We've now seen...

- A few examples of sheaf models for a few problems
- How do we build sheaf models in general?

A differential equation example

- Consider $u' = f(u)$ on the real line
- $C^k(\mathbb{R},\mathbb{R}^d)$ is the space of k-times continuously differentiable functions
- The equation might be expressed diagrammatically:

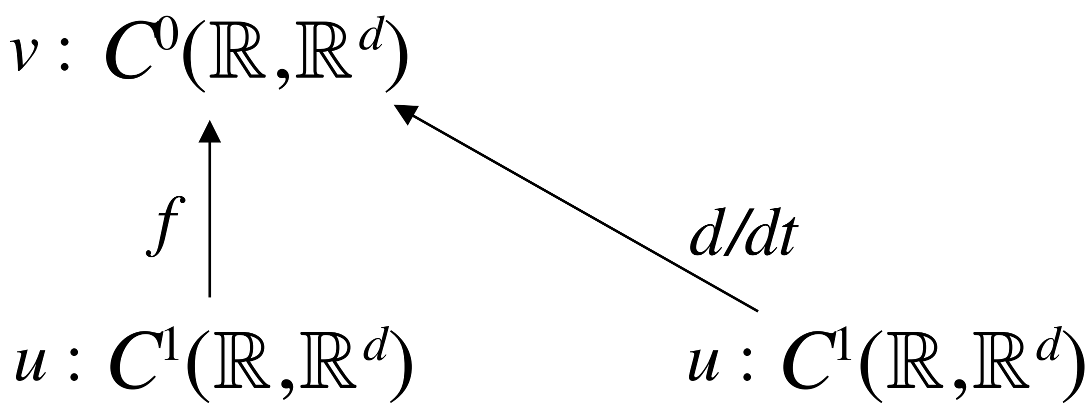
$$\begin{array}{c} u' : C^0(\mathbb{R},\mathbb{R}^d) \\ f \uparrow \quad \uparrow d/dt \\ u : C^1(\mathbb{R},\mathbb{R}^d) \end{array}$$

But wait, diagram does not commute, so it cannot be a sheaf!
It also multiply-defines the would-be restriction map :-/

A differential equation example

A standard trick: replace $u' = f(u)$ with the system:

- $v = f(u)$
- $v = d/dt\,(u)$

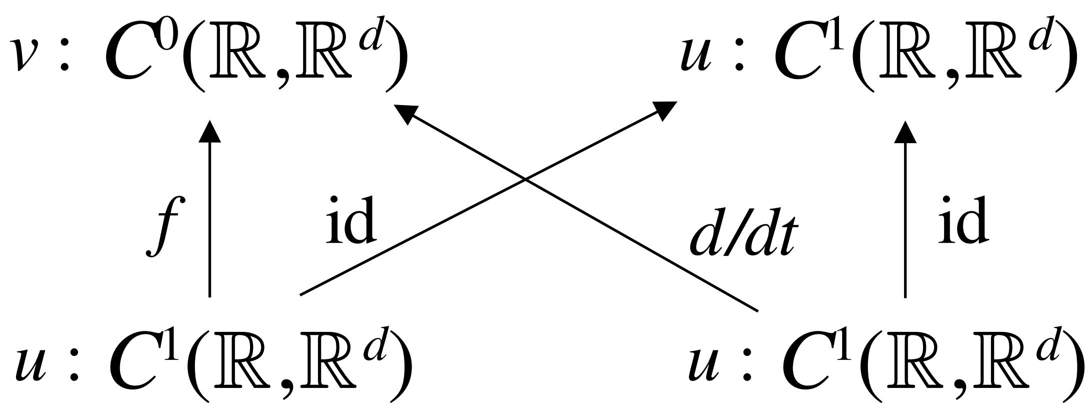


But wait, now the two copies of *u* don't have to agree...

A differential equation example

A standard trick: replace $u' = f(u)$ with the system:

- $v = f(u)$
- $v = d/dt\,(u)$



Sections of this sheaf are solutions to the original equation, because this requires all three copies of u to agree

Multi-equation sheaves

- Theorem: (R.) For every system of equations, there is a sheaf whose global sections are solutions
  - Base poset has two levels: Equations < Variables
  - Stalk over each variable is that variable's set of possible values
  - Stalk over an equation is a subset of the product of the variables involved
  - Restriction maps are projections

Source: M. Robinson, "Sheaf and duality methods for analyzing multi-model systems," arXiv:1604.04647

Goodwin macroeconomic model

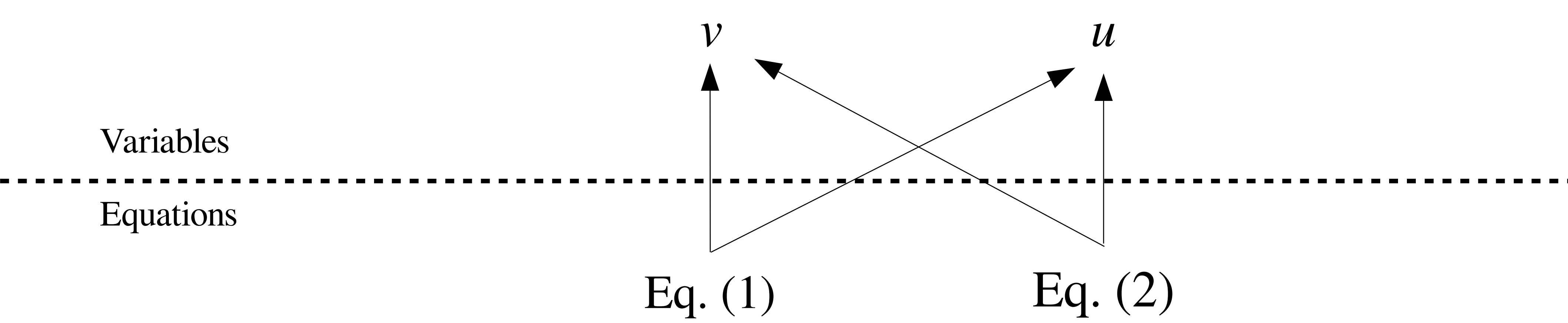
- A simple description of a national economy:

$$\dot{v} = v(t)\left(\frac{1}{\sigma} - (\alpha+\beta) - \frac{u(t)}{\sigma}\right) \quad (1)$$

v = Employment rate

$$\dot{u} = u(t)\left(-(\alpha+\gamma) + (\rho v(t))\right). \quad (2)$$

u = Workers' share of income



Goodwin macroeconomic model

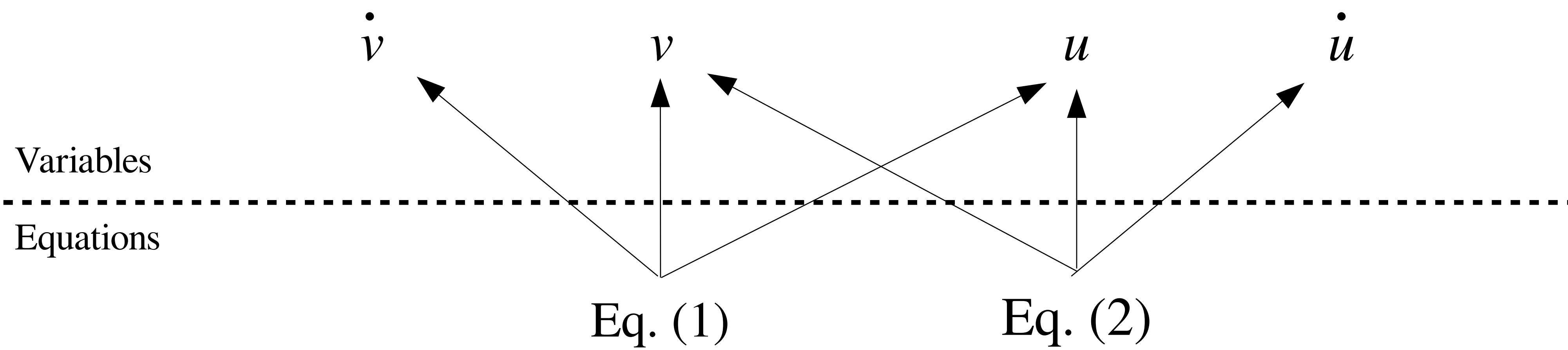
- A simple description of a national economy:

$$\dot{v} = v(t)\left(\frac{1}{\sigma} - (\alpha+\beta) - \frac{u(t)}{\sigma}\right) \quad (1)$$

v = Employment rate

$$\dot{u} = u(t)\left(-(\alpha+\gamma) + (\rho v(t))\right). \quad (2)$$

u = Workers' share of income



Goodwin macroeconomic model

- A simple description of a national economy:

$$\dot{v} = v(t)\left(\frac{1}{\sigma} - (\alpha+\beta) - \frac{u(t)}{\sigma}\right) \quad (1)$$

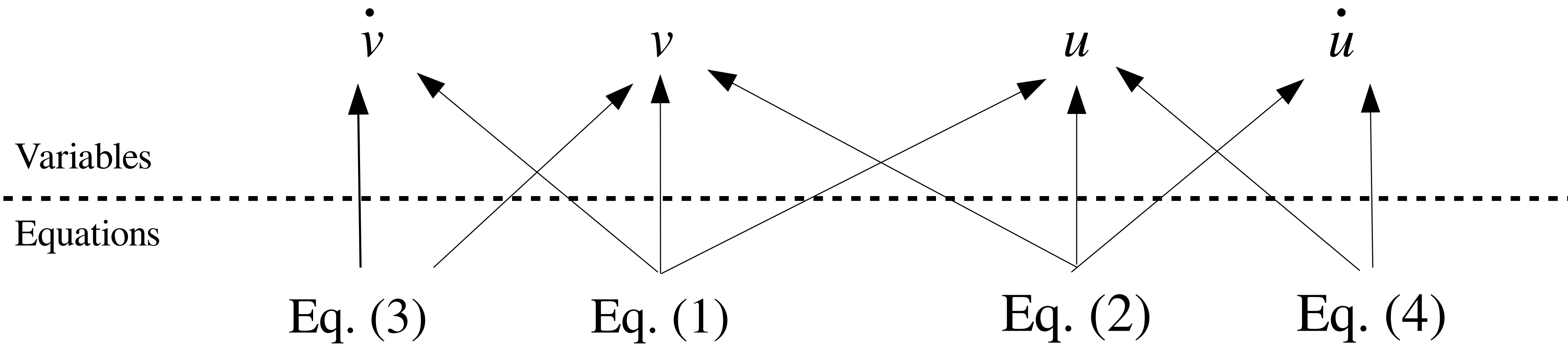
v = Employment rate

$$\dot{u} = u(t)\left(-(\alpha+\gamma) + (\rho v(t))\right). \quad (2)$$

u = Workers' share of income

$$\dot{v} = dv/dt \quad (3)$$

$$\dot{u} = du/dt \quad (4)$$



Goodwin macroeconomic model

- A simple description of a national economy:

$$\dot{v} = v(t)\left(\frac{1}{\sigma} - (\alpha+\beta) - \frac{u(t)}{\sigma}\right) \quad (1)$$

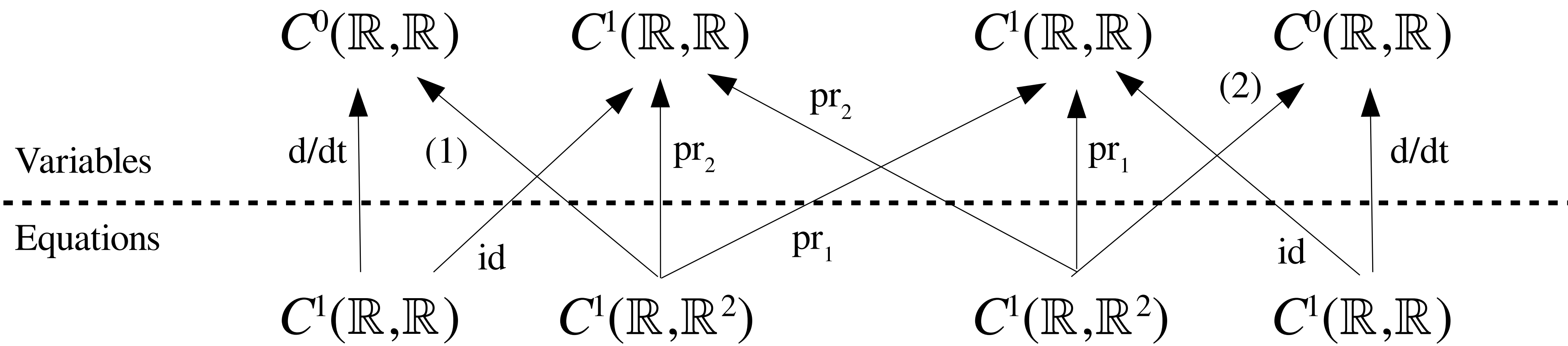
v = Employment rate

$$\dot{u} = u(t)\left(-(\alpha+\gamma) + (\rho v(t))\right). \quad (2)$$

u = Workers' share of income

$$\dot{v} = dv/dt \quad (3)$$

$$\dot{u} = du/dt \quad (4)$$



Bayesian networks (aka "Bayes nets")

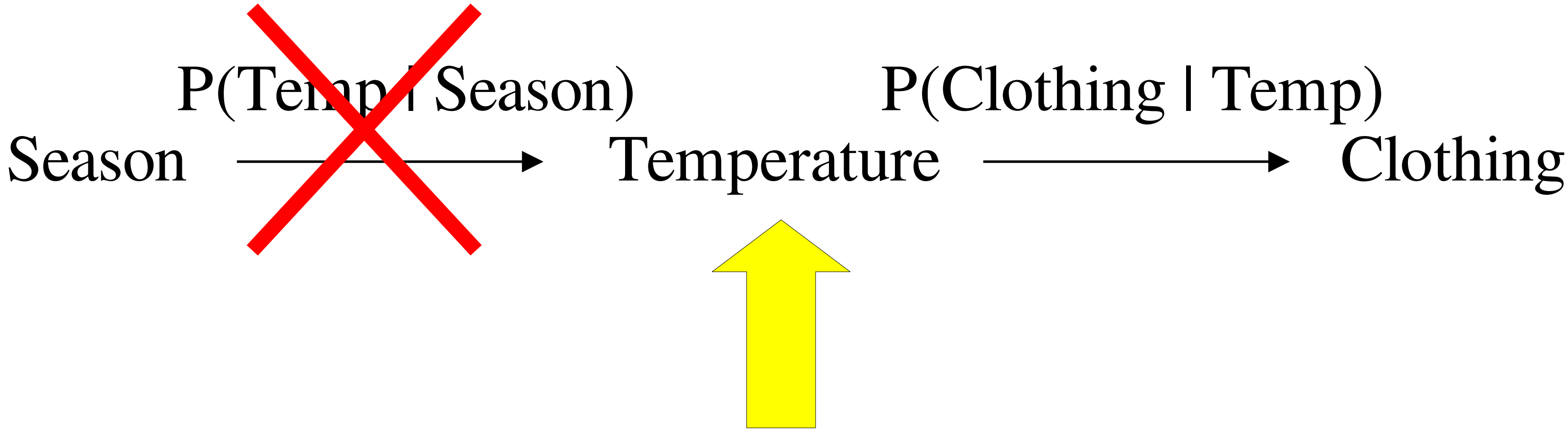
Bayesian networks (aka "Bayes nets")

P(Temp | Season) P(Clothing | Temp)

Season → Temperature → Clothing

Categorical Numeric Categorical

Causal networks



do(Temp = 20 °C)

Let's see what happens when we turn on the A/C;
we don't care about the season any more...

Let's try to write some equations...

Definition: $P(A \mid B)\, P(B) = P(A,B)$

$$\text{Season} \xrightarrow{P(\text{Temp} \mid \text{Season})} \text{Temperature} \xrightarrow{P(\text{Clothing} \mid \text{Temp})} \text{Clothing}$$

$P(\text{Temp},\text{Season}) = P(\text{Temp} \mid \text{Season})\, P(\text{Season})$

$P(\text{Clothing},\text{Temp}) = P(\text{Clothing} \mid \text{Temp})\, P(\text{Temp})$

Let's try to write some equations...

<u>Definition:</u> P(A | B) P(B) = P(A,B)

P(Temp | Season)   P(Clothing | Temp)

Season ⟶ Temperature ⟶ Clothing

P(Temp,Season) = P(Temp | Season) P(Season)
P(Clothing,Temp) = P(Clothing | Temp) P(Temp)

… it seems that we don't have enough equations to fully solve for P(Clothing), say…

… what we're missing are the equations that marginalize out variables from a joint distribution. There are quite a few of these!

But wait, there's more...

Since there are three variable in play, there are many ways to marginalize all the various joints, including those "not in" the Bayes net

P(Temp | Season) P(Clothing | Temp)

Season ⟶ Temperature ⟶ Clothing

$P(\text{Temp},\text{Season}) = P(\text{Temp} \mid \text{Season})\, P(\text{Season})$

$P(\text{Clothing},\text{Temp}) = P(\text{Clothing} \mid \text{Temp})\, P(\text{Temp})$

$P(\text{Season}) = \Sigma_t\, P(\text{Temp} = t, \text{Season})$

$P(\text{Temp}) = \Sigma_s\, P(\text{Temp}, \text{Season} = s)$

$P(\text{Temp}) = \Sigma_c\, P(\text{Clothing} = c, \text{Temp})$

$P(\text{Clothing}) = \Sigma_t\, P(\text{Clothing}, \text{Temp} = t)$

$P(\text{Season}) = \Sigma_c\, P(\text{Clothing} = c, \text{Season})$

$P(\text{Clothing}) = \Sigma_s\, P(\text{Clothing}, \text{Season} = s)$

But wait, there's **even** more...

We forgot the three-way marginals too! (But this is now everything)

$$\text{Season} \xrightarrow{P(\text{Temp} \mid \text{Season})} \text{Temperature} \xrightarrow{P(\text{Clothing} \mid \text{Temp})} \text{Clothing}$$

$P(\text{Temp},\text{Season}) = P(\text{Temp} \mid \text{Season})\, P(\text{Season})$

$P(\text{Clothing},\text{Temp}) = P(\text{Clothing} \mid \text{Temp})\, P(\text{Temp})$

$P(\text{Season}) = \Sigma_t\, P(\text{Temp} = t, \text{Season})$

$P(\text{Temp}) = \Sigma_s\, P(\text{Temp}, \text{Season} = s)$

$P(\text{Temp}) = \Sigma_c\, P(\text{Clothing} = c, \text{Temp})$

$P(\text{Clothing}) = \Sigma_t\, P(\text{Clothing}, \text{Temp} = t)$

$P(\text{Season}) = \Sigma_c\, P(\text{Clothing} = c, \text{Season})$

$P(\text{Clothing}) = \Sigma_s\, P(\text{Clothing}, \text{Season} = s)$

$P(\text{Season},\text{Temp}) = \Sigma_c\, P(\text{Clothing} = c, \text{Temp}, \text{Season})$

$P(\text{Season},\text{Clothing}) = \Sigma_t\, P(\text{Clothing}, \text{Temp} = t, \text{Season})$

$P(\text{Temp},\text{Clothing}) = \Sigma_s\, P(\text{Clothing}, \text{Temp}, \text{Season} = s)$

We'd like to sheafify...

… but things are getting very busy; let's summarize the names

$$X_1 \xrightarrow{P(X_2 \mid X_1)} X_2 \xrightarrow{P(X_3 \mid X_2)} X_3$$

$P(X_1,X_2) = P(X_2 \mid X_1)\, P(X_1)$

$P(X_2,X_3) = P(X_3 \mid X_2)\, P(X_2)$

2 conditional equations

$P(X_1) = \Sigma_t\, P(X_2 = t, X_1)$

$P(X_2) = \Sigma_s\, P(X_2, X_1 = s)$

$P(X_2) = \Sigma_c\, P(X_3 = c, X_2)$

$P(X_3) = \Sigma_t\, P(X_3, X_2 = t)$

$P(X_1) = \Sigma_c\, P(X_3 = c, X_1)$

$P(X_3) = \Sigma_s\, P(X_3, X_1 = s)$

$P(X_1,X_2) = \Sigma_c\, P(X_3 = c, X_2, X_1)$

$P(X_1,X_3) = \Sigma_t\, P(X_3, X_2 = t, X_1)$

$P(X_2,X_3) = \Sigma_s\, P(X_3, X_2, X_1 = s)$

9 Marginal equations

What are the stalks & restrictions?

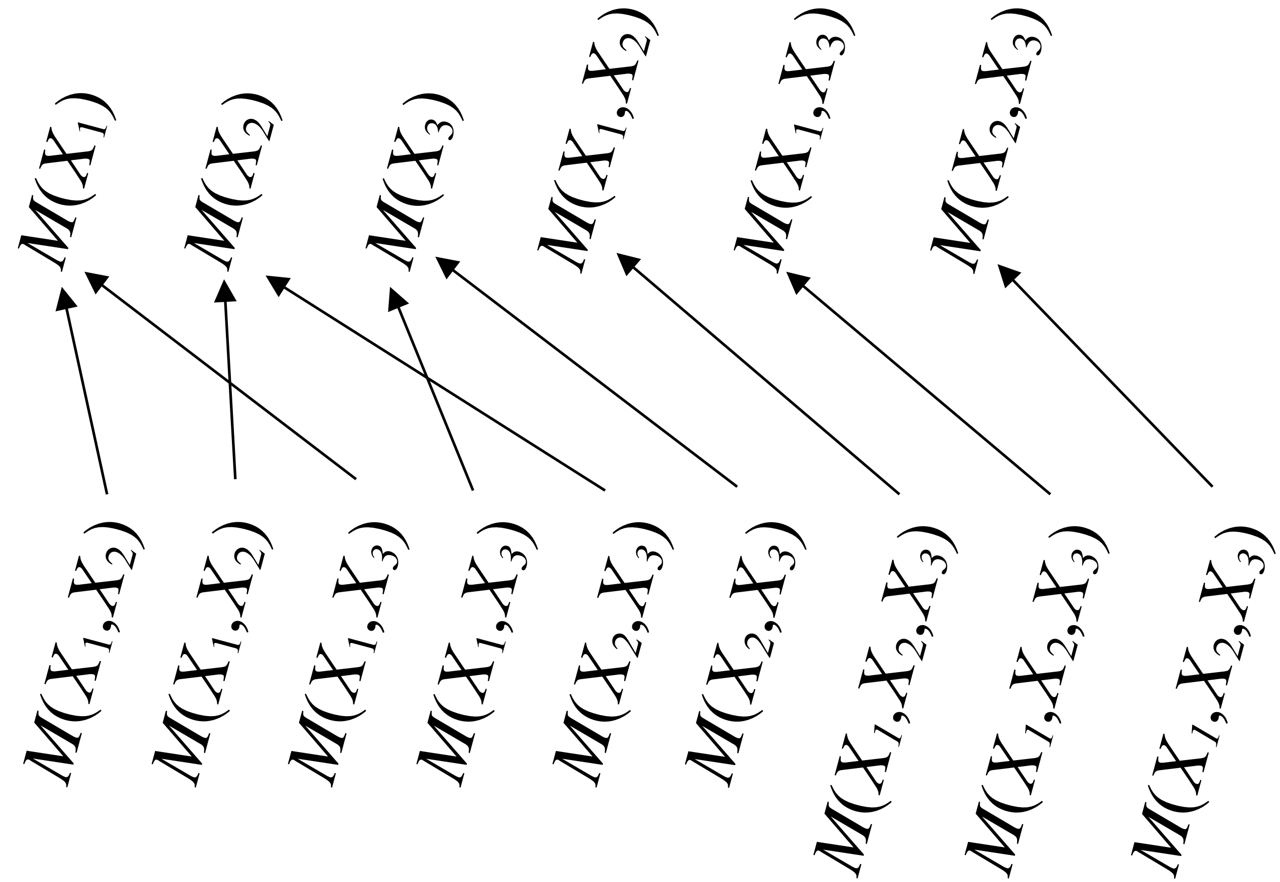
Each "variable" in our system of equations is a probability distribution

<u>Definition</u>: $M(X_1,X_2,X_3)$ is the set of joint probability distributions on X_1,X_2,X_3. (Similar for more/fewer variables)

<u>Equations:</u>	<u>Restriction types:</u>
$P(X_1,X_2) = P(X_2 \mid X_1)\, P(X_1)$	$M(X_1) \to M(X_1,X_2)$
$P(X_2,X_3) = P(X_3 \mid X_2)\, P(X_2)$	$M(X_2) \to M(X_2,X_3)$
$P(X_1) = \Sigma_t\, P(X_2 = t, X_1)$	$M(X_1,X_2) \to M(X_1)$
$P(X_2) = \Sigma_s\, P(X_2, X_1 = s)$	$M(X_1,X_2) \to M(X_2)$
$P(X_2) = \Sigma_c\, P(X_3 = c, X_2)$	$M(X_2,X_3) \to M(X_2)$
$P(X_3) = \Sigma_t\, P(X_3, X_2 = t)$	$M(X_2,X_3) \to M(X_3)$
$P(X_1) = \Sigma_c\, P(X_3 = c, X_1)$	$M(X_1,X_3) \to M(X_1)$
$P(X_3) = \Sigma_s\, P(X_3, X_1 = s)$	$M(X_1,X_3) \to M(X_3)$
$P(X_1,X_2) = \Sigma_c\, P(X_3 = c, X_2, X_1)$	$M(X_1,X_2,X_3) \to M(X_1,X_2)$
$P(X_1,X_3) = \Sigma_t\, P(X_3, X_2 = t, X_1)$	$M(X_1,X_2,X_3) \to M(X_1,X_3)$
$P(X_2,X_3) = \Sigma_s\, P(X_3, X_2, X_1 = s)$	$M(X_1,X_2,X_3) \to M(X_2,X_3)$

Bayesian network as a sheaf

- Marginals… (Always present)



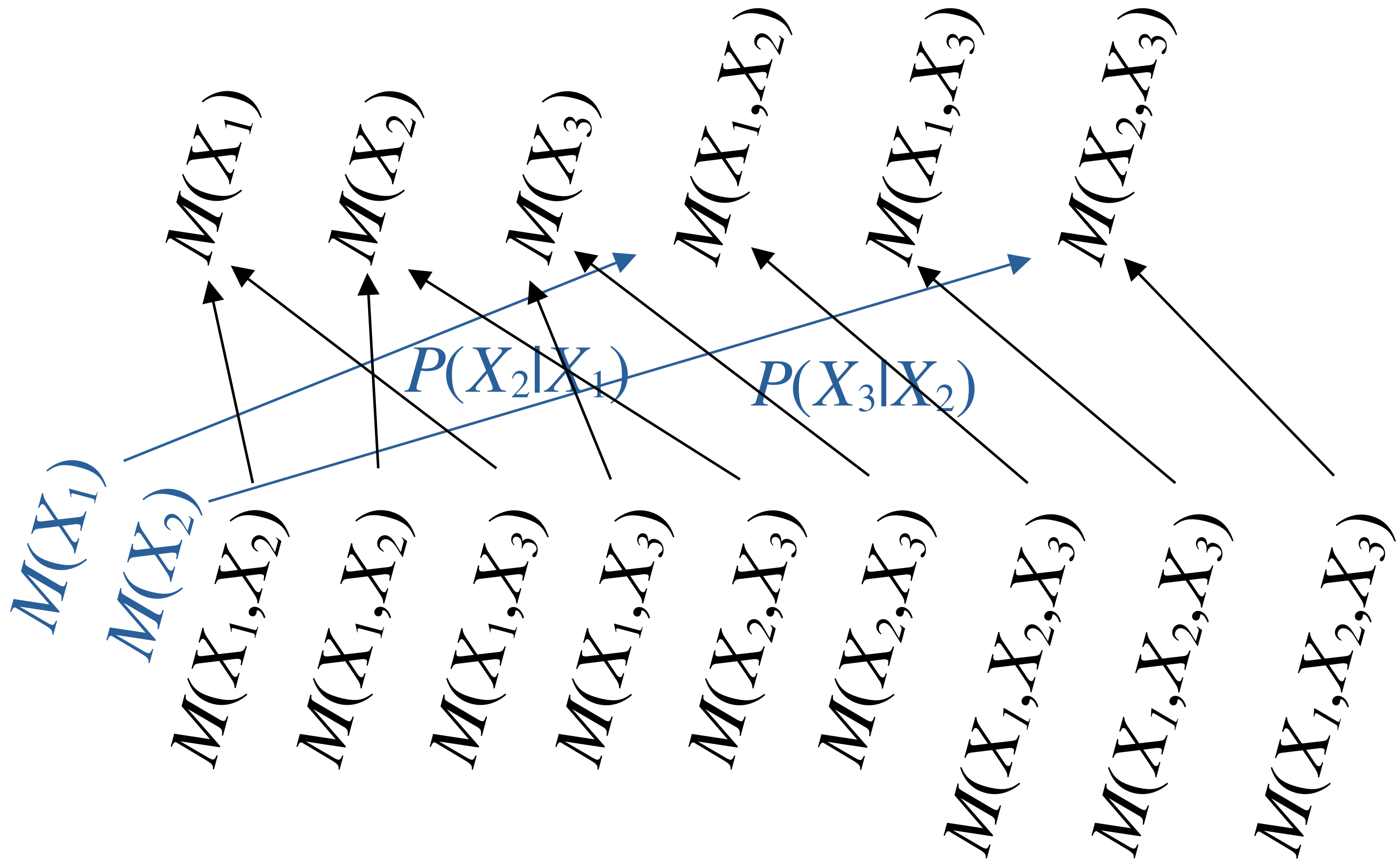
Olivia Chen Michael Robinson

Bayesian network as a sheaf

- … conditionals … (based upon the Bayes net)

$X_1 \rightarrow X_2 \rightarrow X_3$

Bayes net

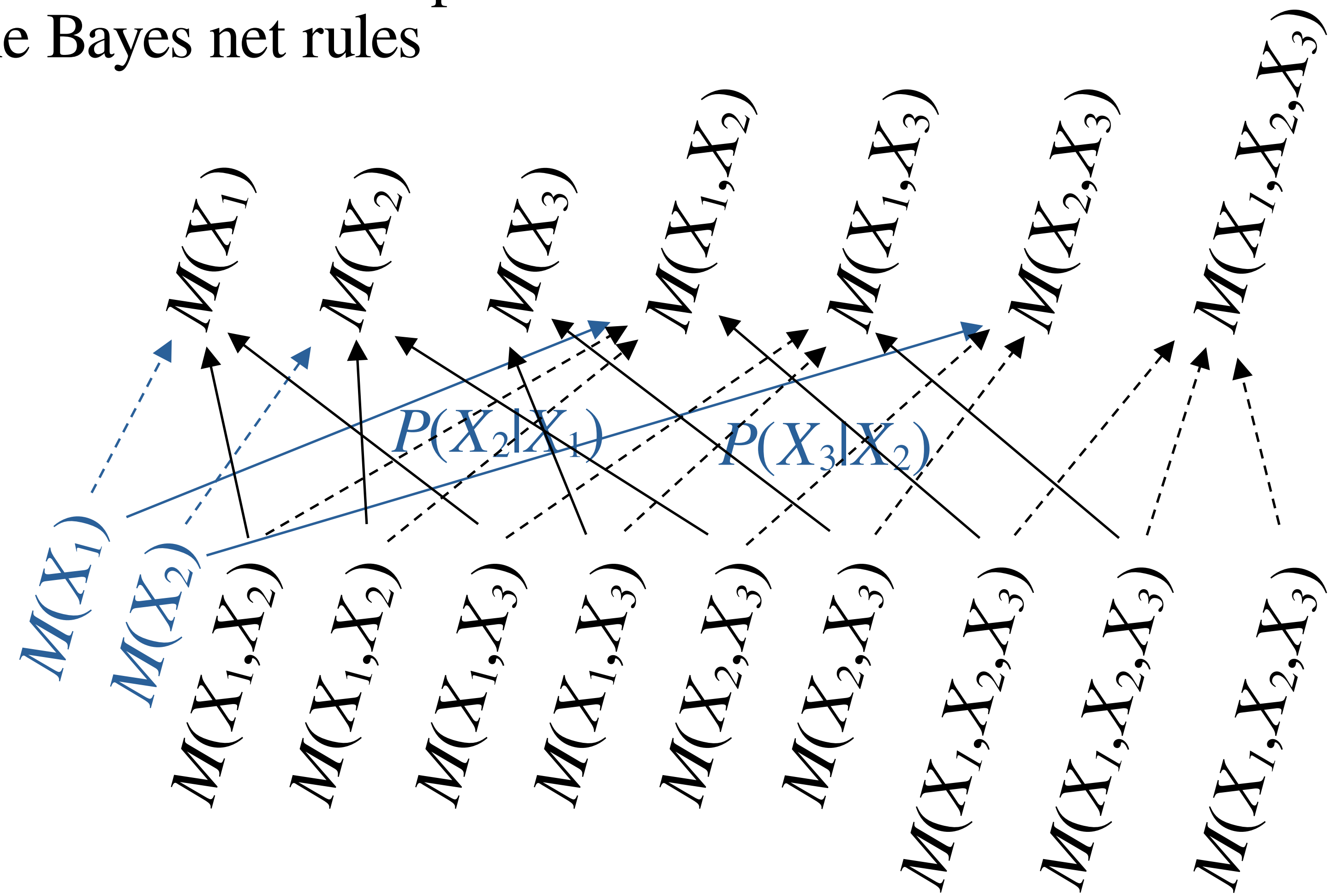


Bayesian network as a sheaf

- … identities! (Added to ensure consistency across copies.)
- <u>Corollary</u>: Global sections are possible sets of distributions that satisfy the Bayes net rules

$X_1 \rightarrow X_2 \rightarrow X_3$

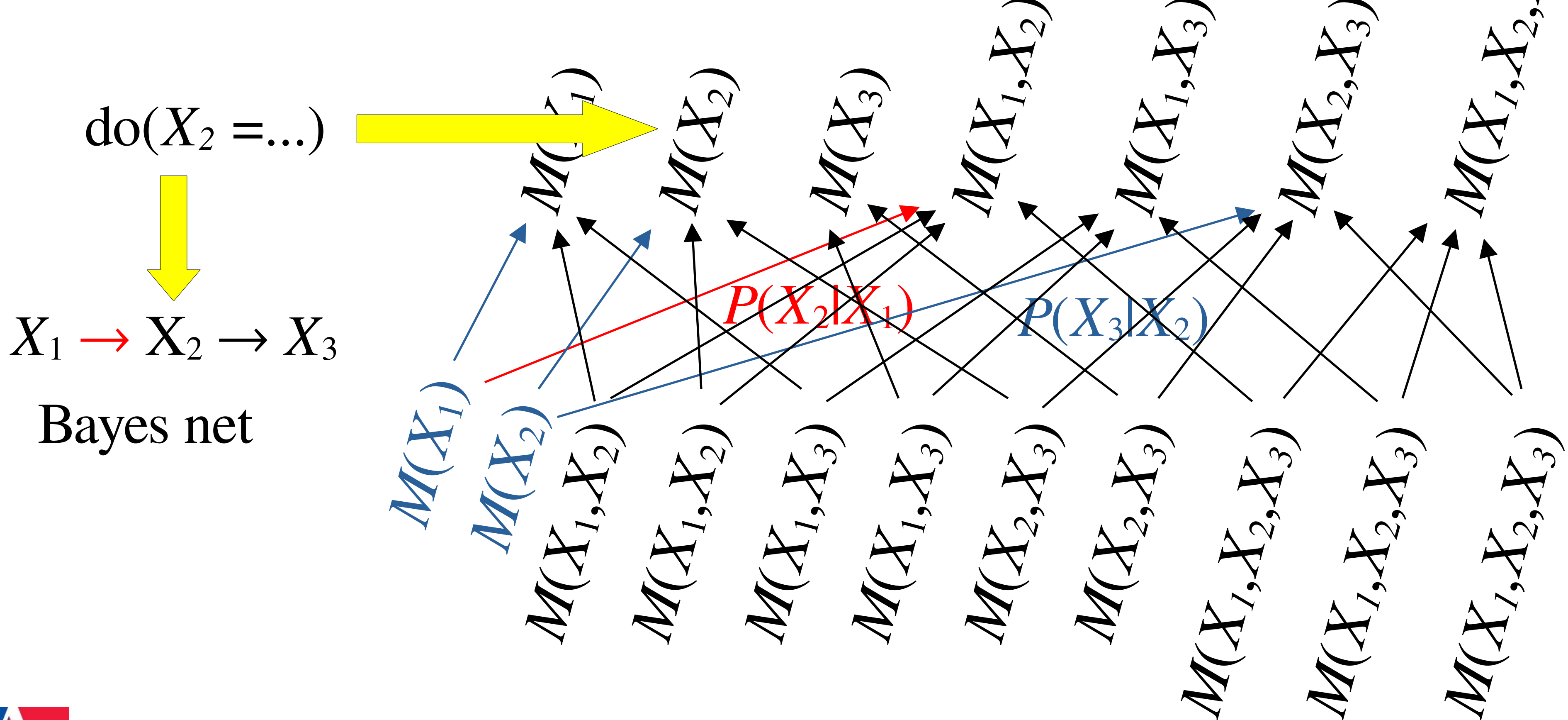
Bayes net



Olivia Chen Michael Robinson

Causal modeling using **do** operator

- Make an assignment to variable in top row with P(desired value) = 1
- Delete the **conditional arrows** (leave the marginals) into that variable
- Minimize consistency radius elsewhere



Consistency: Discretizing correctly

Discretization of functions

$$C^k(X,Y) \longrightarrow \mathbb{R}^n$$

$$f \qquad\qquad (f(x_1),\ldots,f(x_n))$$

Discretization of functions

$$\mathbb{R}^m \longrightarrow C^k(X,Y) \longrightarrow \mathbb{R}^n$$

$$(a_1,\ldots,a_m) \qquad f = \sum a_i f_i(x)$$

Why discretize?

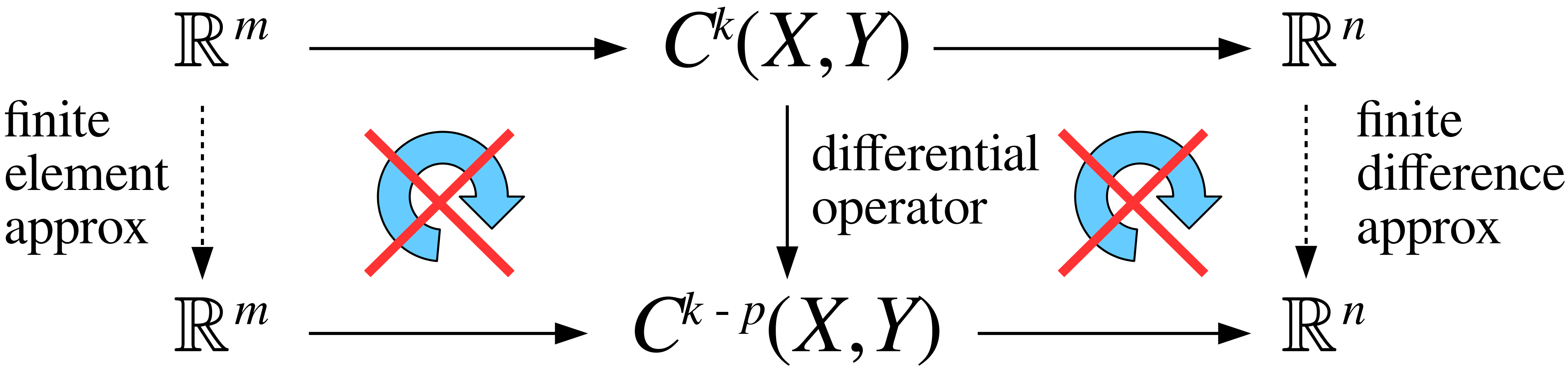
$$\mathbb{R}^m \longrightarrow C^k(X,Y) \longrightarrow \mathbb{R}^n$$

$$C^k(X,Y) \xrightarrow{\text{differential operator}} C^{k-p}(X,Y)$$

Why discretize?

$$\begin{array}{ccccc}
\mathbb{R}^m & \longrightarrow & C^k(X,Y) & \longrightarrow & \mathbb{R}^n \\
 & & \Big\downarrow \text{differential operator} & & \\
\mathbb{R}^m & \longrightarrow & C^{k-p}(X,Y) & \longrightarrow & \mathbb{R}^n
\end{array}$$

Why discretize?

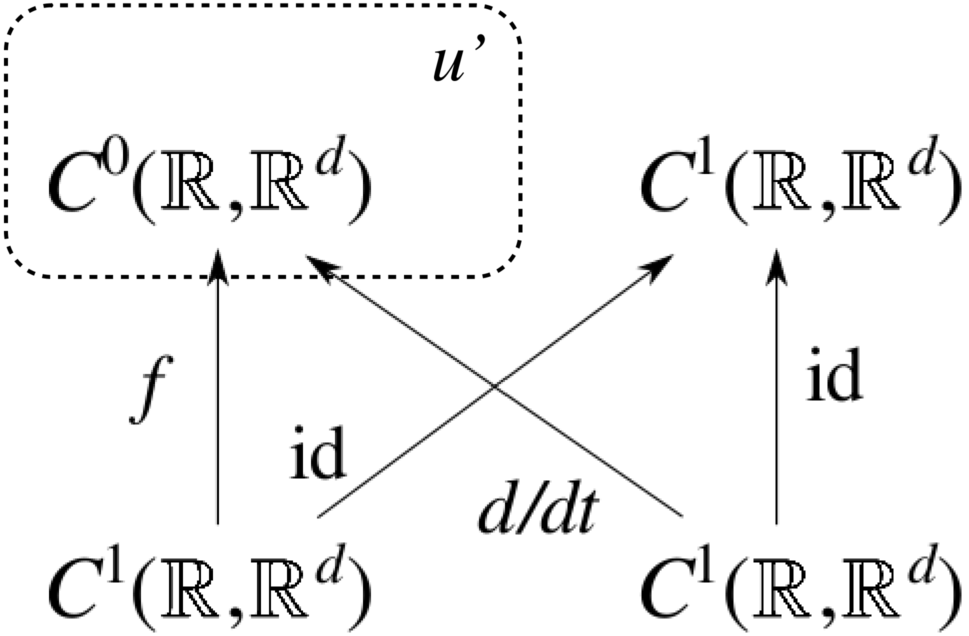


Goals:

1. Make the diagram commute as $m, n \to \infty$ (*consistency* of the approximation)
2. Recover properties of the differential operator from the approximations (*convergence* of the approximation)

Back to our original example

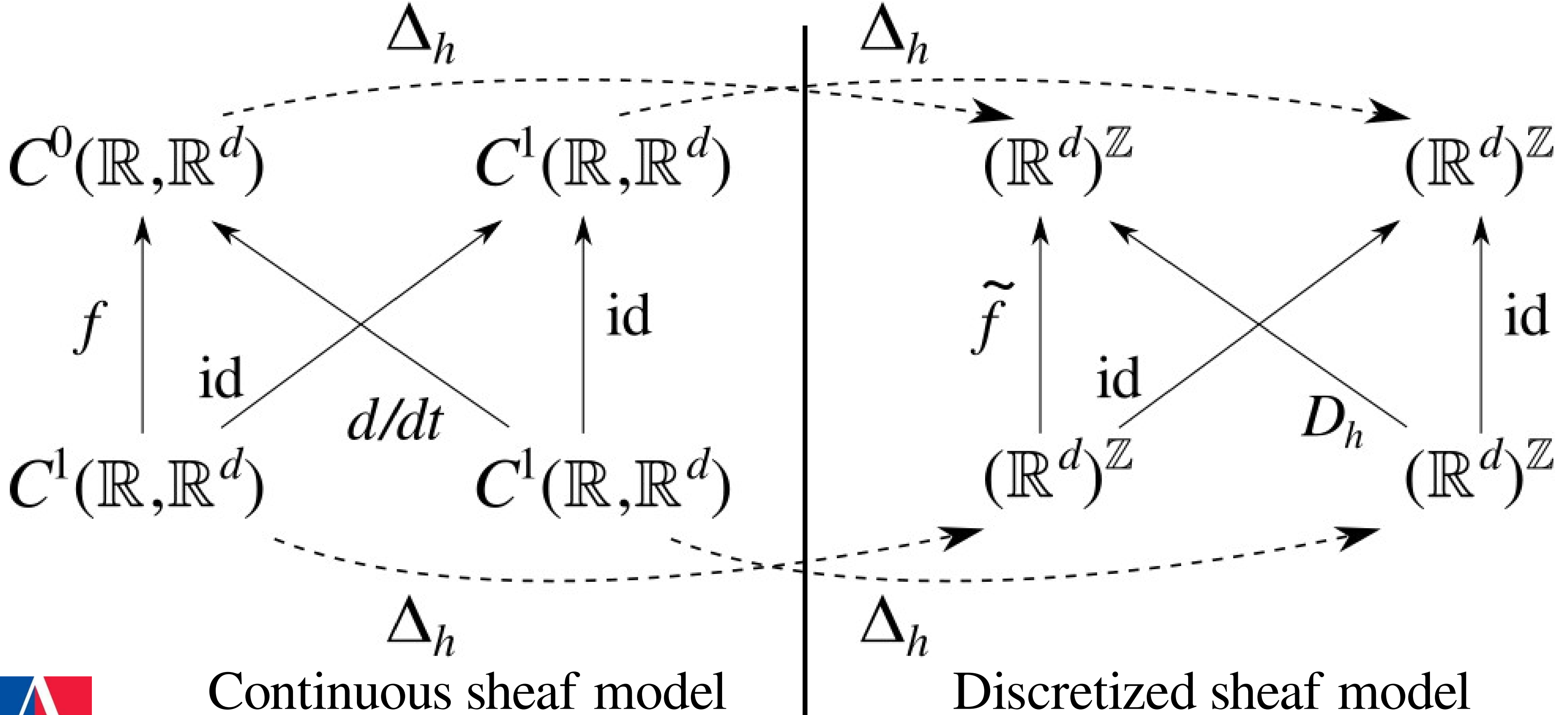
- Consider $u' = f(u)$ on the real line
- This has a sheaf diagram



Finite differences

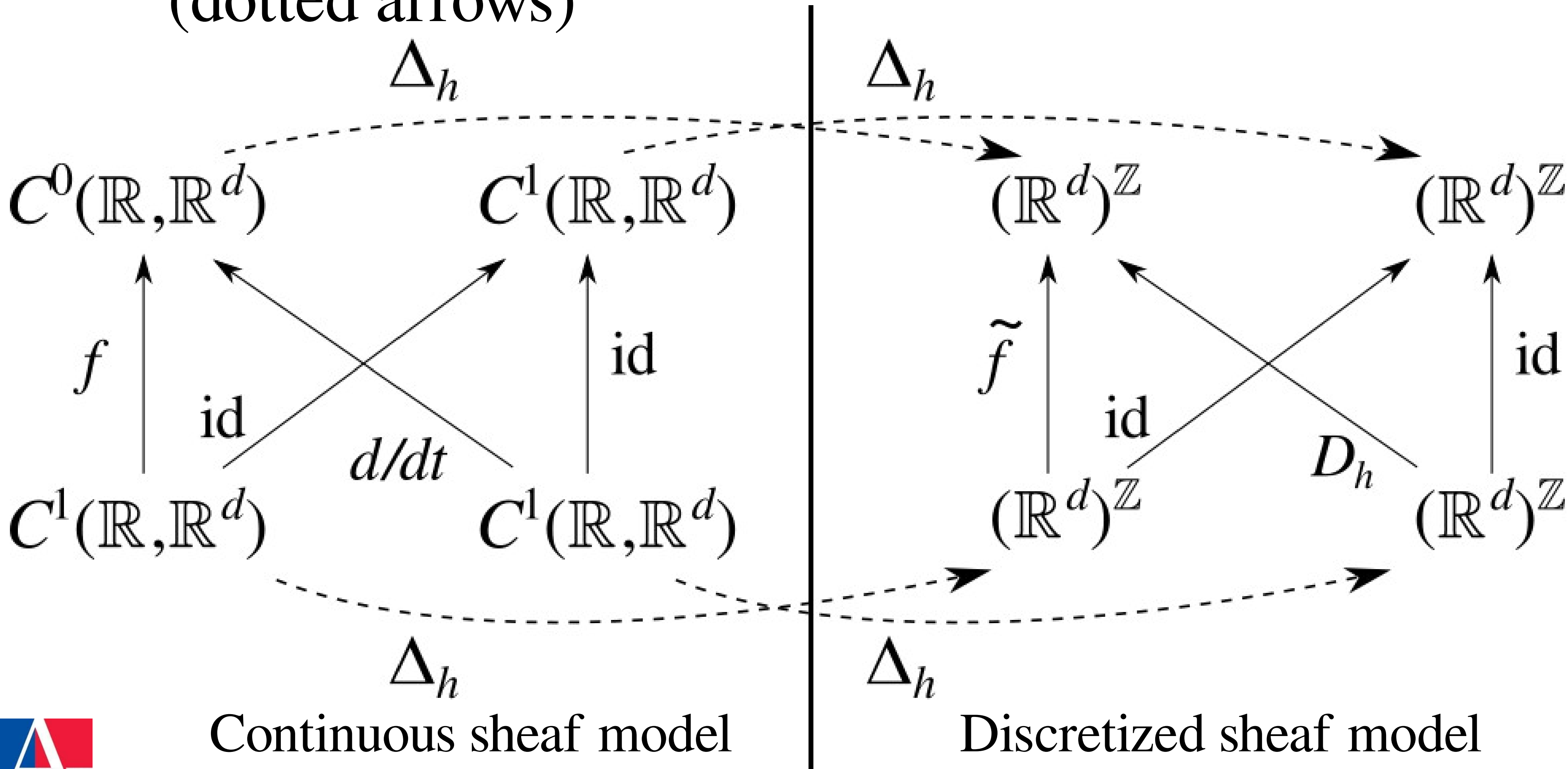
- Discretizing each function space via a fixed step h

$$(\Delta_h u)_n = u(nh)$$



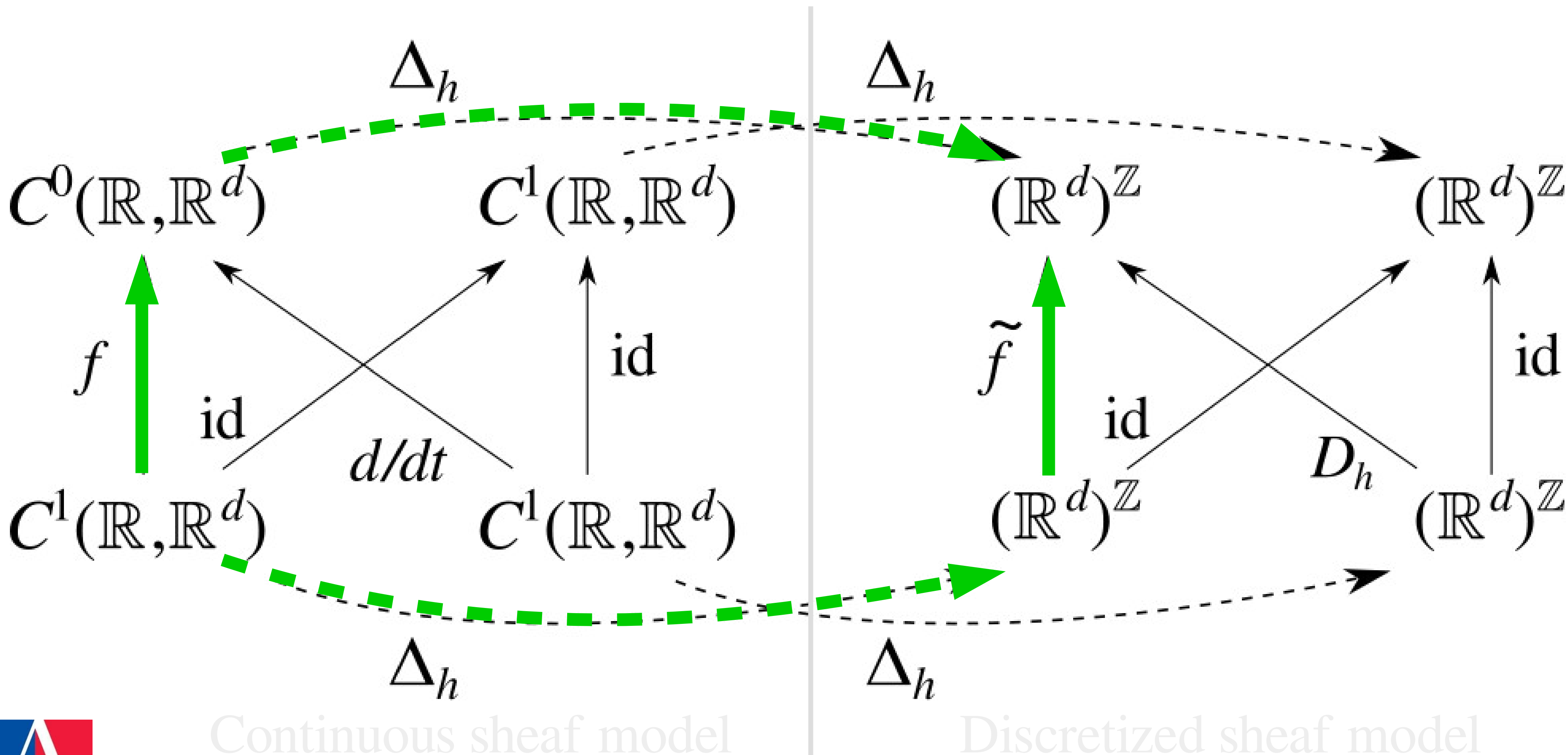
Is it a *sheaf morphism*?

- A *sheaf morphism* is a commutative diagram of maps between stalks of two sheaves… is this one? (dotted arrows)



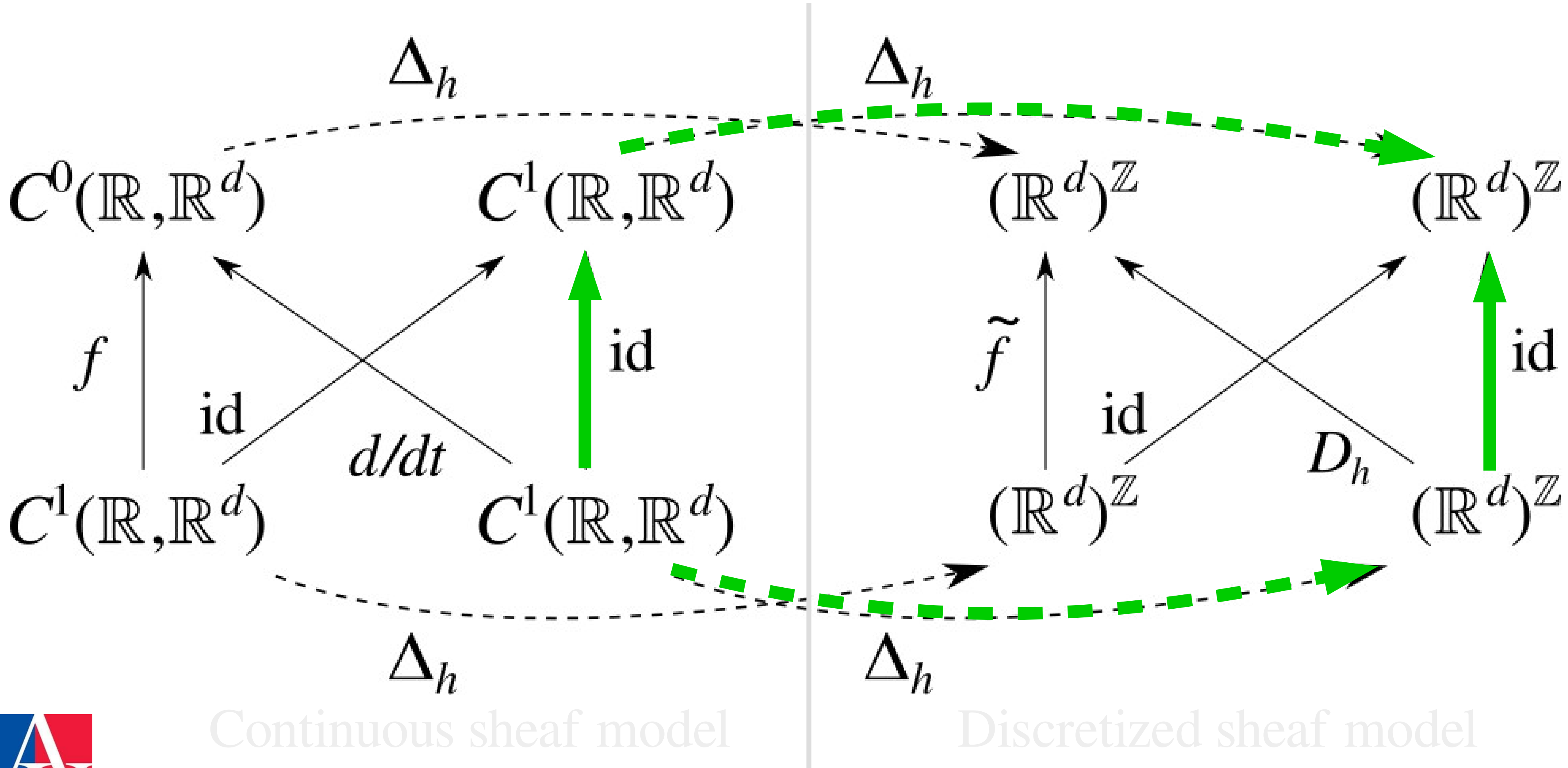
Is it a *sheaf morphism*?

- This square commutes if we pick $\tilde{f}$ correctly...



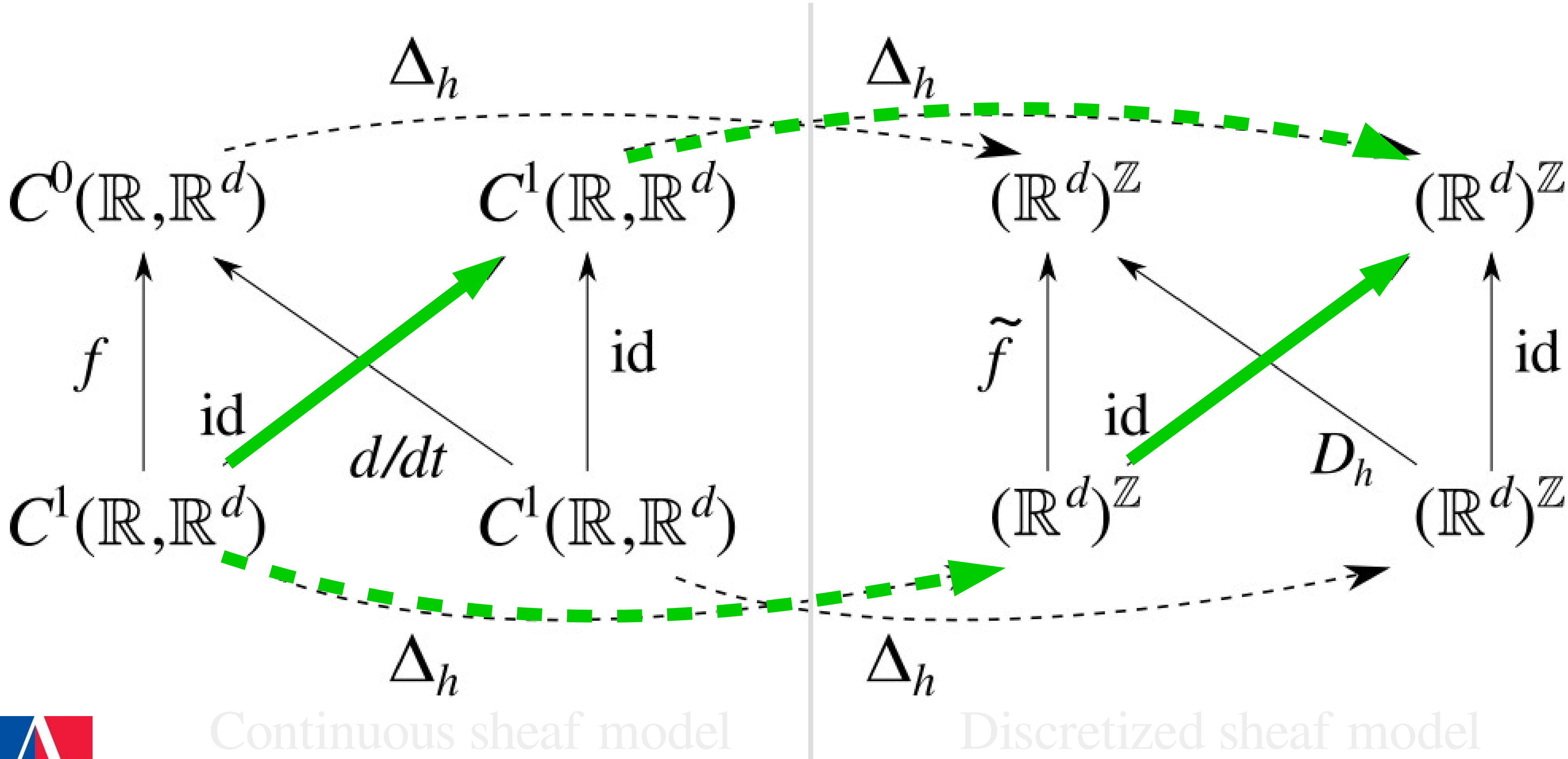
Is it a *sheaf morphism*?

- … this one commutes trivially …



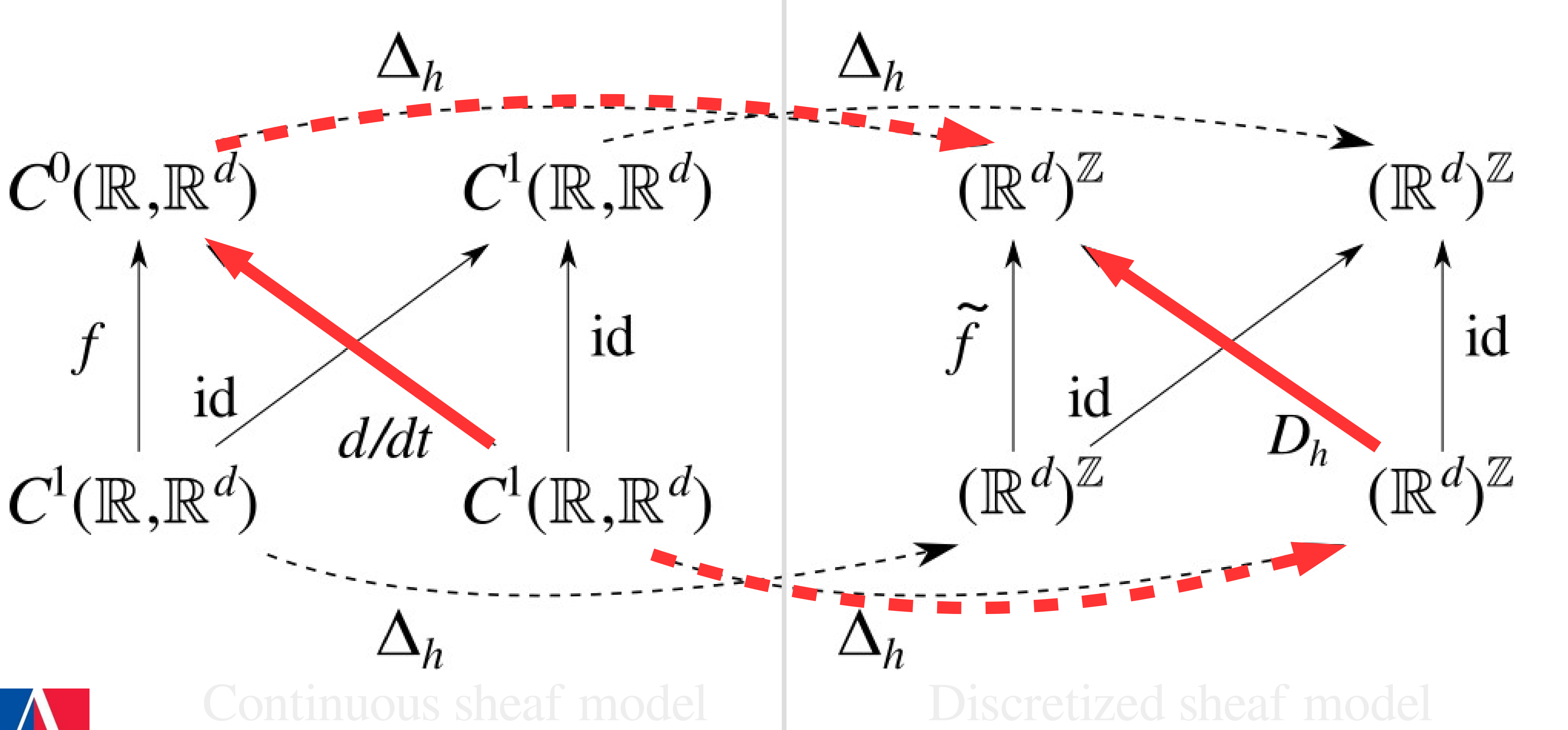
Is it a *sheaf morphism*?

- … this one also commutes trivially …



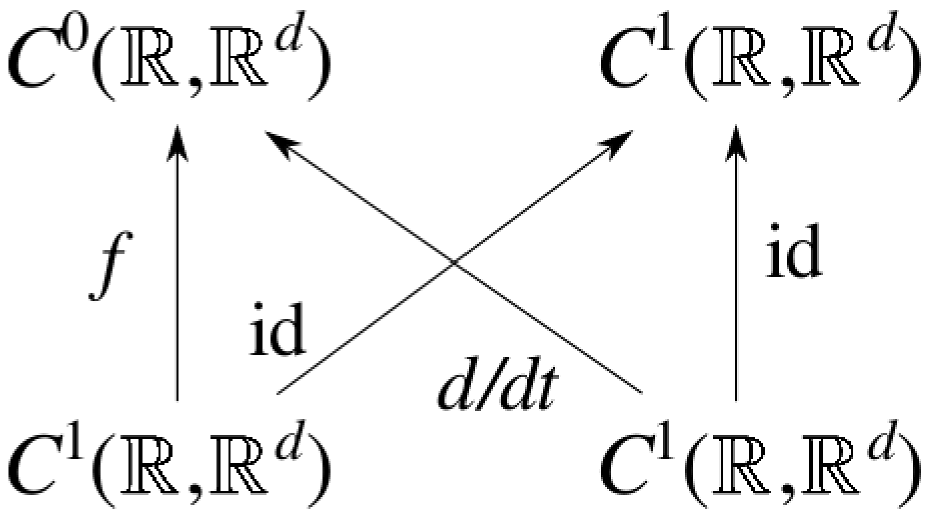
Is it a *sheaf morphism*?

- ...but this asks that $u'(nh) = D_h u_n$, which means discretized version is **exactly correct**. Oops!



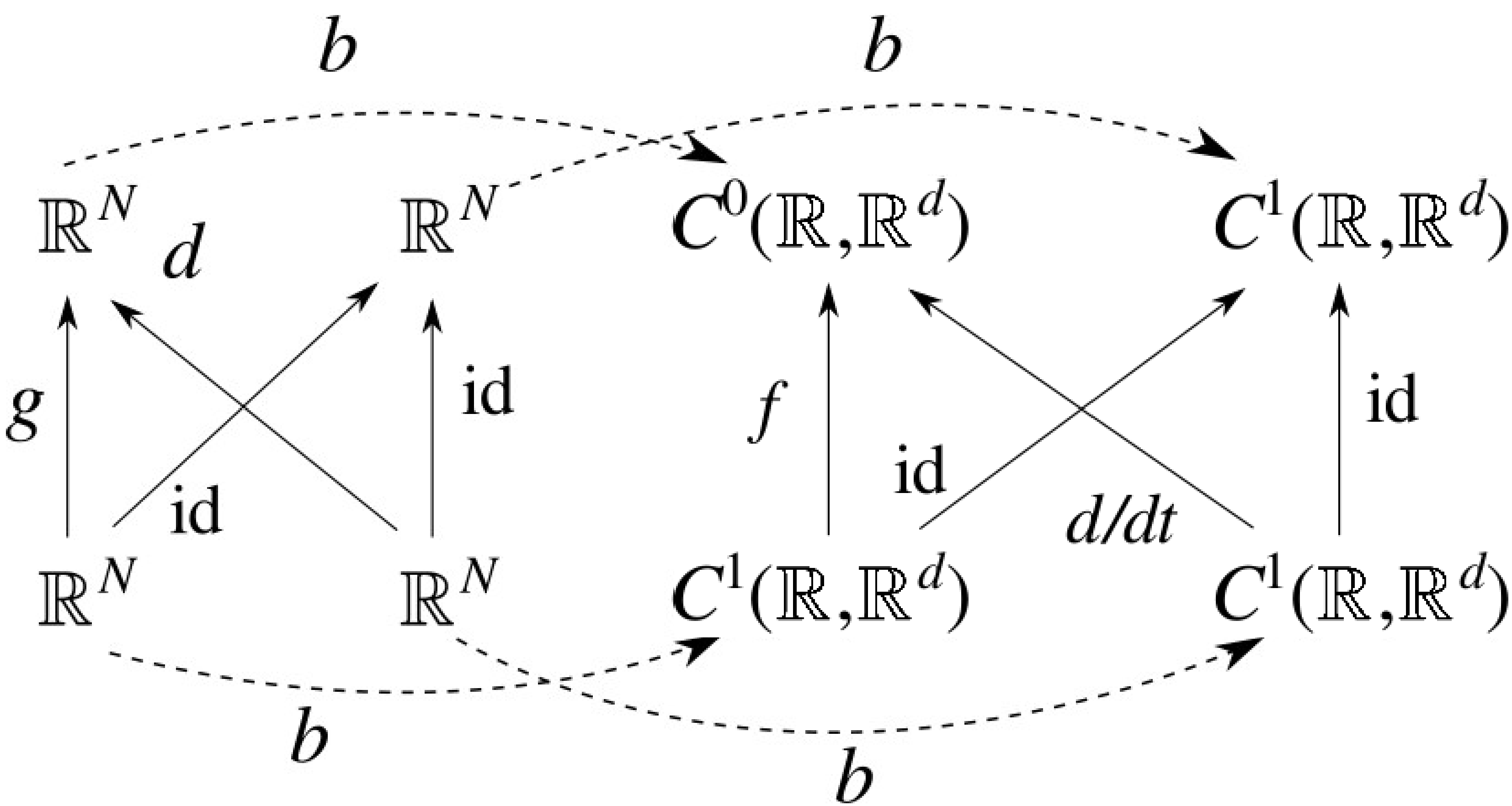
Finite elements

- We can also try to construct a finite elements approximation… from the “other side”
- Again start with the same continuous sheaf model



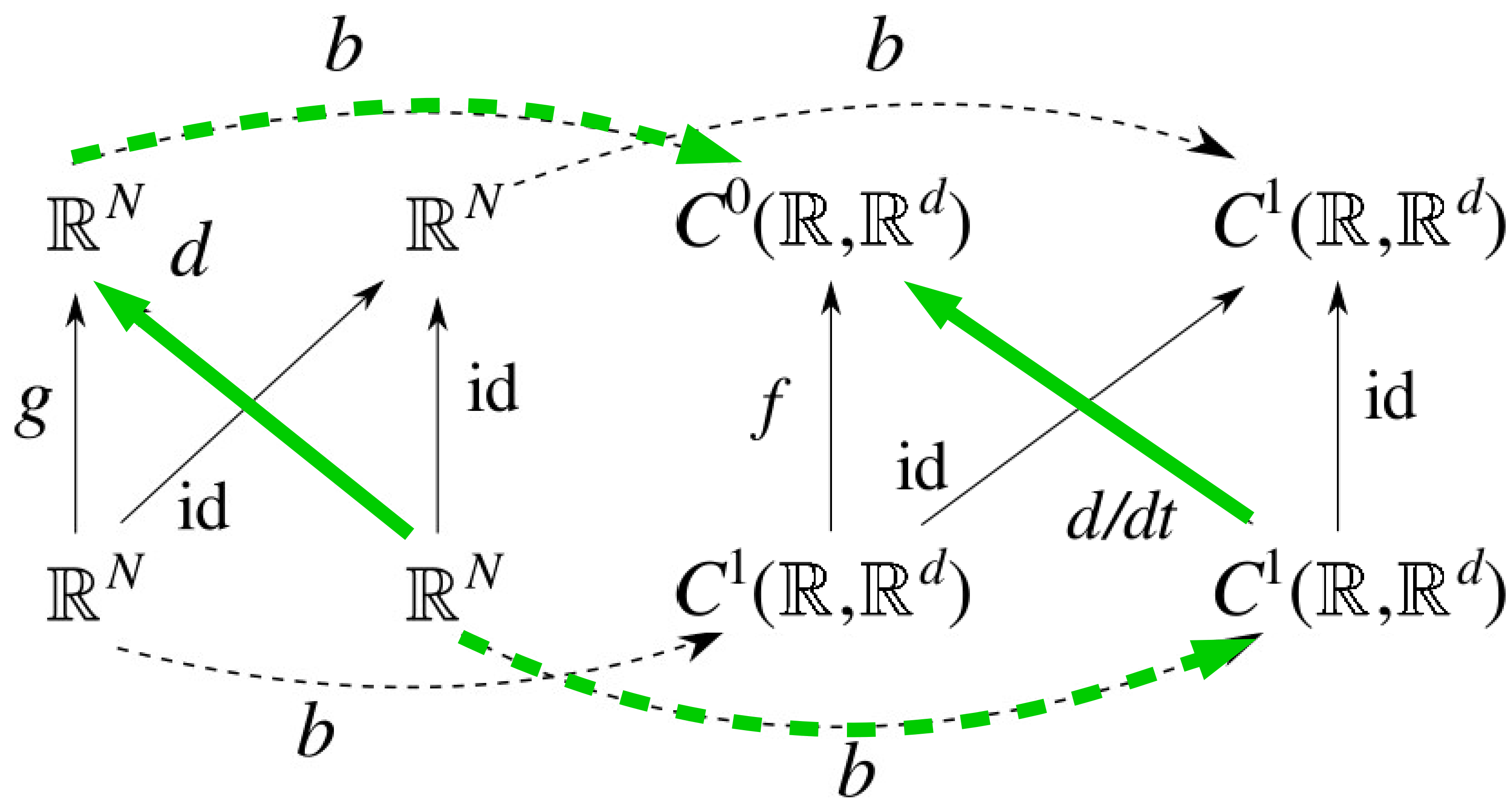
Finite elements sheaf model

- Use an N dimensional subspace of functions with a linear embedding $b : \mathbb{R}^N \to B \subseteq C^1(\mathbb{R}, \mathbb{R}^d)$.



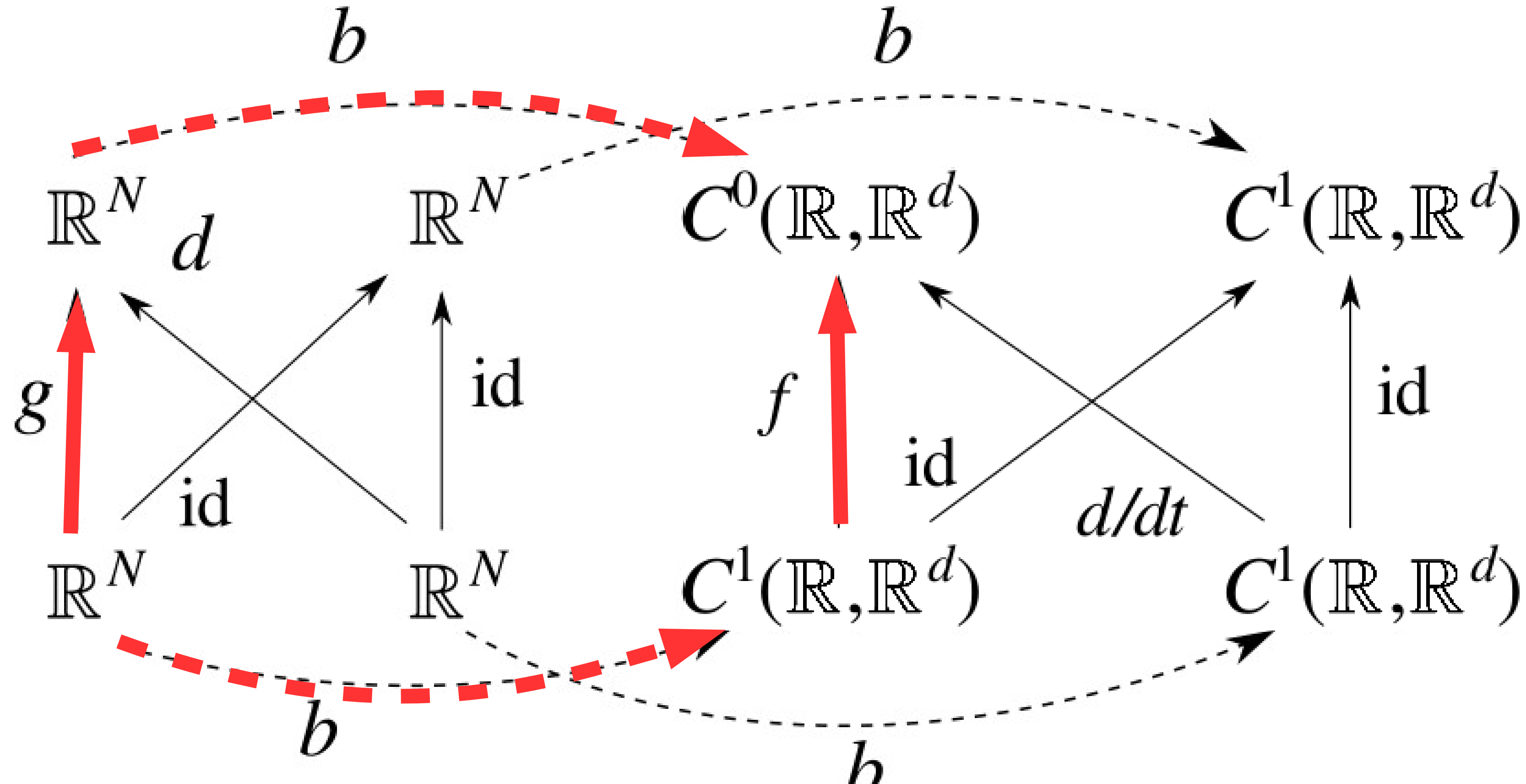
Is it a sheaf morphism?

- Although the derivative approximation can now be corrected by a judicious choice of embedding *b*…

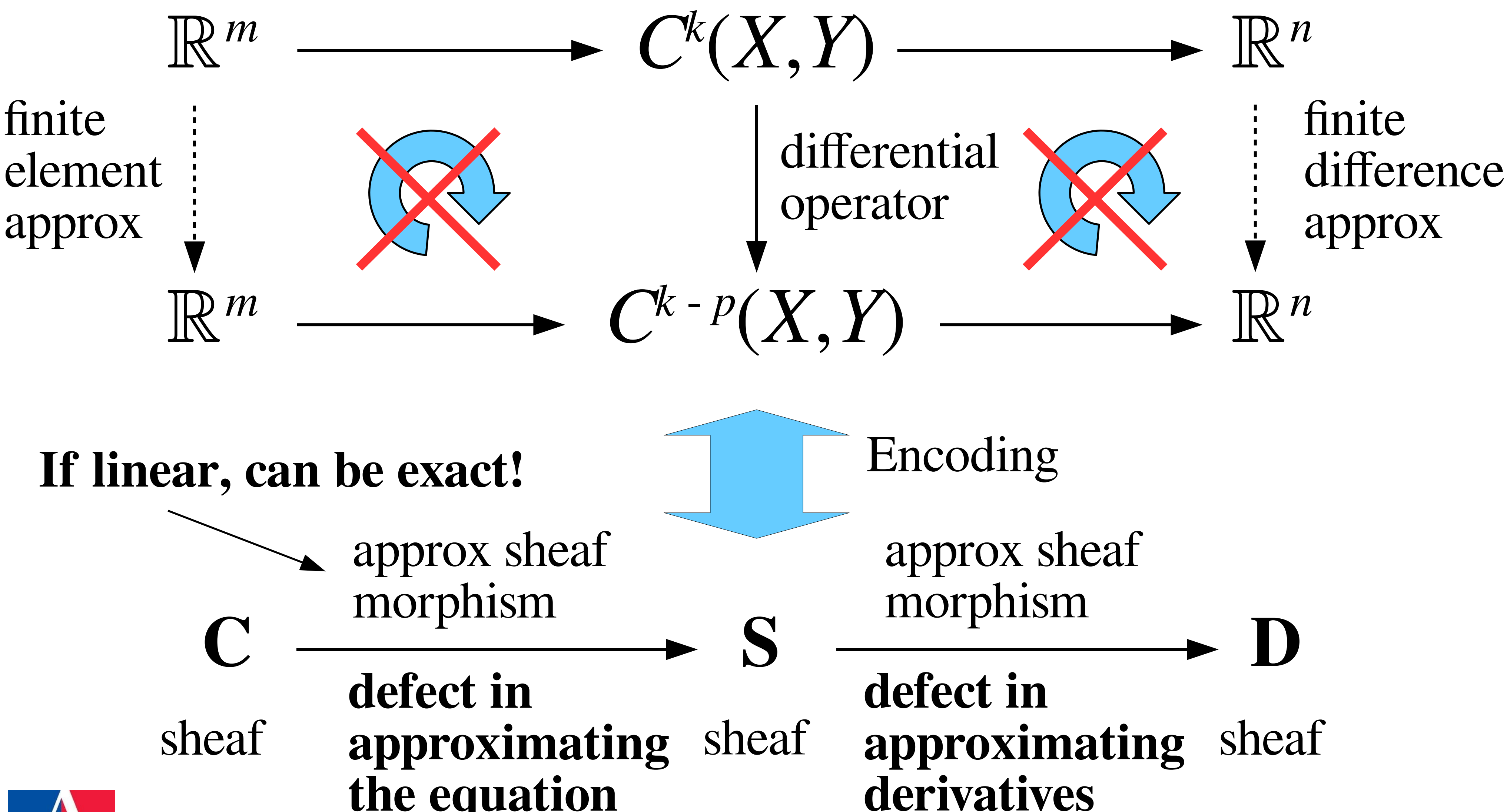


Might be a sheaf morphism...

- …if not linear, now the equation itself fails
- …if linear, we may get a morphism; *Galerkin method*.



Observations about consistency



In summary...

Sheaves capture variable relationships in any system of equations; that's most scientific models!

- Differential equation systems
- Bayes & causal nets
- … basically anything described by equations

Consistency radius estimates:

- Measurement error,
- Data modeling error, and
- Discretization error