^2\right)^2 \qquad (*)$$

Equation $(*)$ is nonlinear, and using it would incur considerable computational cost. To avoid this issue, we proceed with a linearization-like simplification. First, for $\mathcal{H}\neq\bar{h}$, we get

$$(*) \iff 2\mathcal{H}\left(1-\frac{\bar{q}^2\bar{h}}{g\left(\bar{h}^2-\mathcal{H}^2\right)^2}\right)=-\Delta Z.$$

We then choose a “quadratization” of this expression around $\mathcal{H}=[h]$:

$$2\mathcal{H}\left(1-\frac{\bar{q}^2(h_L+h_R)}{2g(\bar{h}^2-[h]^2)}+a\left([h]-\mathcal{H}\right)\right)=-\Delta Z.$$

"Quadratized" relation

$$2\mathcal{H}\left(g\left(\bar{h}^2 - \mathcal{H}^2\right)^2 - \bar{q}^2\bar{h}\right) = -g\Delta Z\left(\bar{h}^2 - \mathcal{H}^2\right)^2 \qquad (*)$$

Equation $(*)$ is nonlinear, and using it would incur considerable computational cost. To avoid this issue, we proceed with a linearization-like simplification. First, for $\mathcal{H} \neq \bar{h}$, we get

$$(*) \iff 2\mathcal{H}\left(1 - \frac{\bar{q}^2\bar{h}}{g\left(\bar{h}^2 - \mathcal{H}^2\right)^2}\right) = -\Delta Z.$$

We then choose a "quadratization" of this expression around $\mathcal{H} = [h]$:

$$2\mathcal{H}\left(1 - \underbrace{\frac{\bar{q}^2(h_L + h_R)}{2g h_L^2 h_R^2}}_{\mathrm{Fr}^2} + a\left([h] - \mathcal{H}\right)\right) = -\Delta Z.$$

“Quadratized” relation

$$2\mathcal{H}\left(g\left(\bar{h}^2-\mathcal{H}^2\right)^2-\bar{q}^2\bar{h}\right)=-g\Delta Z\left(\bar{h}^2-\mathcal{H}^2\right)^2 \qquad (*)$$

Equation $(*)$ is nonlinear, and using it would incur considerable computational cost. To avoid this issue, we proceed with a linearization-like simplification. First, for $\mathcal{H}\neq\bar{h}$, we get

$$(*) \iff 2\mathcal{H}\left(1-\frac{\bar{q}^2\bar{h}}{g\left(\bar{h}^2-\mathcal{H}^2\right)^2}\right)=-\Delta Z.$$

We then choose a “quadratization” of this expression around $\mathcal{H}=[h]$:

$$2\mathcal{H}\left(1-\underbrace{\frac{\bar{q}^2(h_L+h_R)}{2gh_L^2h_R^2}}_{\mathrm{Fr}^2}+a\left([h]-\mathcal{H}\right)\right)=-\Delta Z.$$

In practice, after some testing, we choose $a=\operatorname{sgn}(\Delta Z)\sqrt{\frac{|\Delta Z|}{2[\![h]\!]^3}}$.

Final expression of $\mathcal{H}$

We are left with $\mathcal{H}$ satisfying a quadratic relation.

Solving this quadratic equation for $\mathcal{H}$ leads to

$$\mathcal{H} = \frac{1}{2}\left(E - \operatorname{sgn}(1 - \mathrm{Fr}^2)\operatorname{sgn}(\Delta Z)\sqrt{E^2 + \sqrt{\frac{1}{2}|\Delta Z|[\![h]\!]^3}}\right),$$

$$\text{with } \; E = [h] + \frac{1 - \mathrm{Fr}^2}{2}\operatorname{sgn}(\Delta Z)\sqrt{\frac{[\![h]\!]^3}{2|\Delta Z|}}.$$

We show that, if ΔZ and $1 - \mathrm{Fr}^2$ do not simultaneously vanish:

1. this expression of $\mathcal{H}$ is well-balanced;
2. this expression of $\mathcal{H}$ is consistent, despite the divisions by ΔZ.

Motivation and general objectives

1/ Extending a first-order well-balanced scheme to high-order accuracy

Berthon, Bulteau, Foucher, M'Baye, M.-D., SIAM SISC, 2022

2/ Making any consistent numerical flux fully WB for the shallow water equations

Berthon, M.-D., J. Numer. Math., 2024

The hydrodynamic reconstruction

Suitable expression of $\mathcal{H}$

Numerical experiments

3/ Enhancing Discontinuous Galerkin bases with a prior

Franck, M.-D., Navoret, in revision, 2024

Conclusion and perspectives

Setup

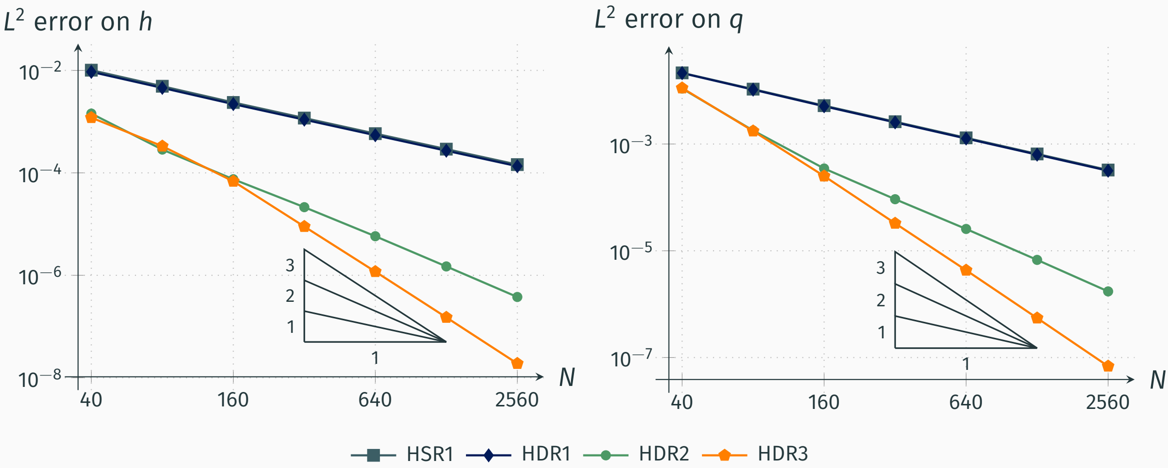
We provide several numerical tests with a finite volume scheme using the HLL flux:

- an order of convergence test,
- three tests of the well-balanced property,
- a dam-break on a dry slope.

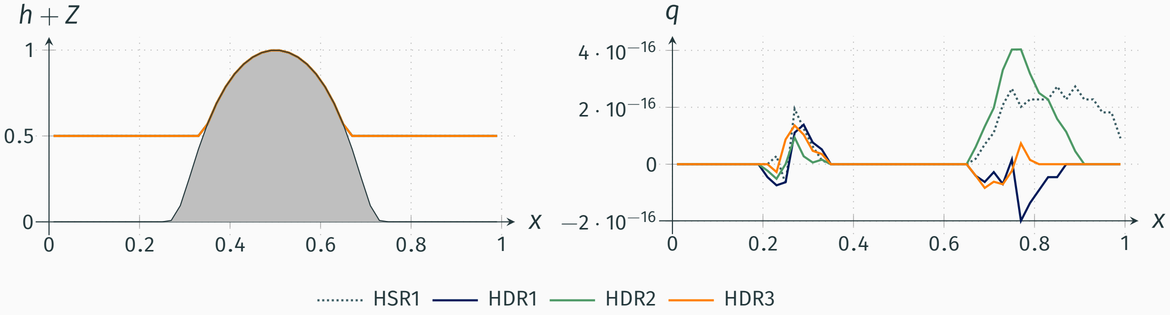
These tests are performed with the **h**ydro**s**tatic **r**econstruction (HSR) and the **h**ydro**d**ynamic **r**econstruction (HDR).

The schemes of order δ are denoted by HSRδ and HDRδ, and they make use of the indicator developed in the previous section.

Order of convergence

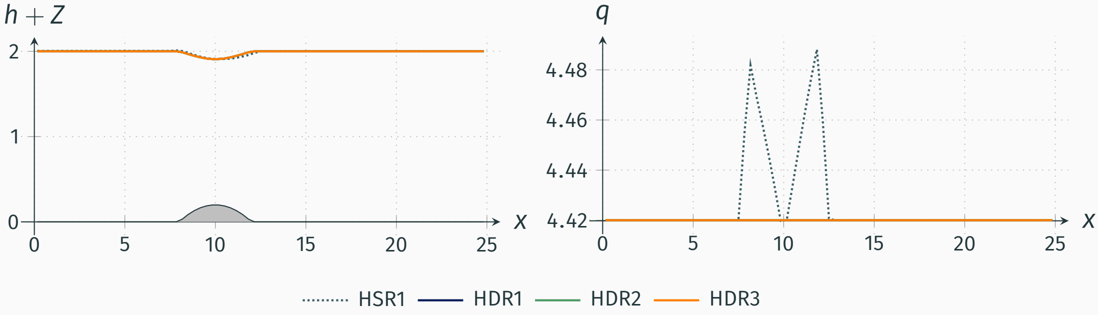


Emerged lake at rest (50 cells)



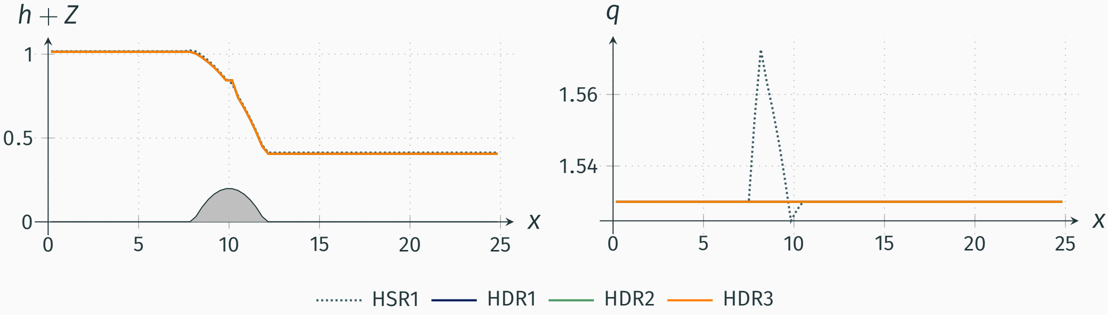
	HSR, $\mathbb{P}_0$	HDR, $\mathbb{P}_0$	HDR, $\mathbb{P}_1$	HDR, $\mathbb{P}_2$
L^2 error on h	$1.85 \cdot 10^{-17}$	$2.75 \cdot 10^{-17}$	$3.07 \cdot 10^{-17}$	$1.32 \cdot 10^{-17}$
L^2 error on q	$1.24 \cdot 10^{-16}$	$5.17 \cdot 10^{-17}$	$1.24 \cdot 10^{-16}$	$3.59 \cdot 10^{-17}$

Subcritical flow (75 cells)



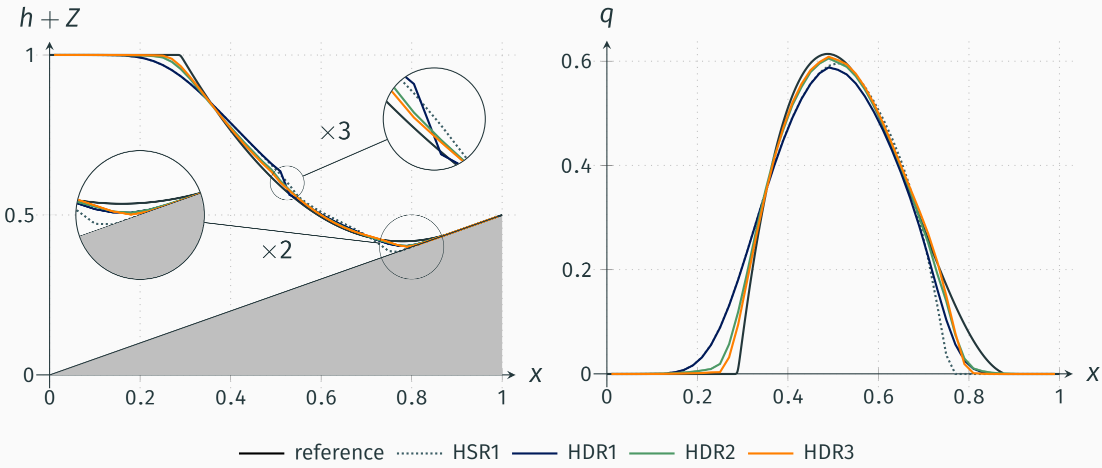
	HSR, $\mathbb{P}_0$	HDR, $\mathbb{P}_0$	HDR, $\mathbb{P}_1$	HDR, $\mathbb{P}_2$
L^2 error on q	$7.73 \cdot 10^{-2}$	$1.06 \cdot 10^{-14}$	$1.31 \cdot 10^{-14}$	$1.30 \cdot 10^{-14}$
L^2 error on B	$1.79 \cdot 10^{-1}$	$2.73 \cdot 10^{-14}$	$3.61 \cdot 10^{-14}$	$2.68 \cdot 10^{-14}$

Transcritical flow (75 cells)



	HSR, $\mathbb{P}_0$	HDR, $\mathbb{P}_0$	HDR, $\mathbb{P}_1$	HDR, $\mathbb{P}_2$
L^2 error on q	$3.74 \cdot 10^{-2}$	$4.73 \cdot 10^{-14}$	$5.15 \cdot 10^{-14}$	$5.21 \cdot 10^{-14}$
L^2 error on B	$1.45 \cdot 10^{-1}$	$4.50 \cdot 10^{-14}$	$5.12 \cdot 10^{-14}$	$5.92 \cdot 10^{-14}$

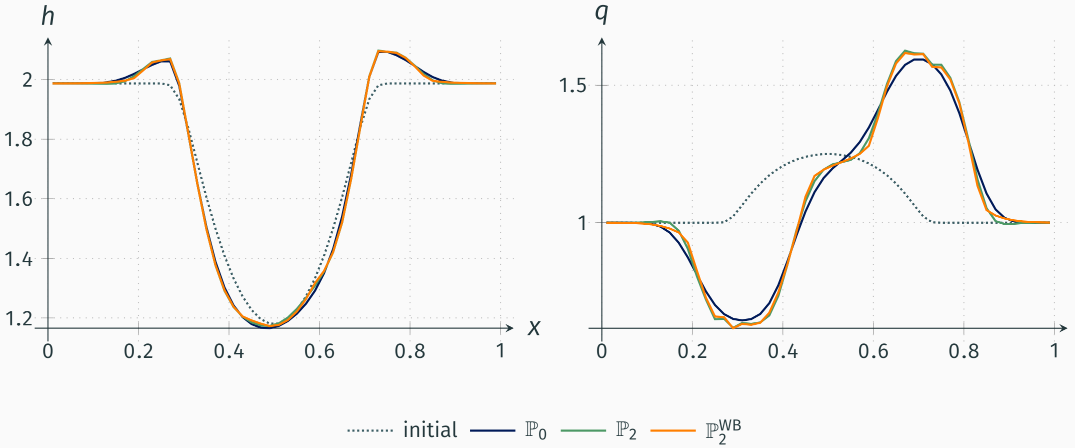
Dam-break on a dry slope (50 cells)



Large perturbation of a steady solution

The initial condition (dotted line) consists in a large perturbation of a steady solution.

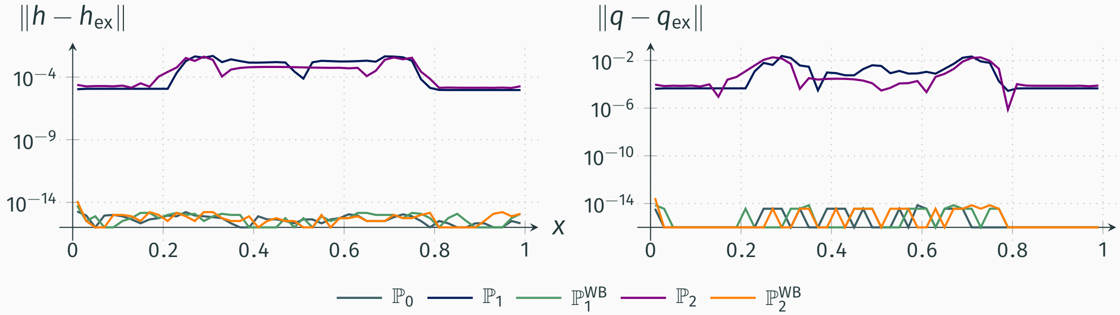
We show the evolution of the perturbation after a short time.



Large perturbation of a steady solution

After a long physical time, the perturbation is dissipated:

- numerical noise remains with the $\mathbb{P}_d$ scheme;
- the unperturbed steady state is exactly recovered with the $\mathbb{P}_d^{\mathrm{WB}}$ scheme.

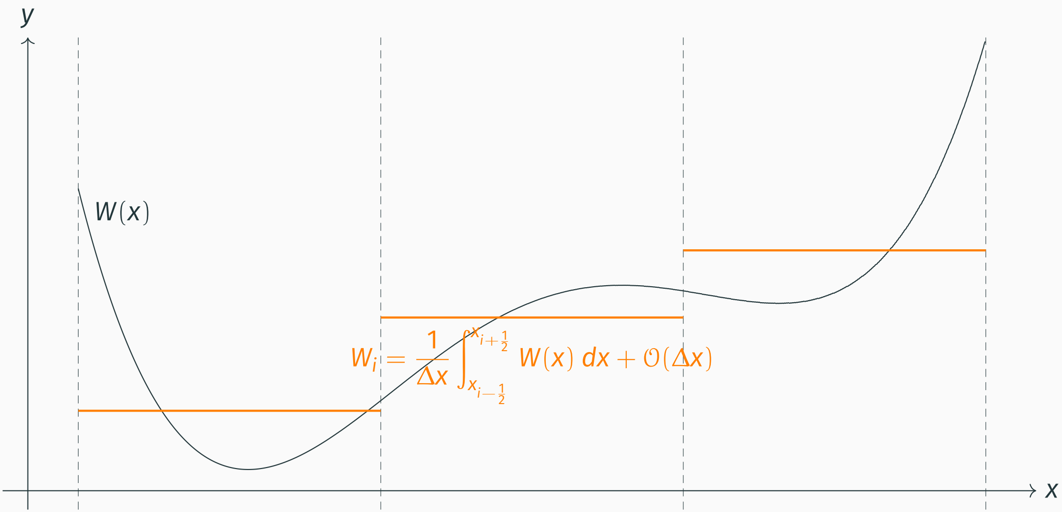


Large perturbation of a steady solution

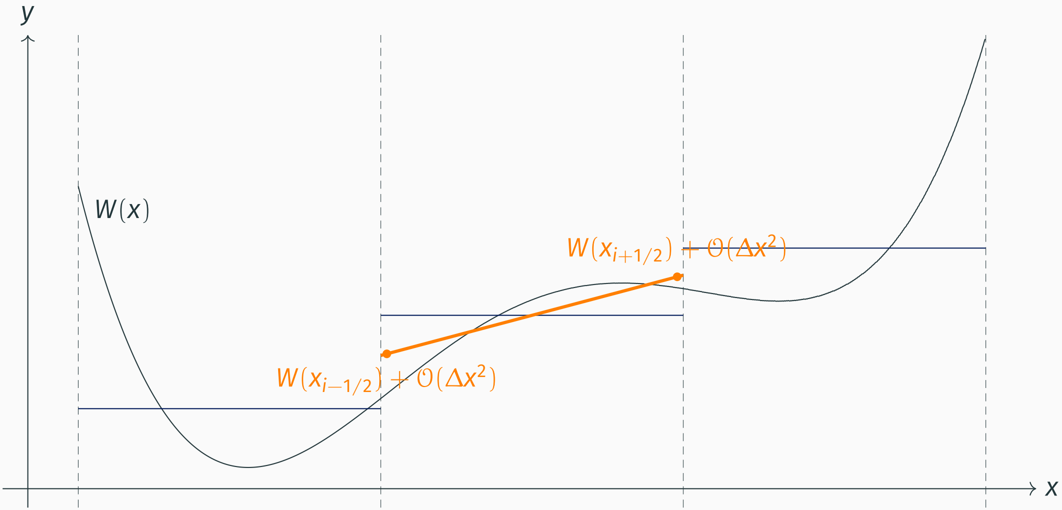
We note that the correction in the $\mathbb{P}_d^{\text{WB}}$ scheme incurs a negligible computational cost, as evidenced in the following table.

	$\mathbb{P}_0$ scheme	$\mathbb{P}_1$ scheme	$\mathbb{P}_1^{\text{WB}}$ scheme	$\mathbb{P}_2$ scheme	$\mathbb{P}_2^{\text{WB}}$ scheme
CPU time (s)	2.91	8.59	9.5	23.78	24

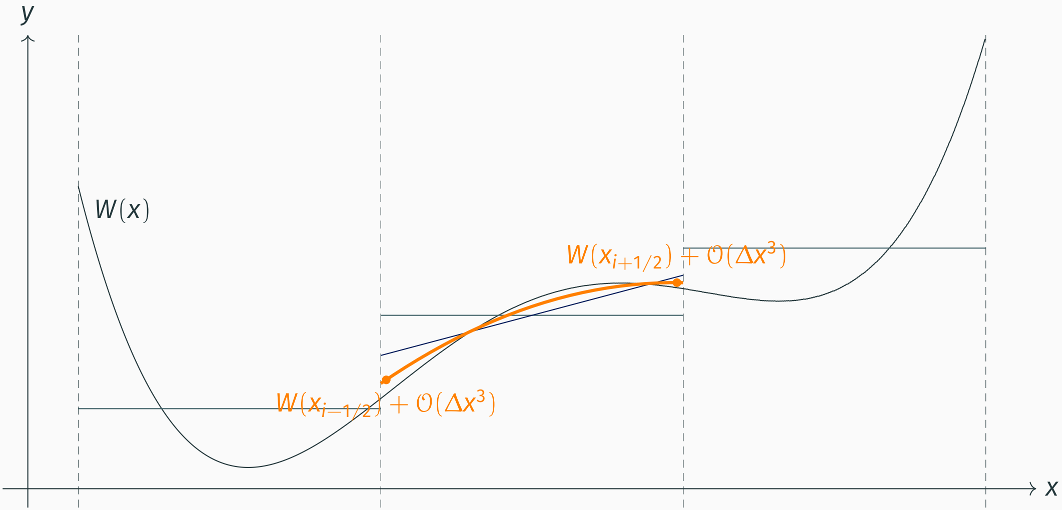
Finite volume method, visualized



Discontinuous Galerkin, visualized



Discontinuous Galerkin, visualized



Discontinuous Galerkin: an example

On the previous slide, the data W is represented by

- a polynomial of degree 2 in each cell (Galerkin approximation),
- which is Discontinuous at interfaces between cells.

Discontinuous Galerkin: an example

On the previous slide, the data W is represented by

- a polynomial of degree 2 in each cell (Galerkin approximation),
- which is Discontinuous at interfaces between cells.

Therefore, in each cell Ω_i, W is approximated by

$$W\big|_{\Omega_i} \simeq W_i^{\text{DG}} := \alpha_0 + \alpha_1 x + \alpha_2 x^2 = \sum_{j=0}^{2} \alpha_j x^j,$$

where the polynomial coefficients α_0, α_1 and α_2 are determined to ensure fitness between the continuous data and its polynomial approximation.

Any polynomial of degree two can be exactly represented this way.

Discontinuous Galerkin: polynomial basis

More generally, we define a polynomial basis $\varphi_0, \dots, \varphi_N$ on each cell Ω_i and approximate the solution in this basis.

A usual example is the following so-called **modal basis**:

$$\forall j \in \{0, \dots, N\}, \quad \varphi_j(x) = x^j.$$

Discontinuous Galerkin: polynomial basis

More generally, we define a polynomial basis $\varphi_0, \dots, \varphi_N$ on each cell Ω_i and approximate the solution in this basis.

A usual example is the following so-called **modal basis**:

$$\forall j \in \{0, \dots, N\}, \quad \varphi_j(x) = x^j.$$

Main takeaway: The DG scheme is exact on every function that can be exactly represented in the basis!

Main idea

Recall that the DG scheme will be exact on every function that can be exactly represented in the DG basis, as soon as it is also a solution to the PDE.

Main idea

Recall that the DG scheme will be exact on every function that can be exactly represented in the DG basis, as soon as it is also a solution to the PDE.

Main idea

Enhance the DG basis by using the steady solution!

⇝ If the **steady solution or an approximation thereof is contained in the basis**, then:

- using the exact steady solution in the basis will make the scheme exactly well-balanced;
- using an approximation of the steady solution will make the scheme approximately well-balanced.

Enhanced DG bases

Assume that you know a **prior** $\overline{W}$ on the steady solution.

It can be the exact steady solution ($\overline{W} = W_{\text{eq}}$), or it can be an approximation ($\overline{W} \simeq W_{\text{eq}}$).

The goal is now to **enhance the modal basis** V using $\overline{W}$:

$$V = \{1, x, x^2, \dots, x^N\}.$$

Enhanced DG bases

Assume that you know a **prior** $\overline{W}$ on the steady solution.

It can be the exact steady solution ($\overline{W} = W_{\text{eq}}$), or it can be an approximation ($\overline{W} \simeq W_{\text{eq}}$).

The goal is now to **enhance the modal basis** V using $\overline{W}$:

$$V = \{1, x, x^2, \dots, x^N\}.$$

First possibility: multiply the whole basis by $\overline{W}$

$$\bar{V}_* = \{\overline{W}, x\,\overline{W}, x^2\,\overline{W}, \dots, x^N\,\overline{W}\}.$$

Enhanced DG bases

Assume that you know a **prior** $\overline{W}$ on the steady solution.

It can be the exact steady solution ($\overline{W} = W_{\text{eq}}$), or it can be an approximation ($\overline{W} \simeq W_{\text{eq}}$).

The goal is now to **enhance the modal basis** V using $\overline{W}$:

$$V = \{1, x, x^2, \dots, x^N\}.$$

First possibility: multiply the whole basis by $\overline{W}$

$$\bar{V}_* = \{\overline{W}, x\,\overline{W}, x^2\,\overline{W}, \dots, x^N\,\overline{W}\}.$$

Second possibility: replace the first element with $\overline{W}$

$$\bar{V}_+ = \{\overline{W}, x, x^2, \dots, x^N\}.$$

Error estimates

We denote by:

- W_{ex} the exact solution,
- W_{DG} the approximate solution without prior,
- $\overline{W_{\text{DG}}}$ the approximate solution with prior $\overline{W}$ and basis $\overline{V}_*$.

For a DG scheme of order $q+1$, we obtain[3] the following error estimates:

$$\left\| W_{\text{ex}} - W_{\text{DG}} \right\| \lesssim \left| W_{\text{ex}} \right|_{H^{q+1}} \Delta x^{q+1},$$

$$\left\| W_{\text{ex}} - \overline{W_{\text{DG}}} \right\| \lesssim \left| \frac{W_{\text{ex}}}{\overline{W}} \right|_{H^{q+1}} \Delta x^{q+1} \left\| \overline{W} \right\|_{L^\infty}.$$

Conclusion of the error estimates: the prior $\overline{W}$ needs to provide a **good approximation of the derivatives** of the steady solution.

[3] Rigorous error estimates are written in terms of the error in the projection onto both bases.

Obtaining a prior

For very simple systems, one can use the exact steady solution as a prior.

However, in many cases, even for some simple and well-known systems, one cannot compute the exact steady solution. Therefore, **an approximation is required**.

How to obtain such an approximation?

Obtaining a prior

For very simple systems, one can use the exact steady solution as a prior.

However, in many cases, even for some simple and well-known systems, one cannot compute the exact steady solution. Therefore, **an approximation is required**.

How to obtain such an approximation?

1. **First possibility**: use a traditional numerical approximation, obtained by classical ODE solvers (e.g. Runge-Kutta schemes).
2. **Second possibility**: use a Physics-Informed Neural Network (PINN), a specifically-trained neural network.

Next step: Present the PINNs, which will be preferred since they are mesh-less and able to approximate solutions to parametric PDEs.

PINNs

Remark: Neural networks are smooth functions of the inputs (provided smooth activation functions are used!).

Since their derivatives are easily computable by automatic differentiation, they are therefore **natural objects to approximate solutions to PDEs or ODEs**.

PINNs

Remark: Neural networks are smooth functions of the inputs (provided smooth activation functions are used!).

Since their derivatives are easily computable by automatic differentiation, they are therefore **natural objects to approximate solutions to PDEs or ODEs**.

Definition: PINN

A PINN is a neural network with input x and trainable weights θ, approximating the solution to a PDE or ODE, and denoted by $W_\theta(x)$.

Hence, the PINN W_θ will approximate the solution to the PDE

$$\mathcal{D}(W, x) = 0,$$

with $\mathcal{D}$ a differential operator.

PINNs: loss function

Ommitting boundary conditions, the problem becomes

$$\text{find } W \text{ such that } \mathcal{D}(W, x) = 0 \text{ for all } x \in \Omega \subset \mathbb{R}^d.$$

Based on this observation, the PINN W_θ should approximately satisfy the above PDE, and the problem becomes:

$$\text{find } \theta_{\text{opt}} \text{ such that } \mathcal{D}(W_{\theta_{\text{opt}}}, x) \simeq 0 \text{ for all } x \in \Omega \subset \mathbb{R}^d.$$

PINNs: loss function

Ommitting boundary conditions, the problem becomes

$$\text{find } W \text{ such that } \mathcal{D}(W, x) = 0 \text{ for all } x \in \Omega \subset \mathbb{R}^d.$$

Based on this observation, the PINN W_θ should approximately satisfy the above PDE, and the problem becomes:

$$\text{find } \theta_{\text{opt}} \text{ such that } \mathcal{D}(W_{\theta_{\text{opt}}}, x) \simeq 0 \text{ for all } x \in \Omega \subset \mathbb{R}^d.$$

The idea behind PINNs training is to find the optimal weights θ_{opt} by **minimizing a loss function built from the ODE residual**:

$$\theta_{\text{opt}} = \underset{\theta}{\operatorname{argmin}} \int_\Omega \|\mathcal{D}(W_\theta, x)\|_2^2 \, dx.$$

The Monte-Carlo method is used for the integrals, which makes the whole approach **mesh-less** and able to deal with **parametric PDEs**.

Parametric PINNs

A parametric PDE is nothing but the following problem:

$$\text{find } W \text{ such that } \mathcal{D}(W, x; \boldsymbol{\mu}) = 0 \text{ for all } x \in \Omega \text{ and } \boldsymbol{\mu} \in \mathbb{P} \subset \mathbb{R}^m.$$

The parametric PINN $W_\theta(x; \boldsymbol{\mu})$ should approximately satisfy the above PDE, and the problem becomes:

$$\text{find } \theta_{\text{opt}} \text{ such that } \mathcal{D}(W_{\theta_{\text{opt}}}, x; \boldsymbol{\mu}) \simeq 0 \text{ for all } x \in \Omega \text{ and } \boldsymbol{\mu} \in \mathbb{P} \subset \mathbb{R}^m.$$

Parametric PINNs

A parametric PDE is nothing but the following problem:

$$\text{find } W \text{ such that } \mathcal{D}(W, x; \mu) = 0 \text{ for all } x \in \Omega \text{ and } \mu \in \mathbb{P} \subset \mathbb{R}^m.$$

The parametric PINN $W_\theta(x; \mu)$ should approximately satisfy the above PDE, and the problem becomes:

$$\text{find } \theta_{\text{opt}} \text{ such that } \mathcal{D}(W_{\theta_{\text{opt}}}, x; \mu) \simeq 0 \text{ for all } x \in \Omega \text{ and } \mu \in \mathbb{P} \subset \mathbb{R}^m.$$

The minimization problem then becomes

$$\theta_{\text{opt}} = \underset{\theta}{\operatorname{argmin}} \int_{\mathbb{P}} \int_{\Omega} \|\mathcal{D}(W_\theta, x; \mu)\|_2^2 \, dx d\mu.$$

Setup: the advection equation

We run experiments on the **advection equation with source term**, with a given initial condition $W_0 : \mathbb{R} \to \mathbb{R}$:

$$\begin{cases} \partial_t W + c\partial_x W = aW + bW^2 & \text{for } x \in (0,1),\ t \in (0,T), \\ W(0,x) = W_0(x) & \text{for } x \in (0,1), \\ W(t,0) = u_0 & \text{for } t \in (0,T). \end{cases}$$

The **steady solution** W_{eq} satisfies the BVP

$$\begin{cases} c\partial_x W_{\text{eq}} - aW_{\text{eq}} - bW_{\text{eq}}^2 = 0 & \text{for } x \in (0,1), \\ W_{\text{eq}}(0) = u_0, \end{cases}$$

whose unique solution is, with parameters $\boldsymbol{\mu} = \{a, b, c, u_0\} \in \mathbb{P} \subset \mathbb{R}^4$:

$$W_{\text{eq}}(x; \boldsymbol{\mu}) = \frac{a u_0}{(a + b u_0) e^{-\frac{ax}{c}} - b u_0}.$$

PINNs as a DG prior: steady solution

We use the DG scheme to solve the advection equation with the **steady solution as initial condition**. We expect the DG scheme with prior:

- to provide a **better approximation of the steady solution** than the classical DG scheme (approximate well-balanced property),
- while converging with the **same order of accuracy**.

We report below some statistics on the gains with 1000 random sets of parameters in $\mathbb{P}$, for a DG scheme of order $q+1$.

q	minimum gain	average gain	maximum gain
0	63.46	735.08	4571.89
1	32.22	149.38	450.74
2	6.20	54.16	118.45
3	1.55	19.54	108.10

PINNs as a DG prior: unsteady solution

We use the DG scheme to solve an unsteady advection problem, without a source term. We expect the DG scheme with prior:

- to provide a **similar approximation of the solution** than the classical DG scheme,
- while converging with the **same order of accuracy**.

The table below shows the gains made by using the prior, for several values of the number n of space cells.

q	gain, $n = 10$	gain, $n = 40$	gain, $n = 160$
0	0.80	0.81	0.81
1	1.00	1.00	1.00
2	1.00	1.00	1.00
3	1.00	1.00	1.00

PINNs as a DG prior: computation time

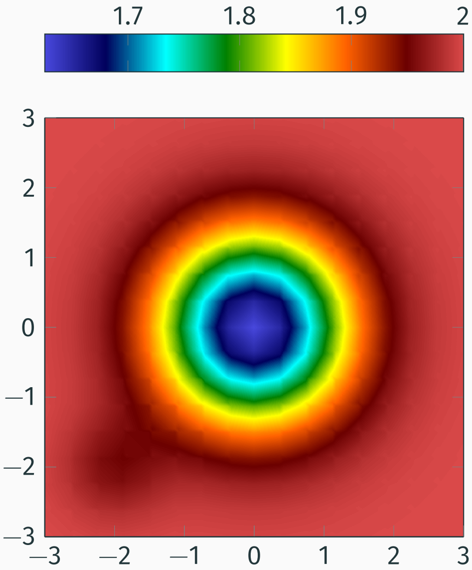
Finally, we compare the computation time in bases V and $\bar{V}_+$. We expect that the prior will:

- **increase** the computation time of the **DG mass matrices**,
- **have no effect on the computation time of the main loop**.

The table below shows the **CPU time increase factor** when using the prior, for several values of the number n of space cells. We observe that the **increase in computation time due to the prior is negligible**.

q	factor, $n = 10$	factor, $n = 40$	factor, $n = 160$
0	1.26	1.07	1.01
1	1.15	1.01	1.00
2	1.04	1.03	1.01
3	1.07	1.00	1.01

Perturbation of a shallow water steady solution



PINN trained on a parametric steady solution, driven by the topography

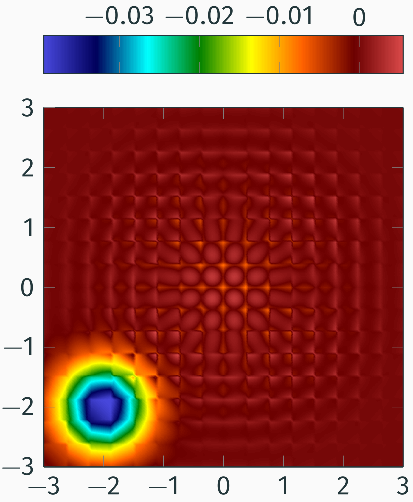
$$Z(x;\mu) = \Gamma \exp\big(\alpha(r_0^2 - \|x\|^2)\big),$$

with physical parameters

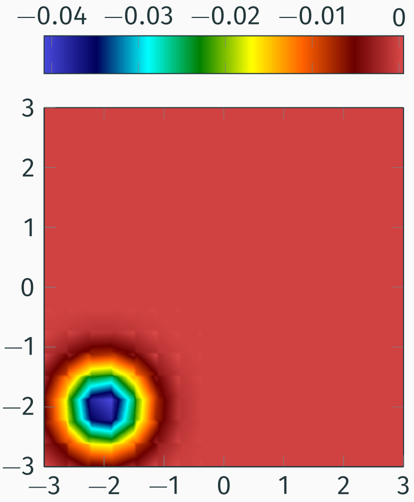
$$\mu \in \mathbb{P} \iff \begin{cases} \alpha \in [0.25, 0.75], \\ \Gamma \in [0.1, 0.4], \\ r_0 \in [0.5, 1.25]. \end{cases}$$

Left plot: initial condition, made of a perturbed steady solution.

Perturbation of a shallow water steady solution

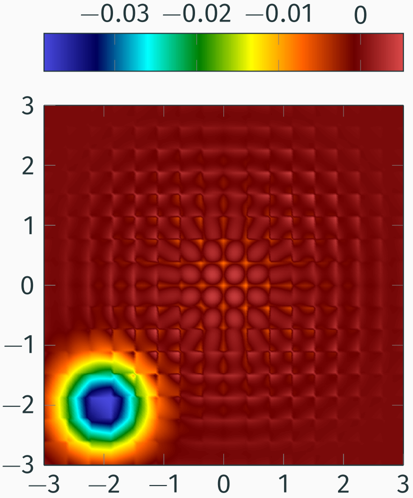


(a) classical basis

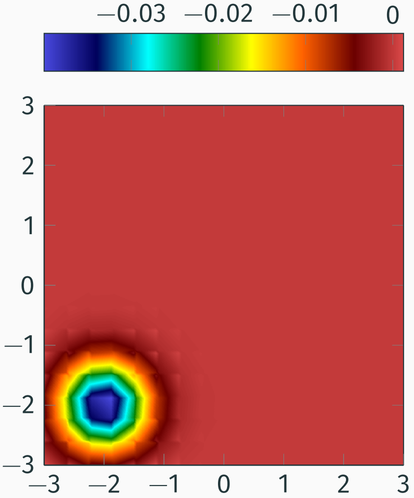


(b) enhanced basis

Perturbation of a shallow water steady solution



(a) classical basis



(b) enhanced basis

Motivation and general objectives

1/ Extending a first-order well-balanced scheme to high-order accuracy

Berthon, Bulteau, Foucher, M'Baye, M.-D., SIAM SISC, 2022

2/ Making any consistent numerical flux fully WB for the shallow water equations

Berthon, M.-D., J. Numer. Math., 2024

3/ Enhancing Discontinuous Galerkin bases with a prior

Franck, M.-D., Navoret, in revision, 2024

Conclusion and perspectives

Conclusion and perspectives

We have obtained:

- an approximately well-balanced DG scheme,
- for parameterized families of steady solutions,
- which works for arbitrary balance laws.

Perspectives include:

- using a space-time DG method and time-dependent priors,
- replacing PINNs with neural operators for added flexibility,
- coding the method in the SciMBA framework.

Related preprint: E. Franck, V. Michel-Dansac and L. Navoret.
"Approximately WB DG methods using bases enriched with PINNs."
git repository: https://github.com/Victor-MichelDansac/DG-PINNs

Thank you for your attention!

An expression of $C^n_{i+1/2}$

To implement the scheme, we need to give an expression of $C = C^n_{i+1/2}$.
We propose $C^0_{i+1/2} = 1$, and, for $n \geqslant 1$:

$$C^n_{i+\frac{1}{2}} = C_\theta \, \frac{1}{2}\left(\frac{\left\| W^n_{i+1} - W^{n-1}_{i+1} \right\|}{\Delta t} + \frac{\left\| W^n_{i} - W^{n-1}_{i} \right\|}{\Delta t} \right),$$

with C_θ a constant parameter.

Note that

$$\theta^n_{i+\frac{1}{2}} = \frac{\varepsilon^n_{i+\frac{1}{2}}}{\varepsilon^n_{i+\frac{1}{2}} + \left(\dfrac{\Delta x}{C^n_{i+\frac{1}{2}}} \right)^\delta} = \frac{\varepsilon^n_{i+\frac{1}{2}} (C^n_{i+\frac{1}{2}})^\delta}{\varepsilon^n_{i+\frac{1}{2}} (C^n_{i+\frac{1}{2}})^\delta + \Delta x^\delta} :$$

we get $\theta^n_{i+\frac{1}{2}} = 0$ if $\varepsilon^n_{i+\frac{1}{2}} = 0$ or $C^n_{i+\frac{1}{2}} = 0$. **Why does this make sense?**

An expression of $C^n_{i+1/2}$ – reasoning

$$\theta^n_{i+\frac{1}{2}} = 0 \text{ if } \varepsilon^n_{i+\frac{1}{2}} = 0 \text{ or } C^n_{i+\frac{1}{2}} = 0$$

$\varepsilon^n_{i+\frac{1}{2}} = 0 \implies$ steady state solution for the equations

$\implies \theta^n_{i+\frac{1}{2}}$ must vanish to preserve the steady state solution

$C^n_{i+\frac{1}{2}} = 0 \implies$ vanishing discrete time derivative

$\implies$ steady state solution for the high-order scheme

$\implies$ not a steady state solution for the equations[4]

$\implies \theta^n_{i+\frac{1}{2}}$ must vanish to perturb the solution

[4]Otherwise, the high-order scheme would be well-balanced.