

Lecture 3: Scalar Advection and Linear Hyperbolic Systems

By

Prof. Dinshaw S. Balsara (dbalsara@nd.edu)

Les Houches Summer School in Computational Astrophysics

<http://www.nd.edu/~dbalsara/Numerical-PDE-Course>



3.1) Introduction

We have seen the need for consistency and stability in FDAs of PDEs. However, for the advection equation, which is linear, the approach failed badly for square pulses. We study why in this chapter.

Our analysis is based on a **pictorial approach for advection**. Concept of *total variation diminishing (TVD) schemes* that is later formalized.

Need two insights for treating systems:

- 1) How to treat *linear hyperbolic systems w/o oscillations* (this chapter).
- 2) How to *deal with non-linearities* in the hyperbolic system (next chapter).

Will also discuss the *Riemann problem* for linear hyperbolic systems.

These topics are *important building blocks* that will be used over and over again later in *scheme design*. Also address *boundary conditions*?

3.2) Qualitative Introduction to Non-Linear Hybridization for Scalar Advection

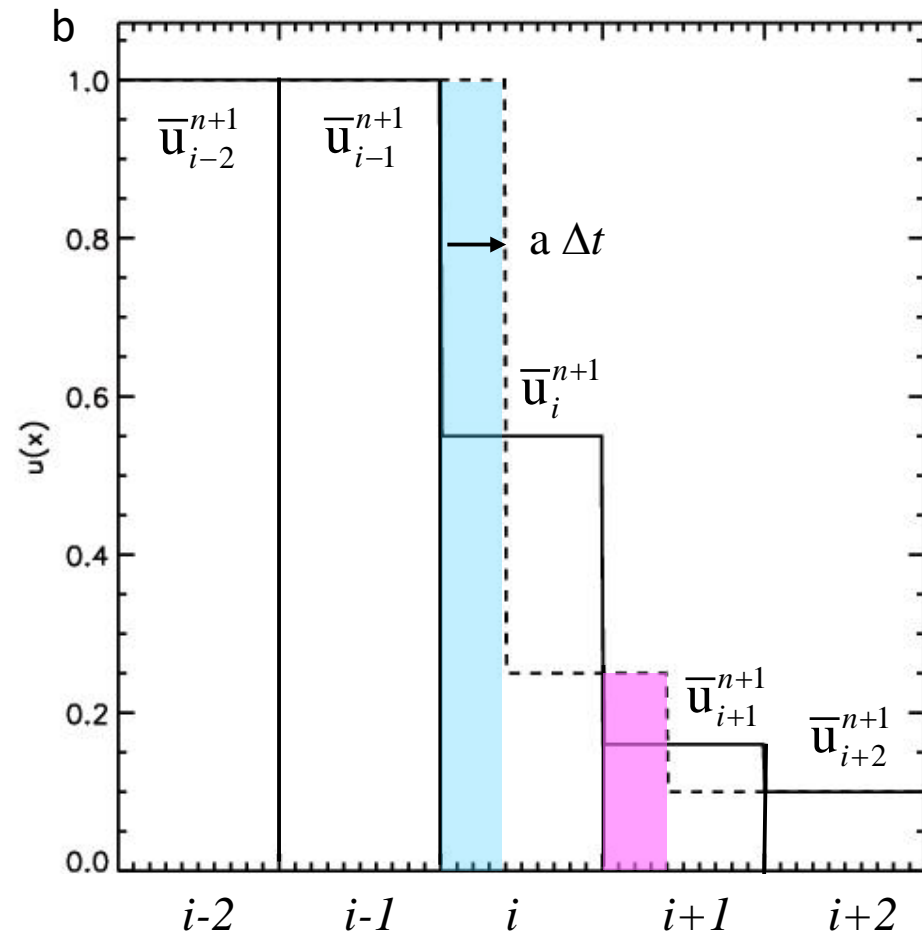
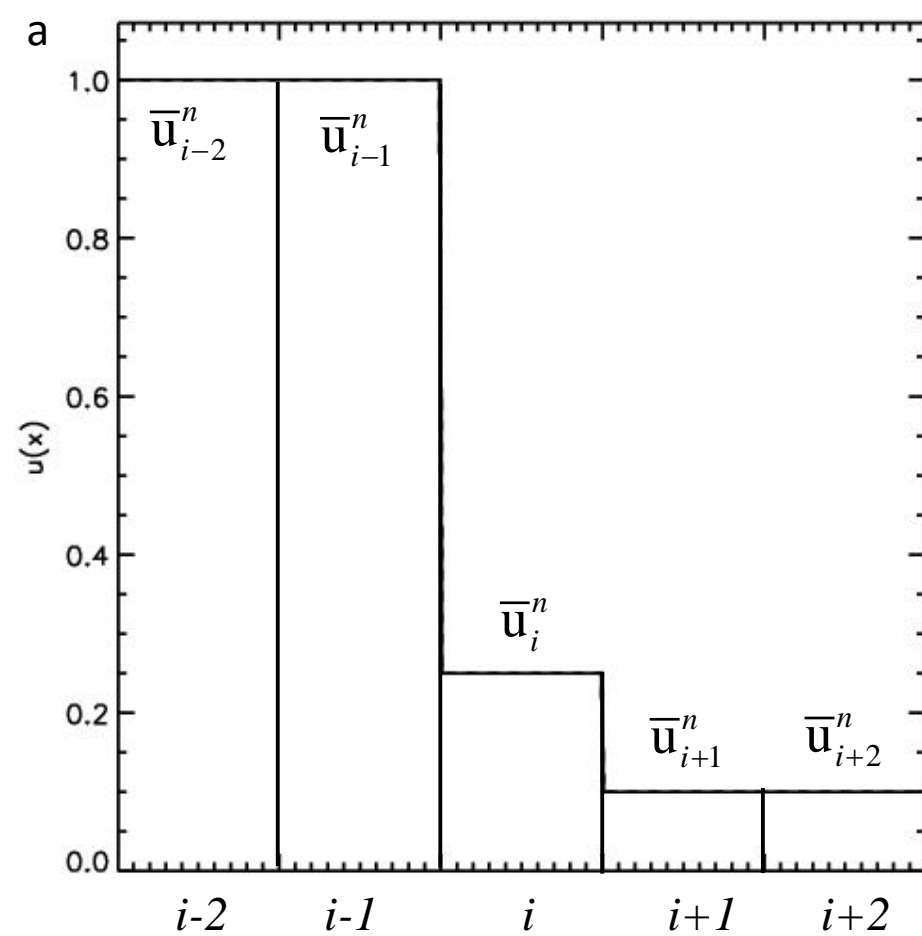
The Dilemma: *Godunov's theorem* : There are no *linear*, second order schemes for treating *linear* advection which would always remain positivity preserving. *Free of oscillations* \Leftrightarrow *monotonicity preserving*.

The Way Out: *van Leer* designed inherently *non-linear* schemes for treating the *linear* advection problem, thereby escaping the clutches of Godunov's theorem!

We follow Godunov's original idea of *literally moving slabs of fluid* in order to show the evolution of fluid from time t^n to $t^{n+1} = t^n + \Delta t$.

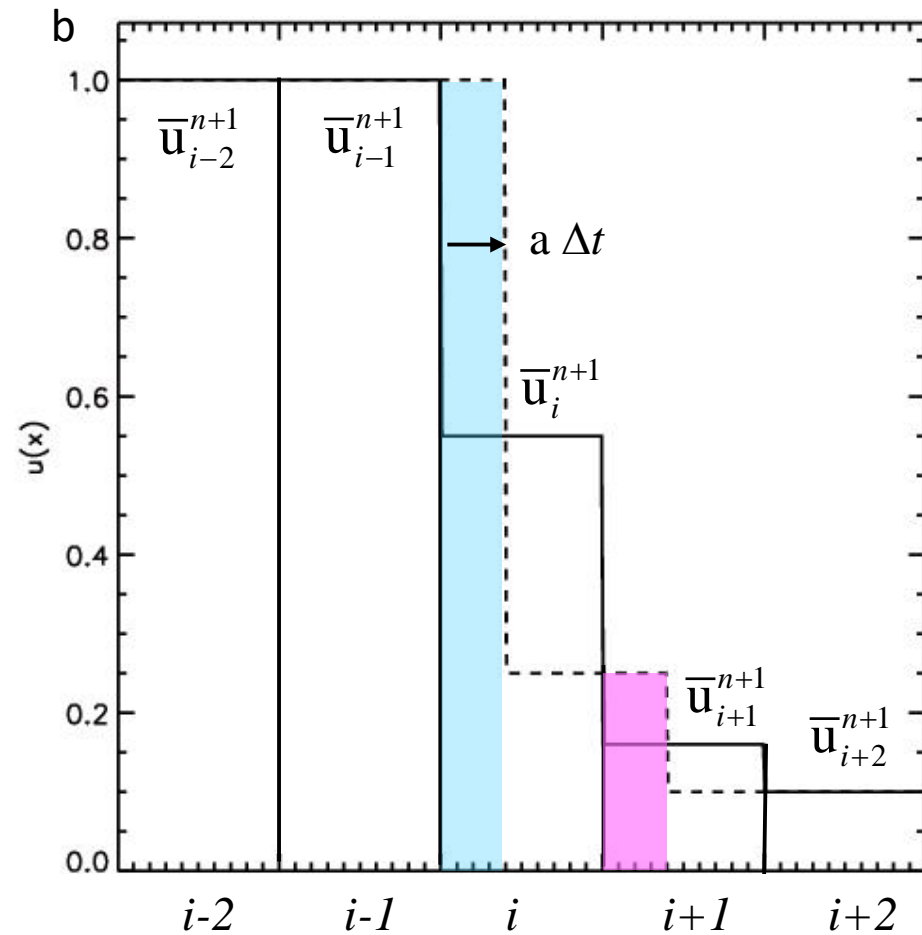
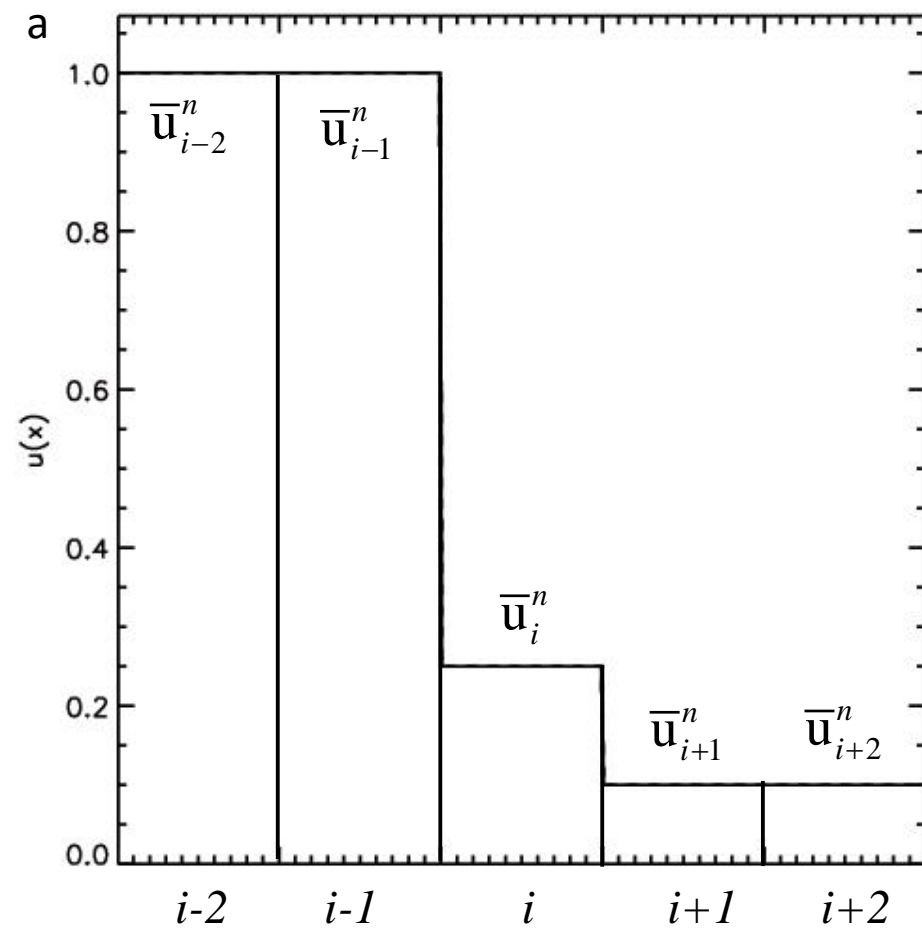
We focus on what happens to the discontinuity in the i^{th} zone when we do that. Consider $\Delta x = 1$ and $\Delta t = 0.4$, i.e. CFL of 0.4.

The *blue* represents the fluid (i.e. flux) that flows *into zone i* . The *pink* represents the flux flowing *out of the same zone i* .



Questions: Can we develop a notion of **monotonicity preserving advection**?

Can we **visually** show that this advection is monotonicity preserving?



Time average of left flux : $\Delta t \bar{f}_{i-1/2}^{n+1/2} = (a \Delta t) \bar{u}_{i-1}^n$

Time average of right flux : $\Delta t \bar{f}_{i+1/2}^{n+1/2} = (a \Delta t) \bar{u}_i^n$

Conservation law in integral form : $\Delta x \bar{u}_i^{n+1} = \Delta x \bar{u}_i^n + \Delta t \bar{f}_{i-1/2}^{n+1/2} - \Delta t \bar{f}_{i+1/2}^{n+1/2}$

$$\Leftrightarrow \bar{u}_i^{n+1} = \bar{u}_i^n - \mu (\bar{u}_i^n - \bar{u}_{i-1}^n) = (1 - \mu) \bar{u}_i^n + \mu \bar{u}_{i-1}^n$$

Question: What can you say about *positivity* for the above scheme?

Time average of left flux : $\Delta t \bar{f}_{i-1/2}^{n+1/2} = (a \Delta t) \bar{u}_{i-1}^n$

Time average of right flux : $\Delta t \bar{f}_{i+1/2}^{n+1/2} = (a \Delta t) \bar{u}_i^n$

Conservation law in integral form :

$$\Delta x \bar{u}_i^{n+1} = \Delta x \bar{u}_i^n + \Delta t \bar{f}_{i-1/2}^{n+1/2} - \Delta t \bar{f}_{i+1/2}^{n+1/2} \quad \Leftrightarrow \quad \bar{u}_i^{n+1} = \bar{u}_i^n - \mu (\bar{u}_i^n - \bar{u}_{i-1}^n) = (1 - \mu) \bar{u}_i^n + \mu \bar{u}_{i-1}^n$$

We see by sheer construction that the *first order upwind* scheme will produce *no new extrema* because each mesh point at t^{n+1} is a *convex* combination of two neighboring mesh points at time t^n .

A scheme with such a property of not generating any new extrema in the solution that were not present initially is called a *monotonicity preserving scheme*.

We wish to explore *monotonicity preserving schemes which are second order accurate extensions* of the first order upwind scheme.

To get to second order accuracy, *endow slabs with slopes*. 3 choices:

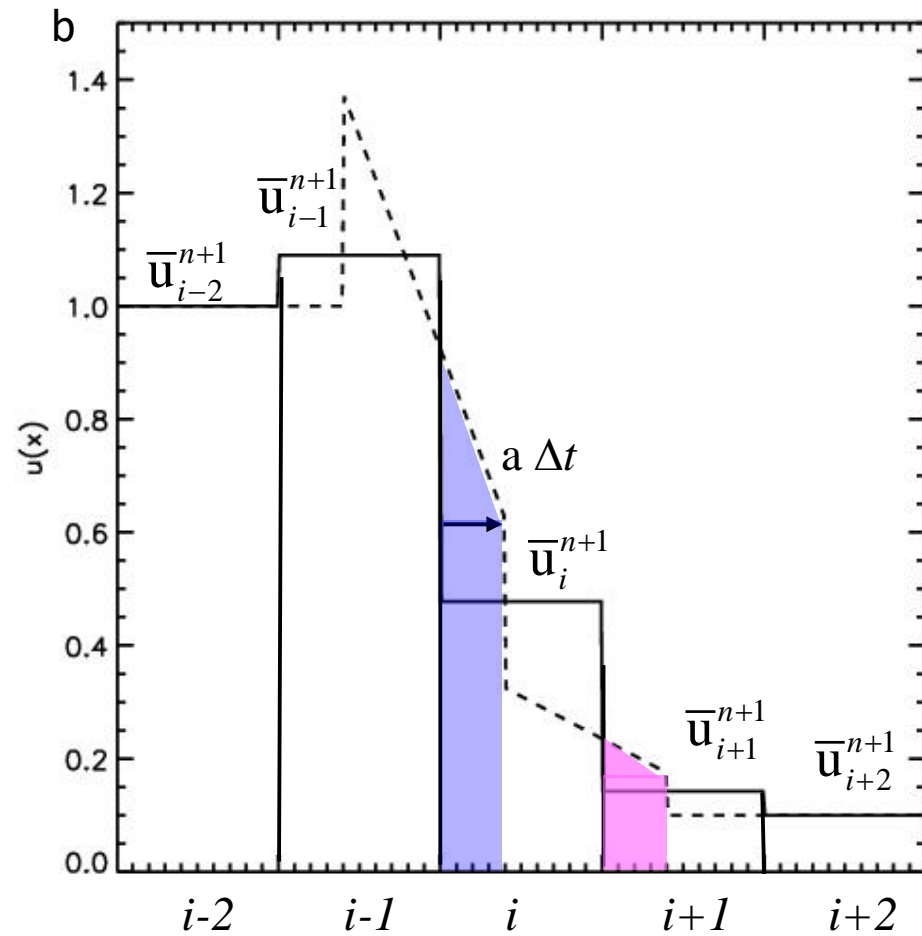
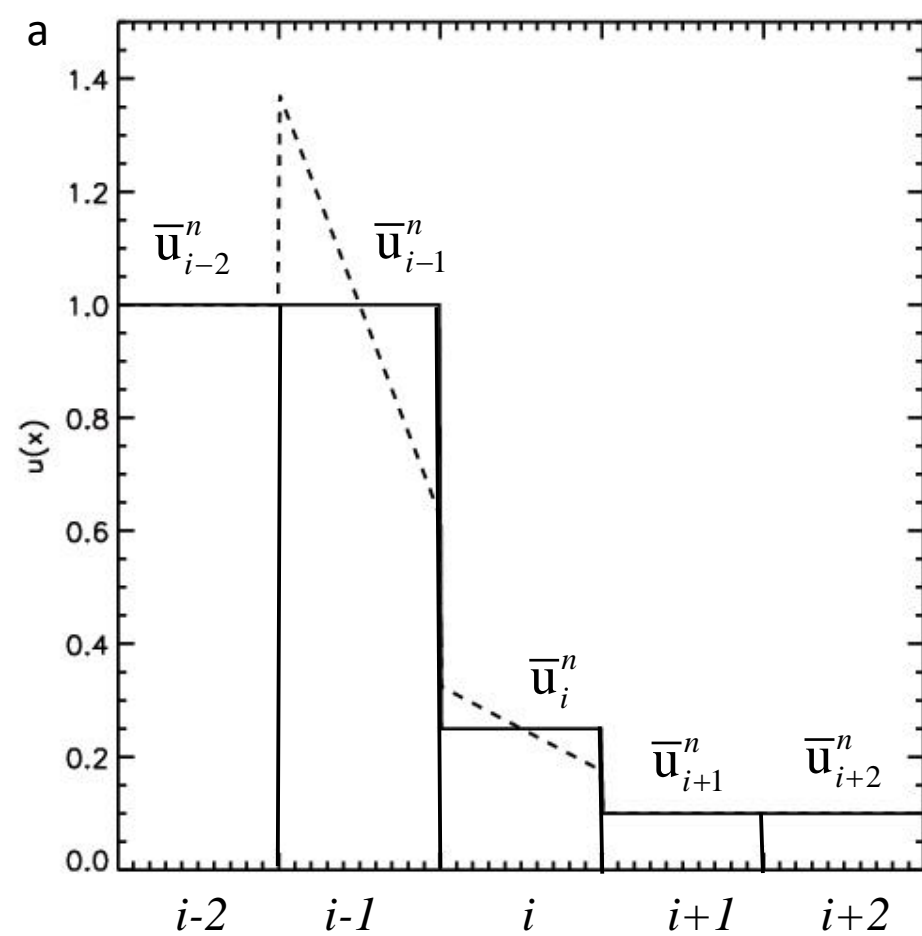
1) Left-biased

2) Central difference

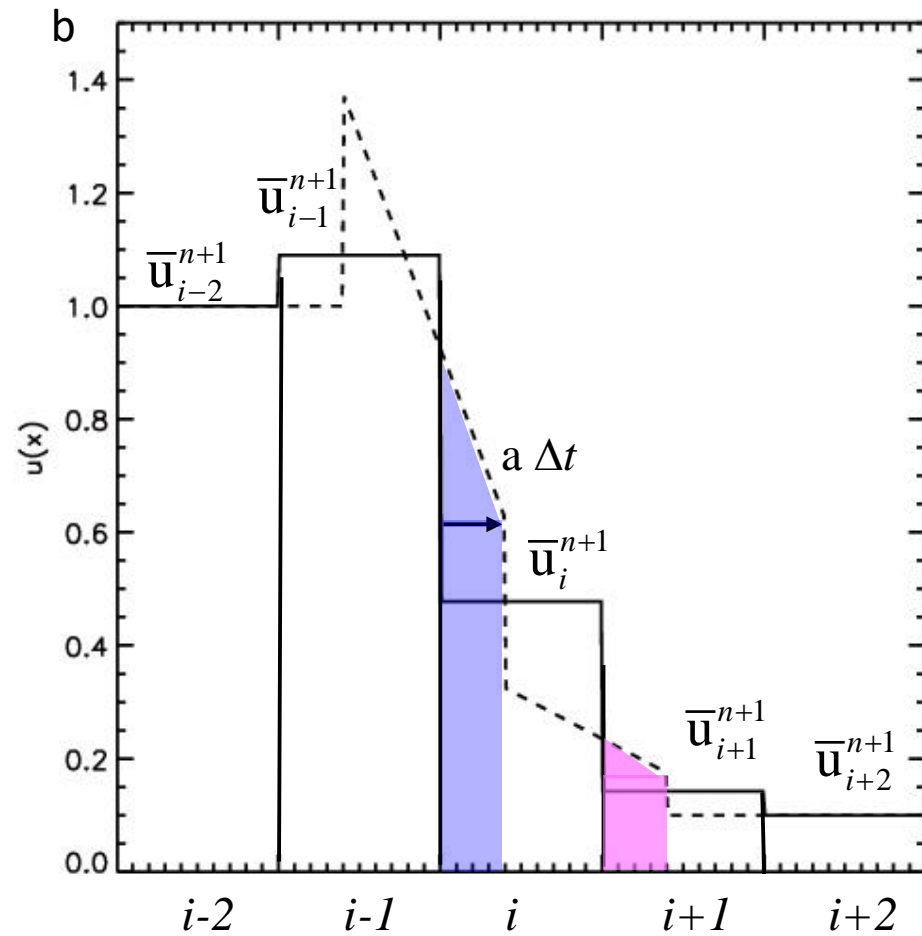
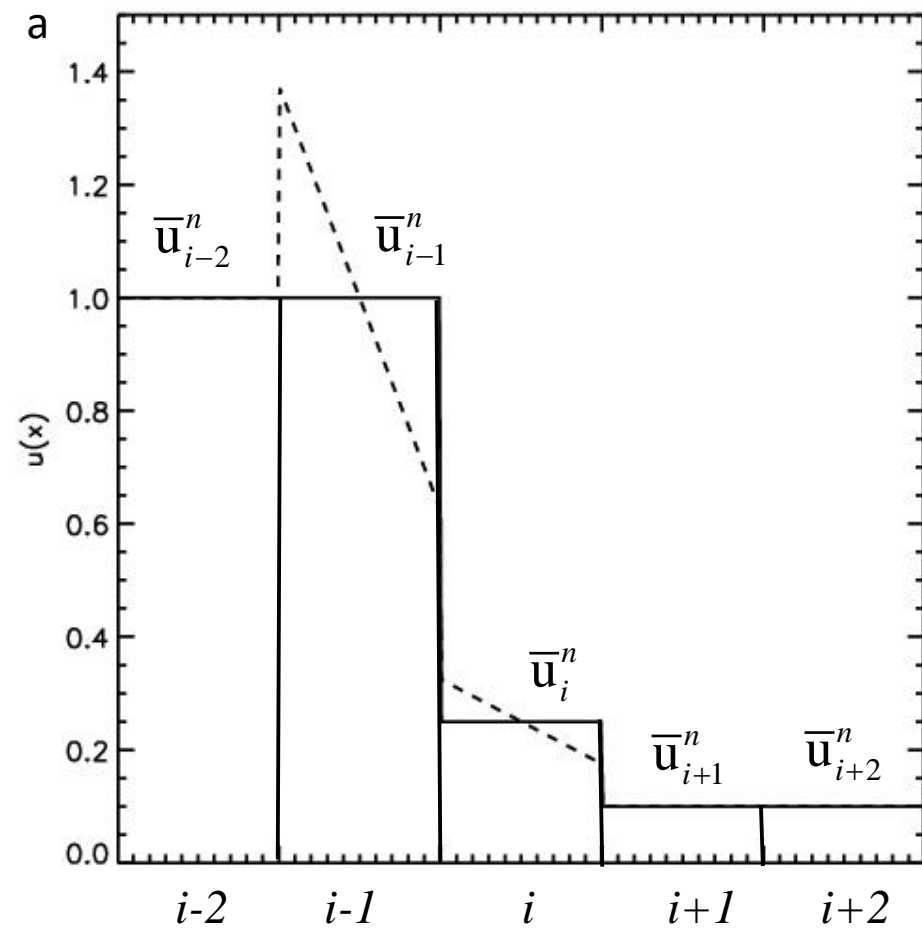
3) Right-biased

Formally, this is called *reconstruction*. We are carrying out *piecewise linear reconstruction*. Reconstruction is an important building block in scheme design. We have:

$$u_i^n(x) = \bar{u}_i^n + \frac{\overline{\Delta u}_i^n}{\Delta x} (x - x_i) \leftarrow \overline{\Delta u}_i^n \text{ is called a "slope" or undivided difference}$$



Questions: Can we **visually** show that this second order advection is not monotonicity preserving?



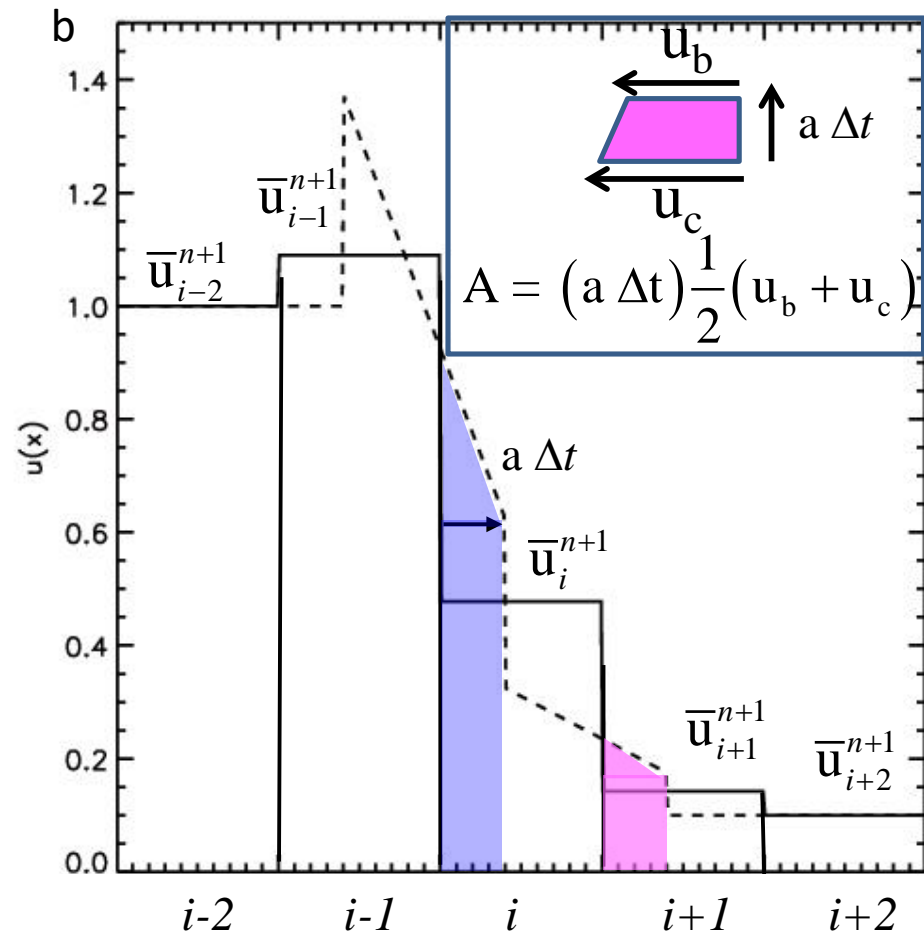
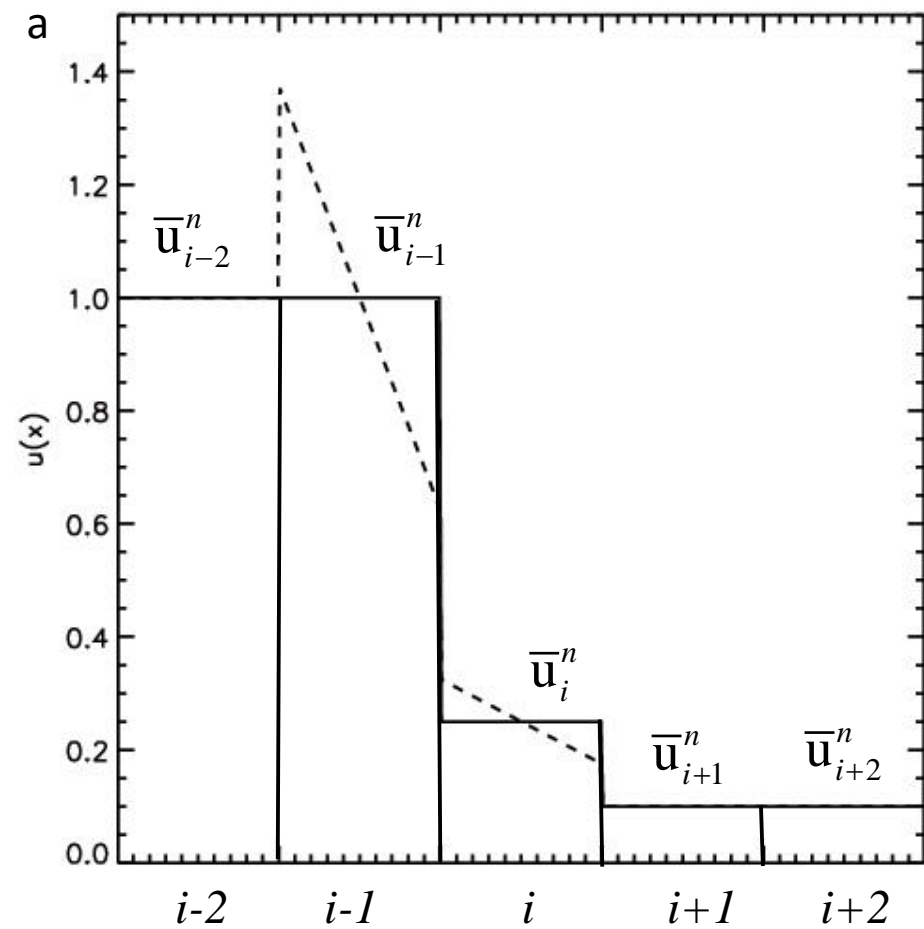
Time average of left flux : $\Delta t \bar{f}_{i-1/2}^{n+1/2} = (a \Delta t) \left[\bar{u}_{i-1}^n + \frac{1}{2}(1-\mu)\overline{\Delta u}_{i-1}^n \right]$

Time average of right flux : $\Delta t \bar{f}_{i+1/2}^{n+1/2} = (a \Delta t) \left[\bar{u}_i^n + \frac{1}{2}(1-\mu)\overline{\Delta u}_i^n \right]$

Conservation law in integral form : $\Delta x \bar{u}_i^{n+1} = \Delta x \bar{u}_i^n + \Delta t \bar{f}_{i-1/2}^{n+1/2} - \Delta t \bar{f}_{i+1/2}^{n+1/2}$

$$\Leftrightarrow \bar{u}_i^{n+1} = \bar{u}_i^n - \mu(\bar{u}_i^n - \bar{u}_{i-1}^n) - \frac{\mu}{2}(1-\mu)(\overline{\Delta u}_i^n - \overline{\Delta u}_{i-1}^n)$$

Question: What can you say about **positivity** for the above scheme?



Time average of left flux : $\Delta t \bar{f}_{i-1/2}^{n+1/2} = (a \Delta t) \left[\bar{u}_{i-1}^n + \frac{1}{2}(1-\mu) \overline{\Delta u}_{i-1}^n \right]$

Time average of right flux : $\Delta t \bar{f}_{i+1/2}^{n+1/2} = (a \Delta t) \left[\bar{u}_i^n + \frac{1}{2}(1-\mu) \overline{\Delta u}_i^n \right]$

Conservation law in integral form : $\Delta x \bar{u}_i^{n+1} = \Delta x \bar{u}_i^n + \Delta t \bar{f}_{i-1/2}^{n+1/2} - \Delta t \bar{f}_{i+1/2}^{n+1/2}$

$$\Leftrightarrow \bar{u}_i^{n+1} = \bar{u}_i^n - \mu(\bar{u}_i^n - \bar{u}_{i-1}^n) - \frac{\mu}{2}(1-\mu)(\overline{\Delta u}_i^n - \overline{\Delta u}_{i-1}^n)$$

We see that we have generated a new extremum.

Right slopes yield Lax-Wendroff scheme; central slopes yield Fromm scheme; left slopes yield Beam-Warming scheme. *All choices are ill!*

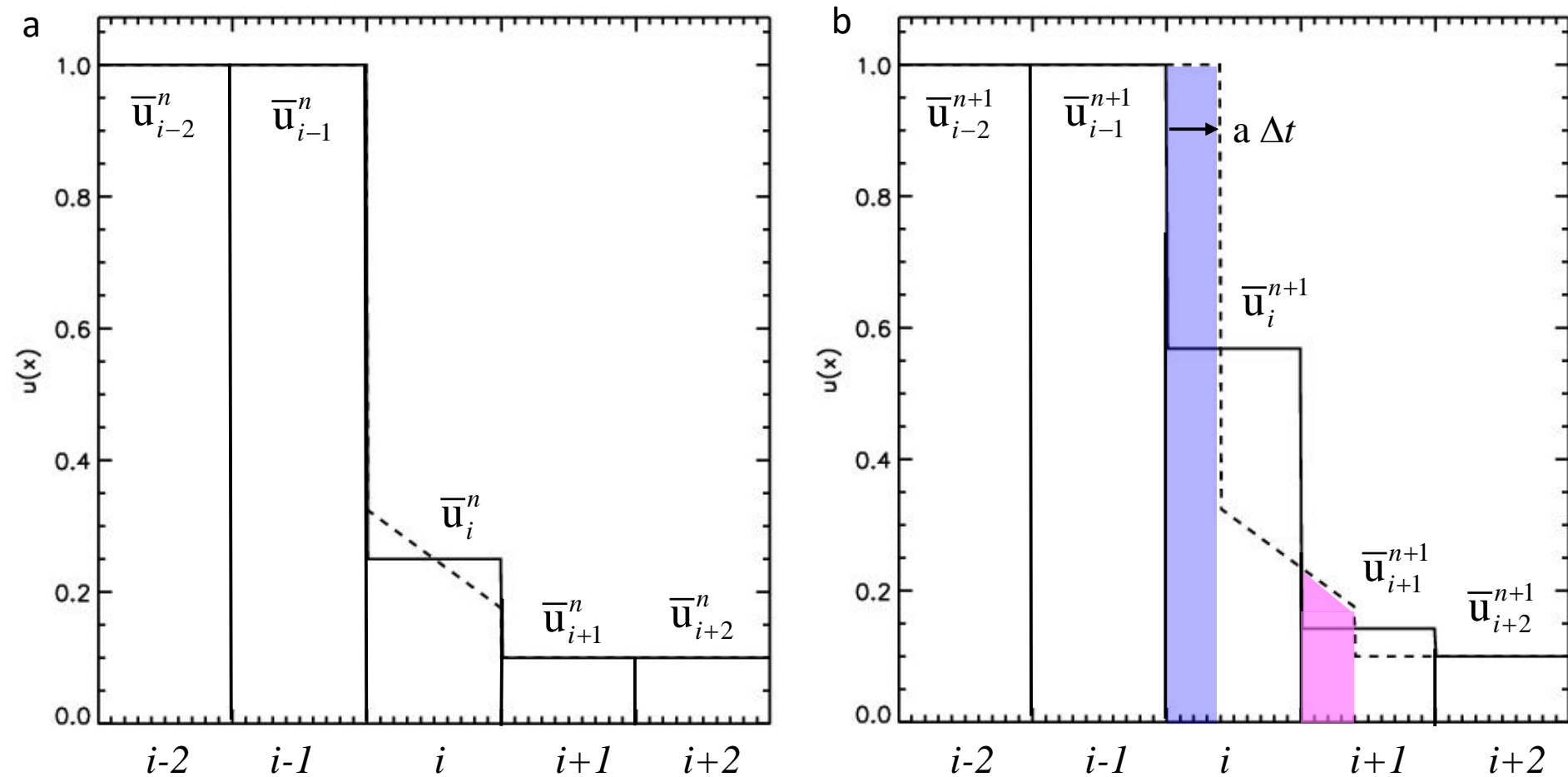
The only solution lies in restricting, i.e. *limiting*, the piecewise linear profile within each zone so that it does not produce any *new* extrema in the advected profile that were not initially present in the original slabs of fluid.

Such limiters are called *slope limiters*. A simple example – the *MinMod limiter* (i.e. choose between the left and right slopes based on which one is smaller; 0 if opposite slopes) :

$$\overline{\Delta u}_i^n = \frac{1}{2} \left(\text{sgn}(\overline{u}_{i+1}^n - \overline{u}_i^n) + \text{sgn}(\overline{u}_i^n - \overline{u}_{i-1}^n) \right) \min \left(|\overline{u}_{i+1}^n - \overline{u}_i^n|, |\overline{u}_i^n - \overline{u}_{i-1}^n| \right)$$

There are several other/better limiters in the text. All limiters are non-linear.

Thus success on linear advection came at the cost of *non-linear hybridization*.

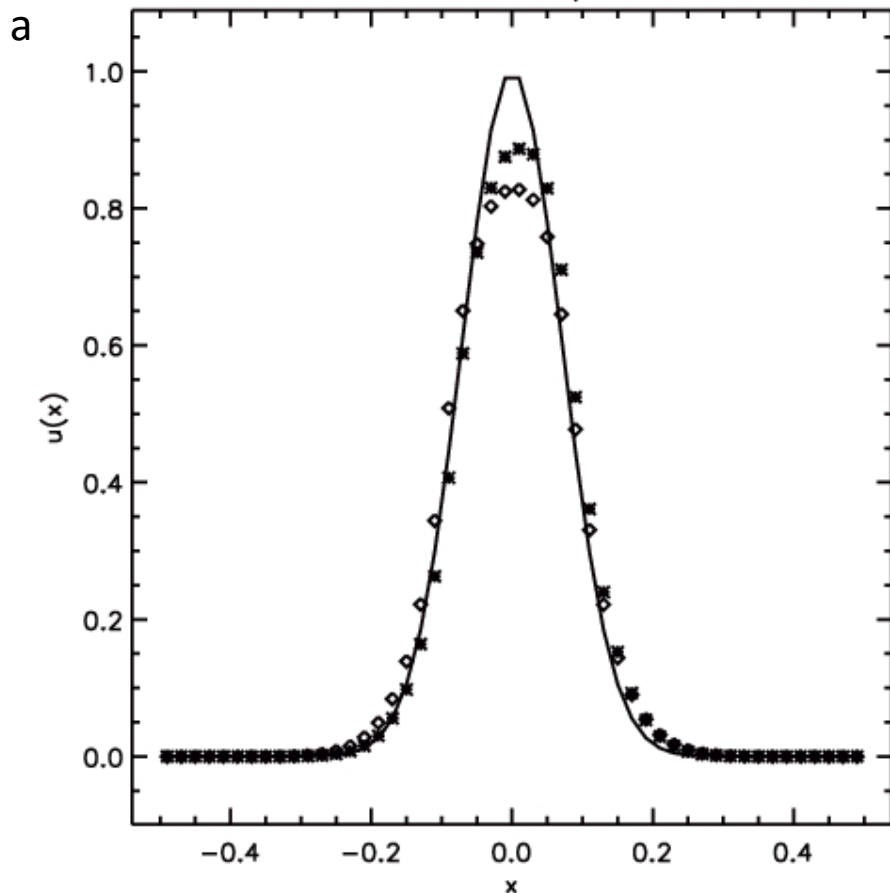


Question: Based on the minmod limiter, justify the slopes in the profile to the left.

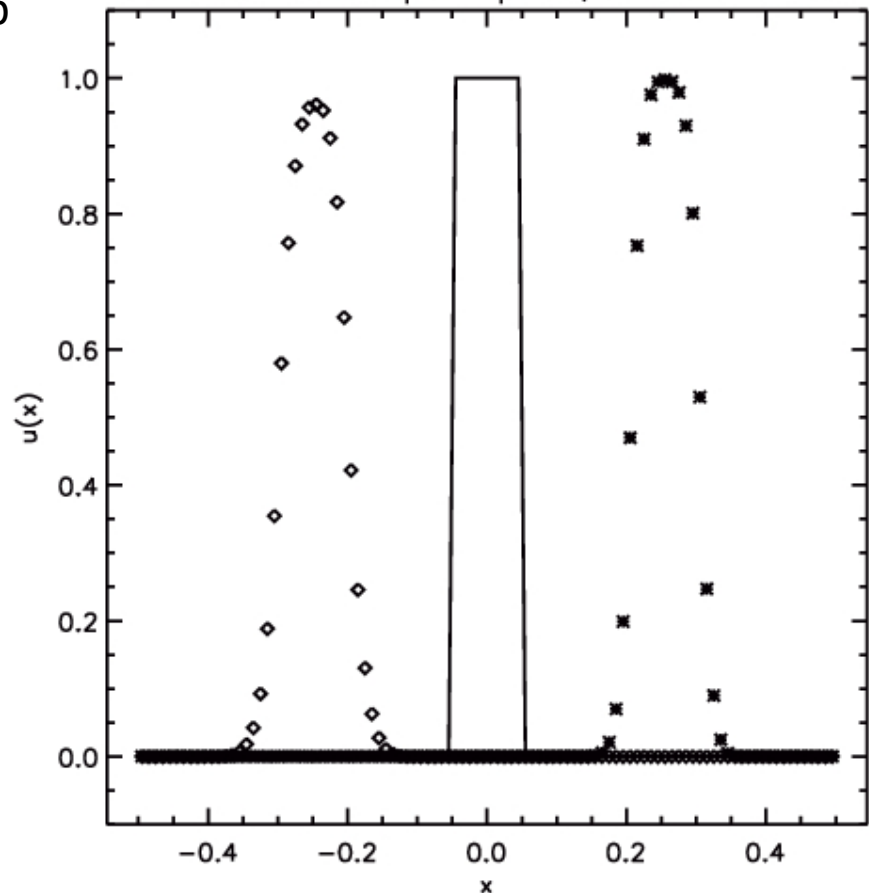
With limiter, the spurious extrema are gone!

This does come at a price: The **order of the method is locally reduced** by a limiter in those zones where it is activated. The limiter achieves its salutary effect by providing strong dissipation (and order reduction) where it is needed to prevent dispersive ripples that would otherwise form in a second order scheme.

Advection Test: Gaussian, minmod Limiter



Advection Test: Square pulse, minmod Limiter



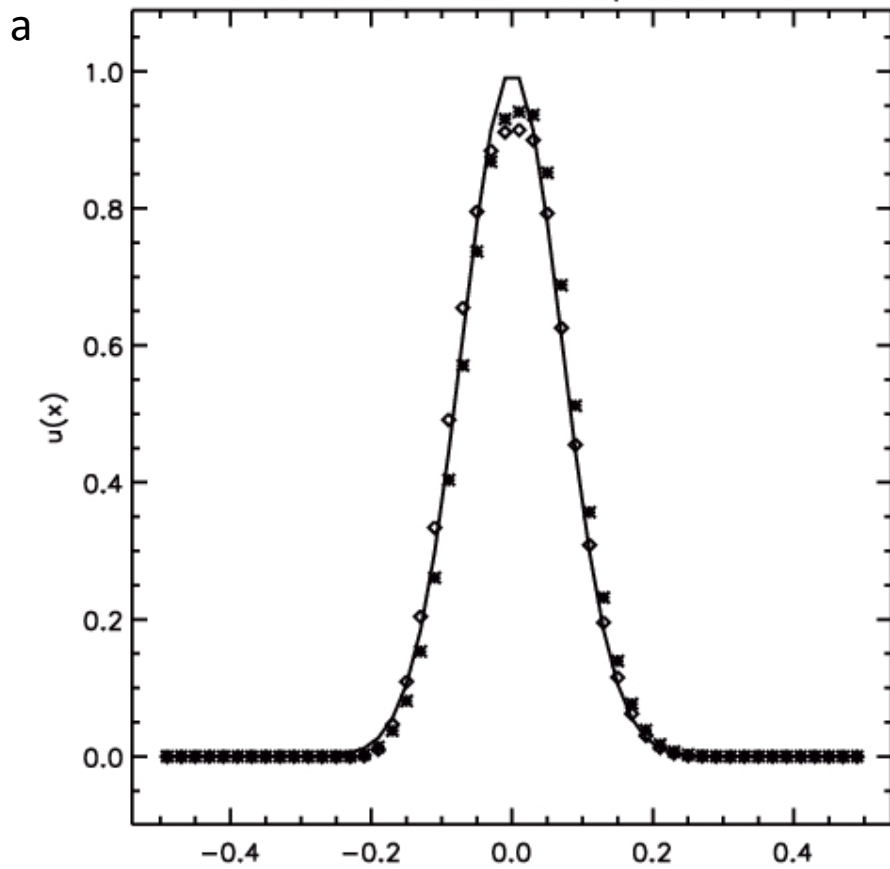
Notice that the square pulse is oscillation – free. However, the top of the Gaussian has been clipped!
 Even the Lax-Wendroff scheme did better than the non-linearly hybridized schemes. The reason :

Limiter only examines the ratio : $\theta_i = \frac{\bar{u}_i^n - \bar{u}_{i-1}^n}{\bar{u}_{i+1}^n - \bar{u}_i^n} = \frac{\text{left slope}}{\text{right slope}} \Rightarrow$ **Can't tell the difference between a physical extremum and a spurious oscillation.**

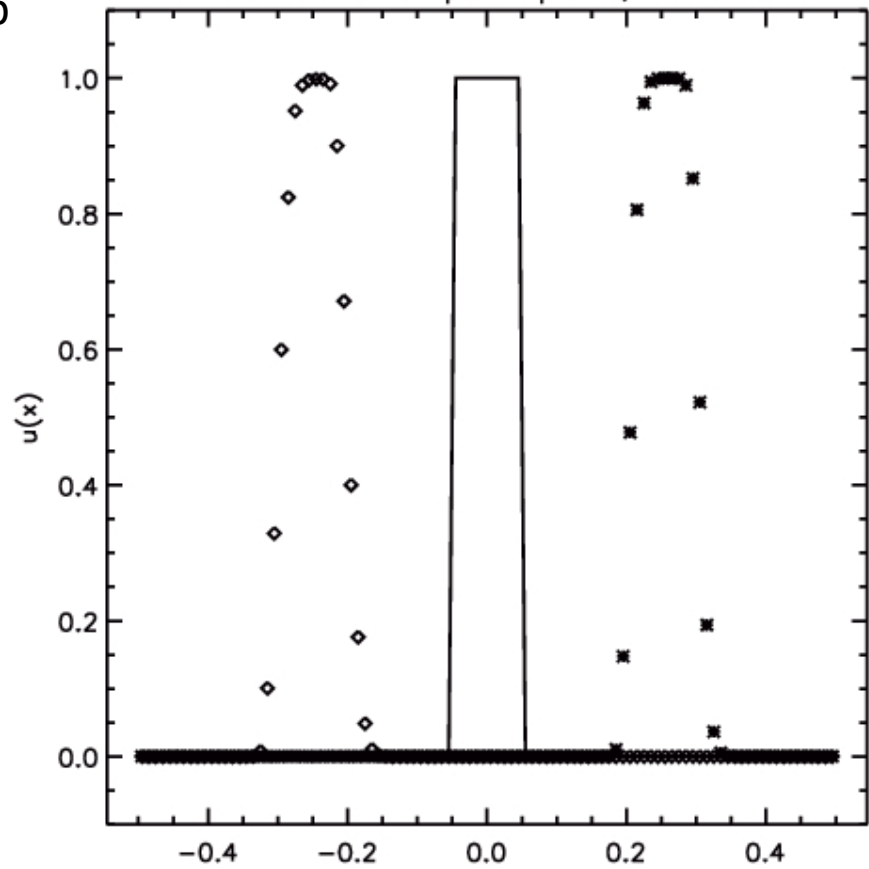
When $\theta_i \rightarrow 1$, the limiter reduces to central difference, which is good and stable.

When $\theta_i \gg 1$ or $\theta_i \ll 1$, slopes are truncated, whether that is good or not.

Advection Test: Gaussian, MC Limiter



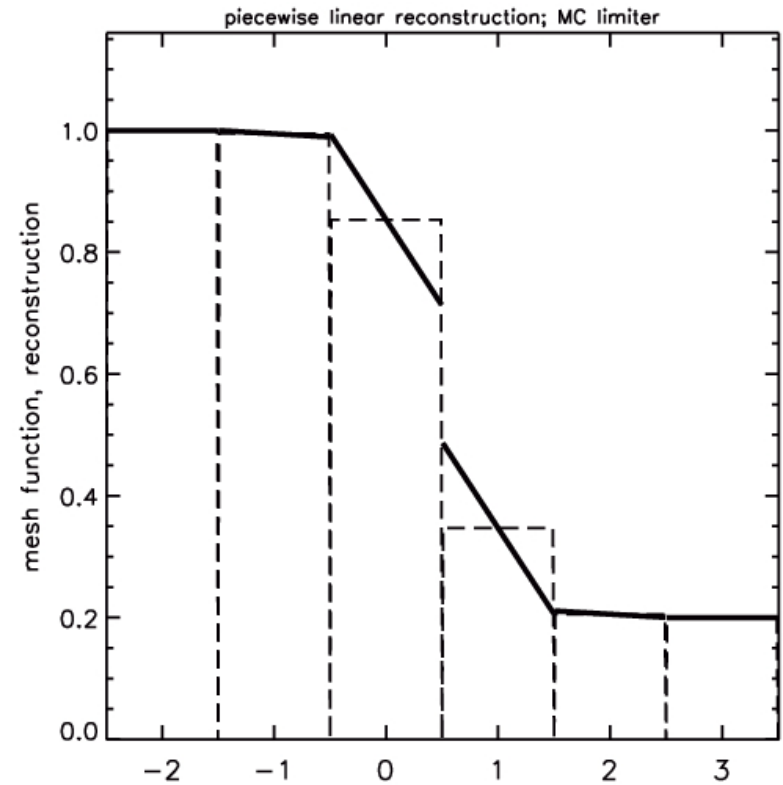
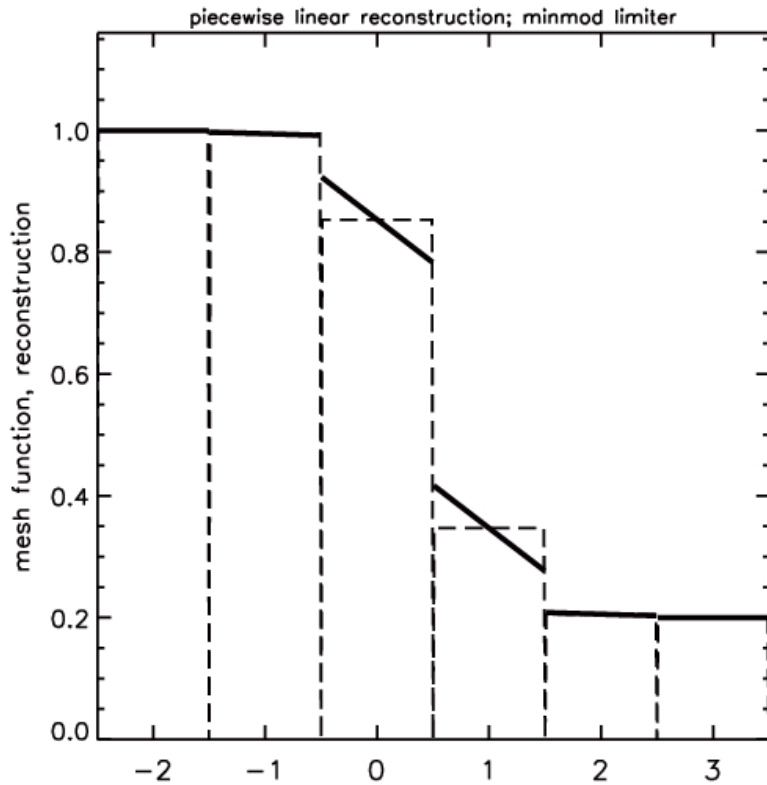
Advection Test: Square pulse, MC Limiter



All limiters clip extrema. The better ones, like the MC limiter shown above, fare better. However, even the MC limiter clips the top of the Gaussian.

Only recourse lies in schemes like WENO or PPM with larger stencils.

We call the MC limiter a *compressive limiter* because it allows the scheme to use larger slopes. As a result, it yields a sharper profile.



Comparing reconstruction by MinMod limiter to MC limiter. Latter produces steeper slopes → sharper profiles.

Questions: Can you justify the slopes to the left and right based on the form of the limiters? *Dissipation* \propto *jumps*. Which limiter dissipates less?

$$\overline{\Delta u}_i^n = \text{minmod}(\overline{u}_{i+1}^n - \overline{u}_i^n, \overline{u}_i^n - \overline{u}_{i-1}^n) = \frac{1}{2}(\text{sgn}(\overline{u}_{i+1}^n - \overline{u}_i^n) + \text{sgn}(\overline{u}_i^n - \overline{u}_{i-1}^n)) \min(|\overline{u}_{i+1}^n - \overline{u}_i^n|, |\overline{u}_i^n - \overline{u}_{i-1}^n|)$$

$$\overline{\Delta u}_i^n = \text{MC}(\overline{u}_{i+1}^n - \overline{u}_i^n, \overline{u}_i^n - \overline{u}_{i-1}^n) = \frac{1}{2}(\text{sgn}(\overline{u}_{i+1}^n - \overline{u}_i^n) + \text{sgn}(\overline{u}_i^n - \overline{u}_{i-1}^n)) \min\left(\frac{1}{2}|\overline{u}_{i+1}^n - \overline{u}_i^n|, 2|\overline{u}_{i+1}^n - \overline{u}_i^n|, 2|\overline{u}_i^n - \overline{u}_{i-1}^n|\right)$$

3.3) The Total Variation Diminishing Property and Understanding the Limiters

Monotonicity preserving property is hard to formulate mathematically; *total variation diminishing (TVD)* is much easier to formulate.

A *TVD scheme is monotonicity preserving* – Harten. Helps with *positivity*.

For a mesh function $\bar{u}^n \equiv \{ \dots, \bar{u}_{i-1}^n, \bar{u}_i^n, \bar{u}_{i+1}^n, \dots \}$ at time t^n we define the *total variation* as:

$$\text{TV}(\bar{u}^n) = \sum_{i=-\infty}^{\infty} |\bar{u}_{i+1}^n - \bar{u}_i^n|$$

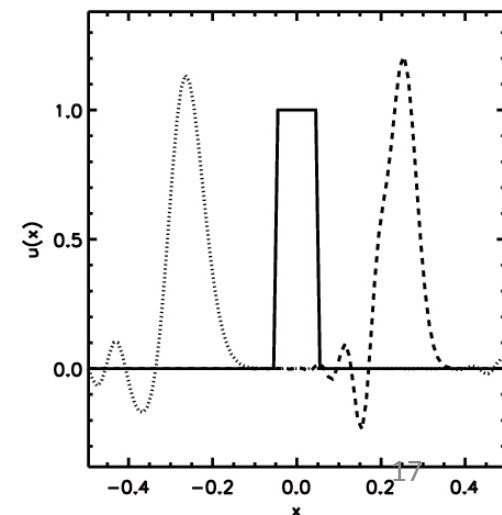
We say that the scheme is *total variation diminishing* (TVD) if we have:

$$\text{TV}(\bar{u}^{n+1}) \leq \text{TV}(\bar{u}^n)$$

Check out the graph to see that the scheme is not TVD. →

If it were TVD, could the wiggles arise?

Question: Can you evaluate TV for the plot to the right?



Harten realized that many update schemes can be written as:

$$\bar{u}_i^{n+1} = \bar{u}_i^n - C_{-,i-1/2} (\bar{u}_i^n - \bar{u}_{i-1}^n) + C_{+,i+1/2} (\bar{u}_{i+1}^n - \bar{u}_i^n)$$

If the scheme is *non-linearly stabilized*, $C_{-,i-1/2}$ and $C_{+,i+1/2}$ may have strongly non-linear dependence on the mesh function.

Harten's Theorem: When $C_{-,i-1/2} \geq 0$, $C_{+,i+1/2} \geq 0$ and $C_{-,i-1/2} + C_{+,i+1/2} \leq 1$ for all zones "i", the scheme is TVD.

The logic of the proof (omitted here) depends on the ratio:-

$$\theta_i = \frac{\bar{u}_i^n - \bar{u}_{i-1}^n}{\bar{u}_{i+1}^n - \bar{u}_i^n} = \frac{\text{left slope}}{\text{right slope}}$$

Studying our schemes in the context of Harten's theorem:

The scheme will be TVD if the fluxes have a special form.

The first order upwind scheme is monotone but has diffusive fluxes.

The fluxes for the second order schemes have extra terms. We call these extra underlined terms the *antidiffusive fluxes*. (Question: Why “anti”?) They reduce diffusion in the 2nd order scheme. But reduce dissipation carefully, or else oscillations appear.

The inclusion of the antidiffusive terms in red (underlined), turn the donor cell fluxes into the Lax-Wendroff fluxes

First order fluxes (Donor cell):

$$\bar{f}_{i-1/2}^{n+1/2} = a \bar{u}_{i-1}^n ;$$
$$\bar{f}_{i+1/2}^{n+1/2} = a \bar{u}_i^n$$

Second order fluxes (Lax-Wendroff):

$$\bar{f}_{i-1/2}^{n+1/2} = a \left[\bar{u}_{i-1}^n + \frac{1}{2}(1-\mu)(\bar{u}_i^n - \bar{u}_{i-1}^n) \right] ;$$
$$\bar{f}_{i+1/2}^{n+1/2} = a \left[\bar{u}_i^n + \frac{1}{2}(1-\mu)(\bar{u}_{i+1}^n - \bar{u}_i^n) \right]$$

The previous second order Lax-Wendroff fluxes, with their anti-diffusive parts, are not TVD!

Since the Limiter only depends on the ratio $\theta_i = \frac{\bar{u}_i^n - \bar{u}_{i-1}^n}{\bar{u}_{i+1}^n - \bar{u}_i^n}$ we write the

TVD fluxes with limiter as (Use $(\bar{u}_i^n - \bar{u}_{i-1}^n) \rightarrow \phi(\theta_{i-1}) (\bar{u}_i^n - \bar{u}_{i-1}^n)$):

$$\bar{f}_{i-1/2}^{n+1/2} = a \left[\bar{u}_{i-1}^n + \frac{1}{2}(1-\mu) \phi(\theta_{i-1}) (\bar{u}_i^n - \bar{u}_{i-1}^n) \right] ;$$

$$\bar{f}_{i+1/2}^{n+1/2} = a \left[\bar{u}_i^n + \frac{1}{2}(1-\mu) \phi(\theta_i) (\bar{u}_{i+1}^n - \bar{u}_i^n) \right]$$

We seek a limiter $\phi(\theta_i)$ with the following attributes:

- 1) $\phi(\theta) = 0$ for $\theta < 0$, i.e. to avoid extrema.
- 2) $\phi(\theta) > 0$ for $\theta > 0$, i.e. slope has correct sign.
- 3) $\phi(\theta) = 1$ for $\theta = 1$, i.e. to be second order accurate for smooth flow.

Harten's Theorem requires $0 \leq C_{-,i-1/2} \leq 1$ which for $0 \leq \mu \leq 1$ implies :

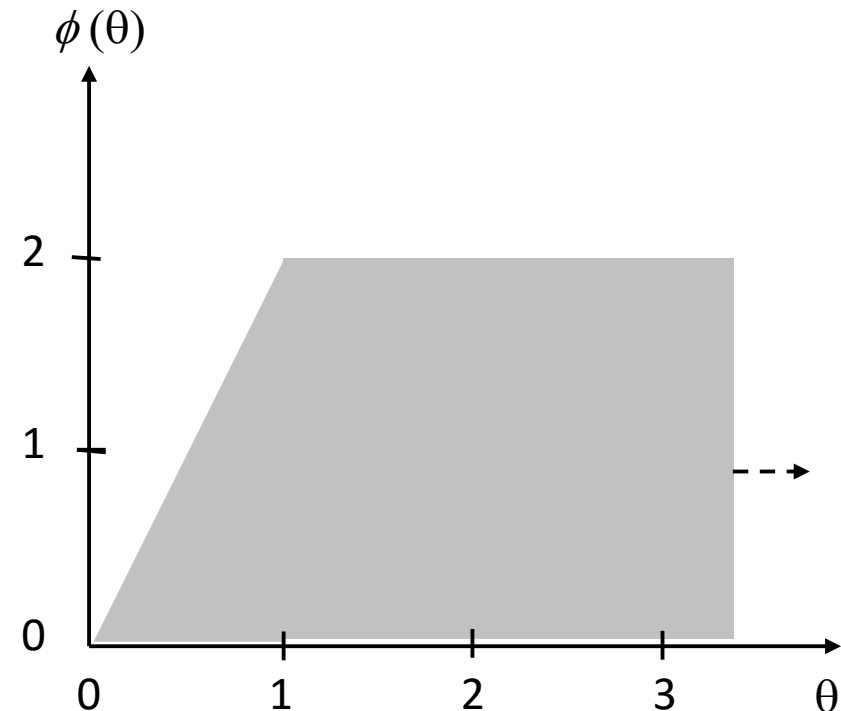
$$-2 \leq \frac{\phi(\theta_i)}{\theta_i} - \phi(\theta_{i-1}) \leq 2$$

Since θ_i and θ_{i-1} are independent, and $\phi(\theta_i)$ and $\phi(\theta_{i-1})$ are +ve, the only way out is to take:

$$0 \leq \frac{\phi(\theta)}{\theta} \leq 2 \quad \text{and} \quad 0 \leq \phi(\theta) \leq 2$$

This gives us the *TVD region*

Questions: Where does the Donor Cell scheme fit in the TVD region?
What about Lax-Wendroff scheme?
What about Beam-Warming scheme?
What about Fromm scheme?



4) The scheme should also be *second order*.

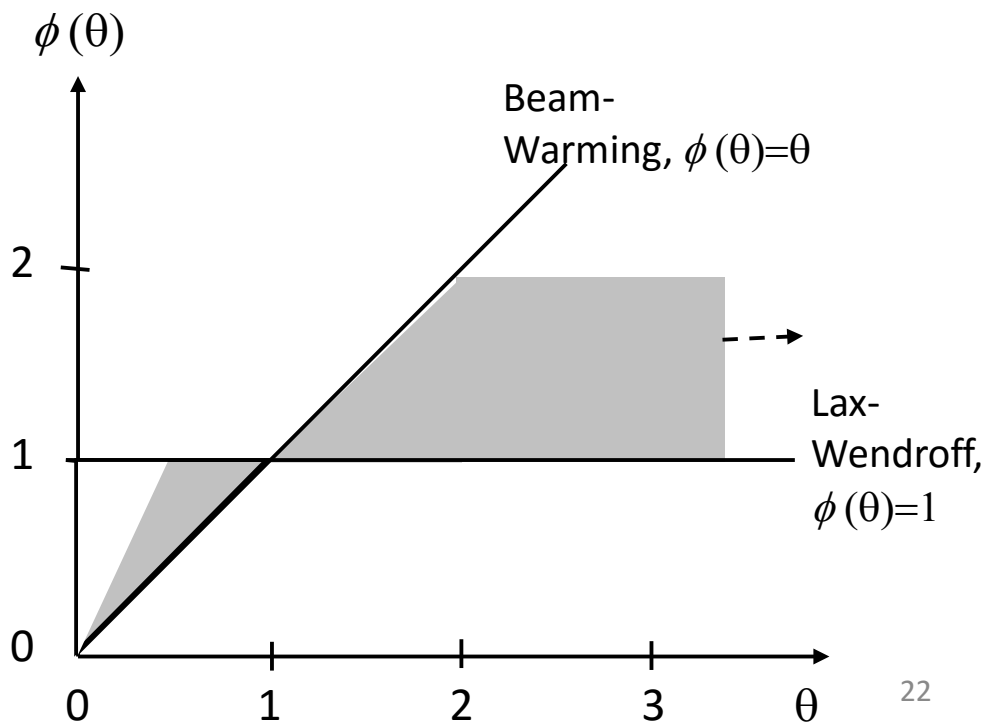
Thus $\phi(\theta)$ should lie within the limits blocked out by :

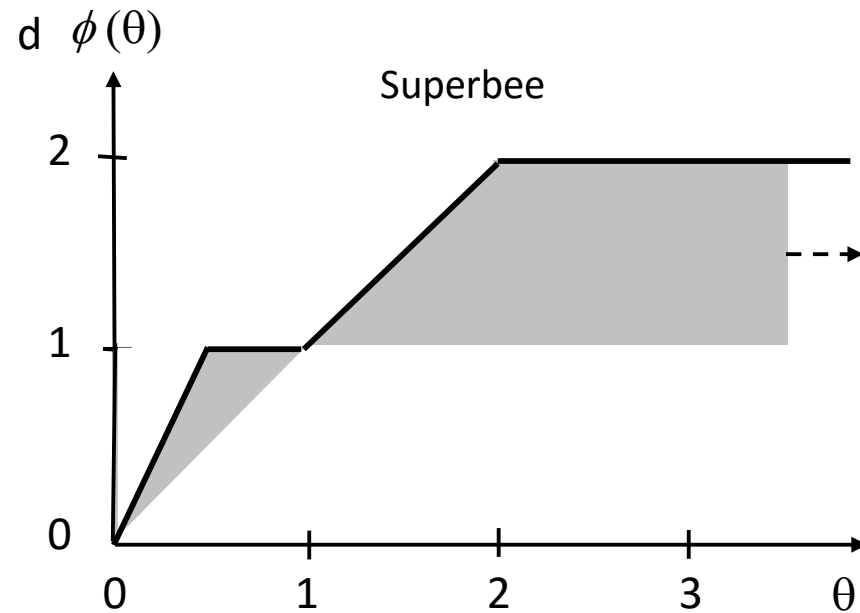
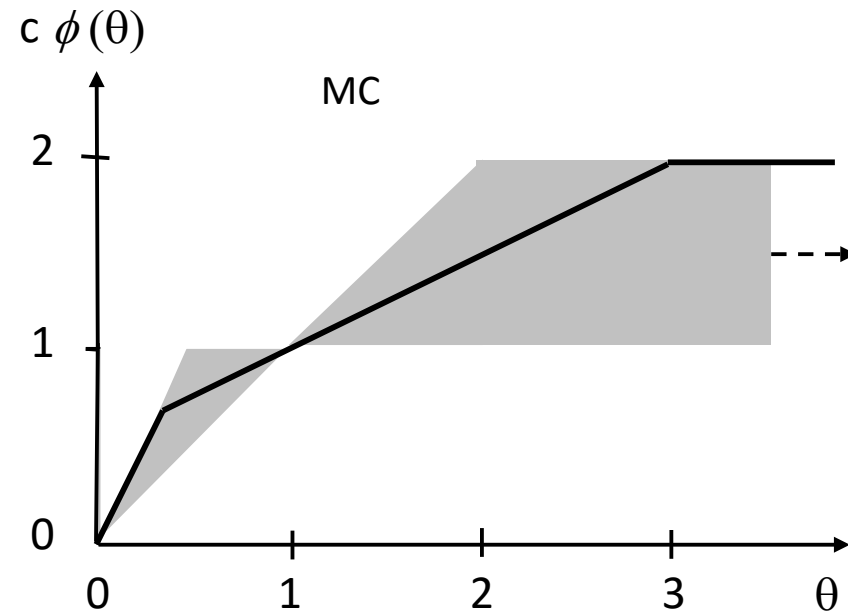
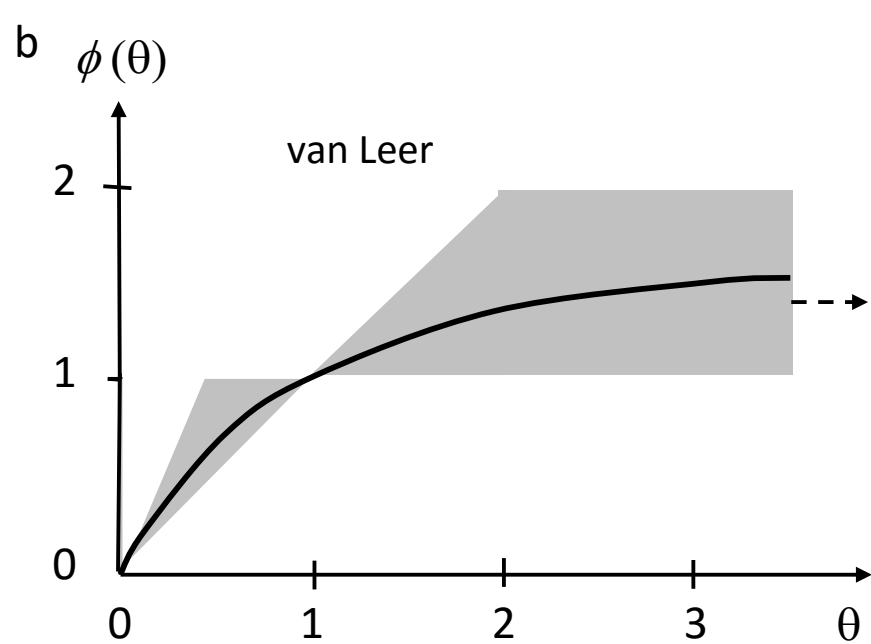
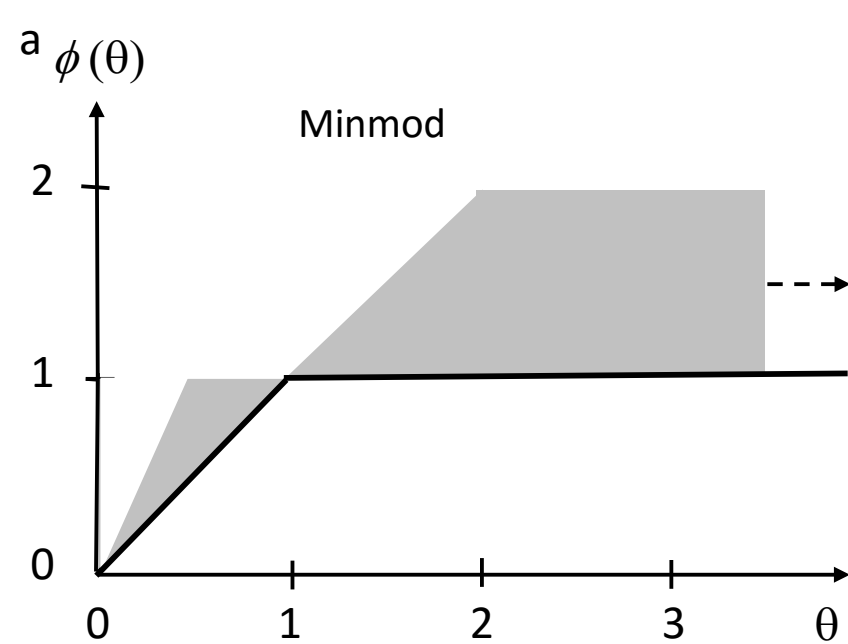
- a) the Lax-Wendroff scheme with its right-biased slopes ($\phi(\theta) = 1$)
- b) the Beam-Warming scheme with its left-biased slopes ($\phi(\theta) = \theta$)

This further restricts the TVD region (*Sweby region*)

5) The limiter should also be *symmetrical*:

$$\phi(1/\theta) = \frac{\phi(\theta)}{\theta}$$





Question: Identify the least and most compressive limiter. What are the consequences?

3.4) Linear Hyperbolic Systems and the Riemann Problem

(Review §1.5)

3.4.1) Solution of Linear Hyperbolic PDEs for Continuous Initial Data

We will need to learn how to classify different hyperbolic PDEs. We take only a first step here.

We have seen that any conservation law $U_t + F_x = 0$ can be linearized as $U_t + A U_x = 0$. For *linear hyperbolic PDEs* we literally have $F = A U$. We focus on such PDEs. A is an $M \times M$ matrix with constant entries.

The $M \times M$ hyperbolic system $U_t + A U_x = 0$ is *strictly hyperbolic* if we have real disjoint eigenvalues : $\lambda^1 < \lambda^2 < \dots < \lambda^M$.

Physically, it means that the waves are well-separated.

Mathematically, it means that we have linearly independent eigenvectors.

→ Sensible solution methods within reach.

Several very important hyperbolic systems are *non-strictly hyperbolic*, i.e. we still have real eigenvalues but : $\lambda^1 \leq \lambda^2 \leq \dots \leq \lambda^M$.

Question: Give an example of such a system.

Physically, under favorable circumstances (i.e. when we have *linearly degenerate eigenvalues*), waves may still remain well-separated.

Mathematically, we have to examine eigenstructure more closely. →
Need to work on solutions for each specific case.

We consider the simple case where “A” is a constant $M \times M$ matrix and has well-separated eigenvalues.

$$A r^m = \lambda^m r^m ; \quad l^m A = \lambda^m l^m \quad m = 1, \dots, M$$

We can build the matrix of right eigenvectors R and left eigenvectors L

as before to get: $A R = R \Lambda ; L A = \Lambda L ; L A R = \Lambda ; A = R \Lambda L$

Recall: $R = \begin{bmatrix} \boxed{r^1} & \boxed{r^2} & \cdot & \boxed{r^M} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} ; L = \begin{bmatrix} \boxed{l^1} \\ \boxed{l^2} \\ \cdot & \cdot & \cdot & \cdot \\ \boxed{l^M} \end{bmatrix} ; \Lambda = \begin{bmatrix} \lambda^1 & 0 & \cdot & 0 \\ 0 & \lambda^2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \lambda^M \end{bmatrix}$

Left-multiply $U_t + A U_x = 0$ by l^m to get : $w_t^m + \lambda^m w_x^m = 0$ for $m = 1, \dots, M$ where $w^m \equiv l^m U$ is called the *eigenweight* or *characteristic variable*.

Notice from $w_t^m + \lambda^m w_x^m = 0$ that the m^{th} characteristic variable moves with speed λ^m along the m^{th} characteristic of the hyperbolic system.

The *characteristic decomposition* described above is one of the standard building blocks used in designing numerical schemes.

Physically, It tells us that *the characteristic variables undergo simple scalar advection with a speed that is given by the *eigenvalue*.*

$$w_t^m + \lambda^m w_x^m = 0$$

It helps us solve the *Cauchy problem for hyperbolic systems*: Given *differentiable* initial conditions on *non-characteristic* surfaces in space-time, the system can be evolved further in time at least for a small time.

Thus given initial conditions $U_0(x)$, we form the characteristic variables $w_0^m(x) = l^m U_0(x)$ for $m = 1, \dots, M$ ← i.e. $w_0^m(x)$ are known functions of x .

Then to evolve the system in time, advect the profiles to get : $w_0^m(x - \lambda^m t)$

The solution at a later time is given by : $U(x, t) = \sum_{m=1}^M w_0^m(x - \lambda^m t) r^m$

3.4.2) Solution of Linear Hyperbolic PDEs for Discontinuous Initial Data: Simple waves and the Riemann Problem

Linear hyperbolic PDEs also admit discontinuous solutions (weak solutions).

For first order *conservation laws*, one can have *discontinuous solutions* even though the *differential form* of the PDE *does not admit them*.

Reason: The *integral form* of the PDE *admits such weak solutions*.

Study *weak solutions* in two stages:

a) *Simple waves*

b) *Riemann Problem*

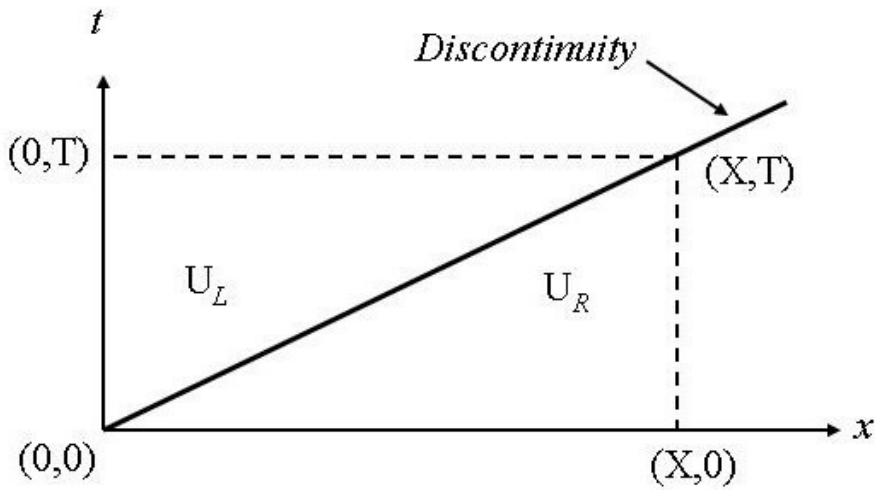


Fig. 3.9 shows the propagating discontinuity in space-time. It propagates a distance X along the x -axis in a time T . The linear hyperbolic equation is integrated over the rectangle shown by the dashed line.

Integral form $\int_{t=0}^{t=T} \int_{x=0}^{x=X} (U_t + A U_x) dx dt = 0$ gives:

$$U_L X - U_R X + A (U_R T - U_L T) = 0 \quad \Leftrightarrow \quad A (U_R - U_L) = \frac{X}{T} (U_R - U_L)$$

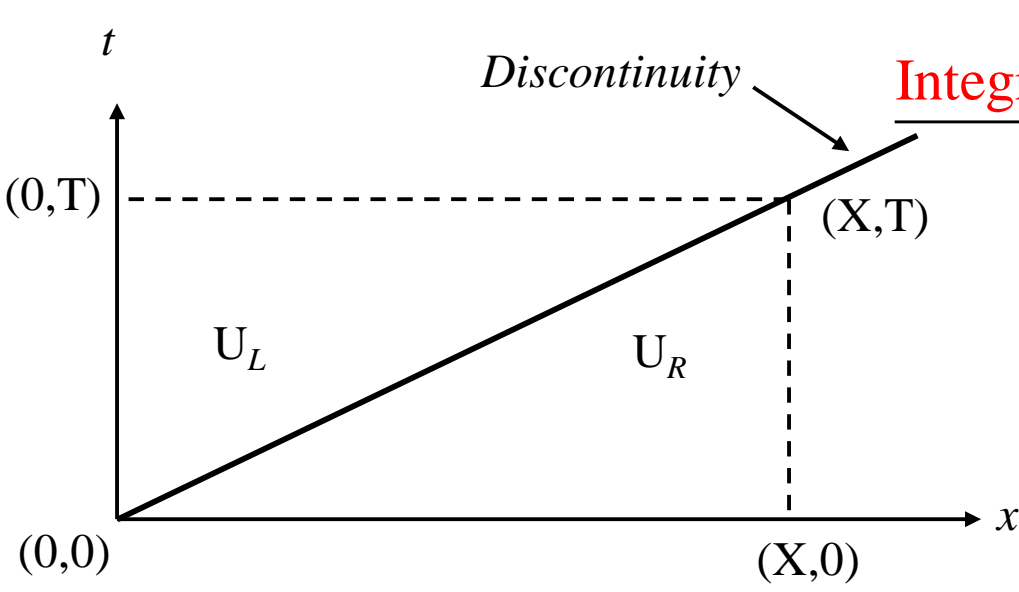
$\frac{X}{T}$ must match up with an eigenvalue of A : $A (U_R - U_L) = \lambda^m (U_R - U_L)$

Thus with initial conditions:

$$U_0(x) = U_L \quad \text{for } x < 0 \quad ; \quad U_0(x) = U_R = U_L + \alpha^m r^m \quad \text{for } x \geq 0$$

The m^{th} *simple wave* is given by:

$$U(x,t) = U_L \quad \text{for } x < \lambda^m t \quad ; \quad U(x,t) = U_R = U_L + \alpha^m r^m \quad \text{for } x \geq \lambda^m t$$



Integral form $\int_{t=0}^{t=T} \int_{x=0}^{x=X} (U_t + A U_x) dx dt = 0$

$A (U_R - U_L) = \frac{X}{T} (U_R - U_L)$ match up with $A r^m = \lambda^m r^m$

Notice: A *simple wave* is a very special wave structure where the jump $U_R - U_L$ is restricted to lie parallel to the m^{th} right eigenvector.

Question: What if we have a very general jump $U_R - U_L$ with no particular arrangement between left and right state?

Answer: Then we have to solve the *Riemann problem*!

As long as the space of right eigenvectors is complete (which is guaranteed for a strictly hyperbolic linear system) we can always make the projection:

$$U_R - U_L = \sum_{m=1}^M \alpha^m r^m \quad \text{where} \quad \alpha^m \equiv l^m (U_R - U_L)$$

α^m is an eigenweight of r^m . But what does it mean *physically*?

It means that we have a set of $M-1$ constant states between U_L and U_R where the m^{th} constant state $U^{(m)}$ lies between $x = \lambda^m t$ and $x = \lambda^{m+1} t$

Physically: We have a *similarity solution* with m simple waves.

The *Riemann problem* is a super-important *building block* for numerical schemes.

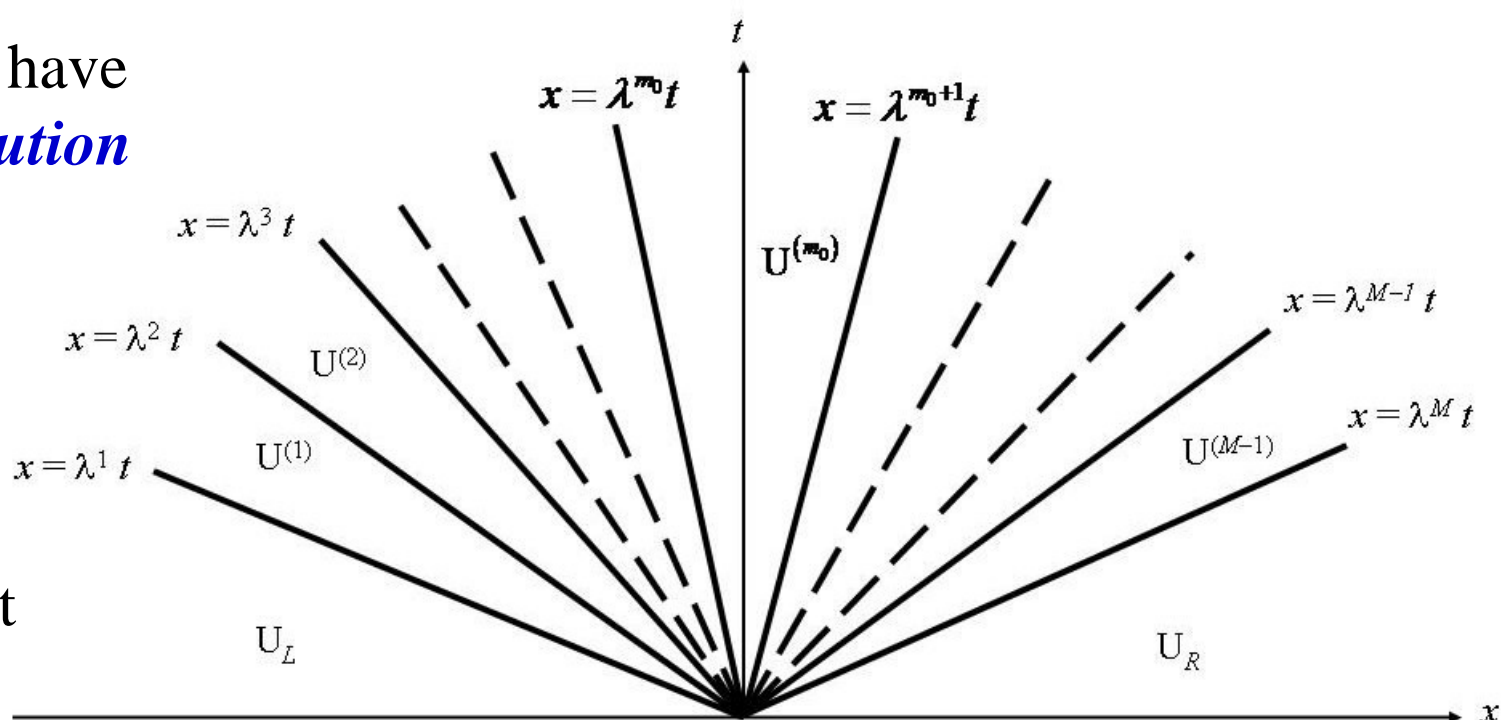
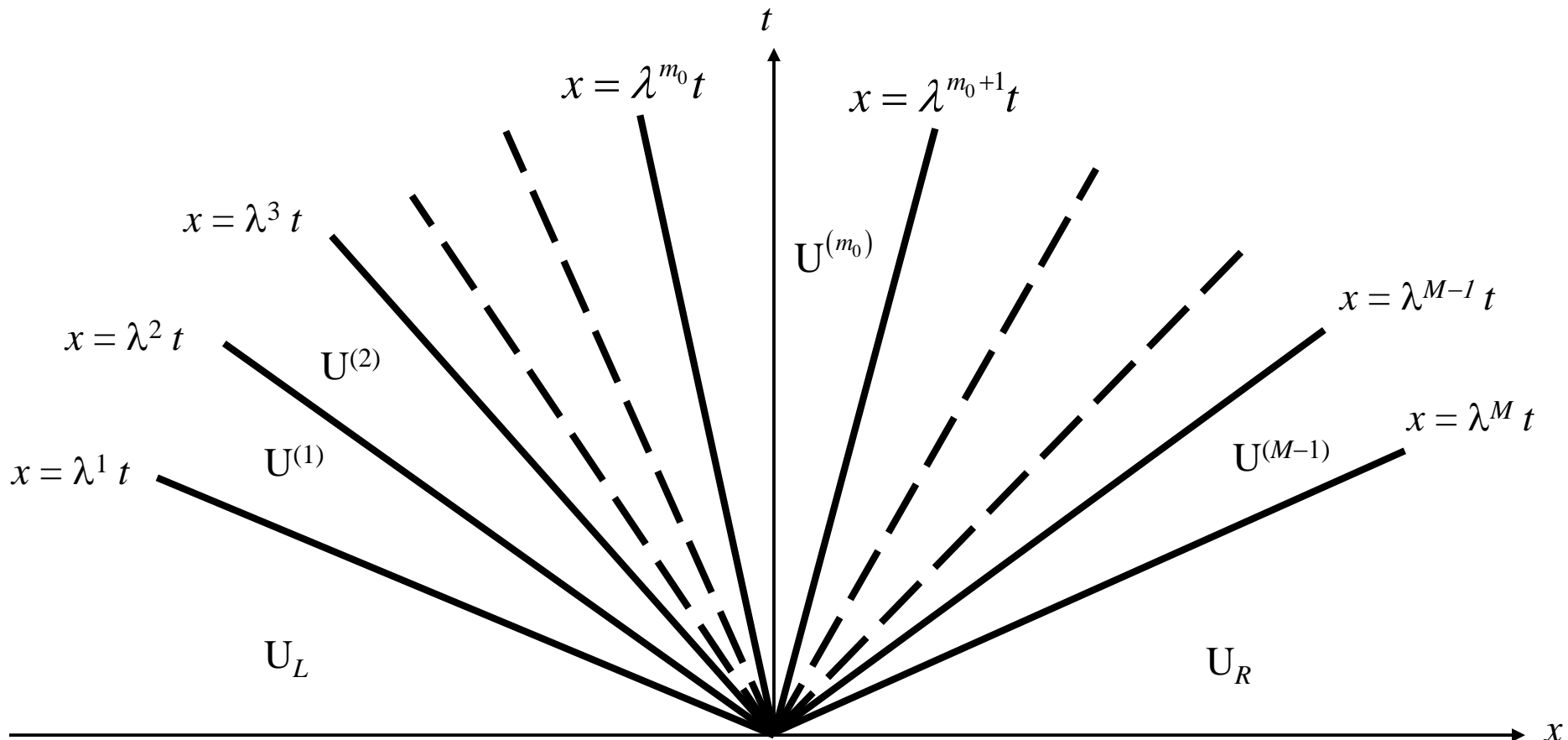


Fig. 3.10 shows the space-time diagram for the propagation of finite amplitude (or infinitesimal) perturbations for an M -component linear hyperbolic system. The solid lines show waves; the dashed lines represent the presence of further waves that may not be explicitly shown here. The left and right states are denoted by U_L and U_R . The resolved state of the Riemann problem is shown as $U^{(RS)} \equiv U^{(m_0)}$.

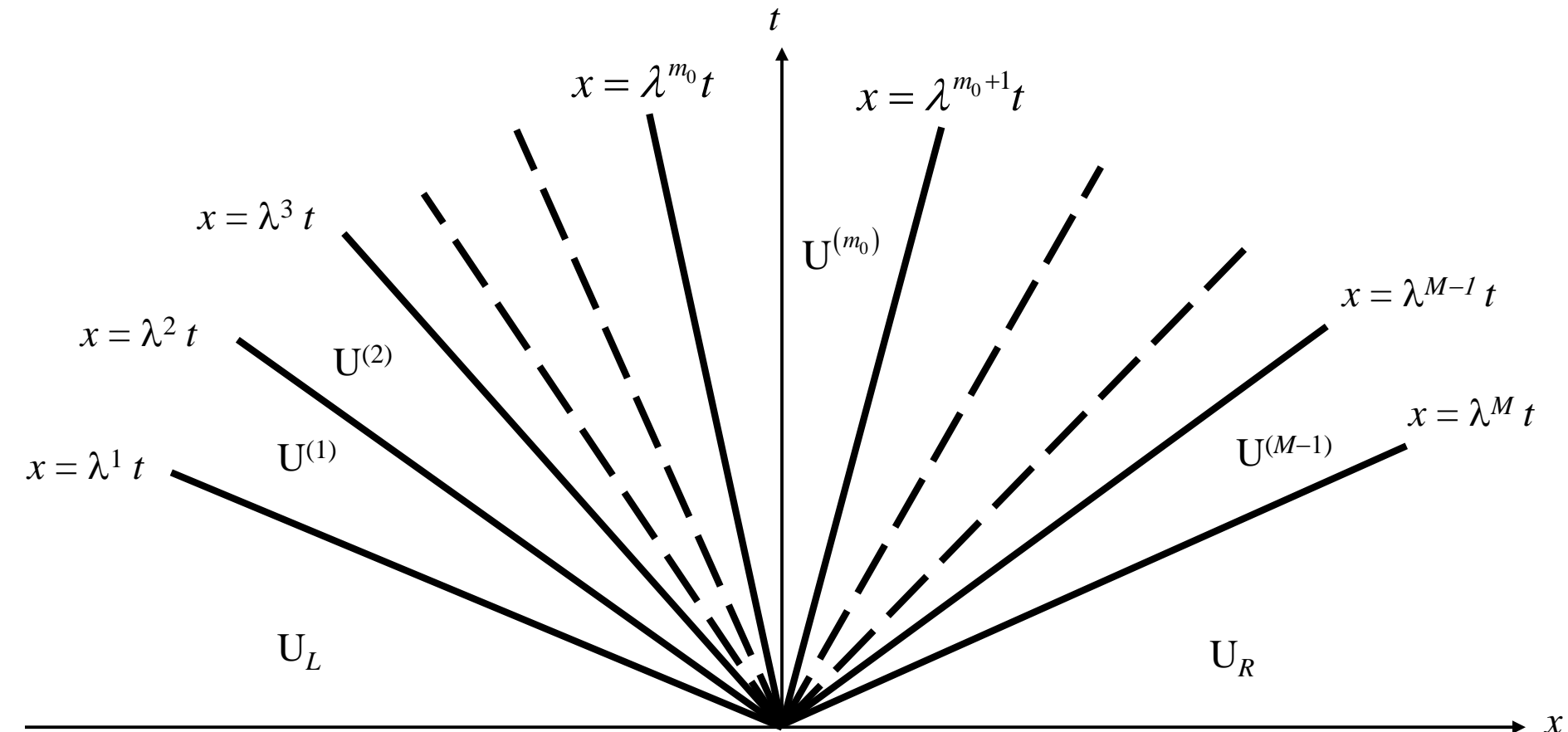
$$\begin{aligned}
 U(x, t) &= U_L && \text{for } \frac{x}{t} < \lambda^1 \\
 &= U^{(m)} \equiv U_L + \sum_{p=1}^m \alpha^p r^p = U_R - \sum_{p=m+1}^M \alpha^p r^p && \text{for } \lambda^m < \frac{x}{t} < \lambda^{m+1}, \quad m = 1, \dots, M-1 \\
 &= U_R && \text{for } \lambda^M < \frac{x}{t}
 \end{aligned}$$



$U^{(1)} =$

$U^{(2)} =$

$U^{(3)} =$



$U^{(M-1)} =$

$U^{(M-2)} =$

$U^{(M-3)} =$

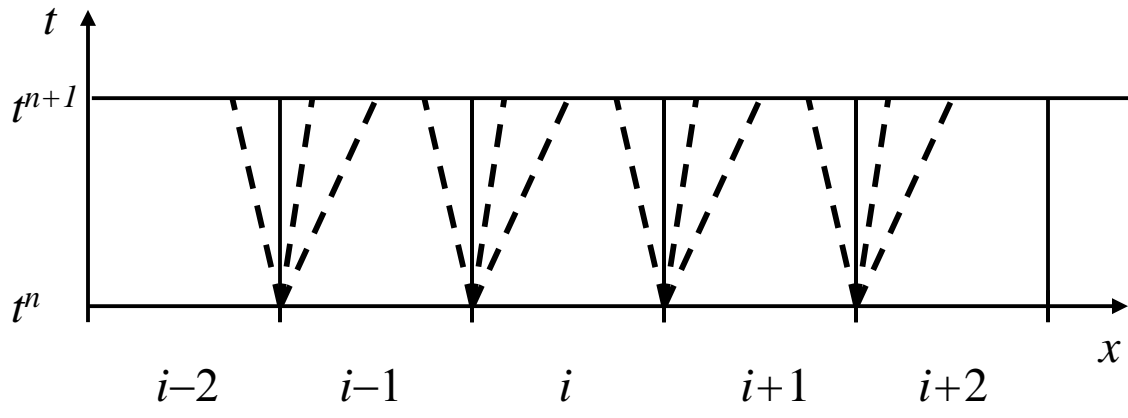
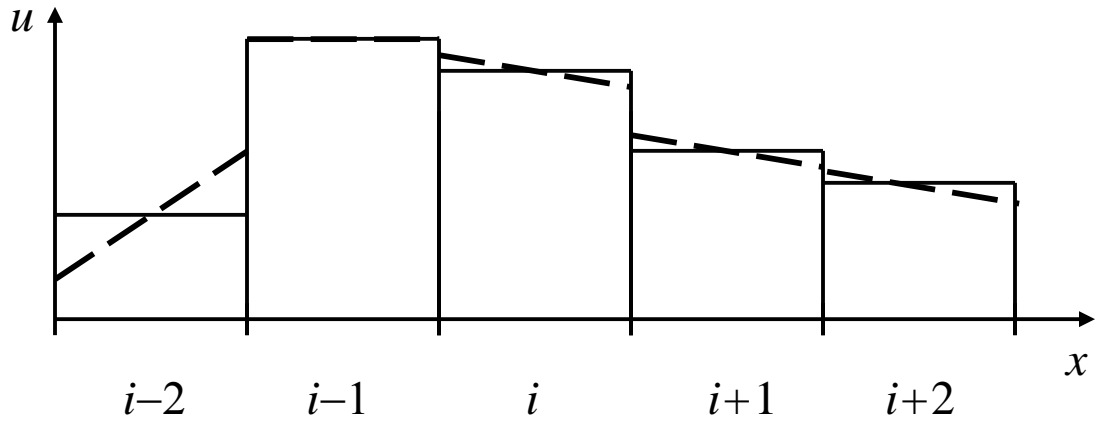
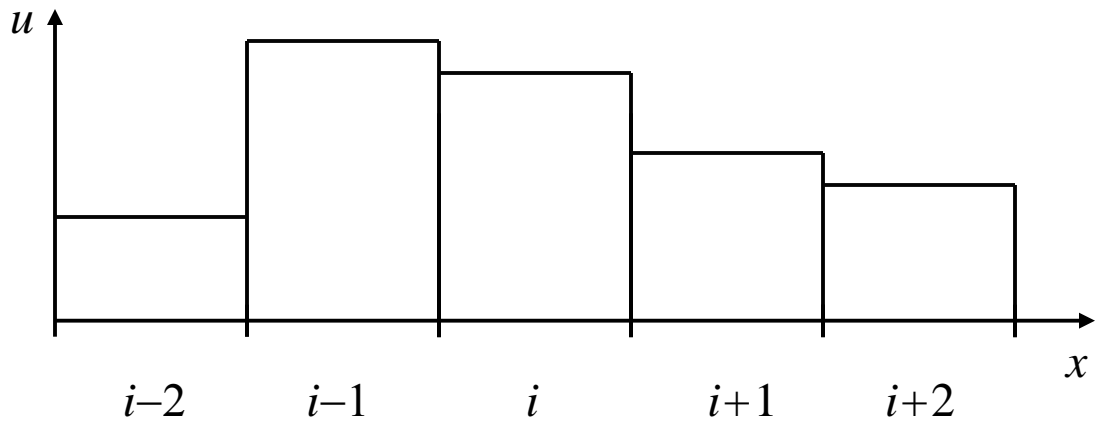
3.4.3) The Riemann Problem as a Building Block for the Numerical Solution of Hyperbolic Systems

Here we focus on formulating a first order (Godunov) scheme for linear hyperbolic systems that can handle discontinuities.

Higher order monotonicity preserving schemes will also produce *jumps at zone boundaries*, albeit the jumps will be smaller. Either way, we have to learn how to handle such jumps for the case of hyperbolic systems.

The Riemann problem from the previous section is well-suited for doing that and comes to our rescue.

The *Riemann solver* is a very important building block for schemes that solve linear and non-linear hyperbolic problems.



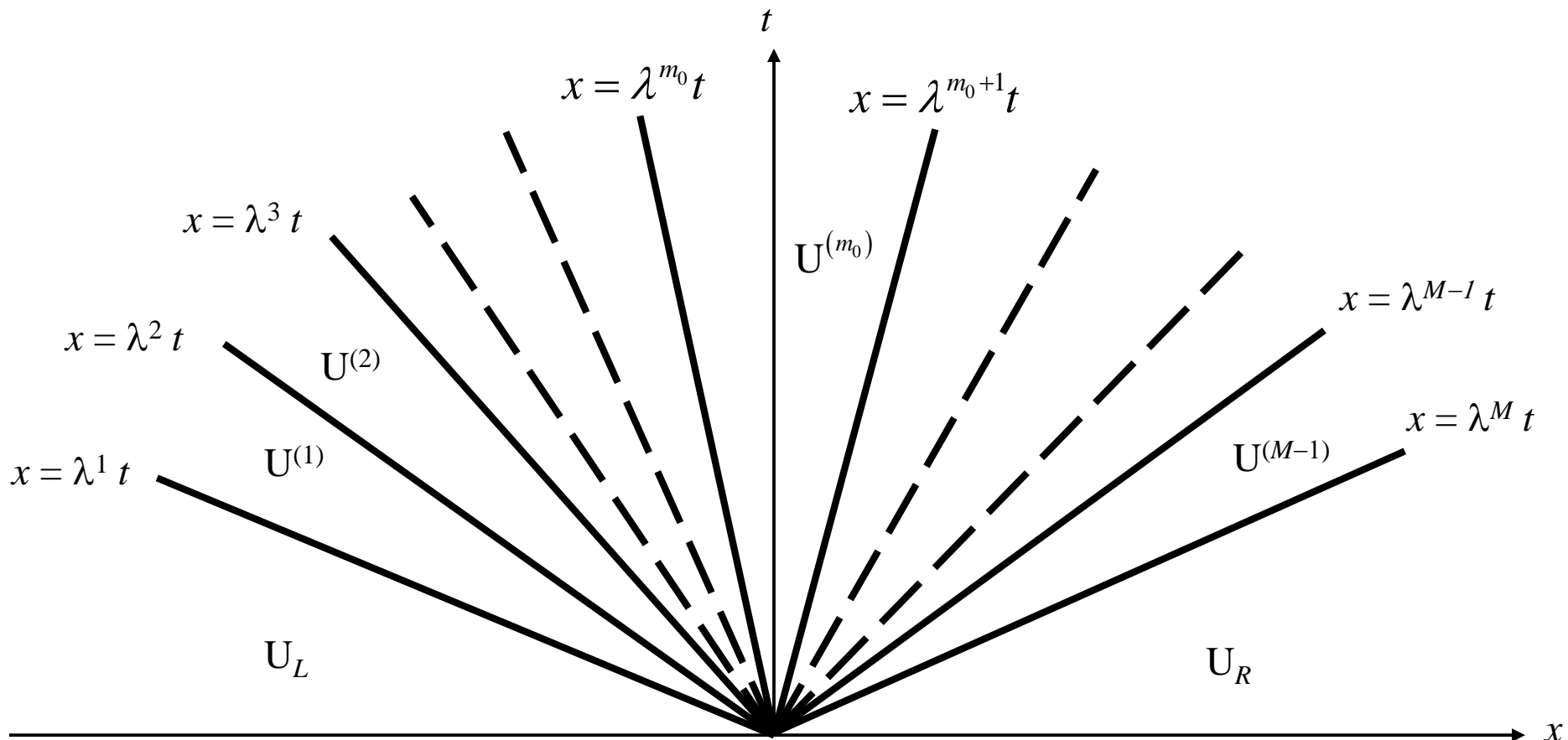
Let us start with a 1d mesh function $\{\bar{U}_i^n\}$ on a mesh with zone size Δx .

For Godunov's method, we wish to evolve the solution from time t^n to $t^n + \Delta t$:

$$\bar{U}_i^{n+1} = \bar{U}_i^n - \frac{\Delta t}{\Delta x} (\bar{F}_{i+1/2}^n - \bar{F}_{i-1/2}^n)$$

The goal is to find properly *upwinded* $\bar{F}_{i+1/2}^n$ at the zone boundaries, i.e. fluxes that build in the realization that there are discontinuities in the solution and incorporate the fact that the discontinuity at each boundary will split into a *Riemann fan* of simple waves that move in different directions..

Shift coordinate system so that the origin lies at the zone boundary of interest.



We seek the **resolved state**, i.e. the solution of the Riemann problem that overlies the zone boundary $x=0$ of interest. It will then be easy to average that state in space and time to get the **numerical** or **resolved flux**. I.e. we want state $\mathbf{U}^{(RS)} \equiv \mathbf{U}^{(m_0)}$ with the property that $\lambda^{m_0} < 0 \leq \lambda^{m_0+1}$ See previous fig. Find state that overlies time axis.

$$\begin{aligned} \mathbf{U}^{(RS)} \equiv \mathbf{U}^{(m_0)} &= \mathbf{U}_L && \text{if } 0 < \lambda^1 \\ &= \mathbf{U}_L + \sum_{p=1}^{m_0} \alpha^p r^p = \mathbf{U}_R - \sum_{p=m_0+1}^M \alpha^p r^p && \text{if } \lambda^{m_0} < 0 \leq \lambda^{m_0+1} \\ &= \mathbf{U}_R && \text{if } \lambda^M \leq 0 \end{aligned}$$

Notice that the resolved state is bounded by the characteristics $x = \lambda^{m_0} t$ and $x = \lambda^{m_0+1} t$. Obtaining the resolved flux requires us to *distinguish* between waves that move to the right and those that move to the left.

$$\mathbf{F}^{(RS)} \equiv \mathbf{A} \mathbf{U}^{(RS)}$$

Because the resolved flux is of *great* importance to us, we obtain compact, computationally efficient expressions for it.

Distinguishing between waves that move to the right and those that move to the left is most easily done by defining:

$$\lambda^{+,m} \equiv \max(\lambda^m, 0) \quad ; \quad \lambda^{-,m} \equiv \min(\lambda^m, 0)$$

The left and right fluxes are also best written as: $F_L \equiv A U_L$; $F_R \equiv A U_R$

Automatic expressions for evaluating the resolved flux are then obtained as:

$$\begin{aligned} F^{(RS)} &= F_L + \sum_{m=1}^M \lambda^{-,m} \alpha^m r^m \\ &= F_R - \sum_{m=1}^M \lambda^{+,m} \alpha^m r^m \\ &= \frac{1}{2}(F_R + F_L) - \frac{1}{2} \sum_{m=1}^M |\lambda^m| \alpha^m r^m \end{aligned}$$

Question: What makes the eqns. above so desirable for *computer implementation*?

Set $\alpha^m \equiv l^m (U_R - U_L)$ in the above formulae to obtain even more compact expressions.

$$U^{(RS)} \equiv U^{(m_0)} = U_L + \sum_{p=1}^{m_0} \alpha^p r^p \quad \text{if } \lambda^{m_0} < 0 \leq \lambda^{m_0+1}$$

To find the flux, do $F^{(RS)} \equiv A U^{(RS)}$

$$F^{(RS)} = F_L + \sum_{m=1}^M \lambda^{-,m} \alpha^m r^m \quad \text{with} \quad \lambda^{-,m} \equiv \min(\lambda^m, 0)$$

$$U^{(RS)} \equiv U^{(m_0)} = U_R - \sum_{p=m_0+1}^M \alpha^p r^p \quad \text{if } \lambda^{m_0} < 0 \leq \lambda^{m_0+1}$$

To find the flux, do $F^{(RS)} \equiv A U^{(RS)}$

$$F^{(RS)} = F_R - \sum_{m=1}^M \lambda^{+,m} \alpha^m r^m \quad \text{with} \quad \lambda^{+,m} \equiv \max(\lambda^m, 0)$$

$$\begin{aligned}
\mathbf{F}^{(RS)} &= \mathbf{F}_L + \sum_{m=1}^M \lambda^{-,m} \alpha^m r^m \\
&= \mathbf{F}_R - \sum_{m=1}^M \lambda^{+,m} \alpha^m r^m
\end{aligned}$$

$$\mathbf{F}^{(RS)} = \frac{1}{2}(\mathbf{F}_R + \mathbf{F}_L) - \frac{1}{2} \sum_{m=1}^M |\lambda^m| \alpha^m r^m$$

Our final *computer-friendly expressions for the resolved flux* are:

$$\begin{aligned} \mathbf{F}^{(RS)} &= \mathbf{F}_L + \mathbf{A}^- (\mathbf{U}_R - \mathbf{U}_L) \\ &= \mathbf{F}_R - \mathbf{A}^+ (\mathbf{U}_R - \mathbf{U}_L) \\ &= \frac{1}{2}(\mathbf{F}_R + \mathbf{F}_L) - \frac{1}{2}|\mathbf{A}|(\mathbf{U}_R - \mathbf{U}_L) \end{aligned}$$

with the definitions:

$$\Lambda^+ \equiv \text{diag} \{ \lambda^{+,1}, \lambda^{+,2}, \dots, \lambda^{+,M} \} \quad ; \quad \Lambda^- \equiv \text{diag} \{ \lambda^{-,1}, \lambda^{-,2}, \dots, \lambda^{-,M} \} \quad ;$$

$$|\Lambda| \equiv \text{diag} \{ |\lambda^1|, |\lambda^2|, \dots, |\lambda^M| \} \quad ;$$

$$\mathbf{A}^+ \equiv \mathbf{R}\Lambda^+\mathbf{L} \quad ; \quad \mathbf{A}^- \equiv \mathbf{R}\Lambda^-\mathbf{L} \quad ; \quad |\mathbf{A}| \equiv \mathbf{R}|\Lambda|\mathbf{L}$$

The matrices \mathbf{A}^+ , \mathbf{A}^- and $|\mathbf{A}|$ are evaluated once and for all for a given linear hyperbolic system!

$$\mathbf{F}^{(RS)} = \mathbf{F}_L + \sum_{m=1}^M \lambda^{-,m} \alpha^m r^m \quad \text{with} \quad \alpha^m = l^m (\mathbf{U}_R - \mathbf{U}_L)$$

$$\mathbf{F}^{(RS)} = \mathbf{F}_L + \mathbf{A}^- (\mathbf{U}_R - \mathbf{U}_L) \quad \text{with} \quad \mathbf{A}^- = R\Lambda^-L$$

$$\mathbf{F}^{(RS)} = \mathbf{F}_R - \sum_{m=1}^M \lambda^{+,m} \alpha^m r^m \quad \text{with} \quad \alpha^m = l^m (\mathbf{U}_R - \mathbf{U}_L)$$

$$\mathbf{F}^{(RS)} = \mathbf{F}_R - \mathbf{A}^+ (\mathbf{U}_R - \mathbf{U}_L) \quad \text{with} \quad \mathbf{A}^+ = R\Lambda^+L$$

Notice that we can now show *consistency* of our scheme.

$F^{(RS)} \rightarrow F(\bar{U})$ is guaranteed when $U_L \rightarrow \bar{U}$ and $U_R \rightarrow \bar{U}$.

As a result, the FDA $\bar{U}_i^{n+1} = \bar{U}_i^n - \frac{\Delta t}{\Delta x} (\bar{F}_{i+1/2}^n - \bar{F}_{i-1/2}^n)$ will converge to the PDE $U_t + A U_x = 0$.

The present Riemann solver goes through with small modifications for non-linear hyperbolic systems, hence its *great* importance.

Example: The linearized Euler system

$$\frac{\partial}{\partial t} \begin{pmatrix} \rho \\ v_x \\ P \end{pmatrix} + \begin{pmatrix} v_{x0} & \rho_0 & 0 \\ 0 & v_{x0} & 1 \\ 0 & \rho_0 c_0^2 & v_{x0} \end{pmatrix} \frac{\partial}{\partial x} \begin{pmatrix} \rho \\ v_x \\ P \end{pmatrix} = 0$$

Recall the *characteristic analysis* $\Rightarrow \lambda^1 = v_{x0} - c_0$; $\lambda^2 = v_{x0}$; $\lambda^3 = v_{x0} + c_0$

$$r^1 = \begin{pmatrix} \rho_0 \\ -c_0 \\ \rho_0 c_0^2 \end{pmatrix} ; \quad r^2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} ; \quad r^3 = \begin{pmatrix} \rho_0 \\ c_0 \\ \rho_0 c_0^2 \end{pmatrix}$$

$$l^1 = \begin{pmatrix} 0 & \frac{-1}{2c_0} & \frac{1}{2\rho_0 c_0^2} \end{pmatrix} ; \quad l^2 = \begin{pmatrix} 1 & 0 & -\frac{1}{c_0^2} \end{pmatrix} ; \quad l^3 = \begin{pmatrix} 0 & \frac{1}{2c_0} & \frac{1}{2\rho_0 c_0^2} \end{pmatrix}$$

Now consider the Riemann problem with $U_L = (\rho_L, v_{xL}, P_L)^T$ and $U_R = (\rho_R, v_{xR}, P_R)^T$.

Our *eigenweights* are given by: $\alpha^1 \equiv \frac{-1}{2c_0}(v_{xR} - v_{xL}) + \frac{1}{2\rho_0 c_0^2}(P_R - P_L)$;

$$\alpha^2 \equiv (\rho_R - \rho_L) - \frac{1}{c_0^2}(P_R - P_L) ; \quad \alpha^3 \equiv \frac{1}{2c_0}(v_{xR} - v_{xL}) + \frac{1}{2\rho_0 c_0^2}(P_R - P_L)$$

$$l^1 = \begin{pmatrix} 0 & \frac{-1}{2 c_0} & \frac{1}{2 \rho_0 c_0^2} \end{pmatrix} ; \quad l^2 = \begin{pmatrix} 1 & 0 & -\frac{1}{c_0^2} \end{pmatrix} ; \quad l^3 = \begin{pmatrix} 0 & \frac{1}{2 c_0} & \frac{1}{2 \rho_0 c_0^2} \end{pmatrix}$$

Now consider the Riemann problem with $U_L = \begin{pmatrix} \rho_L \\ v_{xL} \\ P_L \end{pmatrix}$ and $U_R = \begin{pmatrix} \rho_R \\ v_{xR} \\ P_R \end{pmatrix}$.

Our *eigenweights* are given by: $\alpha^1 \equiv \frac{-1}{2 c_0} (v_{xR} - v_{xL}) + \frac{1}{2 \rho_0 c_0^2} (P_R - P_L) ;$

$\alpha^2 \equiv (\rho_R - \rho_L) - \frac{1}{c_0^2} (P_R - P_L) ; \quad \alpha^3 \equiv \frac{1}{2 c_0} (v_{xR} - v_{xL}) + \frac{1}{2 \rho_0 c_0^2} (P_R - P_L)$

The solution to the *Riemann Problem* is then given by:

$$\begin{aligned}
 U(x, t) &= U_L && \text{for } \frac{x}{t} < v_{x0} - c_0 \\
 &= U^{(1)} \equiv U_L + \alpha^1 r^1 = U_R - \alpha^2 r^2 - \alpha^3 r^3 && \text{for } v_{x0} - c_0 < \frac{x}{t} < v_{x0} \\
 &= U^{(2)} \equiv U_L + \alpha^1 r^1 + \alpha^2 r^2 = U_R - \alpha^3 r^3 && \text{for } v_{x0} < \frac{x}{t} < v_{x0} + c_0 \\
 &= U_R && \text{for } v_{x0} + c_0 < \frac{x}{t}
 \end{aligned}$$

Taking the transonic case $0 < v_{x0} < c_0$ we get $m_0=1$.

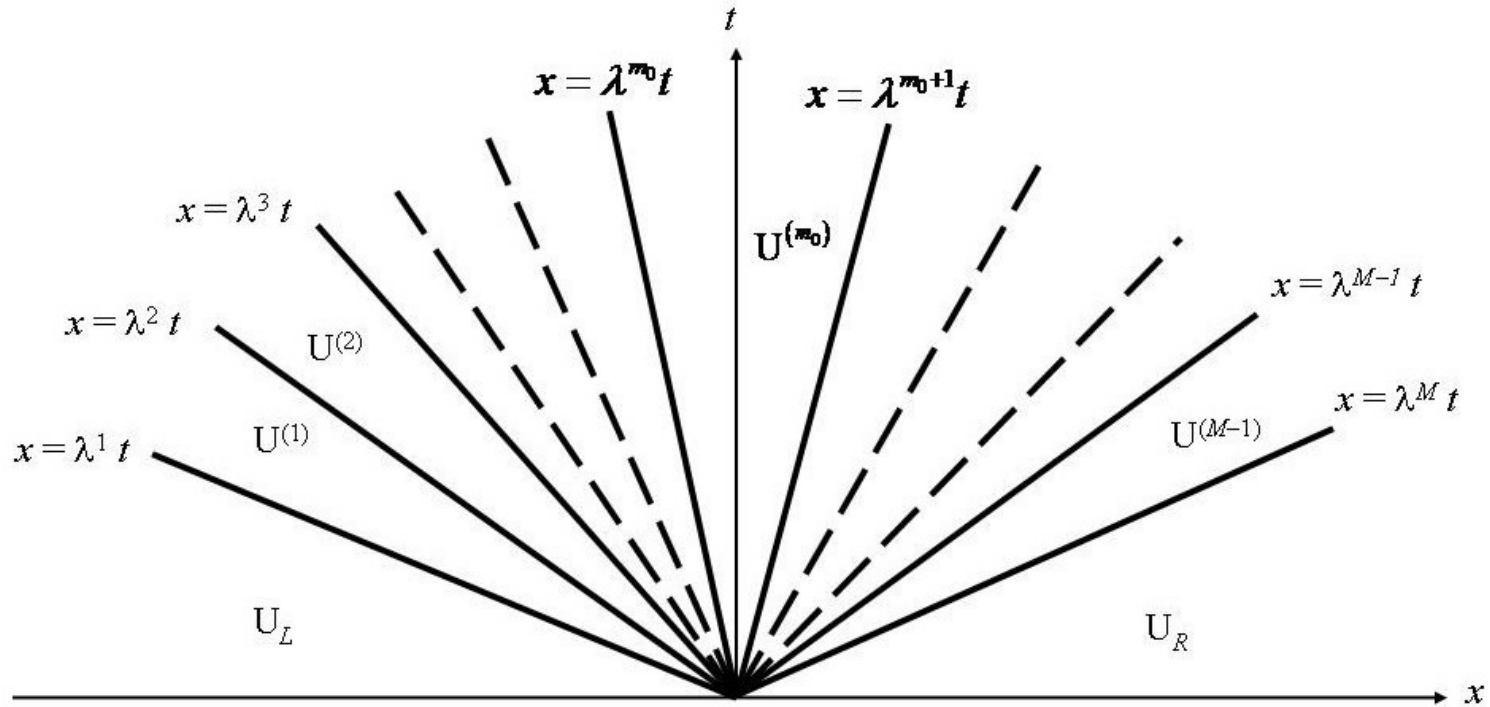
The *Resolved Flux* is now given by:

$$\begin{aligned}
 F^{(RS)} &= A U^{(1)} = A U_L + \lambda^1 \alpha^1 r^1 \\
 &= A U_R - \lambda^2 \alpha^2 r^2 - \lambda^3 \alpha^3 r^3 \\
 &= \frac{1}{2} (A U_R + A U_L) - \frac{1}{2} (|\lambda^1| \alpha^1 r^1 + |\lambda^2| \alpha^2 r^2 + |\lambda^3| \alpha^3 r^3)
 \end{aligned}$$

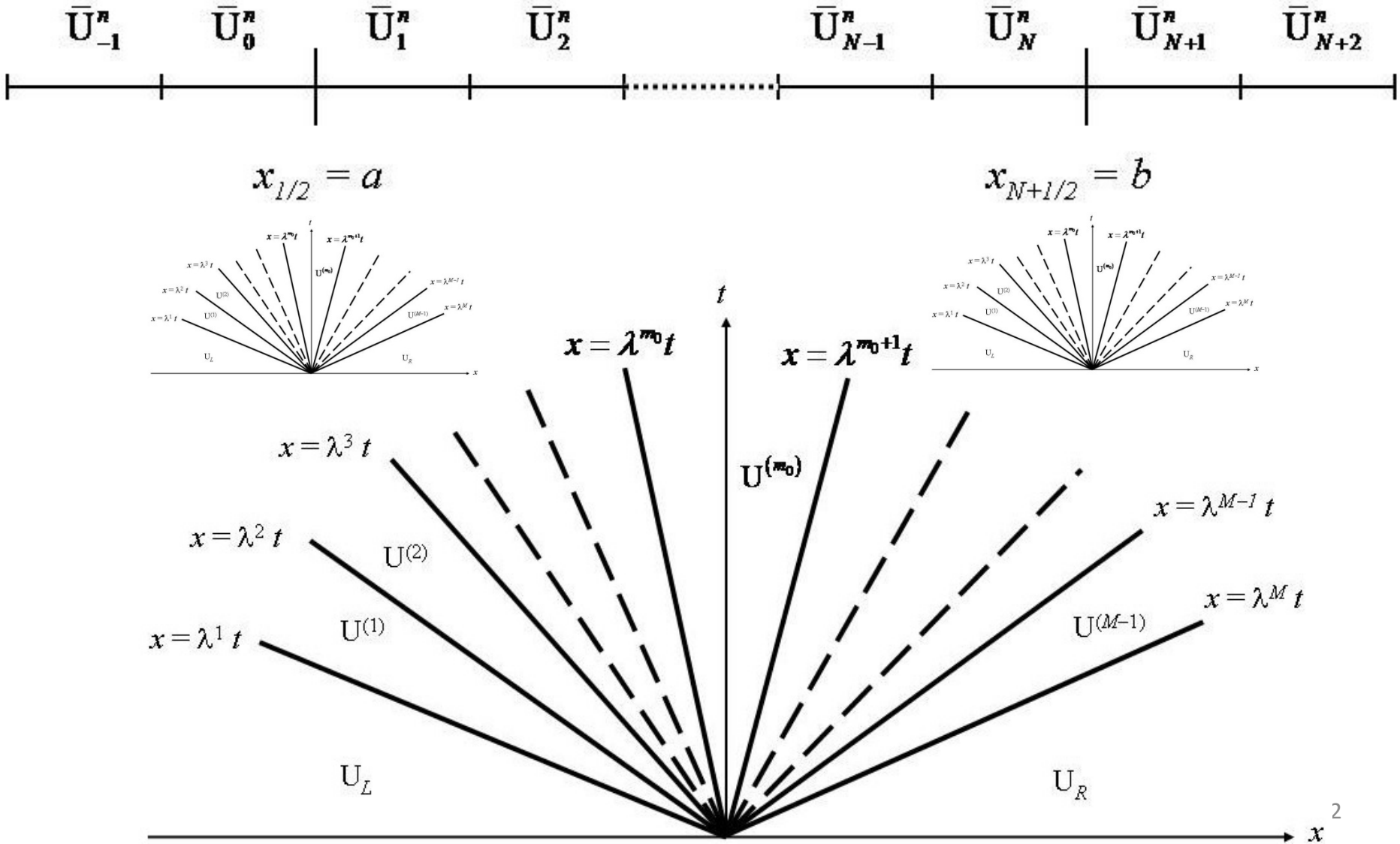
3.5) Numerical Boundary Conditions for Linear Hyperbolic Systems

Recall that for scalar advection $u_t + a u_x = 0$ with $a > 0$, we only need to specify the boundary conditions at the left boundary. Specifying them at the right boundary would overspecify the problem.

We want to understand how it goes for the “ M ” component linear hyperbolic system $U_t + A U_x = 0$.



Because we apply limiters to obtain the slopes, we will need two zones out from each physical boundary. Called *ghost zones*. Question: Why two?



We classify boundary conditions as follows:

Boundary conditions that permit a wave to leave the computational domain without generating a back-reaction are called *radiative or non-reflective boundary conditions*.

Boundary conditions that specify the amplitude of a wave that should flow into a computational domain are called *inflow boundary conditions*.

Our philosophy in developing the boundary conditions is that we should *apply the same numerical algorithm*, if this is at all possible, to all the zones of the mesh.

For *periodic boundary conditions* we set :

$$\bar{U}_0^n = \bar{U}_N^n, \bar{U}_{-1}^n = \bar{U}_{N-1}^n, \bar{U}_{N+1}^n = \bar{U}_1^n, \bar{U}_{N+2}^n = \bar{U}_2^n$$

For *outflow boundary conditions* at left boundary we set :

$$\bar{U}_{-1}^n = \bar{U}_0^n = \bar{U}_1^n$$

3.6) Second Order Upwind Schemes for Linear Hyperbolic Systems

Two Schemes:

A) Two-stage Runge-Kutta

B) Predictor-Corrector

Each of the two schemes described here can be *extended to non-linear hyperbolic systems*.

The latter two are also relatively easy to *extend to higher orders*.

Consider the same mesh over the x-domain, [a,b], that we used for describing *boundary conditions*.

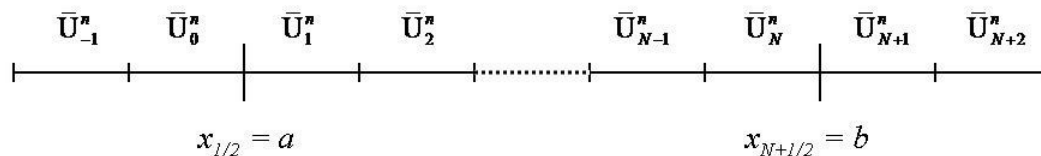


Fig. 3.12 shows a one-dimensional mesh with “N” zones covering [a,b]. The values of the zone-centered mesh function are shown at time t^n . The ghost zones, at which the boundary conditions are specified, are also shown.

3.6.2) Second Order Accurate, Two-Stage Runge-Kutta Scheme with Limiters

The methods described here are amongst the most popular ones.

Reason: They have a very appealing *plug-and-chug* quality.

Each stage of a Runge-Kutta method looks exactly like the other. While each stage only needs to attain the desired spatial accuracy, the whole scheme will have the desired spatial and temporal accuracy. Thus RK schemes are called *method of lines* or *semi-discrete* methods. $U_t = -F_x$

It is possible, though, to obtain schemes that are more efficient than Runge-Kutta. Two possible second order accurate Runge Kutta schemes:

Modified Euler Approximation:

$$\bar{U}_i^{n+1/2} = \bar{U}_i^n - \frac{\Delta t}{2\Delta x} \left(\bar{F}_{i+1/2}^n(\bar{U}^n) - \bar{F}_{i-1/2}^n(\bar{U}^n) \right)$$

$$\bar{U}_i^{n+1} = \bar{U}_i^n - \frac{\Delta t}{\Delta x} \left(\bar{F}_{i+1/2}^{n+1/2}(\bar{U}^{n+1/2}) - \bar{F}_{i-1/2}^{n+1/2}(\bar{U}^{n+1/2}) \right)$$

Improved Euler Approximation:

$$\bar{U}_i^{(1)} = \bar{U}_i^n - \frac{\Delta t}{\Delta x} \left(\bar{F}_{i+1/2}^n(\bar{U}^n) - \bar{F}_{i-1/2}^n(\bar{U}^n) \right)$$

$$\bar{U}_i^{n+1} = \frac{1}{2} \left(\bar{U}_i^n + \bar{U}_i^{(1)} \right) - \frac{\Delta t}{2\Delta x} \left(\bar{F}_{i+1/2}^{(1)}(\bar{U}^{(1)}) - \bar{F}_{i-1/2}^{(1)}(\bar{U}^{(1)}) \right)$$

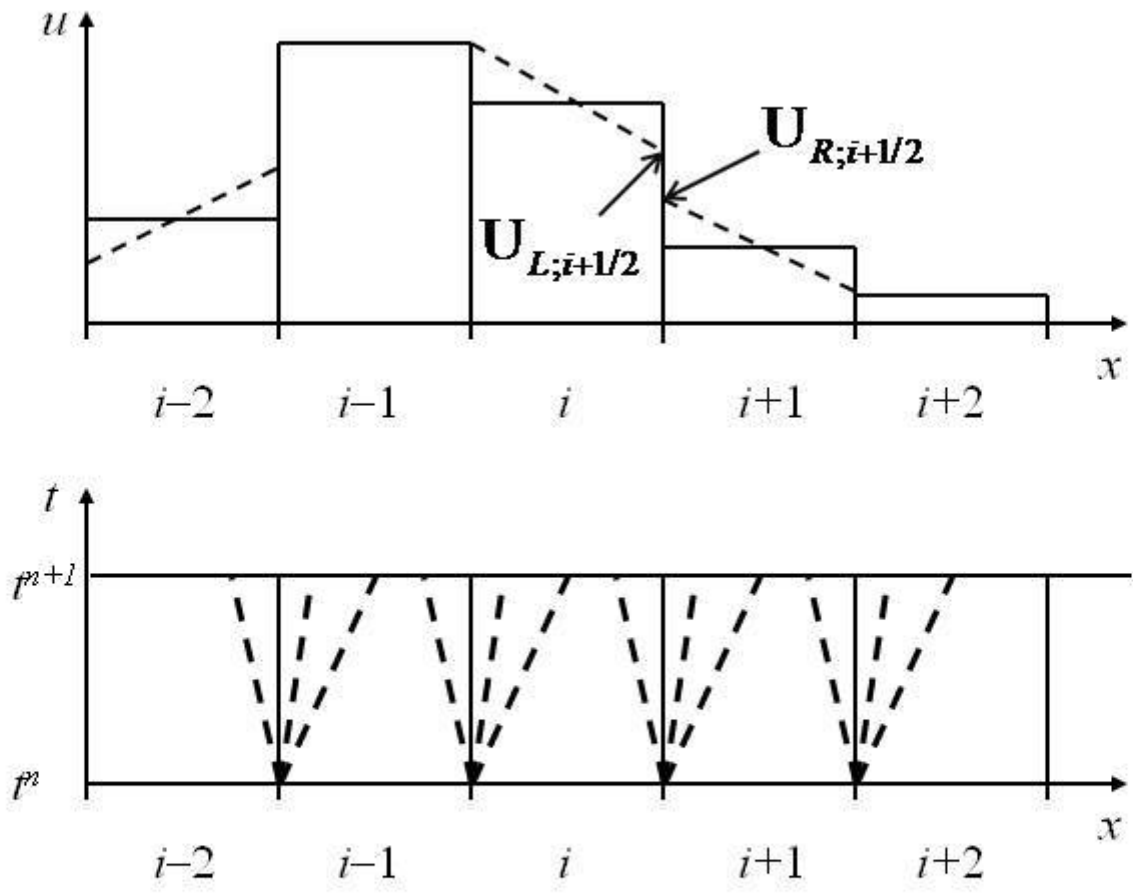


Fig. 3.13 schematically shows one stage of the second order Runge-Kutta method on a one-dimensional mesh. Here each zone is represented as a slab of fluid (solid lines) with a piecewise linear profile (dashed lines) in the top panel. The bottom panel shows the evolution in space and time of the Riemann problems at zone boundaries.

We only need to describe one stage, since they all look alike:

Step 1: We have to obtain the *undivided differences* of the conserved variables. This can be done using two different styles of limiting:

Limiting on the Characteristic Variables:

a) Obtain characteristic variables : $w_i^m = l^m \bar{U}_i$

b) Limit characteristic variables : $\Delta w_i^m = \text{Limiter} \left(w_{i+1}^m - w_i^m, w_i^m - w_{i-1}^m \right)$

c) Project back to obtain undivided differences in conserved

variables : $\Delta \bar{U}_i = \sum_{m=1}^M \Delta w_i^m r^m$

Limiting on the Conserved Variables:

a) Write conserved variables as vector : $\bar{U}_i \equiv \left(\bar{u}_i^1, \bar{u}_i^2, \dots, \bar{u}_i^M \right)^T$

b) Limit each of the components of the vector:

$\overline{\Delta u}_i^m = \text{Limiter} \left(\bar{u}_{i+1}^m - \bar{u}_i^m, \bar{u}_i^m - \bar{u}_{i-1}^m \right)$ gives $\Delta \bar{U}_i \equiv \left(\overline{\Delta u}_i^1, \overline{\Delta u}_i^2, \dots, \overline{\Delta u}_i^M \right)^T$

Question: What are the strengths and weaknesses of each limiting strategy?

Step 2: Obtain the *left and right states at the zone boundary*:

(i.e. since we know the variation within the zone, we can obtain the values of the solution at any point in the zone, including the face-centers.)

$$U_{L;i+1/2} \equiv \bar{U}_i + \frac{1}{2} \overline{\Delta U}_i \quad ; \quad U_{R;i+1/2} \equiv \bar{U}_{i+1} - \frac{1}{2} \overline{\Delta U}_{i+1}$$

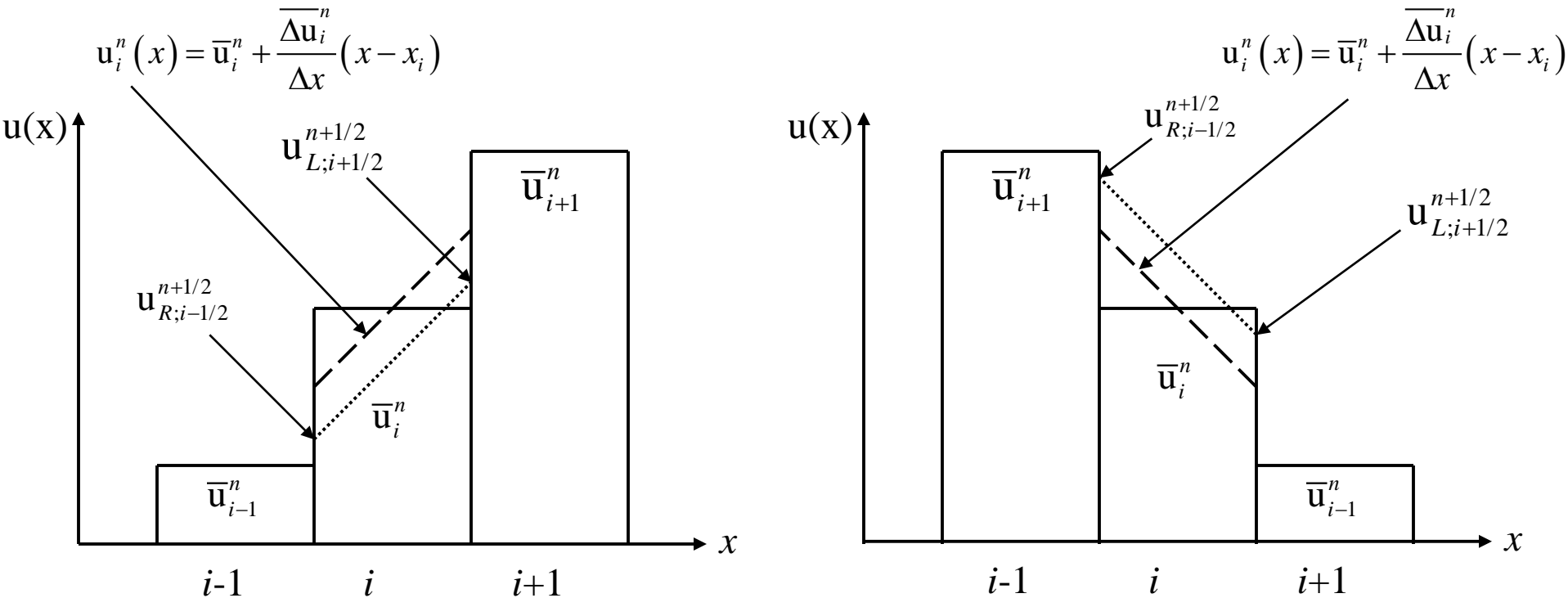
Step 3: Treat the Riemann solver as a machine that accepts two states and spits out a flux. Feed the above left and right states into the Riemann solver and obtain a properly *upwinded flux*.

$$\bar{F}_{i+1/2} = F_{RS} \left(U_{L;i+1/2}, U_{R;i+1/2} \right)$$

Step 4: These fluxes can be used in each of the two stages, as needed, to obtain the full update.

3.6.3) Predictor-Corrector Formulation

It is based on the idea that if we have the piecewise linear slopes within a zone, we can also predict the time-evolution of the hyperbolic system for at least a small time interval within that zone.



The panel to the left shows a mesh function that is not decreasing with x . The panel to the right shows a mesh function that is not increasing with x . We solve the advection equation $u_t + a u_x = 0$ with $a > 0$. In each case the dashed line shows the piecewise linear reconstructed profile in x at time t^n while the dotted line shows the same profile at time $t^n + \Delta t / 2$.

Step 1: We have to obtain the *undivided differences* of the conserved variables; same as before.

Step 2: Realize that we can obtain the time evolution within a zone

because:
$$\frac{\Delta U_i}{\Delta t} + A \frac{\overline{\Delta U}_i^n}{\Delta x} = 0$$

Use this to directly obtain time-centered left and right states at the zone boundaries. This time-centering makes the scheme second order.

$$U_{L;i+1/2}^{n+1/2} \equiv \bar{U}_i^n + \frac{1}{2} \overline{\Delta U}_i^n + \frac{1}{2} \Delta t \left(\frac{\Delta U_i}{\Delta t} \right) \leftarrow \text{Taylor expansion in space \& time; zone "i"}$$

$$= \bar{U}_i^n + \frac{1}{2} \overline{\Delta U}_i^n - \frac{1}{2} \frac{\Delta t}{\Delta x} A \overline{\Delta U}_i^n \leftarrow \text{i.e. we substitute } \frac{\Delta U_i}{\Delta t} = -A \frac{\overline{\Delta U}_i^n}{\Delta x}$$

$$U_{R;i+1/2}^{n+1/2} \equiv \bar{U}_{i+1}^n - \frac{1}{2} \overline{\Delta U}_{i+1}^n + \frac{1}{2} \Delta t \left(\frac{\Delta U_{i+1}}{\Delta t} \right) \leftarrow \text{Taylor expansion in space-time; zone "i+1"}$$

$$= \bar{U}_{i+1}^n - \frac{1}{2} \overline{\Delta U}_{i+1}^n - \frac{1}{2} \frac{\Delta t}{\Delta x} A \overline{\Delta U}_{i+1}^n \leftarrow \text{i.e. we substitute } \frac{\Delta U_{i+1}}{\Delta t} = -A \frac{\overline{\Delta U}_{i+1}^n}{\Delta x}$$

This is the *predictor step*.

Step 3: Treat the Riemann solver as a machine that accepts two states and spits out a flux. Feed the above *time-centered* left and right states into the Riemann solver and obtain a properly *upwinded flux*.

$$\bar{F}_{i+1/2} = F_{RS} \left(U_{L;i+1/2}^{n+1/2}, U_{R;i+1/2}^{n+1/2} \right)$$

Step 4: Use the fluxes to obtain the update step:

$$\bar{U}_i^{n+1} = \bar{U}_i^n - \frac{\Delta t}{\Delta x} \left(F_{RS} \left(U_{L;i+1/2}^{n+1/2}, U_{R;i+1/2}^{n+1/2} \right) - F_{RS} \left(U_{L;i-1/2}^{n+1/2}, U_{R;i-1/2}^{n+1/2} \right) \right)$$

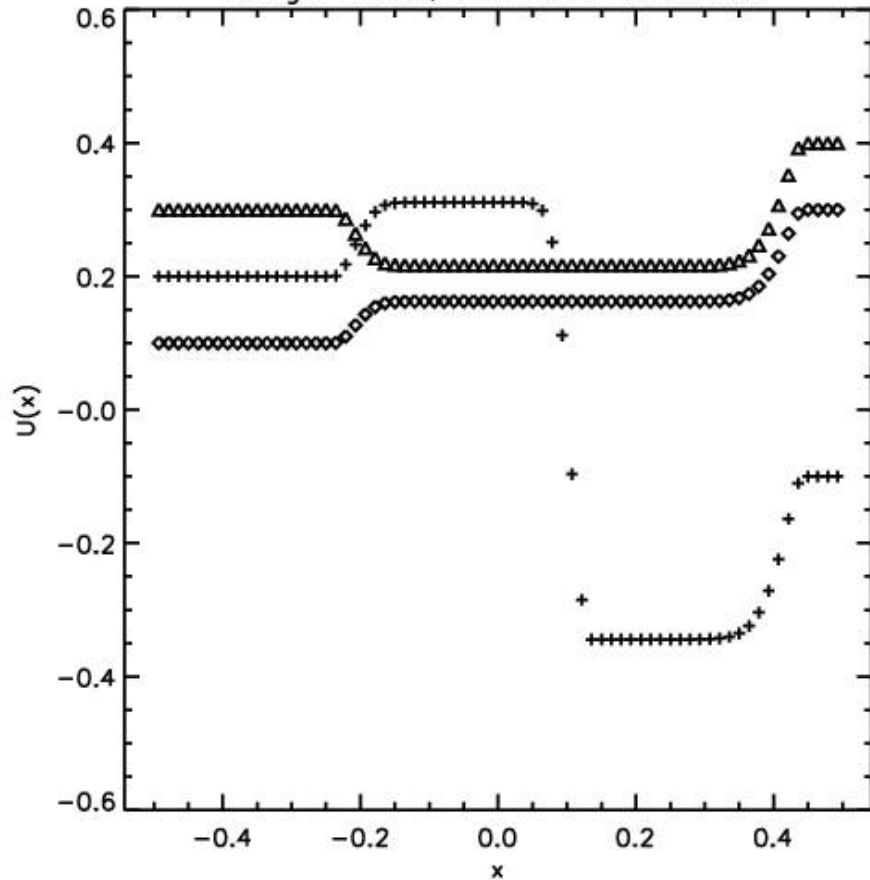
This is the *corrector step*.

Questions: How would the speed of this method compare to that of the Runge-Kutta method? Which steps are repeated in the Runge-Kutta method? Which steps dominate the cost?

3.6.4) Numerical Results from the Previous Two Schemes

Question: Compare and contrast the results of the two schemes:

Runge-Kutta, Linearized Acoustics



Predictor-Corrector, Linearized Acoustics

