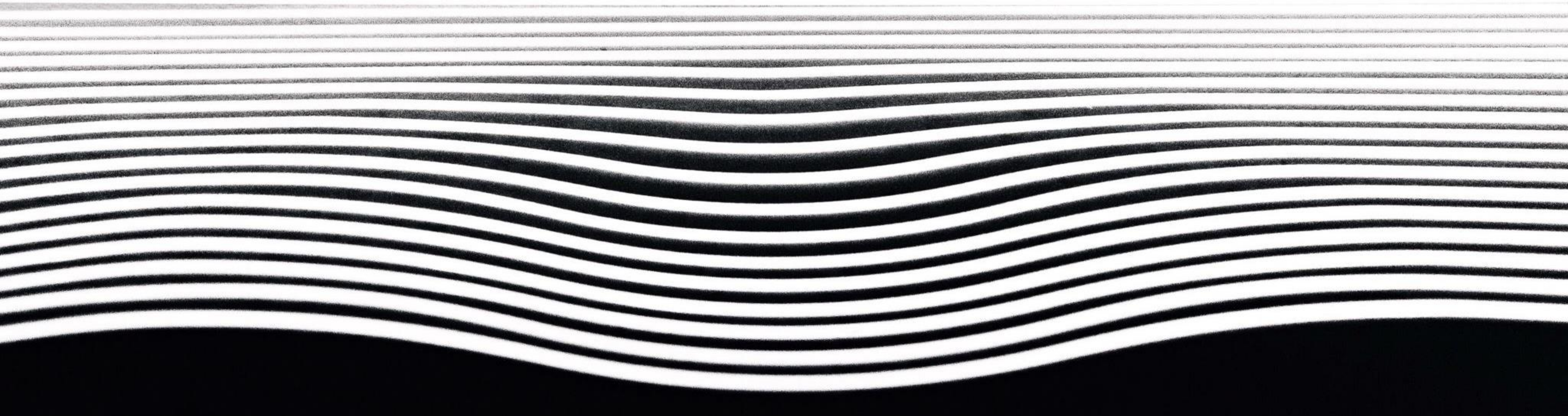
Numerical methods for conservation laws
17: Discontinuous Galerkin methods



Recall

- We have discussed (weighted) essentially non-oscillatory schemes, which can be seen as an adaptive generalization of linear higher-order schemes. Their behavior is very nonlinear and they use a cartesian mesh.

- We now discuss another approach to numerical methods for conservation laws: discontinuous Galerkin methods. (DG methods)

- DG methods share ideas with finite element methods

- Remark: as we have seen, in many simple cases, the differences between the different methods (FD,FV,DG) are only notational.

Recap: Continuous Galerkin methods (finite element mini-recap)

The model problem for Galerkin methods in numerical PDE is the Poisson problem with homogeneous Dirichlet boundary conditions:

$$
\begin{gathered}
-u''(x) = g(x), \qquad x \in (0,1) \\
u(0) = 0, \qquad u(1) = 0.
\end{gathered}
$$

How to develop numerical methods for those? First reformulate the problem.

Note that u is a solution only if for all test function $\phi \in C_c^\infty(0,1)$

$$
-\int_0^1 u''(x)\phi(x)dx = \int_0^1 g(x)\phi(x)dx
$$

We integrate by parts: u is a solution only if for all test function $\phi \in C_c^\infty(0,1)$ the weak formulation holds:

$$\int_0^1 u'(x)\phi'(x)dx = \int_0^1 g(x)\phi(x)dx \qquad (*)$$

We let $H_0^1(0,1)$ be the space of square-integrable functions whose weak derivative is square-integrable and satisfies $u(0) = u(1) = 0$ at the boundary.

Weak formulation: suppose that $g \in L^2(0,1)$. We seek $u \in H_0^1(0,1)$ such that for all $\phi \in H_0^1(0,1)$ we have $(*)$ satisfied.

Theorem:

Suppose that $g \in L^2(0,1)$. There exists $u \in H_0^1(0,1)$ satisfying the weak formulation.

Weak formulations work well for the theoretical understanding of partial differential equations. How can we use them to develop numerical methods?

Idea: u is sought within a trial space (here, $H_0^1(0,1)$) and then tested against functions from a test space (again, $H_0^1(0,1)$).

What if use the weak formulation but the trial and test spaces are subspaces?

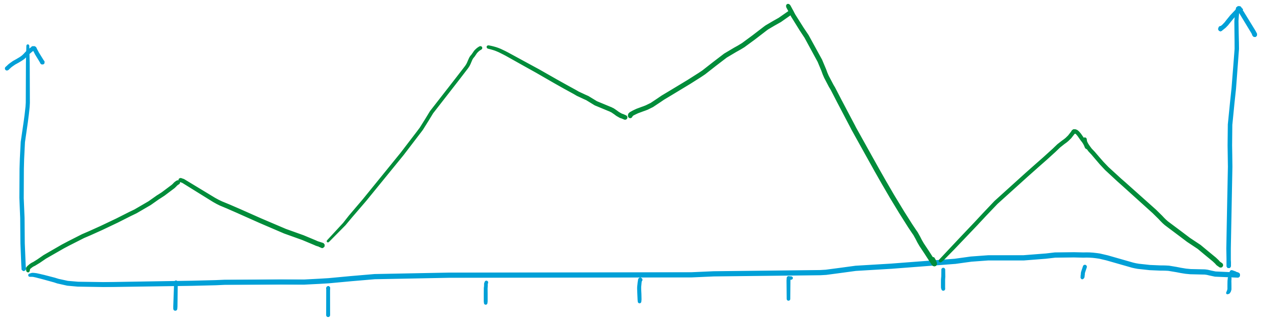
Consider some subspaces of $H_0^1(0,1)$:

- The space of degree m polynomials, $P^m([0,1])$ satisfying the boundary conditions.

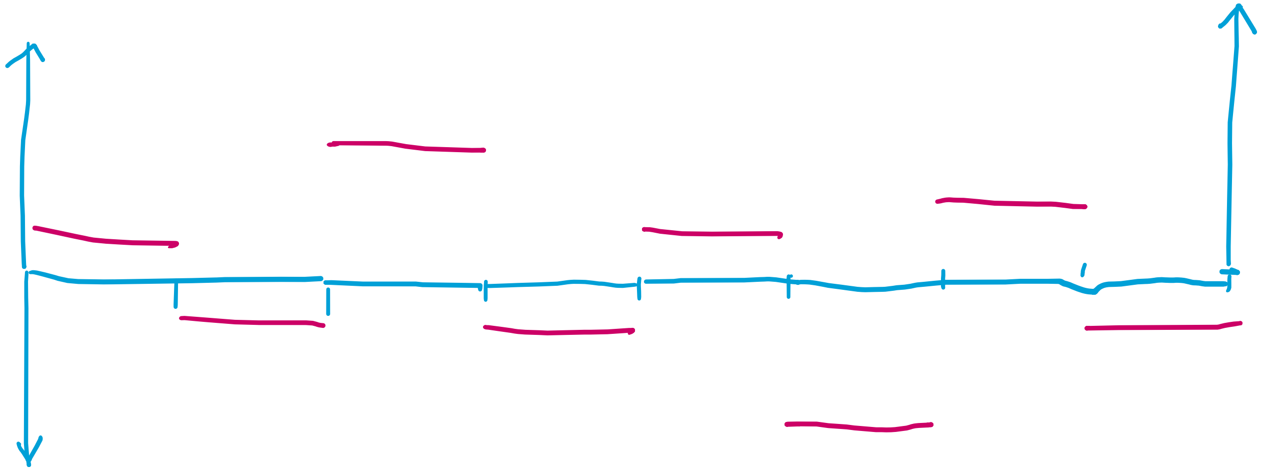
- The span of the first m sine modes: $\sin(2\,\pi\,k\,x\,)$, $1 \leq k \leq m$

- The space of continuous piecewise linear functions with respect to some partition of the interval and satisfying the boundary conditions.

The last example is the most important one: it is the Lagrange space and leads to the finite element method.



Their spatial derivative is piecewise constant



Obviously, these functions constitute a finite-dimensional subspace of $H_0^1(0,1)$

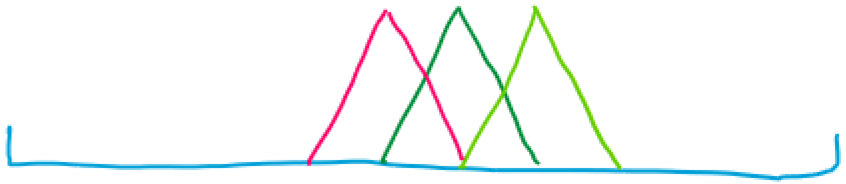
A basis of the Lagrange space is given by the hat functions Φ_i corresponding to the inner nodes.

We can write every $v_h \in V_h$ as a linear combination of those hat functions.

$$v_h(x) = \sum_i v(x_i)\, \Phi_\mathrm{i}(x)$$

hat function

coefficient



Suppose we have fixed a subspace $V_h \subseteq H_0^1(0,1)$.

Galerkin formulation: we seek $u_h \in V_h$ such that for all $\phi_h \in V_h$:

$$\int_0^1 u_h'(x)\phi_h'(x)dx = \int_0^1 g_h(x)\phi_h(x)dx$$

Theorem:

Suppose that $f \in L^2(0,1)$. There exists $u_h \in V_h$ satisfying the Galerkin formulation.

What about finite element methods for conservation laws? Consider the viscous problem

$$\partial_t u + \partial_x F(u) = \epsilon \partial_{xx}^2 u$$

over the interval $(0,1)$ and with homogeneous boundary conditions.

A weak formulation reads: for all $\phi \in C_c^\infty(0,1)$

$$\int_0^1 \partial_t u(x,t)\phi(x) - F(u)\partial_x \phi(x,t)\ dx = -\epsilon \int_0^1 u'(x,t)\phi'(x) dx$$

- We can set up the weak formulation with trial and test space $H_0^1(0,1)$
- Similarly, we can set up the Galerkin formulation, say, using the Lagrange space.

Galerkin formulation: find $u_h \in V_h$ such that for all $\phi_h \in V_h$

$$\int_0^1 \partial_t u_h(x,t)\phi_h(x) - F(u_h)\partial_x \phi_h(x,t)\, dx = -\epsilon \int_0^1 u_h'(x,t)\phi_h'(x) dx$$

- It suffices to only test against the hat functions Φ_i since those are a basis of the trial space
- We can express the solution in terms of the test functions

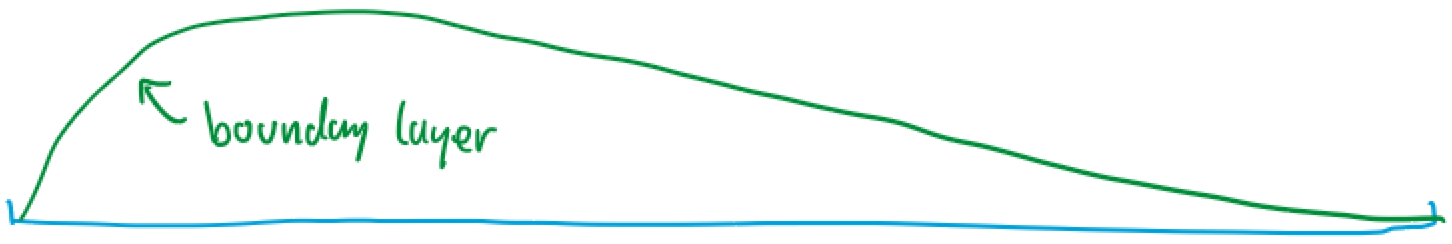
$$u(x,t) = \sum_i v_i(t)\Phi_h(x)$$

- *In combination with some time stepping scheme, we only need to solve a finite (nonlinear) system in finitely many unknowns.*

As the viscous regularization ϵ goes to zero, the original conservation law emerges.

However, there are issues with using the Galerkin formulation:

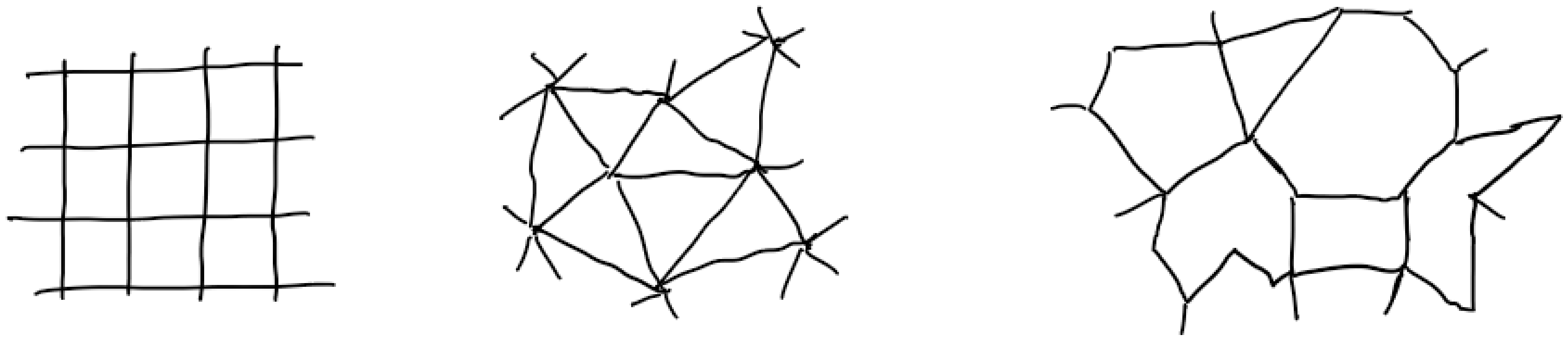
- The integrals against $F(u_h)$ can only be computed approximately for nonlinear flux F (using quadrature).
- For vanishing viscosity, the solution may develop steep gradients. That happens even in the linear case $F(u) = cu$. These boundary layers are difficult to approximate unless the mesh size gets very small.



When the solution of conservation laws develops steep gradients and jumps, then continuous Galerkin methods will struggle to approximate the solution.

However, Galerkin methods are well-understood from a theoretical point of view and are easily adapted to different geometries.

Can we leverage those advantages?



The finite element method emerges when using piecewise polynomial continuous functions in the trial and test spaces. It is a special example of a Galerkin method.

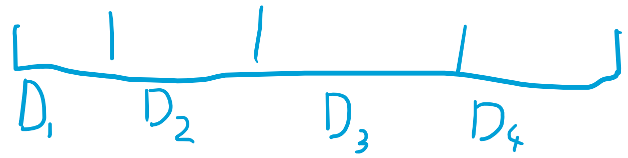
In a discontinuous Galerkin method, we use **discontinuous** piecewise polynomials for the trial and test spaces. Of course, that requires some tweaking to make sure the method works well…

The same idea is used for conservation laws! We use a discontinuous trial and test space, and tweak the formulation.

Step 1) We discretize the domain Ω into N non-overlapping cells:

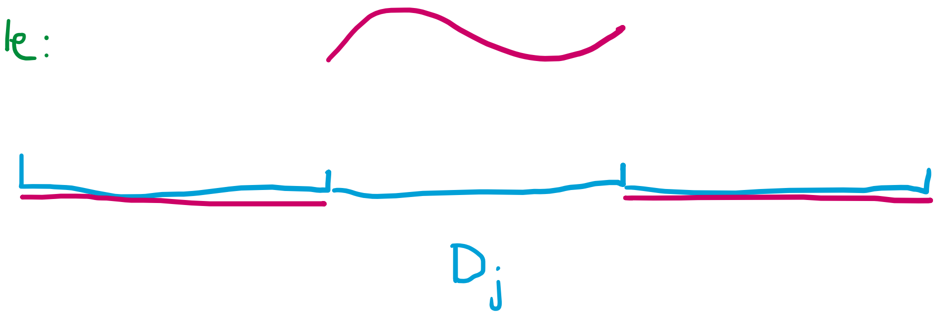
$$\Omega = \cup_{j=1}^{N} D_j .$$

Example:



Step 2) Over each cell D_j we fix a basis $\{\Phi_j^i\}_{i=0}^{m}$ of the vector space $V_h^j = P_m(D_j)$, functions that are degree m polynomials over D_j and vanish outside of D_j.

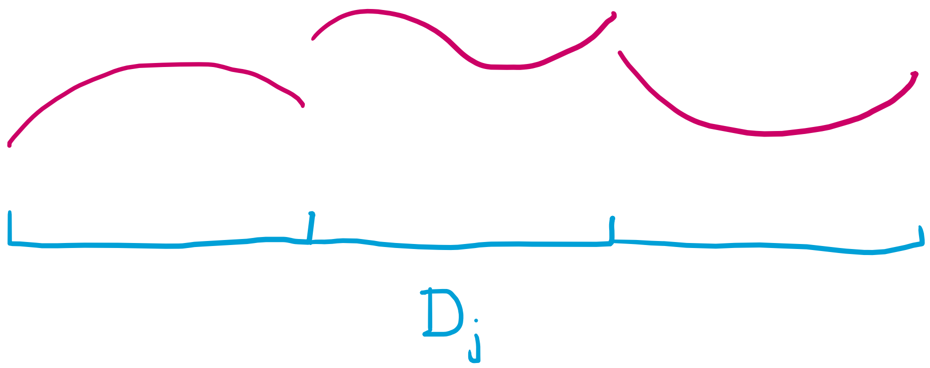
Example:



Step 3) We define the space of functions that are piecewise polynomial over the mesh.

$$\mathrm{V_h} = \bigoplus_{j=1}^{N} V_h^j$$

This is also known as the space of **broken** polynomials. These functions may be discontinuous across cell boundaries.



Step 4) We define the global solution as the sum of cellwise global solutions:

$$\underbrace{U_h(x,t) = \sum_{j=1}^{N} U_h^j(x,t)}_{V_h}, \qquad \underbrace{U_h^j(x,t) = \sum_{i=0}^{m} U_j^i(t)\,\Phi_j^i(x)}_{V_h^j}$$

The basis functions depend only on position x, their coefficients depend only on time t

Step 5) We define the method by a Galerkin formulation. On each cell D_j:

$$\int_{D_j} \partial_t U_h^j v_h + div\, F\big(U_h^j\big) v_h \, dx \ = 0, \qquad\qquad v_h \in V_h^j$$

Using the basis, we use

$$\int_{D_j} \partial_t U_h^j \Phi_j^i(x) + div\, F\big(U_h^j\big) \Phi_j^i(x) dx \ = 0, \qquad\qquad 0 \le i \le m$$

We perform integration by parts in space:

$$\int_{D_j} \partial_t U_h^j \Phi_j^i(x) - F\big(U_h^j\big) \nabla \Phi_j^i(x) dx \ = -\oint_{D_j} \Phi_j^i(x) F\big(U_h^j\big) \cdot \vec{n}\, ds\,, \qquad 0 \le i \le m$$

The problems on each cell are uncoupled.

We need to couple them between cells.

Baisc idea:

we replace the fluxes at the boundary integral on the right-hand side by a numerical flux. At each cell interface we will incorporate data from the neighboring cells.

A local conservation of mass holds on every cell. Taking a constant test function, we find

$$\int_{D_j} \partial_t U_h^j \, dx = -\oint_{D_j} F\left(U_h^j\right) \cdot \vec{n} \, ds \,, \qquad 0 \leq i \leq m$$

For simplicity we consider a one-dimensional setting:

$$D_j = [x_{j-1/2}, x_{j+1/2}]$$

The formulation reads:

$$\int_{D_j} \partial_t U_h^j \Phi_j^i(x)\,dx - \int_{D_j} F\left(U_h^j\right)\nabla\Phi_j^i(x)\,dx = -\Phi_j^i \cdot F\left(U_h^j\right)\Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

We introduce a **numerical flux** F^*, incorporating data from neighboring cells and thus providing the desired coupling:

$$\Phi_j^i \cdot F\left(U_h^j\right)\Big|_{x_{j-1/2}}^{x_{j+1/2}} = \Phi_j^i \cdot F^*\Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

This is called the weak formulation of the DG method.

$$\int_{D_j} \partial_t U_h^j \Phi_j^i(x)\,dx - \int_{D_j} F\left(U_h^j\right)\nabla\Phi_j^i(x)\,dx = -\Phi_j^i \cdot F^*\Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

We integrate the left-hand side in space again.

We also have the strong formulation of the DG method:

$$\int_{D_j} \partial_t U_h^j \Phi_j^i(x)dx + \int_{D_j} div\, F\left(U_h^j\right)\Phi_j^i(x)dx = -\Phi_j^i \cdot \left(F^* - F\left(U_h^j\right)\right) \Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

Both formulations are equivalent, and each has benefits in the theoretical analysis.

We define a few matrices that will be important in the discussion

Mass matrix

$$M_{kl}^{j} = \int_{D_j} \Phi_j^k(x)\Phi_j^l(x)\, dx\,, \qquad S_{kl}^{j} = \int_{D_j} \Phi_j^k(x)\partial_x\Phi_j^l(x)\, dx$$

We approximate $F\left(U_h^j\right)$ by its (approximate) projection onto the local space over D_j

$$F_h^j(x,t) = \sum_{i=0}^{m} F_j^i(t)\, \Phi_j^i(x), \qquad x \in D_j,$$

which is uniquely defined by requiring for all $0 \le p \le m$:

$$\int_{D_j} F_h^j(x,t)\Phi_j^p(x)dx \equiv \sum_{i=0}^{m} F_j^i(t) \int_{D_j} \Phi_j^i(x)\Phi_j^p(x)dx = \int_{D_j} F\left(U_h^j(x,t)\right)\Phi_j^p(x)dx\,.$$

With those matrices we can rewrite the local weak formulation in terms of coefficients.

$$M^j \partial_t U^j - \left(S^j\right)^t F^j = -\Phi^j \cdot F^* \Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

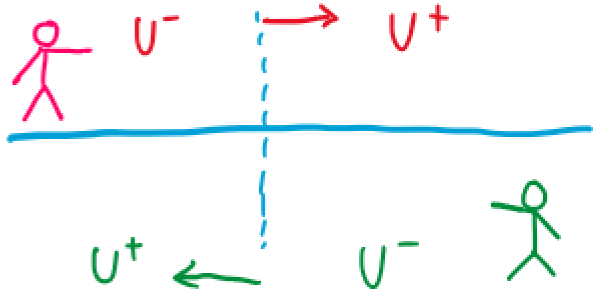
where we use the vectors

$$U^j = \left(U_j^0, U_j^1, \cdots, U_j^m\right)^t, \qquad F^j = \left(F_j^0, F_j^1, \cdots, F_j^m\right)^t,$$
$$\Phi^j = \left(\Phi_j^0, \Phi_j^1, \cdots, \Phi_j^m\right)^t \ at \ \ x_{j\pm 1/2}$$

Similarly, the strong formulation can be expressed locally as:

$$M^j \partial_t U^j + S^j F^j = -\Phi^j \cdot (F^* - F)\Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

Example (Lax-Friedrichs flux)



$$F^*(U^-,U^+) = \frac{F^*(U^-) - F^*(U^+)}{2} - \frac{\alpha}{2}\vec{n}(U^+ - U^-)$$

In the case of the 1D transport equation with speed c

$$F^*(U^-,U^+) = c\,\frac{U^- - U^+}{2} - \frac{\alpha}{2}\vec{n}(U^+ - U^-)$$

We now address some theoretical properties.

Consider the weak formulation

$$\int_{D_j} \partial_t U_h^j \Phi_j^i(x)dx - \int_{D_j} F\left(U_h^j\right)\nabla\Phi_j^i(x)dx = -\Phi_j^i \cdot F^*\Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

Suppose that $\Psi(x,t)$ is a smooth test function and consider its L^2 projection $\Psi_h(x,t)$ onto the global space V_h of broken piecewise polynomials

$$\Psi(x,t) \approx \Psi_h(x,t) = \sum_{j=1}^{N} \Psi_h^j(x,t)$$

Here, the local L^2 projection is defined via

$$\int_{D_j} \Psi_h^j(x,t)\Phi_j^i(x)dx = \int_{D_j} \Psi\,\Phi_j^i(x)dx$$

Consider its coefficient vector in the j-th cell:

$$\Psi^j = \left(\Psi_j^0, \Psi_j^1, \cdots, \Psi_j^m\right)^t$$

We multiply the weak formulation by the test function

$$\Psi^j \; M^j \partial_t U^j - \Psi^j \left(S^j\right)^t F^j = -\Psi^j \cdot F^* \Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

$$\int_{D_j} \Psi_h^j(x) \cdot \partial_t U_h^j(x)\, dx, \qquad \int_{D_j} \partial_x \Psi_h^j(x) \cdot F_h^j(x) dx$$

Observation 1: connection to finite volume methods

Assume that the test function $\Psi = 1/h$ is constant. Then so is $\Psi_h = 1/h$.

When we use that test function in the weak form, then this simplifies to the local conservation of mass:

$$\frac{1}{h}\partial_t \int_{D_j} U_h^j \, dx = -\frac{1}{h}\left(F_{j+1/2}^* - F_{j-1/2}^*\right)$$

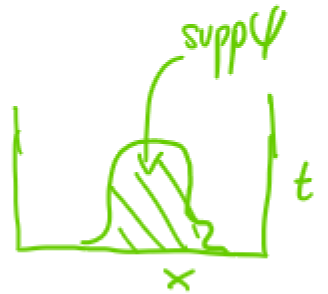
We thus recover the basic form of a finite volume scheme. Summing up leads to

$$\partial_t \sum_{j=1}^{N} \bar{U}_h^j = -\frac{1}{h}\sum_{j=1}^{N}\left(F_{j+1/2}^* - F_{j-1/2}^*\right) = F_{N+1/2}^* - F_{0-1/2}^*$$

From the local conservation of mass we get global conservation of mass.

Typically, the boundary fluxes are the same (periodic BC) or they are part of the problem data.

Observation 2: the limit of discrete solutions is a weak solution



Let us assume that Ψ is compactly supported on Ω and vanishes for large $t > 0$.

We integrate over the time interval and integrate by parts in time. Over each cell this leads to

$$-\int_0^T \int_{D_j} U_h^j \, \partial_t \Psi_h^j + F_h^j \nabla \Psi_h^j \;\, dx\, dt + \int_{D_j} \Psi_h^j(x,0) U_h^j(x,0)\, dx = -\int_0^T \Psi_h^j F^* \Big|_{x_{j-1/2}}^{x_{j+1/2}} \, dt$$

We study the right-hand side. As h goes to zero, we have in all relevant metrics

$$\Psi_h \;\to \Psi$$

But then we have (because of the conservation property)

$$\sum_{j=1}^{N} \int_0^T \Psi_h^j F^* \Big|_{x_{j-1/2}}^{x_{j+1/2}} \, dt \to 0$$

We observe

$$-\sum_j \int_0^T \int_{D_j} U_h^j \, \partial_t \Psi_h^j + F_h^j \nabla \Psi_h^j \; dx \, dt + \int_{D_j} \Psi_h^j(x,0) U_h^j(x,0) \, dx$$

$$= -\int_0^T \int_\Omega U_h \;\; \partial_t \Psi_h \; + F_h \, \nabla \Psi_h \;\; dx \, dt + \int_\Omega \Psi_h \; (x,0) U_h \; (x,0) \, dx$$

Where the gradient $\nabla \Psi_h$ is to be understood piecewise.

If U_h converges to some function U and $F(U_h)$ converges $F(U)$, then in the limit we get

$$-\int_0^T \int_\Omega U \, \partial_t \Psi + F(U) \nabla \Psi \;\; dx \, dt + \int_\Omega \Psi(x,0) U(x,0) \, dx = 0$$

This is the weak form of the conservation law.

That is, if the limit U exists, then it is a weak solution of the conservation law.

Observation 3: there is an energy preservation principle (this takes some effort)

Consider the strong formulation:

$$\int_{D_j} \partial_t U_h^j \Phi_j^i(x)dx \ + \int_{D_j} div\ F\big(U_h^j\big)\Phi_j^i(x)dx \ = -\,\Phi_j^i \cdot \Big(F^* - F\big(U_h^j\big)\Big) \ \Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

We test this formulation with the solution U_h itself. Thus

$$\int_{D_j} \partial_t U_h^j U_h^j(x)dx \ + \int_{D_j} div\ F\big(U_h^j\big)U_h^j(x)dx \ = -\,U_h^j \cdot \Big(F^* - F\big(U_h^j\big)\Big) \ \Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

We test this formulation with the solution U_h itself. Thus

$$\frac{1}{2}\partial_t \|U_h\|_{L^2(D_j)}^2 + \int_{D_j} div\ F\big(U_h^j\big)U_h^j(x)dx \ = -\,U_h^j \cdot \Big(F^* - F\big(U_h^j\big)\Big) \ \Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

Can we control the energy term $||U_h||^2_{L^2(D_j)}$ over time?

For simplicity, we assume

$$F\left(U_h^j(x,t)\right) = F_h^j(x,t),$$

that is, the approximation of the flux is the flux of the approximation. For example, that is true if the flux itself is a polynomial, such as in Burgers' equation.

Otherwise, if that identity is not true, the difference will need to be estimated too.

With that assumption, and working in 1D, we arrive at

$$\int_{D_j} U_h^j \cdot DF_h^j dx = \int_{D_j} U_h^j \cdot DF\left(U_h^j\right) DU_h^j \, dx$$

Suppose that $\tilde{F}$ is a function with derivative $\mathrm{D}\tilde{F}\left(U_h^j\right) = DF\left(U_h^j\right) U_h^j$. Thus, with the chain rule

$$\int_{D_j} U_h^j \cdot DF_h^j dx = \int_{D_j} \mathrm{D}\tilde{F}\left(U_h^j\right) \cdot DU_h^j dx = \int_{D_j} \mathrm{D}\left(\tilde{F}\left(U_h^j\right)\right) dx = \tilde{F}\left(U_h^j\right)\Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

Together with that, we find

$$\frac{1}{2}\partial_t \|U_h\|^2_{L^2(D_j)} = -\tilde{F}\left(U_h^j\right) - U_h^j \cdot \left(F^* - F\left(U_h^j\right)\right) \Bigg|_{x_{j-1/2}}^{x_{j+1/2}}$$

Summing over all cells leads to

$$\frac{1}{2}\partial_t \|U_h\|^2_{L^2(\Omega)} = \sum_j -\tilde{F}\left(U_h^j\right) - U_h^j \cdot \left(F^* - F\left(U_h^j\right)\right) \Bigg|_{x_{j-1/2}}^{x_{j+1/2}}$$

We will reorder the sum and take a look at the interface values. If the values at each interface sum to a non-positive term, then we can conclude

$$\frac{1}{2}\partial_t \|U_h\|^2_{L^2(\Omega)} \leq 0$$

This can be shown when the flux is monotone (increasing in first variable, decreasing in second variable), see textbook (Theorem 12.8)

We now devote some attention to the accuracy of DG schemes.

We mimic the error analysis for FD/FV schemes. Let us assume we have a linear scalar problem

$$\partial_t U + LU = 0$$

We define the error as

$$\epsilon_h(x,t) = U(x,t) - U_h(x,t)$$

Suppose that L_h approximates L. We then have

$$\partial_t \, \epsilon_h(x,t) + L\,U \; - L_h U_h = 0$$

$$\partial_t \, \epsilon_h(x,t) + L_h \, U \; - L_h U_h = L_h U - \; L\,U$$

$$\partial_t \, \epsilon_h(x,t) + L_h \epsilon_h(x,t) = (L_h - L)\, U =: T(U)$$

This expresses the error as the solution of a differential equation with a source term

$$\partial_t \, \epsilon_h(x,t) + L_h \epsilon_h(x,t) = (L_h - L)\, U =: T(U)$$

The solution can be looked up from any ODE textbook:

$$\epsilon_h(x,t) = \exp(\,-L_h t)\epsilon_h(x,0) + \int_0^t \exp\big(\,L_h(s-t)\big)\, T(U(s))\, ds$$

With that definition in mind, we get the estimate

$$\|\epsilon_h(x,t)\| \le \|\exp(\,-L_h t)\| \cdot \|\epsilon_h(x,0)\| + \int_0^t \big\|\exp\big(\,L_h(s-t)\big)\big\| \cdot \|T(U(s))\|\, ds$$

For convergence we need

1. $\|\epsilon_h(x,0)\| \to 0$
2. $\|T(U(s))\| \to 0$
3. $\left\|\exp\left(L_h(s-t)\right)\right\| \to 0$

This rewrites the error as the sum of two parts:

1) Approximation of the initial data:

$$\|\epsilon_h(x,0)\| \leq ch^{m+1}$$

2) Accumulation of the truncation error:

Complicated, see the book

In total, one can show that

1D: $\|U - U_h\| \leq ch^{m+1/2}$ (sub-optimal)

1D: $\|U - U_h\| \leq ch^{m+1}$ (for linear problems with strict upwinding)

2D+3D: $\|U - U_h\| \leq ch^{m+1/2}$ (sub-optimal)

Consider the weak formulation over the j-th cell:

$$M^j \partial_t U^j - \left(S^j\right)^t F^j = -\Phi^j F^* \Big|_{\partial\Omega_{\mathrm{j}}}$$

This is still a semi-discrete formulation. What does the time discretization look like?

Relevant question in the choice of the time discretization include:

1) What is the accuracy, that is, the local discretization error?
2) Does the time discretization have oscillations?
3) Do conserved quantities stay conserved?

Let $k > 0$ be the size of the time step, as before.

We review a few examples.

Example : $\boxed{\text{Explicit Euler}}$ (Forward time)

$$M^j \frac{U_{n+1}^j - U_n^j}{k} - (S^j)^T F_n^j = - \Phi^j F_n^* \Big|_{\partial\Omega_j}$$

Isolate U_{n+1}^j :

$$U_{n+1}^j = U_n^j + k\,(M^j)^{-1} \left[(S^j)^T F_n^j - \Phi^j F_n^* \Big|_{\partial\Omega_j} \right]$$

We need to invert a local matrix at each step.

The local truncation error is $\mathcal{O}(k^2)$ but this method suffers from numerical instability for large time steps.

Example : $\boxed{\text{Implicit Euler}}$

$$M^j \frac{U^j_{n+1} - U^j_n}{k} - (S^j)^T F^j_{n+1} = - \Phi^j F^*_{n+1} \Big|_{\partial\Omega_j}$$

Fixpoint formulation:

$$U^j_{n+1} = U^j_n + k\,(M^j)^{-1} \left[(S^j)^T F^j_{n+1} - \Phi^j F^*_{n+1} \Big|_{\partial\Omega_j} \right]$$

At each time step, we need to solve the global system above, which is a fixpoint equation.

The local truncation error is $O(k^2)$. It is stable even for large time steps. Main disadvantage: computational costs.

Example : $\boxed{\text{Crank-Nicolson method}}$

$$M^j \frac{U^j_{n+1} - U^j_n}{k} - \frac{1}{2}(S^j)^T F^j_{n+1} - \frac{1}{2}(S^j)^T F^j_n$$

$$= -\frac{1}{2}\Phi^j F^*_{n+1} - \frac{1}{2}\Phi^j F^*_n \Big|_{\partial\Omega_j}$$

Similar properties as implicit Euler but often with energy conservation.

Example: $\boxed{\theta\text{- method}}$ $\quad 0 \le \theta \le 1$

$$M^j \frac{U^j_{n+1} - U^j_n}{k} - \theta (S^j)^T F^j_{n+1} - (1-\theta)(S^j)^T F^j_n$$
$$= -\theta \Phi^j F^*_{n+1} - (1-\theta) \Phi^j F^*_n \Big|_{\partial\Omega_j}$$

special case includes Crank-Nicolsen $\theta = \frac{1}{2}$

We now pay some attention to the implementation of the matrices associated to each cell.

We use a map from a reference cell $D = [-1,1]$ onto each physical cell $D_j = \left[x_{j-1/2}, x_{j+1/2}\right]$:

$$x_j \colon [-1,1] \to \left[x_{j-1/2}, x_{j+1/2}\right], \qquad r \mapsto x_{j-1/2} + \frac{1+r}{2} h_j,$$

where h_j denotes the cell diameter. With that we have on the j-th cell:

$$\Phi_j^i\big(x(r)\big) = \Phi^i(r) \qquad \Leftrightarrow \qquad \Phi^i\big(x^{-1}(x)\big) = \Phi_j^i(x)$$

Typically, we define the physical basis functions as transforms of reference basis functions.

Thus we can rewrite the solution on each cell via

$$U_h^j(x(r),t) = \sum_{i=0}^{m} U_j^i(t)\Phi_j^i\big(x(r)\big) = \sum_{i=0}^{m} U_j^i(t)\Phi^i(r)$$

Similarly, we can define reference matrices such that $M^j = \frac{h_j}{2} M$ and $S = S^j$. We have:

$$\int_{D_j} \Phi_j^p(x)\Phi_j^q(x) = \int_{D_j} \Phi^p\left(x_j^{-1}(x)\right)\Phi^q\left(x_j^{-1}(x)\right) = \frac{h_j}{2}\int_D \Phi^p(r)\Phi^q(r)$$

$$\int_{D_j} \Phi_j^p(x)\partial_x\Phi_j^q(x) = \int_{D_j} \Phi^p\left(x_j^{-1}(x)\right)\partial_x\left(\Phi^q\left(x_j^{-1}(x)\right)\right) = \int_{D_j} \Phi^p\left(x_j^{-1}(x)\right)\partial_r\Phi^q\left(x_j^{-1}(x)\right)\frac{2}{h_j} = \int_D \Phi^p(r)\partial_r\Phi^q(r)$$

The choice of the reference basis determines the physical matrices and their algebraic properties.

What are possible choices for the reference basis?

Option 1: monomial basis

$$\phi^i(\,r\,) \coloneqq r^i, \qquad\qquad 0 \le i \le m$$

The L^2 projection U_h of any function U onto the approximation space is calculated by solving (local) linear systems of equations involving the local mass matrix. We want

$$\int_D U_h(r)\ \phi^i(r) = \int_D U(r)\ \phi^i(r)$$

When we write

$$U_h(r) = \sum_{l=0}^{m} U^l \phi^l(r)$$

Then we get the linear system

$$\sum_{l=0}^{m} U^l \int_D \phi^l(r)\ \phi^i(r) = \int_D U(r)\ \phi^i(r)$$

However, the mass matrix for the monomial basis resembles the Hilbert matrix,

$$M_{k,l} = \frac{1 + (-1)^{k+l}}{k + l - 1}$$

And one can show that it’s properties are not conductive (bad conditioning).

Option 2: orthonormal

With some additional effort, we can construct an L^2 - orthonormal reference basis, using the Gram-Schmidt process.

The resulting mass matrix is diagonal.

This connects to the theory of orthogonal polynomials. We have

$$\phi^0(r) = \frac{1}{\sqrt{2}}, \qquad \phi^1(r) = \frac{\sqrt{3}}{\sqrt{2}} r, \qquad r\phi^i = \alpha_i \phi^{i-1} + \alpha_{i+1} \phi^{i+1}$$

The reference matrices can be computed when the bases are fixed in advance.

We still need to integrate U against the polynomial basis in order to determine the right-hand side of the system. One possibility is numerical quadrature.

We can choose the quadrature points and weights to get the Gauss quadrature formula that is exact up to degree $2m + 1$.

quadrature weights

$$\int_{-1}^{1} U(r)\phi^i(r)dx \approx \sum_{i=0}^{m} U(r_i)\phi^i(r_i)\, \omega_i$$

quadrature nodes

This is an identity whenever $U \in P_m$.

We have determined the approximations using integral identities. We call the coefficients the **modal** coefficients:

$$U_h(r) = \sum_{i=0}^{m} \widehat{U}^i \phi^i(r)$$

Alternatively, we can express the approximation in terms of the Lagrange polynomials with respect to some nodal points $\tilde{r}_0, \tilde{r}_1, \dots, \tilde{r}_m$.

$$U_h(r) = \sum_{i=0}^{m} \widetilde{U}^i \, l_i(r)\,, \qquad \widetilde{U}^i = U_h(\tilde{r}_i)$$

The coefficients are generally not the same, but they can be translated into each other (change of basis).

Moreover, computing the approximation via integral identities is generally not the same as computing the approximation via nodal interpolation. However, we can show approximation estimates of order $m+1$ for both.

Lastly, we discuss a discrete entropy condition satisfied by discontinuous Galerkin methods. Consider the weak formulation over the cell D_j

$$\int_{D_j} \partial_t U_h^j \Phi_j \quad dx \; - \int_{D_j} F\left(U_h^j\right) \nabla \Phi_j \quad dx \; = - \, \Phi_j \quad \cdot F^* \Big|_{x_{j-1/2}}^{x_{j+1/2}}$$

We plug in $\Phi_j = U^j$, and *the drop the index h for simplicity.*

$$\int_{D_j} \partial_t U^j U^j \, dx \; - \int_{D_j} F\left(U^j\right) \nabla U^j dx \; = -F_{j+1/2}^* U^j\left(x_{j+1/2}\right) + F_{j-1/2}^* U^j\left(x_{j-1/2}\right)$$

Let $\tilde{F}(u)$ be an antiderivative of the flux F. Then we can write

$$\int_{D_j} \partial_t \frac{\left(U^j\right)^2}{2} \, dx - \tilde{F}\left(U^j\left(x_{j+\frac{1}{2}}\right)\right) + \tilde{F}\left(U^j\left(x_{j-\frac{1}{2}}\right)\right) + F^*_{j+\frac{1}{2}} U^j\left(x_{j+\frac{1}{2}}\right) - F^*_{j-\frac{1}{2}} U^j\left(x_{j-\frac{1}{2}}\right) = 0$$

Consider the last few terms:

$$-\tilde{F}\left(U^j\left(x_{j+\frac{1}{2}}\right)\right) + F^*_{j+\frac{1}{2}} U^j\left(x_{j+\frac{1}{2}}\right) + \tilde{F}\left(U^j\left(x_{j-\frac{1}{2}}\right)\right) - F^*_{j-\frac{1}{2}} U^j\left(x_{j-\frac{1}{2}}\right)$$

We rewrite this:

$$-\tilde{F}\left(U^j\left(x_{j+\frac{1}{2}}\right)\right) + F^*_{j+\frac{1}{2}} U^j\left(x_{j+\frac{1}{2}}\right) + \tilde{F}\left(U^{j-1}\left(x_{j-\frac{1}{2}}\right)\right) - F^*_{j-\frac{1}{2}} U^{j-1}\left(x_{j-\frac{1}{2}}\right)$$
$$-\tilde{F}\left(U^{j-1}\left(x_{j-\frac{1}{2}}\right)\right) + F^*_{j-\frac{1}{2}} U^{j-1}\left(x_{j-\frac{1}{2}}\right) + \tilde{F}\left(U^j\left(x_{j-\frac{1}{2}}\right)\right) - F^*_{j-\frac{1}{2}} U^j\left(x_{j-\frac{1}{2}}\right)$$

We define

$$\hat{F}_{j+1/2} = -\tilde{F}\left(U^j\left(x_{j+\frac{1}{2}}\right)\right) + F^*_{j+\frac{1}{2}} U^j\left(x_{j+\frac{1}{2}}\right)$$

Thus the above term becomes

$$\hat{F}_{j+1/2} - \hat{F}_{j-1/2} - \tilde{F}\left(U^{j-1}\left(x_{j-\frac{1}{2}}\right)\right) + F^*_{j-\frac{1}{2}} U^{j-1}\left(x_{j-\frac{1}{2}}\right) + \tilde{F}\left(U^j\left(x_{j-\frac{1}{2}}\right)\right) - F^*_{j-\frac{1}{2}} U^j\left(x_{j-\frac{1}{2}}\right)$$

We further inspect

$$\Theta_{\mathrm{j}-1/2} := -\tilde{F}\left(U^{j-1}(x_{j-\frac{1}{2}})\right) + F^*_{j-\frac{1}{2}} U^{j-1}(x_{j-\frac{1}{2}}) + \tilde{F}\left(U^{j}(x_{j-\frac{1}{2}})\right) - F^*_{j-\frac{1}{2}} U^{j}(x_{j-\frac{1}{2}})$$

There exists ξ between $U^{j-1}\left(x_{j-\frac{1}{2}}\right)$ and $U^{j}\left(x_{j-\frac{1}{2}}\right)$ such that

$$\tilde{F}\left(U^{j}(x_{j-\frac{1}{2}})\right) - \tilde{F}\left(U^{j-1}(x_{j-\frac{1}{2}})\right) = \tilde{F}'(\xi)\left(U^{j}\left(x_{j-\frac{1}{2}}\right) - U^{j-1}\left(x_{j-\frac{1}{2}}\right)\right)$$

And so we can write

$$\Theta_{\mathrm{j}-1/2} = \left(\tilde{F}'(\xi) - F^*_{j-\frac{1}{2}}\right)\left(U^{j}\left(x_{j-\frac{1}{2}}\right) - U^{j-1}\left(x_{j-\frac{1}{2}}\right)\right) = \left(F(\xi) - F^*_{j-\frac{1}{2}}\right)\left(U^{j}\left(x_{j-\frac{1}{2}}\right) - U^{j-1}\left(x_{j-\frac{1}{2}}\right)\right)$$

If the numerical flux F^* is the consistent with the flux F, and if the numerical flux is monotone (increasing in left value, decreasing in right value), then the last expressions is ≤ 0.

To summarize, we have shown that

$$\int_{D_j} \partial_t \frac{\left(U^j\right)^2}{2}\, dx + \hat{F}_{j+1/2} - \hat{F}_{j-1/2} + \Theta_{\mathrm{j}-1/2} = 0$$

Writing $\eta(u) = \frac{1}{2}u^2$, we observe a discrete entropy condition:

$$\int_{D_j} \partial_t \eta(u)\, dx + \hat{F}_{j+1/2} - \hat{F}_{j-1/2} \leq 0$$

We observe that

$$\partial_u \hat{F}(u,u) = -\partial_u\, \tilde{F}(u) + \partial_u (F^*(u,u)\, u) = -F(u) + F'(u)u + F(u)$$

whereas the entropy flux corresponding to η equals

$$\Psi(u) = \int \eta'(u) F'(u) du = \int F'(u) u\, du$$

A few words about parallelization in the context of DG methods.

In a parallel computing environment, we have N computers that can send messages to each other.

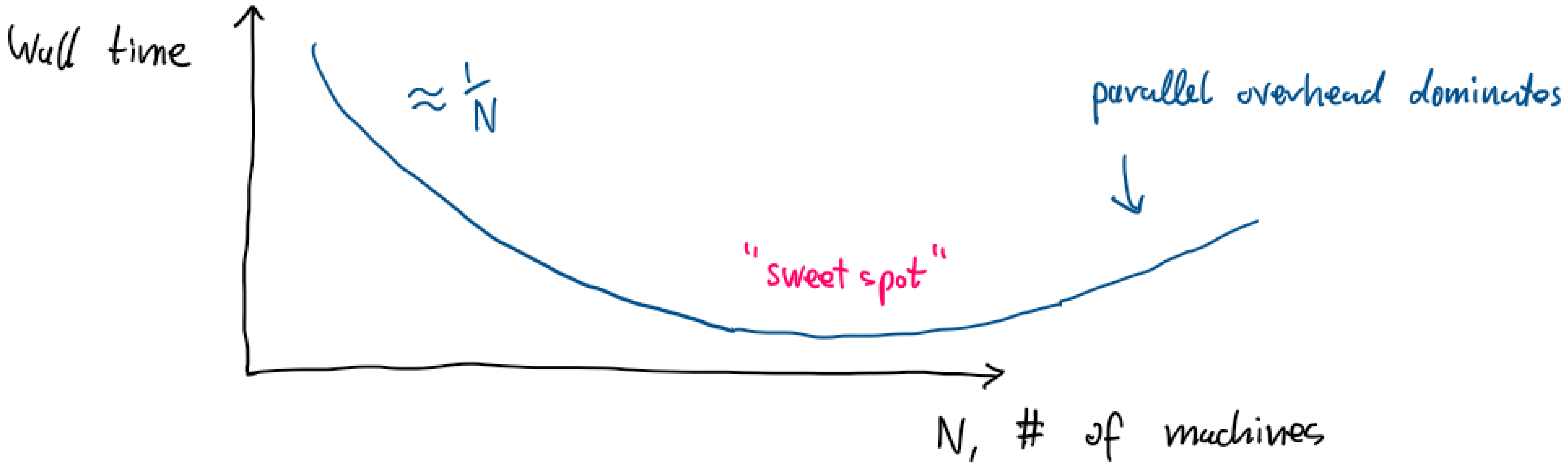
Typically sending messages requires more wall-clock time than the the local computations. Hence in such a computing environment:

We want to **few recipients, few messages, and short messages.**

If that can be insured, we often observe a speed up in comparison to a serial implementation.

Ideally the computing time is antiproportional to the number of machines.

In practice, the parallel overhead takes over at some point.

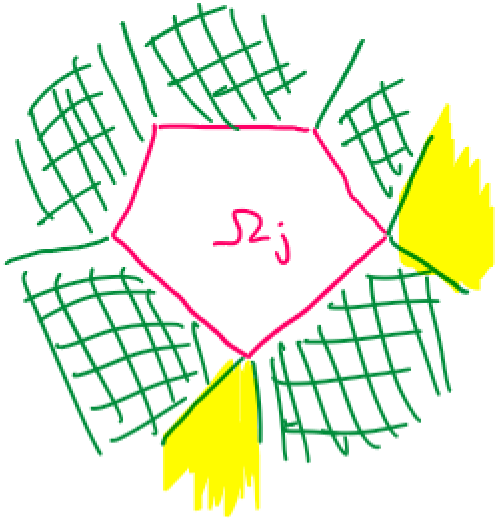


Application to numerical PDE:

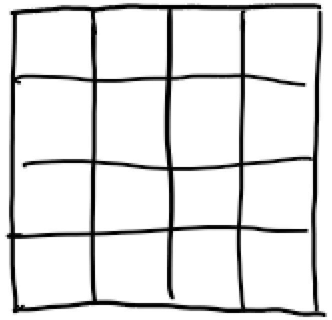
We distribute the mesh over several computers. At each time step, the evolution over a local cell requires information from the neighboring cells, which might be managed on another computer.

We want to keep the parallel overhead low and the local computation load balanced.

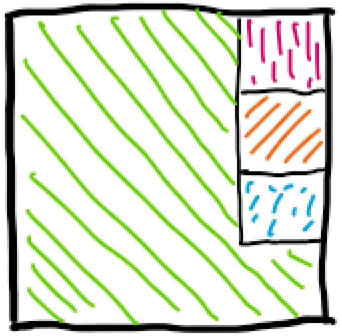
In the DG method for conservation laws, the numerical flux F^* depends on the values in the two neighboring cells. So its computation at each time step requires values from cells that might be managed on another computer.



Example: 4 machines managing a 4x4 grid



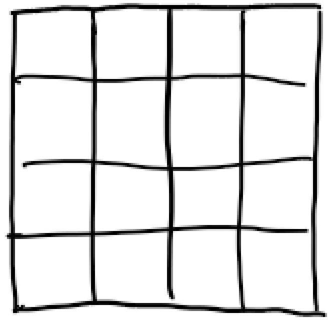
How to distribute these 16 cells to 4 machines? (4 regions)



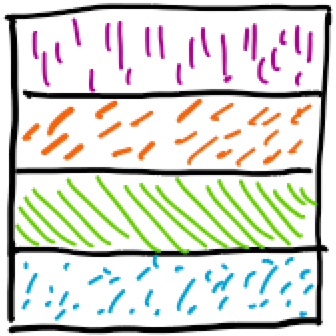
Bad choice: very unbalanced load.

(The 4 regions should have the same volume)

Example: 4 machines managing a 4x4 grid



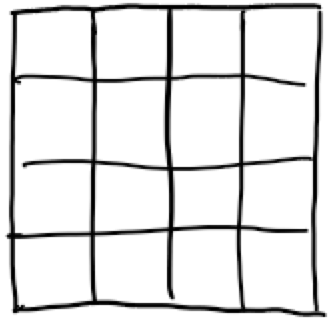
How to distribute these
16 cells to 4 machines?
(4 regions)



So-so choice:

balanced load
but 12 interfaces between regions

Example: 4 machines managing a 4x4 grid



How to distribute these 16 cells to 4 machines?
(4 regions)

Good choice:
balanced load,
8 interfaces between regions

Typically, the distribution is heuristically optimized before the computation begins in a preprocessing step.

Example heuristic: minimize interface surface while keeping the volumes comparable.

If local mesh refinement is performed, then the load may become unbalanced over time, and redistribution becomes reasonably.

That’s all for DG methods