A gentle introduction to

Encoding prior knowledge in image- and data analysis

J. Lellmann

Heidelberg Laureate Forum 2015
24 August 2015

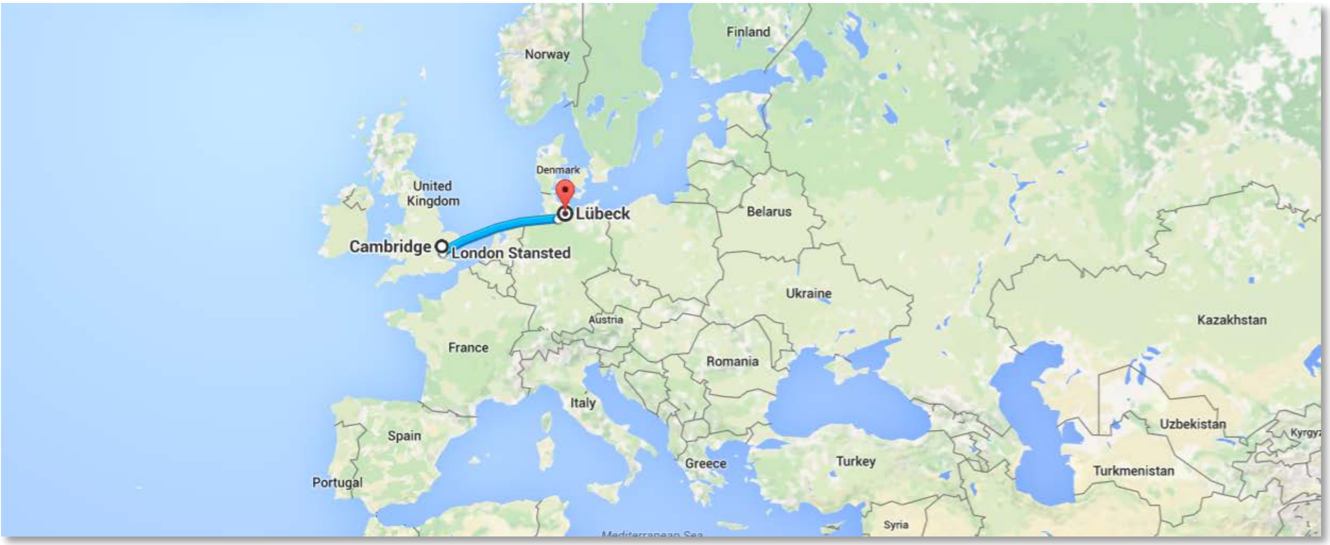
J.Lellmann@damtp.cam.ac.uk

Cambridge Image Analysis
University of Cambridge

www.damtp.cam.ac.uk/research/cia

Mathematical Image Computing
University of Lübeck

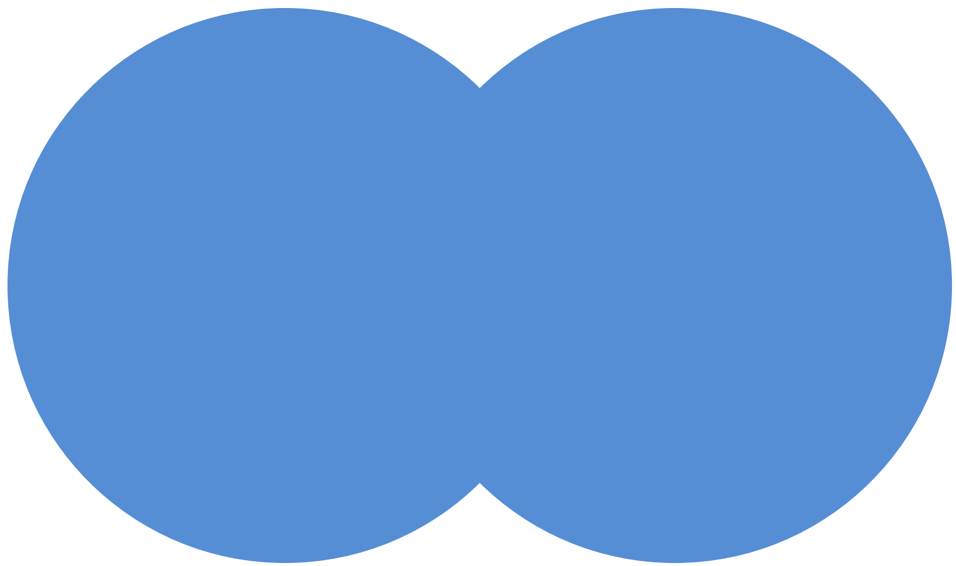
www.mic.uni-luebeck.de

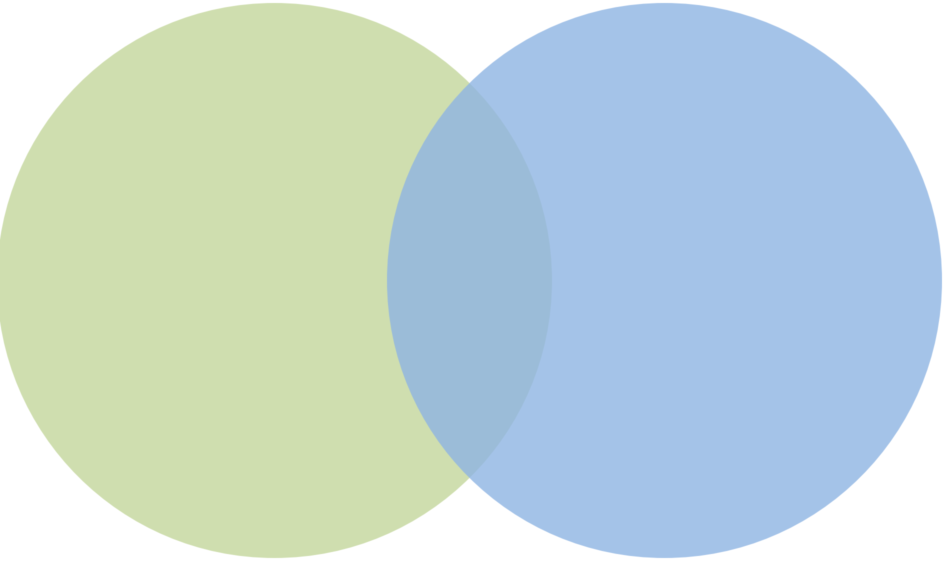


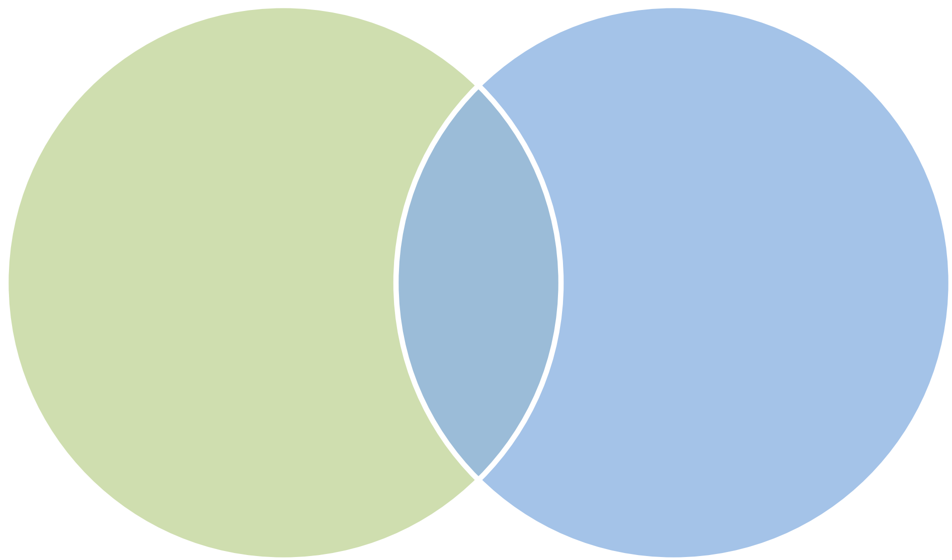
$$A + B = 4$$

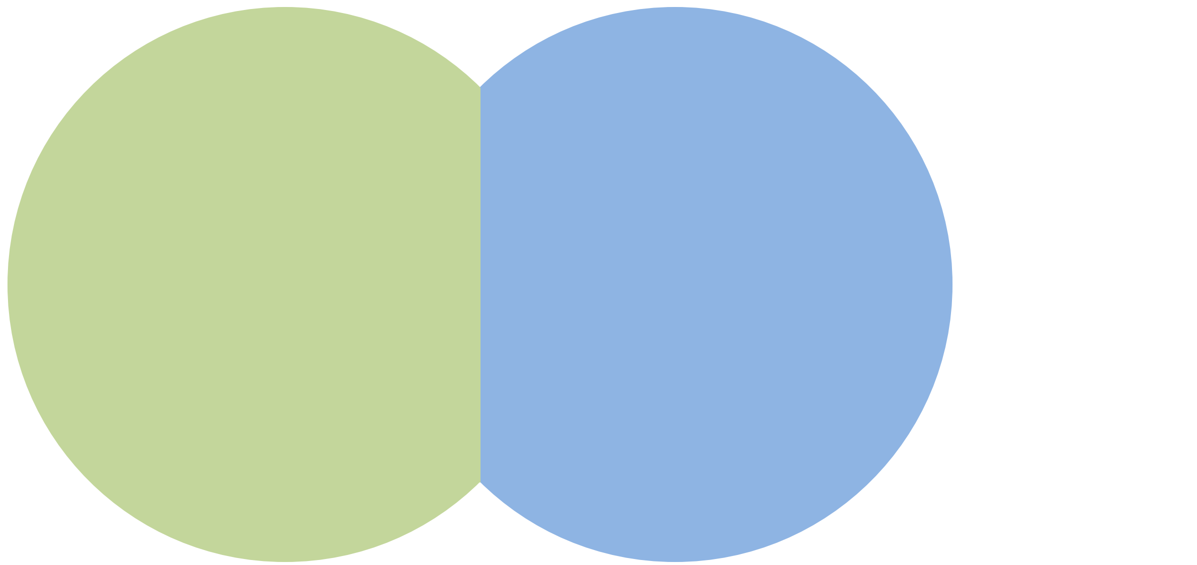
$$A = ? \quad B = ?$$

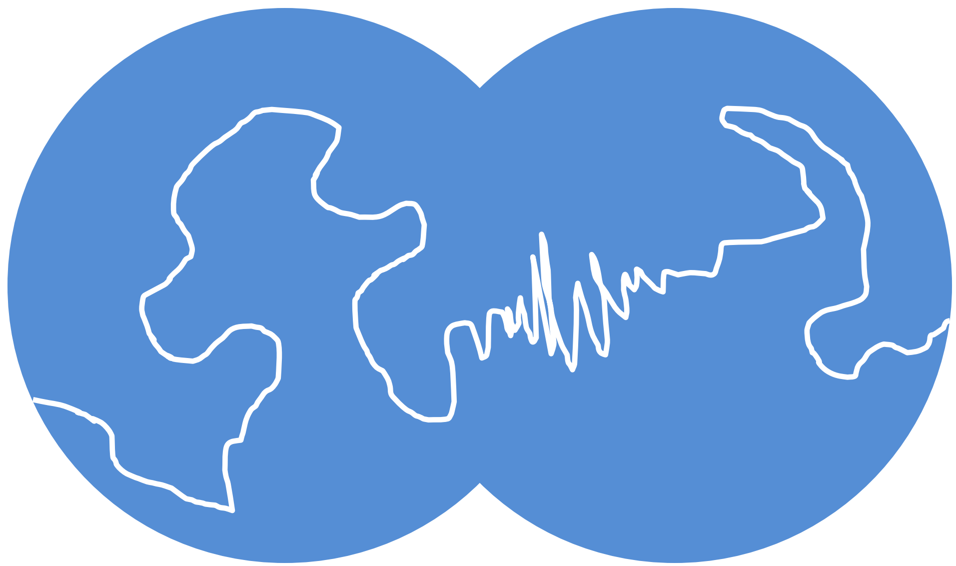
C. Moler '90, M. Teboulle '15

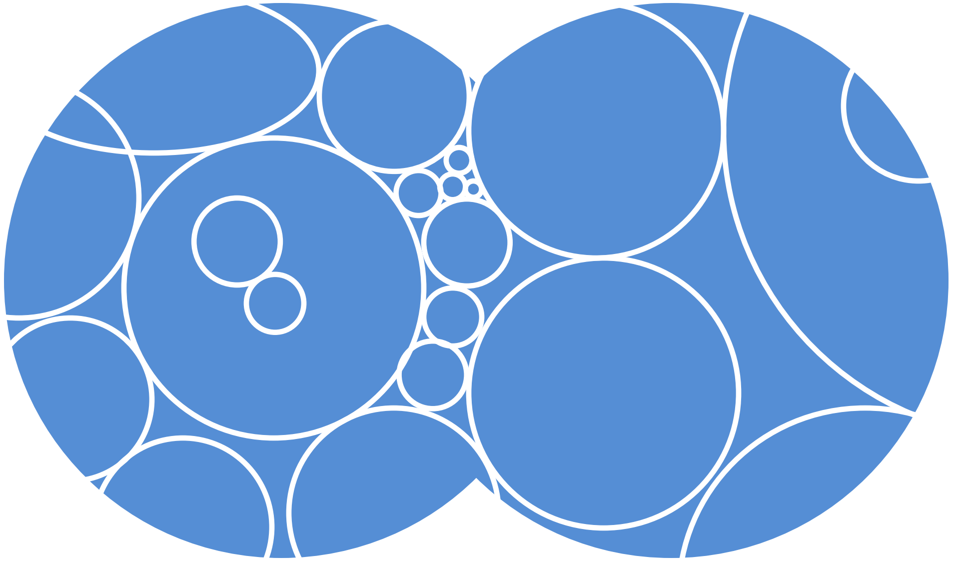










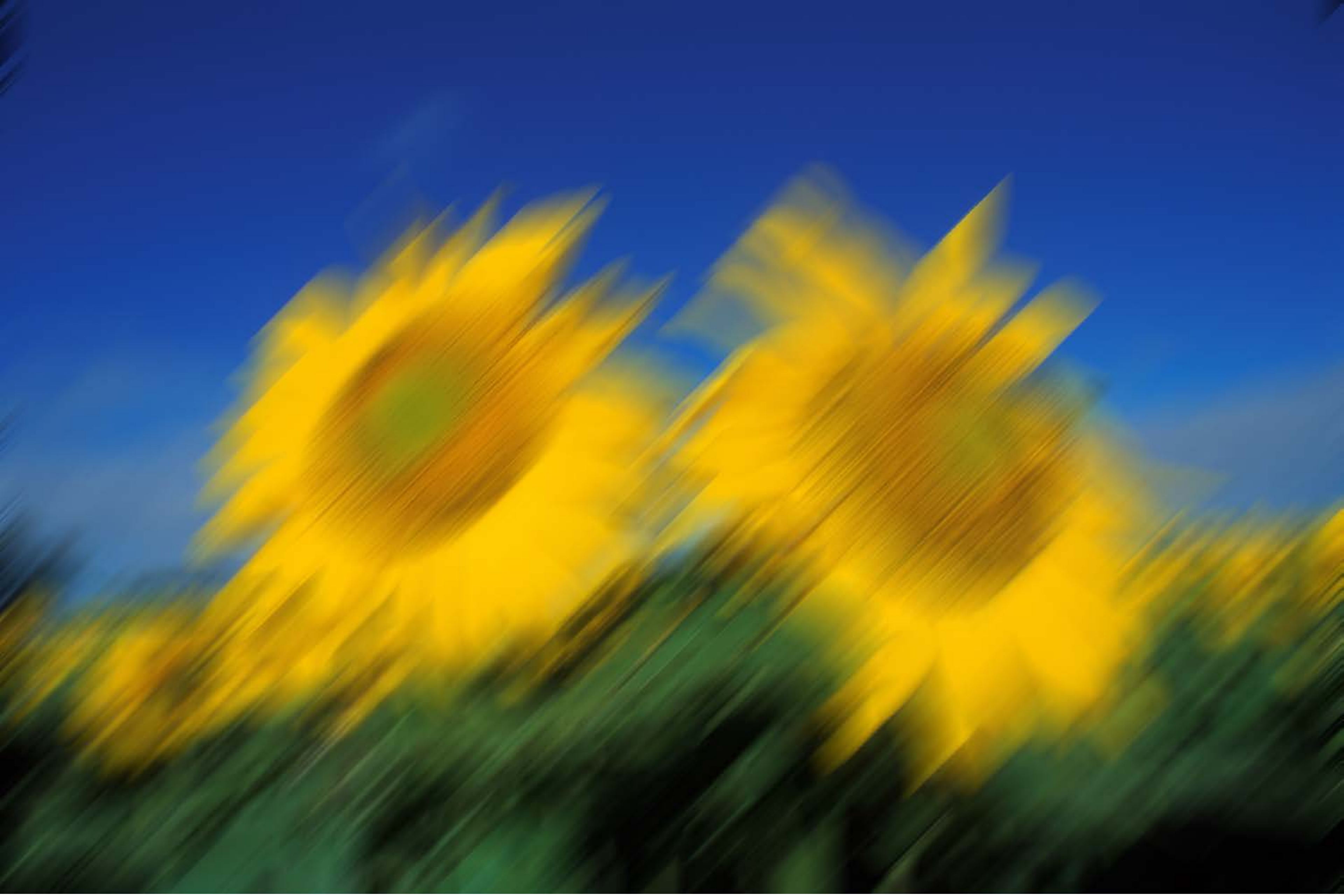


Regularity

A natural object (and data in general) is often

- composed of few objects,
- each of which is "simple" (e.g., geometrically)

u



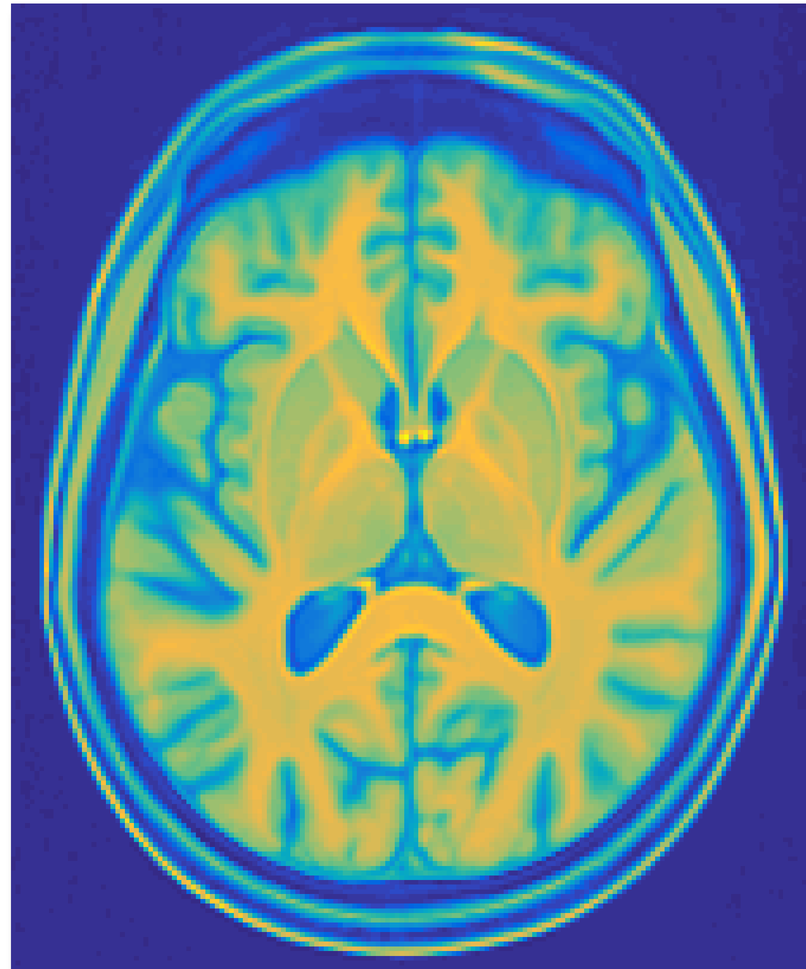
$$b = T(u)$$

$$b = T(u) + n$$

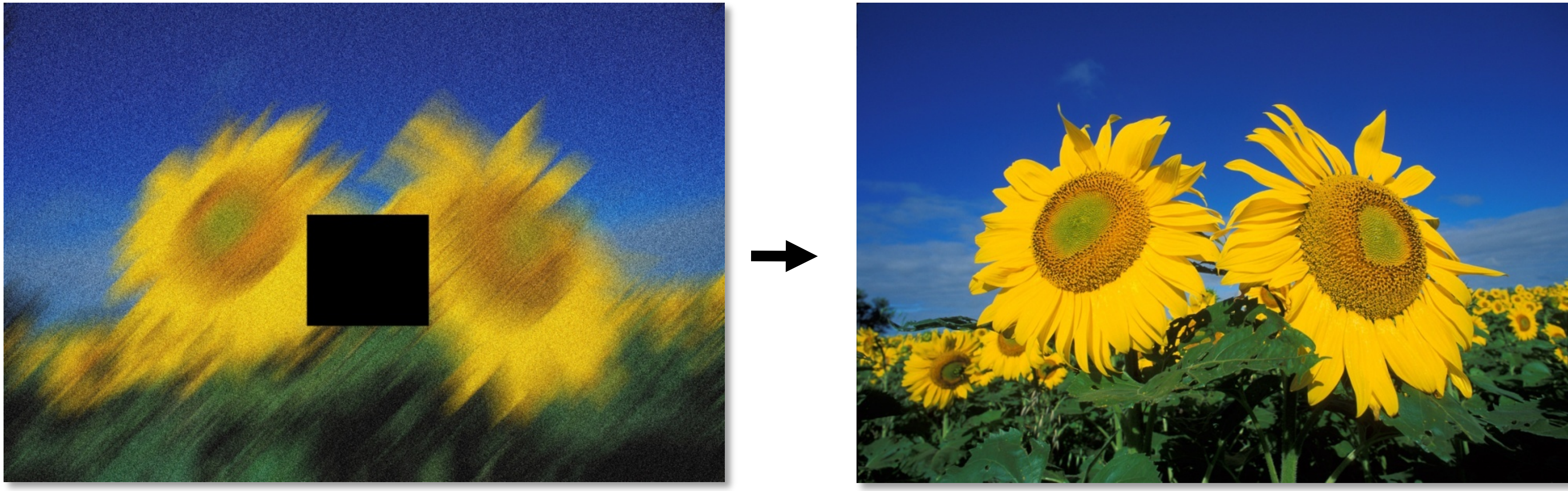
$$b = T'(u) + n'$$

Missing data

R. Hocking



J. Acosta-Cabronero



Given measurements **b**, find image data **u** so that

$$\mathbf{b} = \mathbf{T}(\mathbf{u}) + \mathbf{n}$$

T structural operator, **n** random noise

Often the direct reconstruction is not unique, not stable, or not deterministic – we need prior knowledge

Variational methods

We reconstruct the unknown data u from the measurements b by minimizing the energy

$$\min_u \{D(T(u); b) + R(u)\}$$

Advantages:

- Intuitive – we specify what the results should look like
- Often statistical motivation – maximum a posteriori estimate
- Modular, reusable components

Tikhonov '63

b

$$\int_\Omega \|u(x) - b(x)\|_2^2 dx + \lambda \int_\Omega \|\nabla u(x)\|^2 dx$$

Trade-offs

Model complexity vs. **computability**
Local minimizers vs. **global minimizers**

Top-down approach

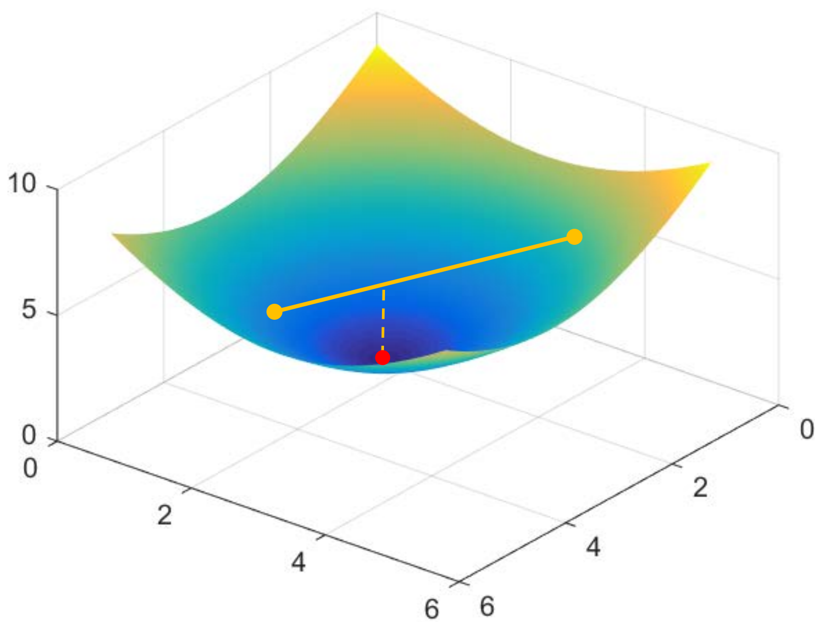
- Physically/biologically motivated
- Advantages:
  - Very specific to the problem
  - Model parameters have meaning

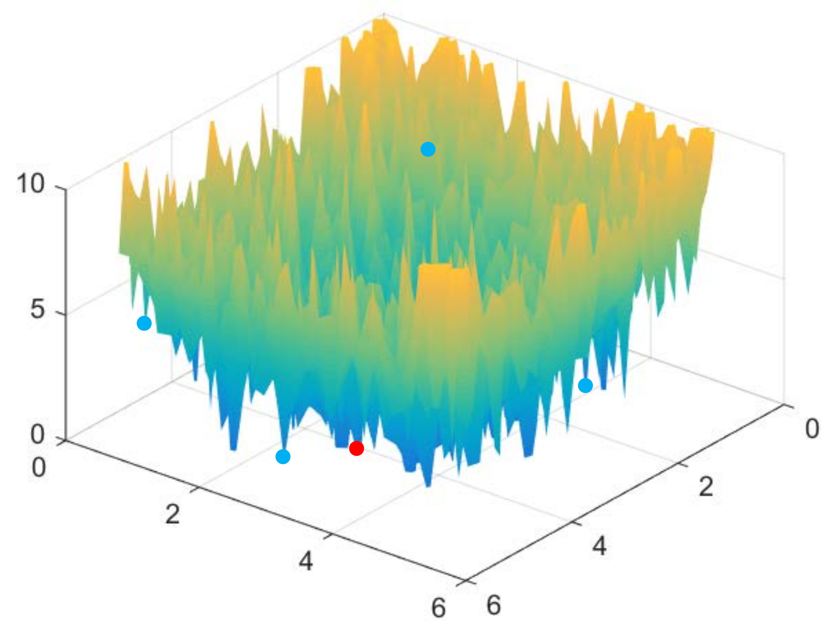
Bottom-up approach

- Built from simple, well-understood components
- Advantages:
  - Mathematical analysis
  - Efficient and/or global optimization often possible

Convexity

Convexity assures that every local minimizer of the energy is also a global minimizer



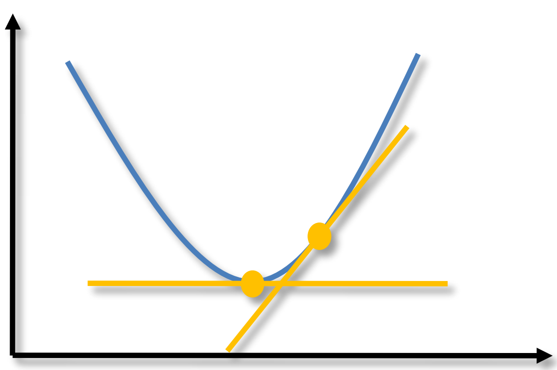


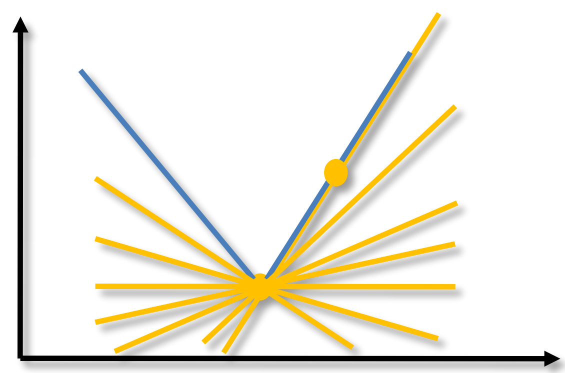
Non-smoothness

Non-smoothness often allows perfect recovery.

If f is convex but non-differentiable, then (Fermat):

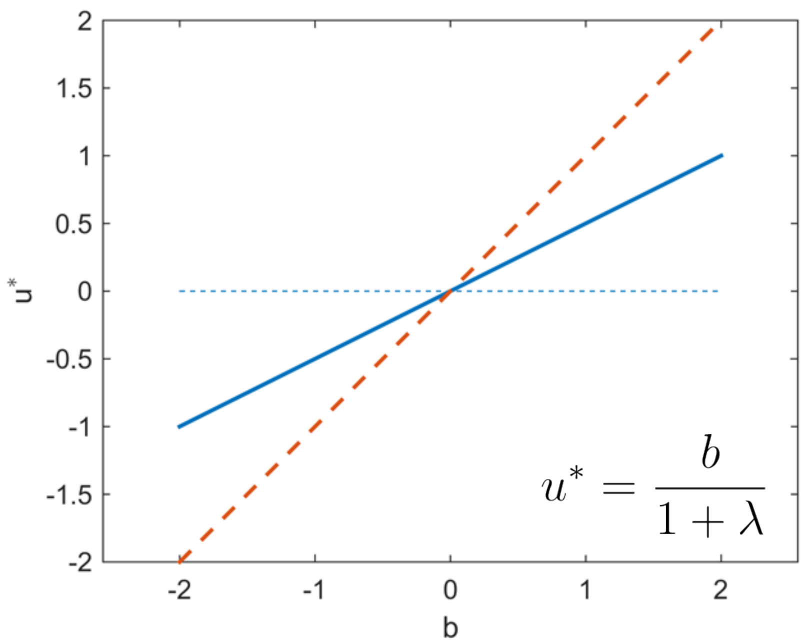
$$0 \in \partial f(u) \Rightarrow u \text{ is a global minimizer}$$





A simple example

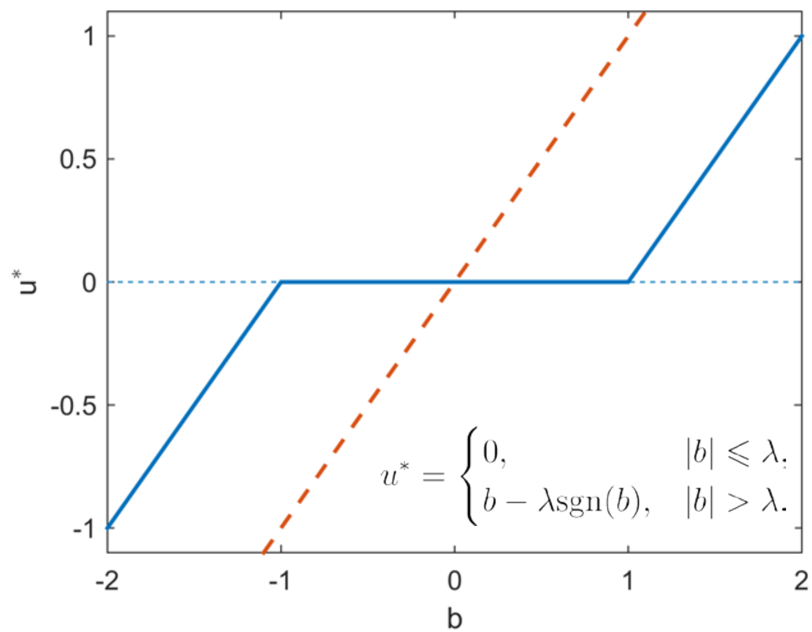
Assume u is scalar and our (perfect!) prior knowledge is that u is zero. We measure $b = u + n$ and try:



$$\min_{u \in \mathbb{R}} \frac{1}{2}(u - b)^2 + \frac{\lambda}{2} u^2$$

A simple example

Assume u is scalar and our (perfect!) prior knowledge is that u is zero. We measure $b = u + n$ and try:



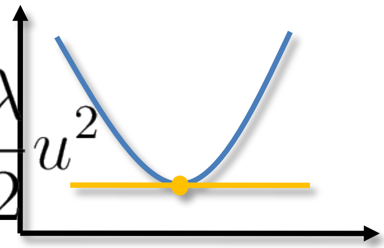
The solution is exact for a wide range of measurements!

$$\min_{u \in \mathbb{R}} \frac{1}{2}(u - b)^2 + \lambda |u|$$

Smooth:

$$\min_{u\in\mathbb{R}} \frac{1}{2}(u-b)^2 + \frac{\lambda}{2}u^2$$

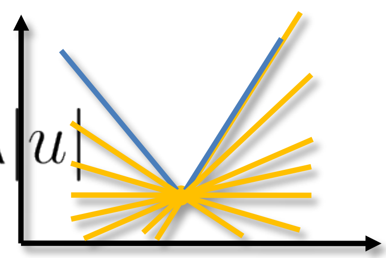
$u = 0$ optimal $\Leftrightarrow b \in$



Nonsmooth:

$$\min_{u\in\mathbb{R}} \frac{1}{2}(u-b)^2 + \lambda|u|$$

$u = 0$ optimal $\Leftrightarrow b \in$



Basis Pursuit

Given an overcomplete dictionary *A* and a vector *u* with few non-zero components, recover *u* from $b = A\,u$:

$$\min_{u \in \mathbb{R}^m} \|u\|_0 \text{ subject to } Au = b$$

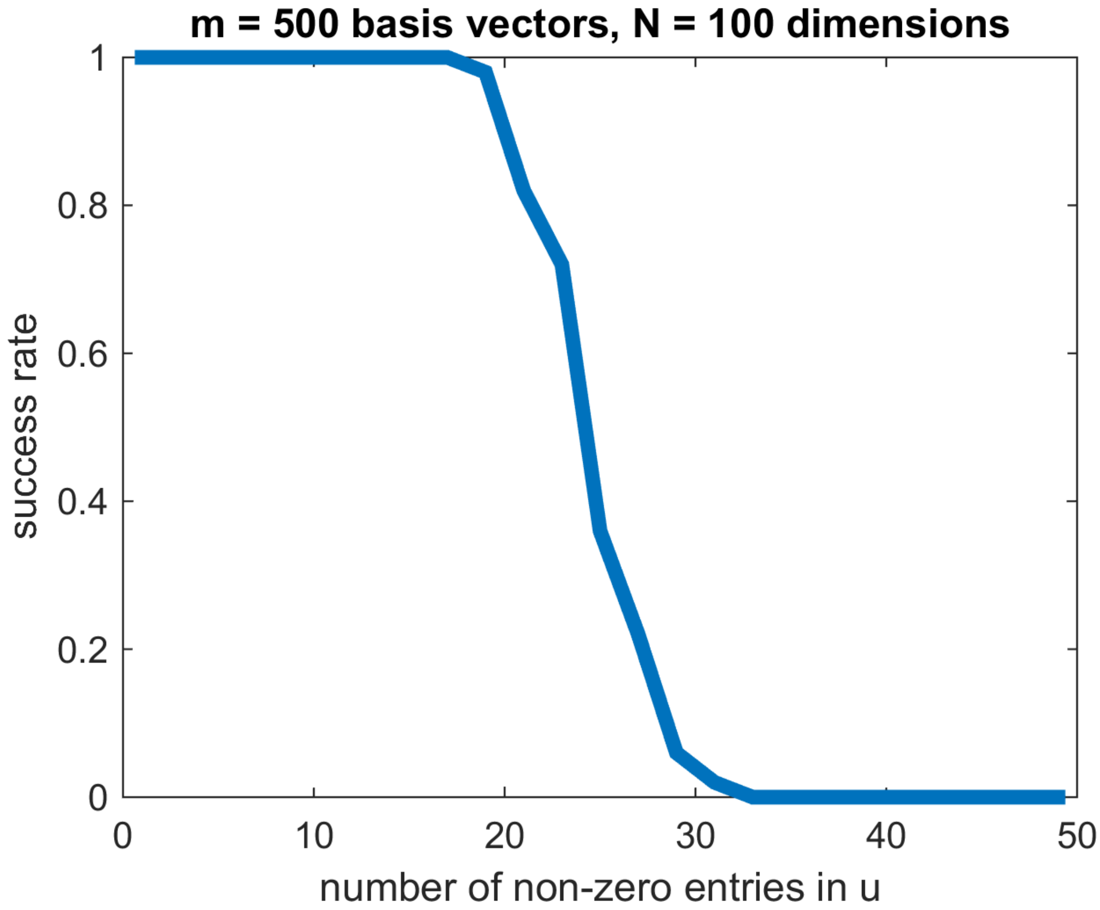
?

Applications in compression, data separation, machine learning, approximation theory,…

Compressive Sensing, Dictionary Learning consider how to choose *A* optimally.

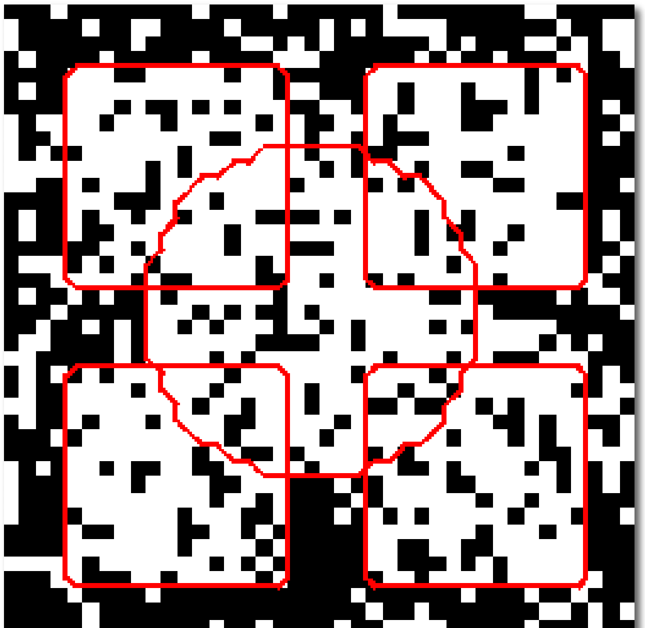
Candes, Romberg, Tao '06; Donoho '06
Overview: Foucart, Rauhut '10

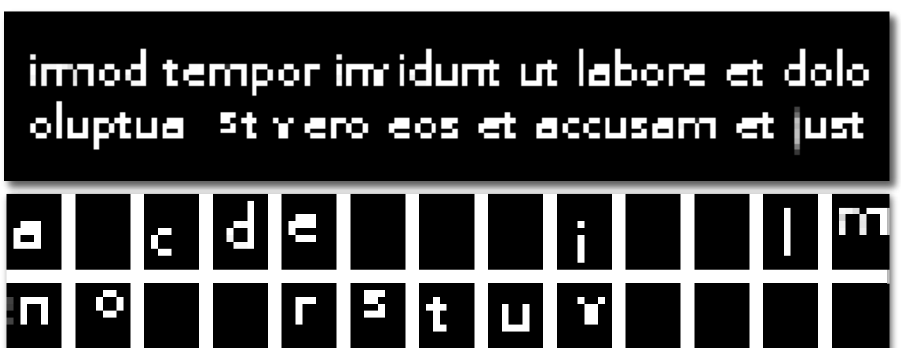
Phase transition



This gives the exact (sparsest) solution in many cases!

Sparse Shape Decomposition





Lellmann, Breitenreicher, Schnörr '11

b

$$\int_\Omega \|u(x) - b(x)\|_2^2 dx + \lambda \int_\Omega \|\nabla u(x)\|^2 dx$$

$$\int_\Omega \|u(x) - b(x)\|_2 dx + \lambda \int_\Omega d\|Du\|_2$$

Back to the roots

Gradient Descent

To minimize *f(u)*, follow the gradient downwards:

$$u_t = -\nabla f(u) \quad \longrightarrow \quad \frac{\boldsymbol{u^{k+1} - u^k}}{\boldsymbol{t^k}} \in -\boldsymbol{\partial f}(u^{k+1})$$

This backward step is unique and can be computed explicitly for many "simple" convex functions.

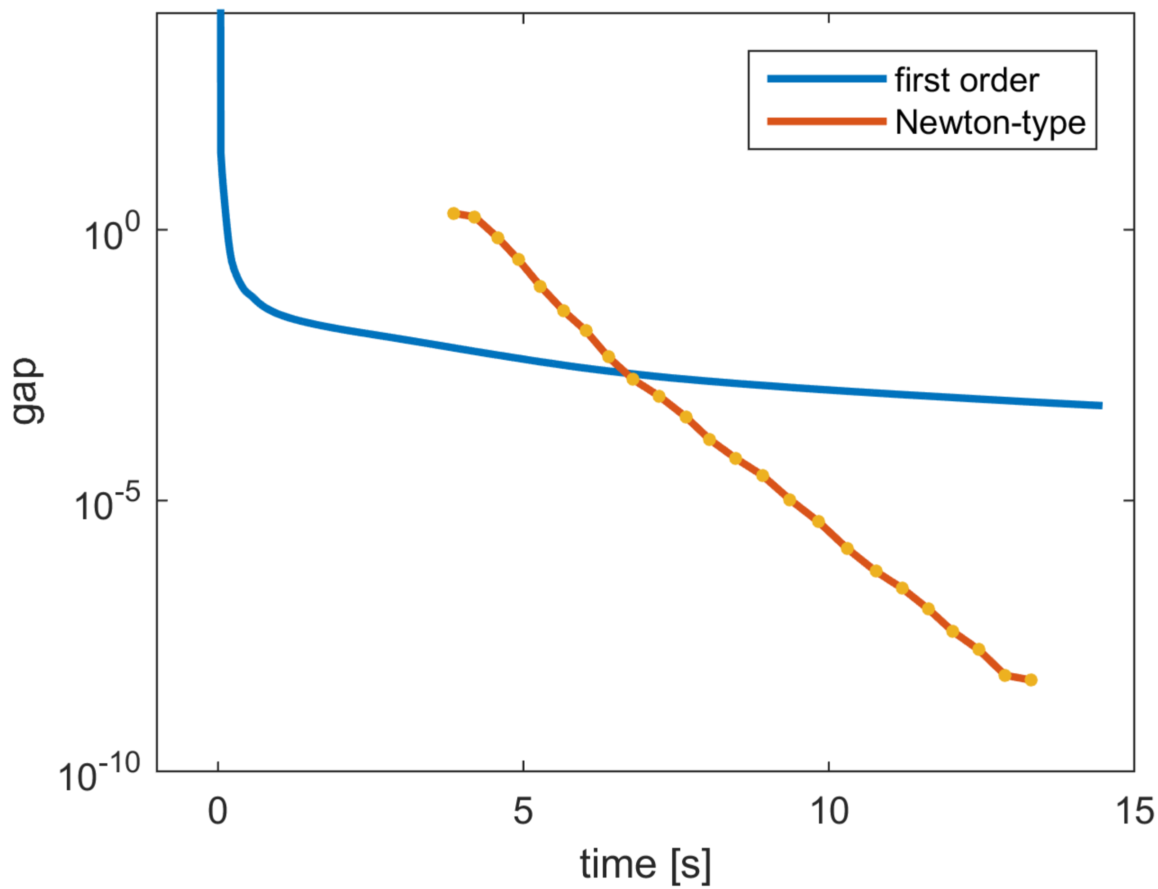
Then apply backward steps to saddle-point form

$$\inf_u \sup_v F(u) + v^\top A u - G(v)$$

This is slow: *O(1/N)*, linear if strongly convex. ***BUT:***

Rockafellar '76; Lions, Mercier '79; Glowinski, Marocco '75; Gabay, Mercier '78; Bertsekas, Tsitsiklis '89
Alvarez, Attouch '01; Nesterov '04; Pock et al. '09; Combettes, Pesquet '11; Shefi, Teboulle '14

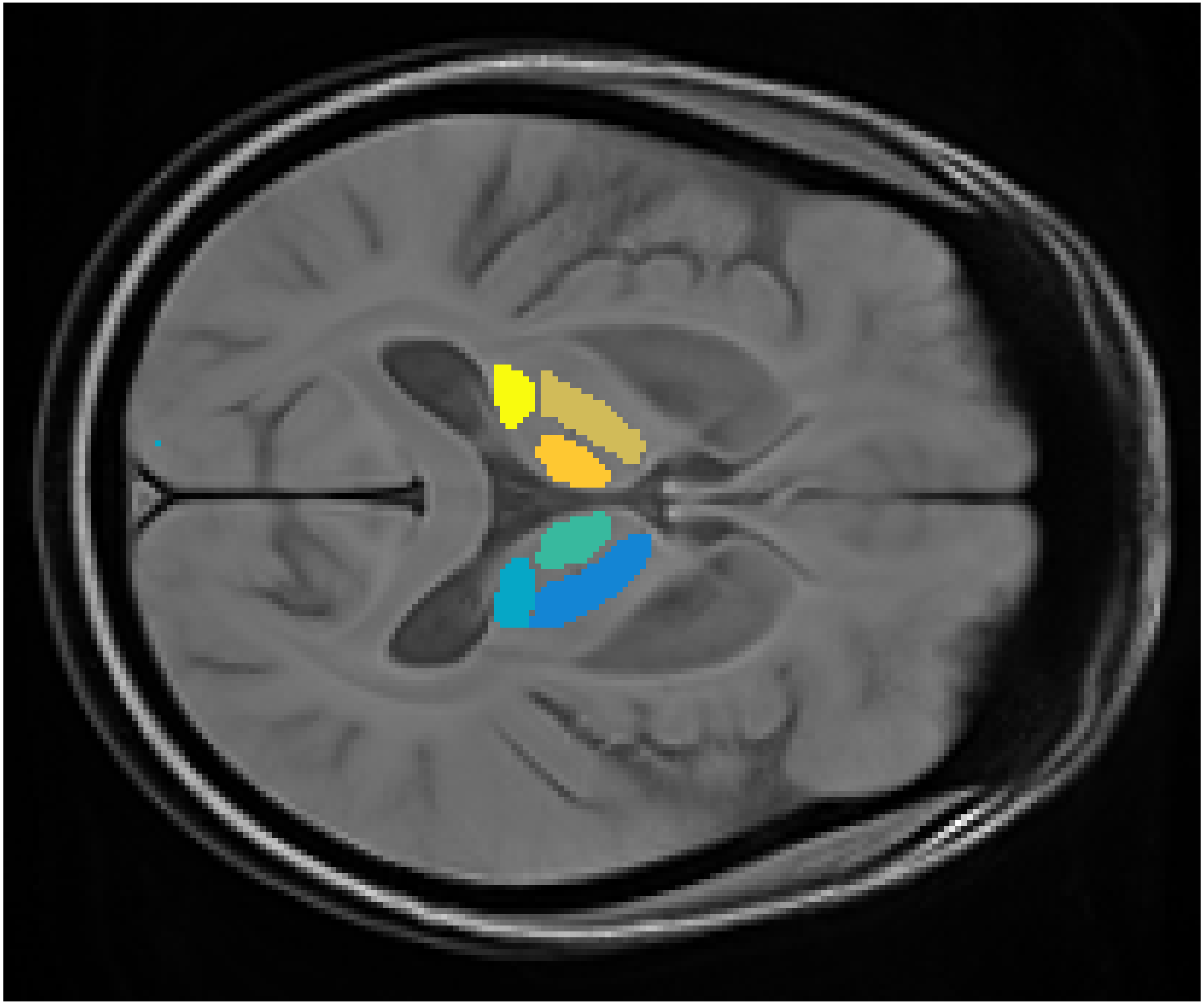
The 80-20 rule



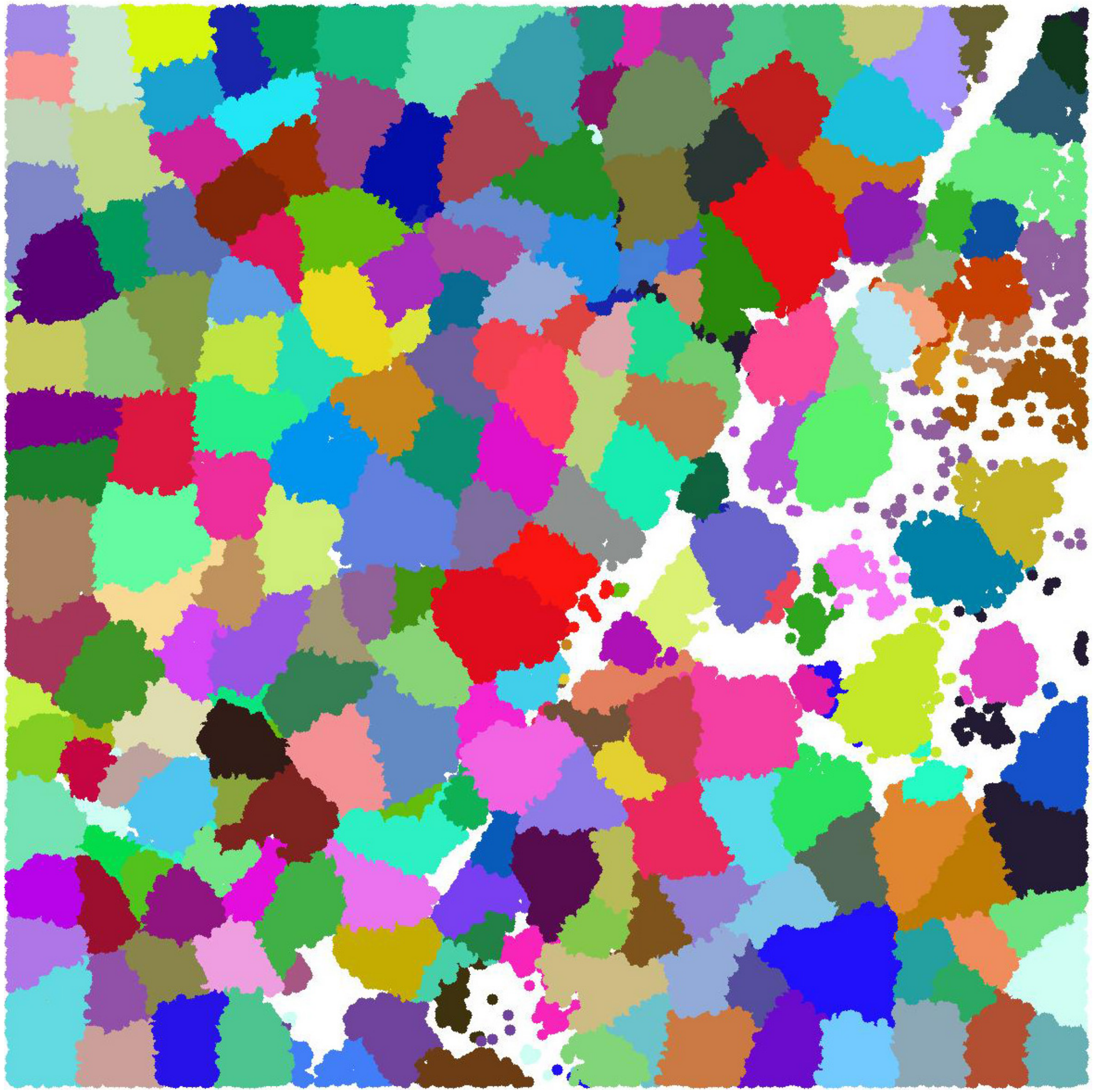
Sometimes quantity (speed) beats quality (accuracy)!

Demo

But what if…

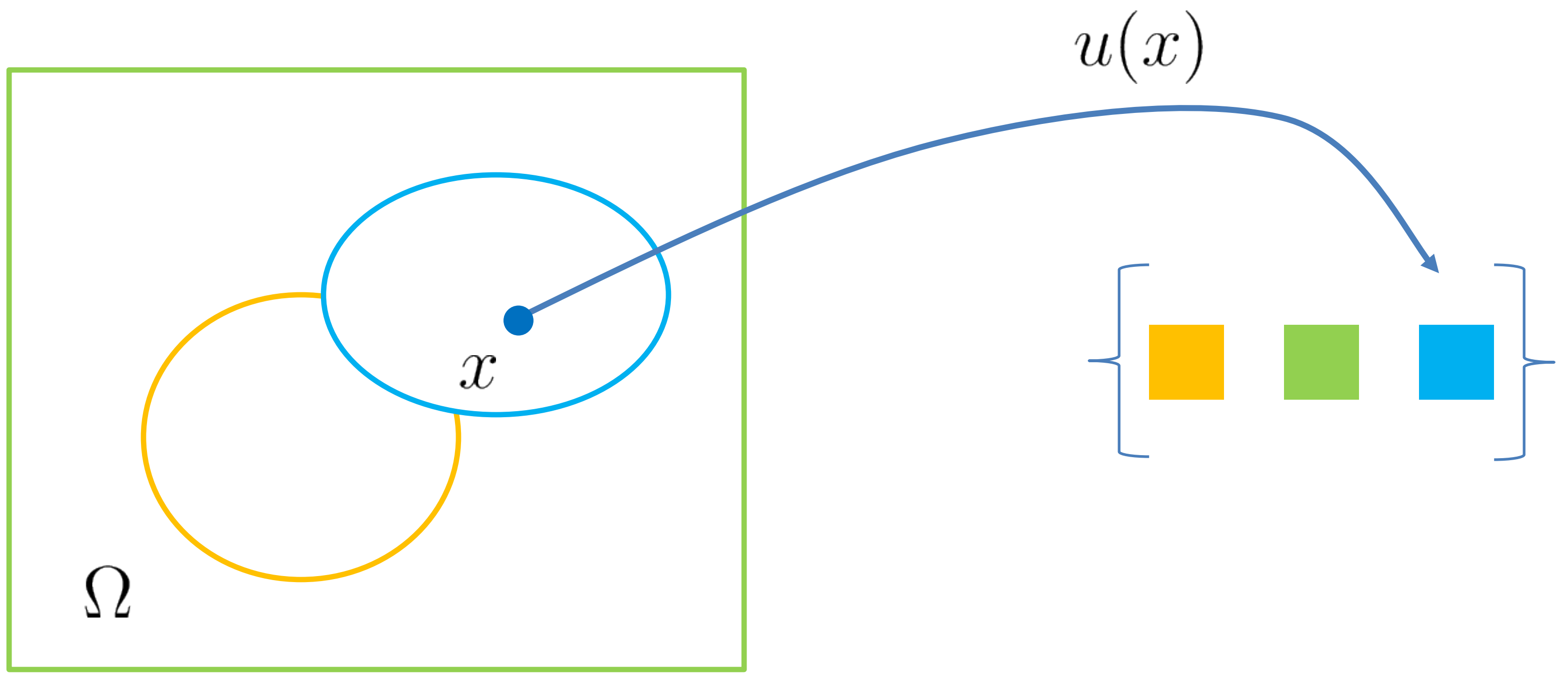


With: V. Corona, C. Schönlieb, J. Acosta-Cabronero, P.~Nestor/DZNE Magdeburg



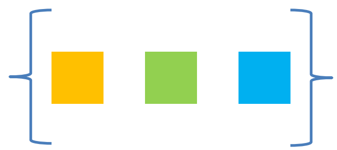
With: J. Lee, D. Coomes, C. Schönlieb

Labeling problems



First approach:

$$\min_{u:\Omega\to X} f(u) := D(u; I) + R(u)$$

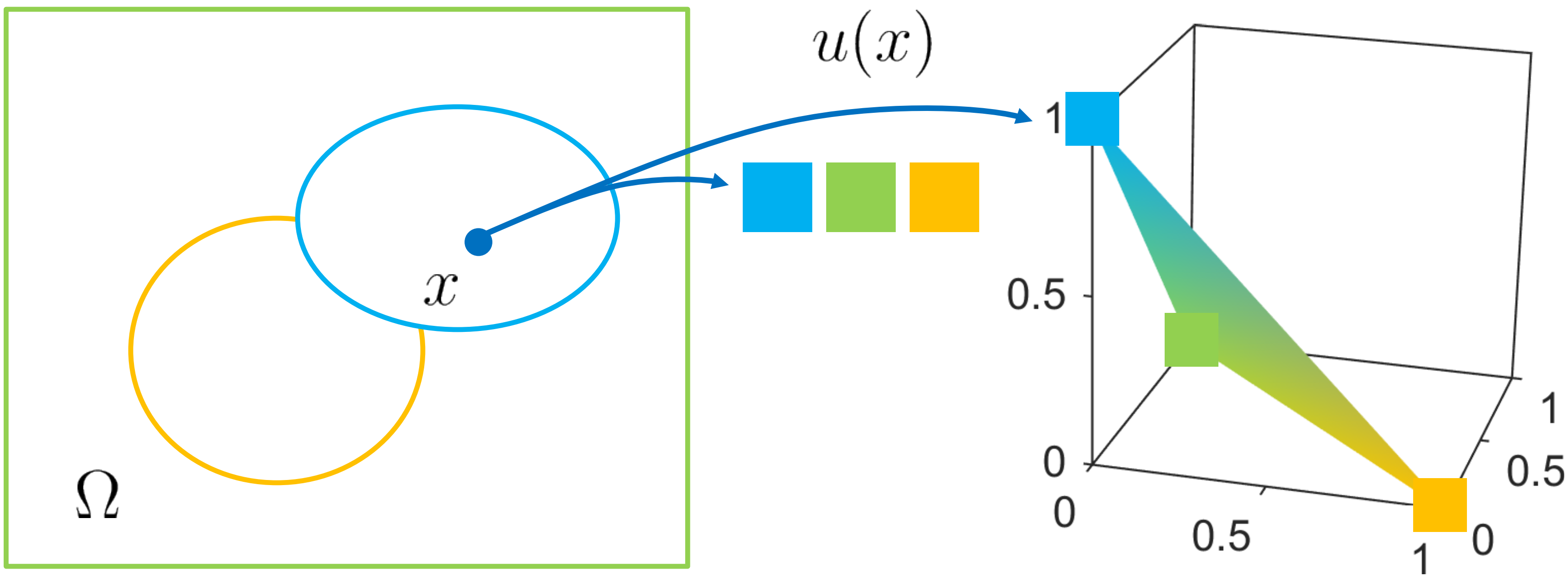
$X :=$ 

This is a combinatorial problem and generally very hard:

- *X* does not have an additive structure - no gradients,
- in particular there is no convexity.

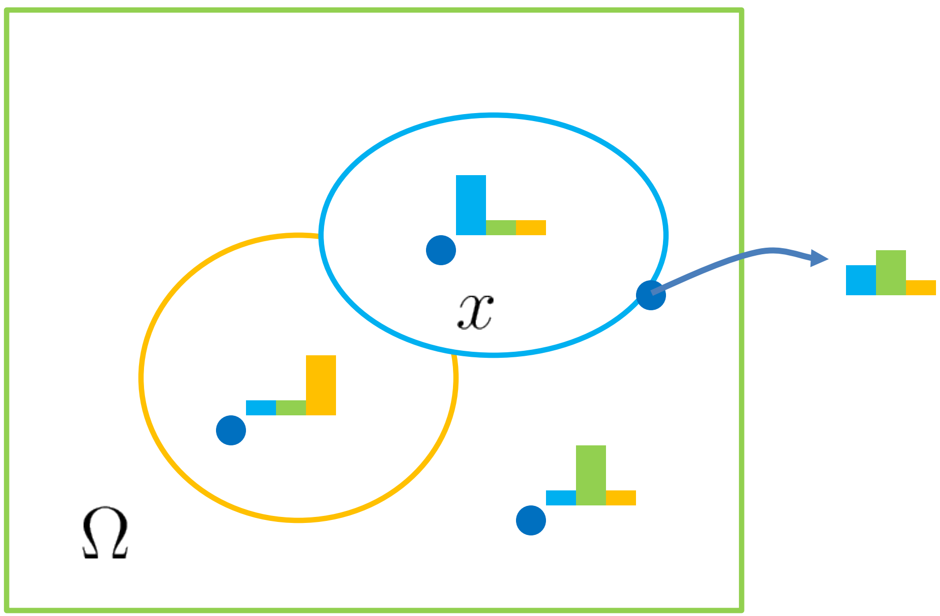
Let’s replace it!

Relaxation



Potts '52; Boykov et al. '98, '01; Kleinberg, Tardos '01
Zach et al. '08; Lellmann, Becker, Schnörr '09; Chambolle, Cremers, Pock '11

Relaxation



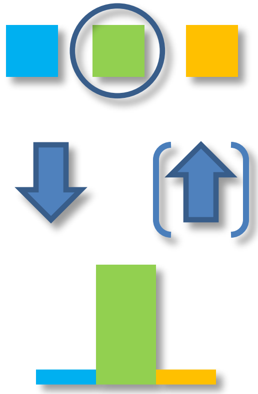
Hard decisions are replaced by soft "probabilities"

Potts '52; Boykov et al. '98, '01; Kleinberg, Tardos '01
Zach et al. '08; Lellmann, Becker, Schnörr '09; Chambolle, Cremers, Pock '11

Relaxation

We would like to extend the problem

$$\min_{u':\Omega\to X} f'(u')$$

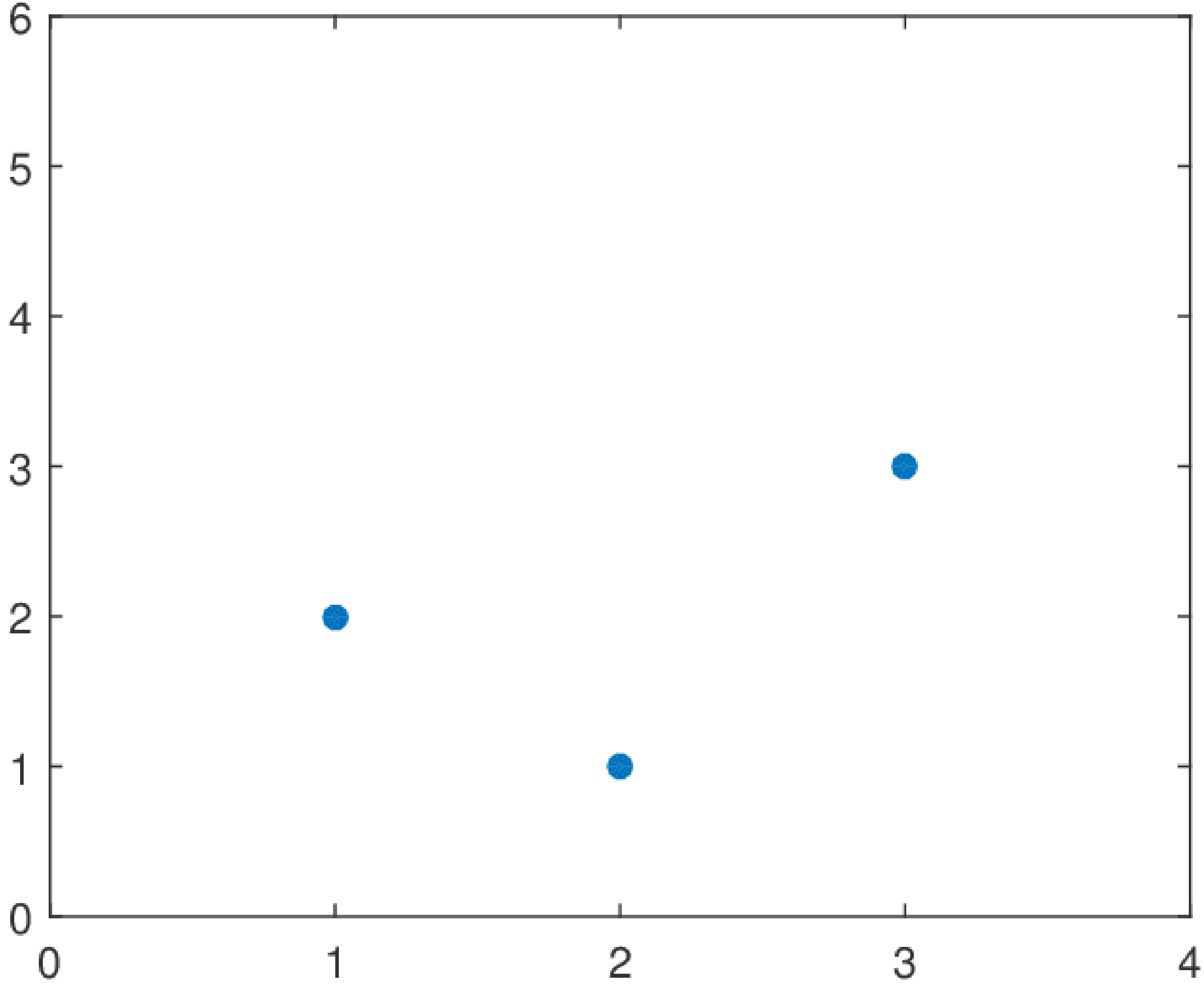


to the probability measures:

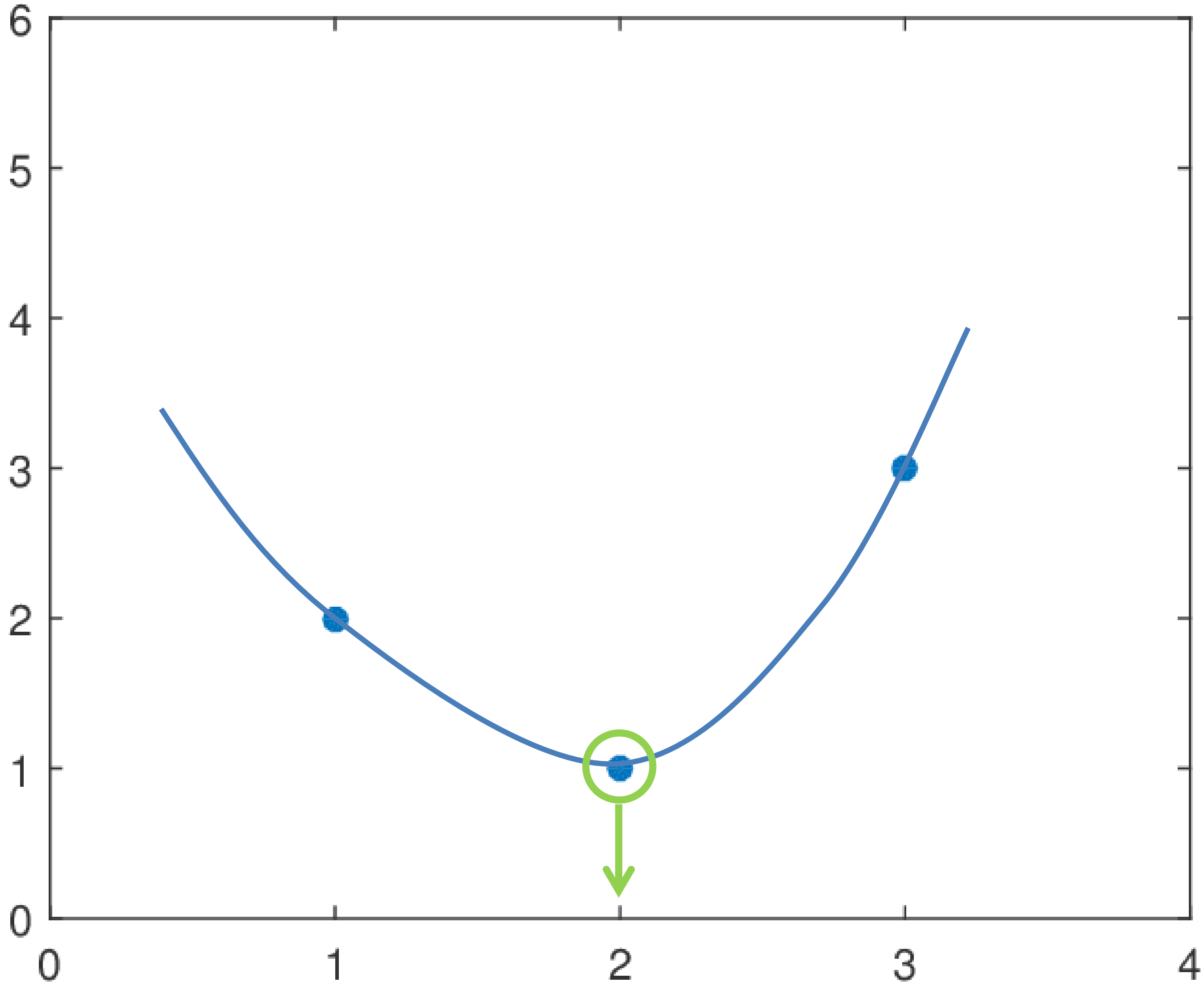
$$\min_{u:\Omega\to\mathcal{P}(X)} f(u)$$

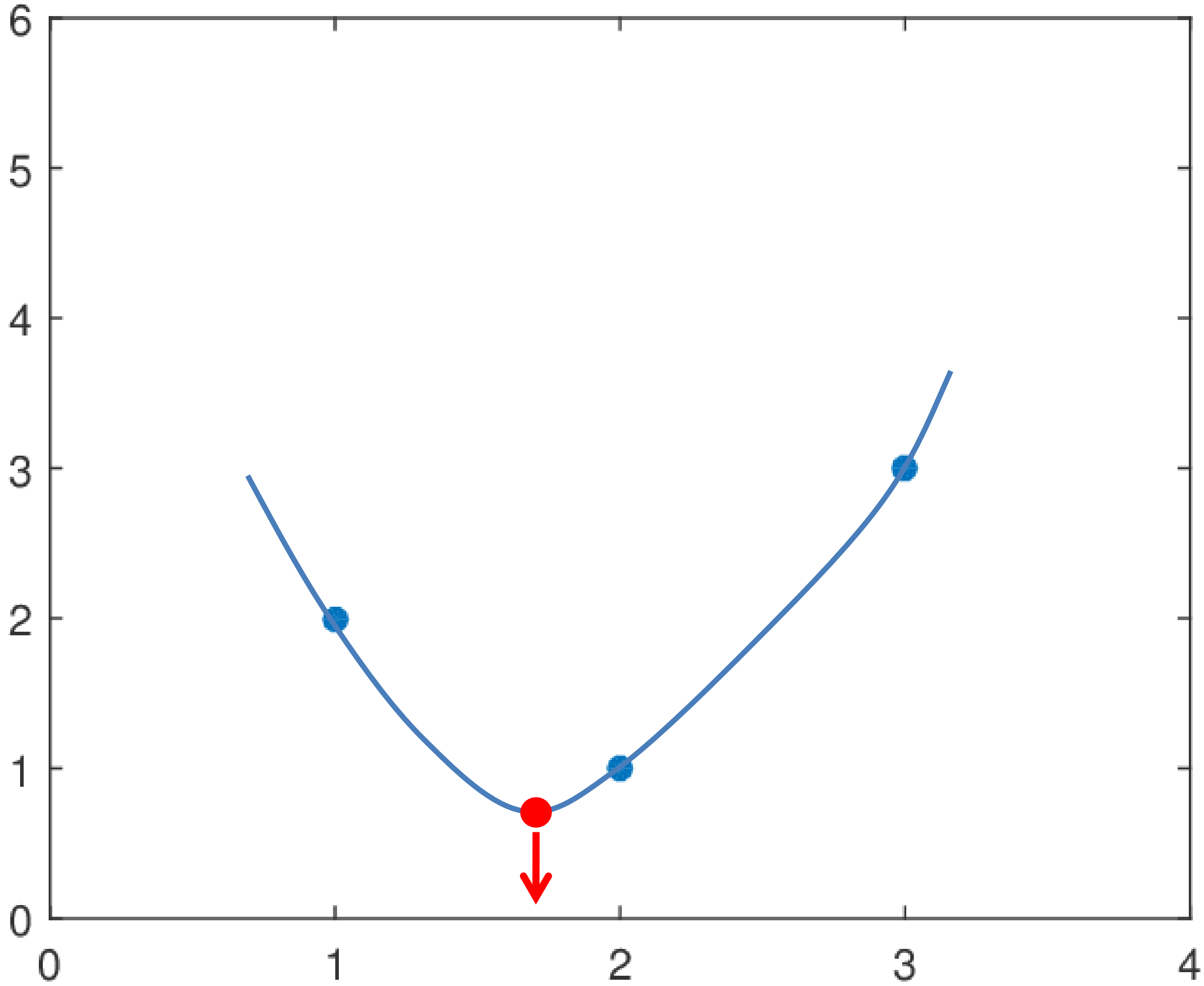
How to define *f*?

- If *u* is integral at every point, *f* should agree with *f′*
- Otherwise, *f* should not create artificial minimizers

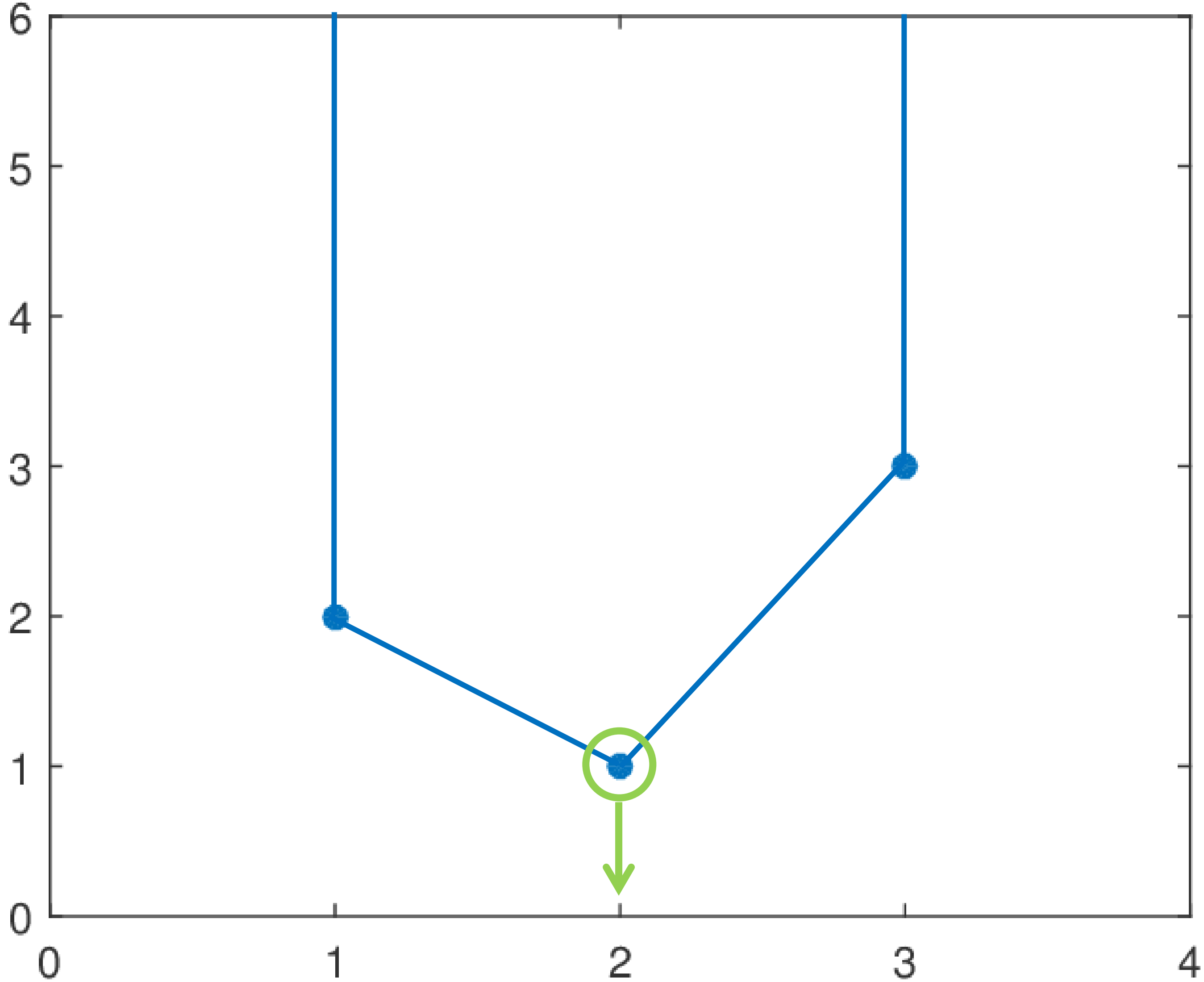


0
1
2
3
4
5
6
0
1
2
3
4

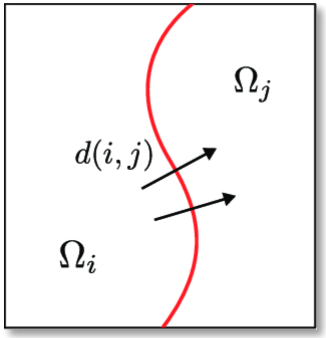

6
5
4
3
2
1
0
0
1
2
3
4



0
1
2
3
4
5
6
0
1
2
3
4


6
5
4
3
2
1
0
0
1
2
3
4

Relaxing segmentation



Assume

- assigning label *i* to point x costs $s_i(x)$.
- boundary between label *i* and *j* costs *d(i,j)*.

Then a possible local relaxation is

$$\min_{u \in \mathrm{BV}(\Omega, \mathcal{P}(X))} f(u) := \int_\Omega \langle u(x), s(x) \rangle dx + \int_\Omega d\Psi(Du)$$

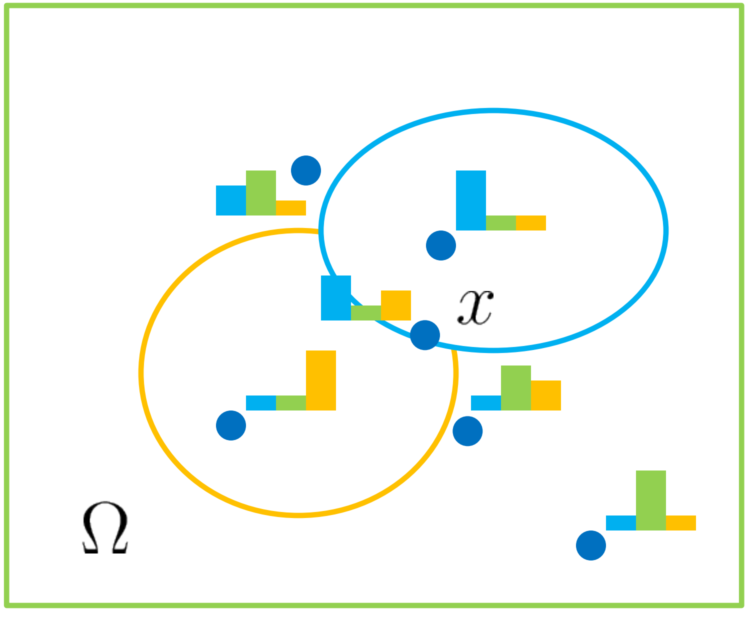
$$\Psi(z) = \sup_{v \in \mathcal{D}} \langle v, z \rangle$$

$$\mathcal{D} = \{(v^1, \ldots, v^L) \in \mathbb{R}^{d \times L} | \|v^i - v^j\| \leqslant d(i,j), \sum_k v^k = 0\}$$

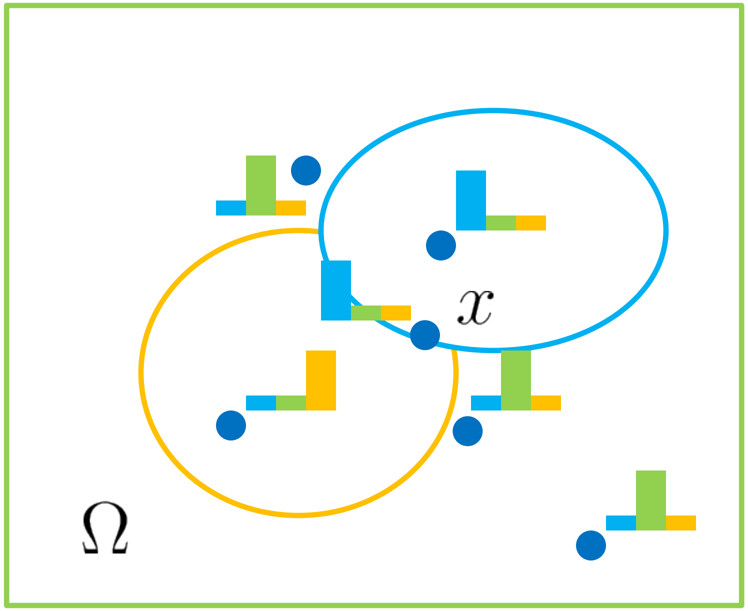
Chambolle, Cremers, Pock '11

Lellmann, Schnörr, SIAM J. Imaging Sci. '10

Is it optimal?



Is it optimal?



Proving optimality

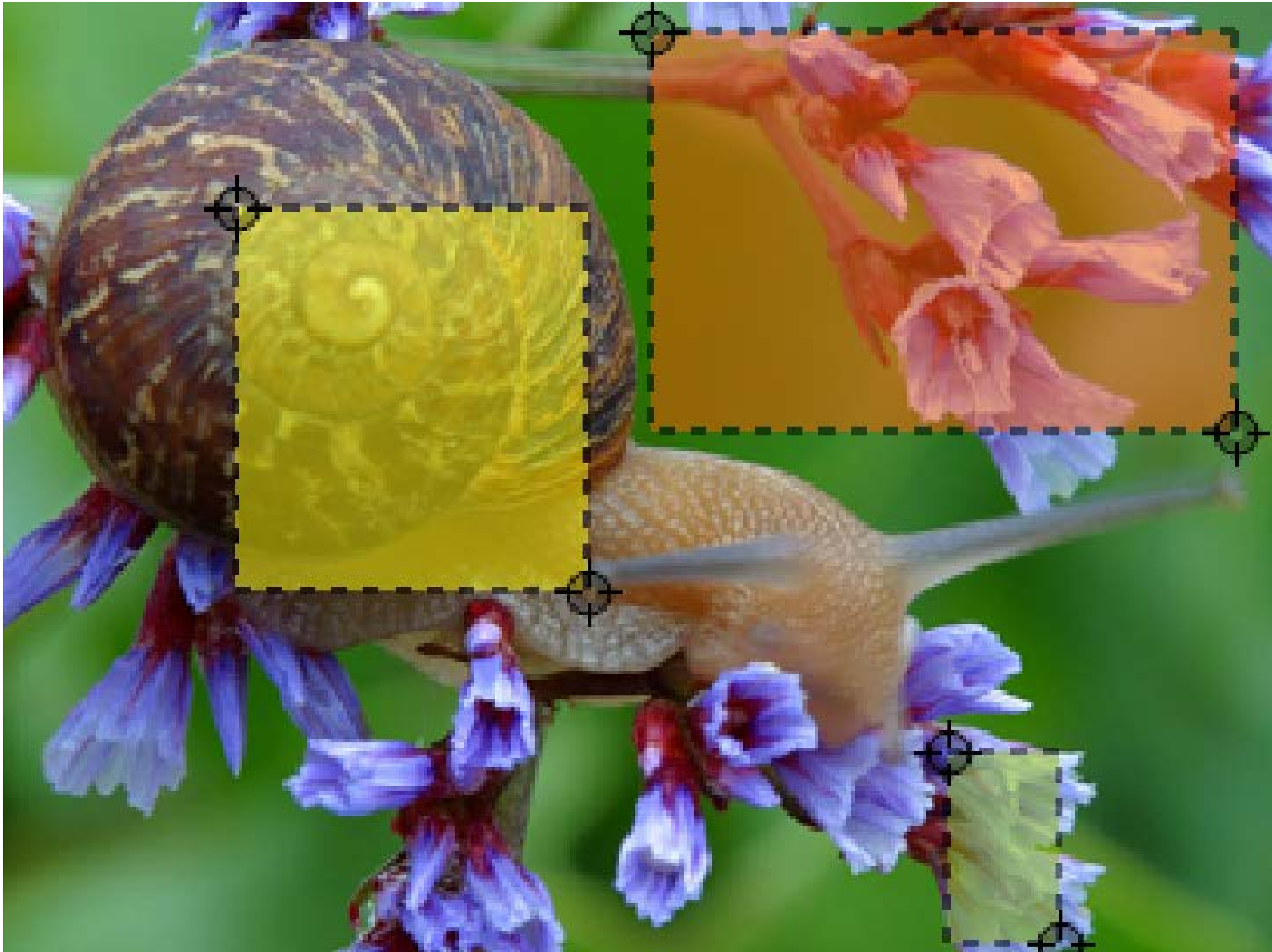
For two classes, recovery is exact:

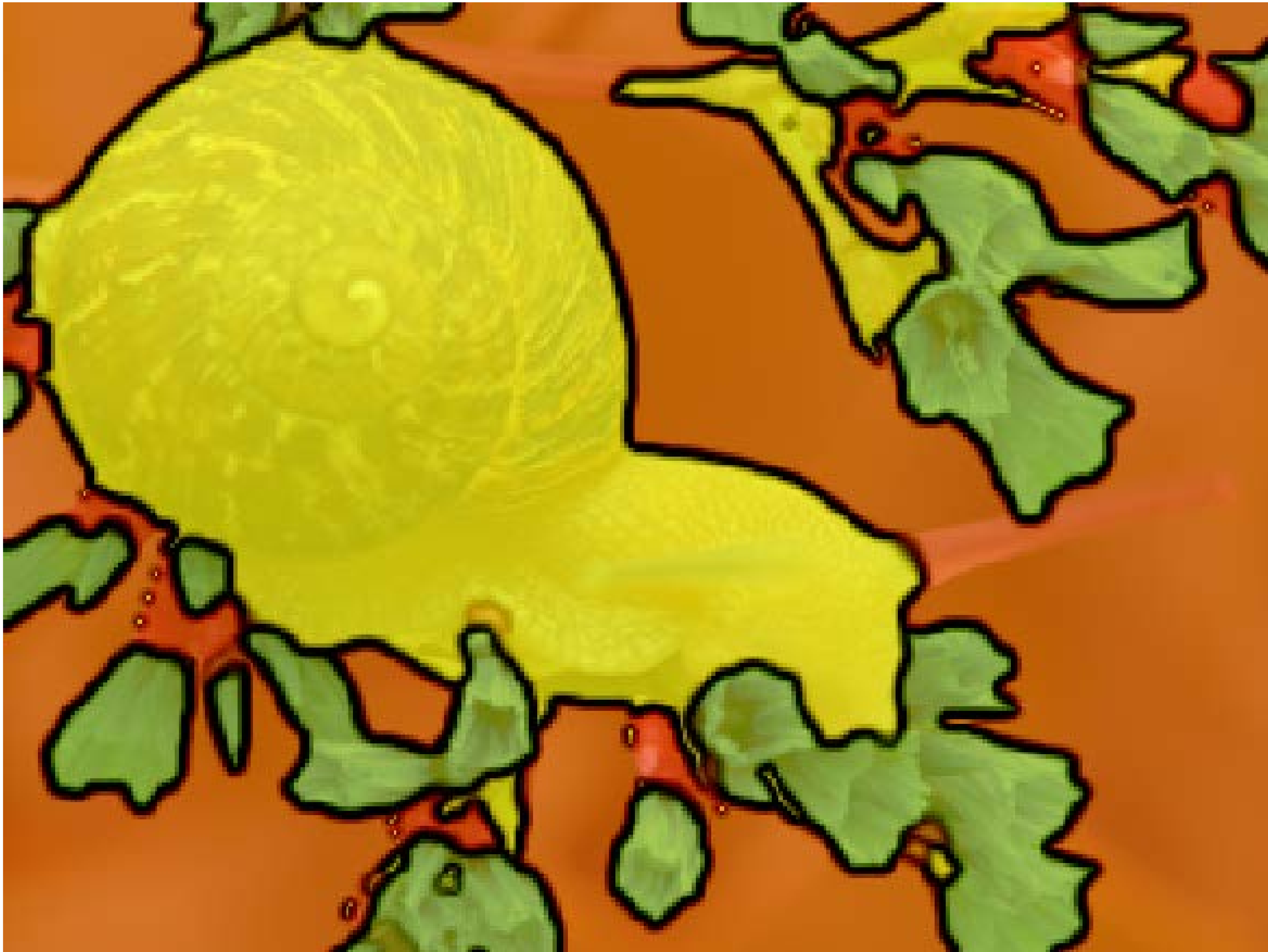
$$f(\mathrm{round}(u^*_{\mathrm{relaxed}})) = f(u^*_{\mathrm{integer}})$$

For $n > 2$ classes, the discrete problem is NP-hard. But:

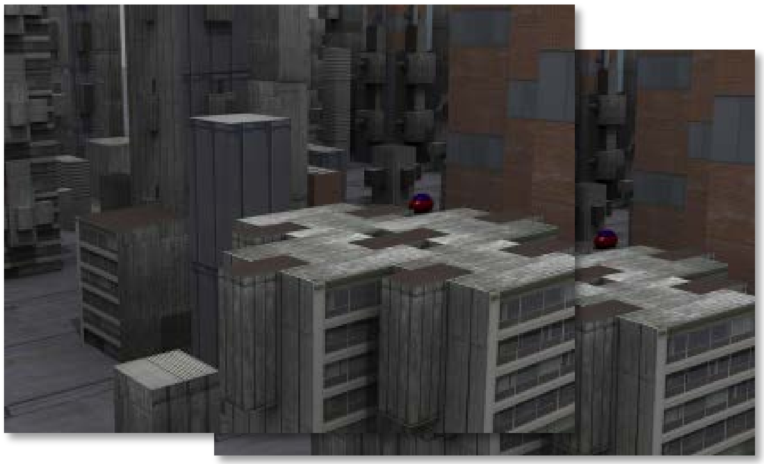
$$\mathbb{E}_\gamma f(\mathrm{round}_\gamma(u^*_{\mathrm{relaxed}})) \leqslant 2\frac{\max_{i\neq j} d(i,j)}{\min_{i\neq j} d(i,j)} f(u^*_{\mathrm{integer}})$$

Two-class: Strang '83; Chan, Esedoglu, Nikolova '06; Zach et al. 09, Olsson et al. 09
Finite-dimensional: Dahlhaus et al. '94, Kleinberg, Tardos '99, Boykov et al. '01, Komodakis Tziritas '07
Multi-class: Lellmann, Lenzen, Schnörr, J. Math. Imag. Vis. 2013

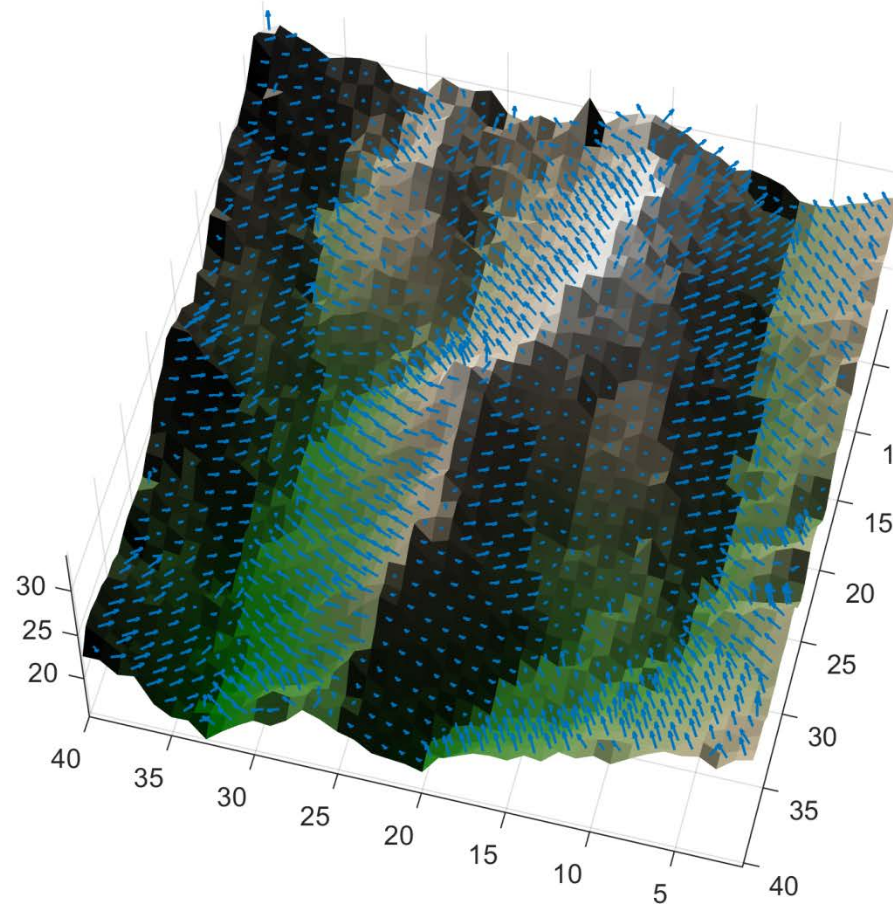




Non-convex functions

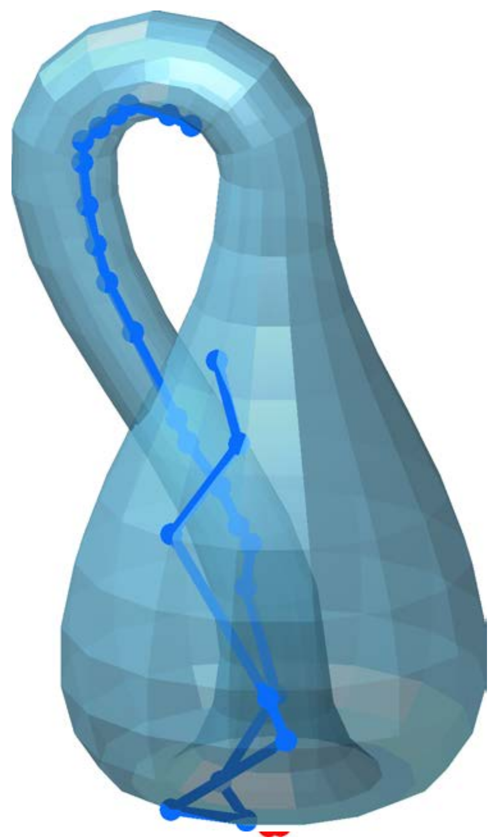


Manifold-valued data

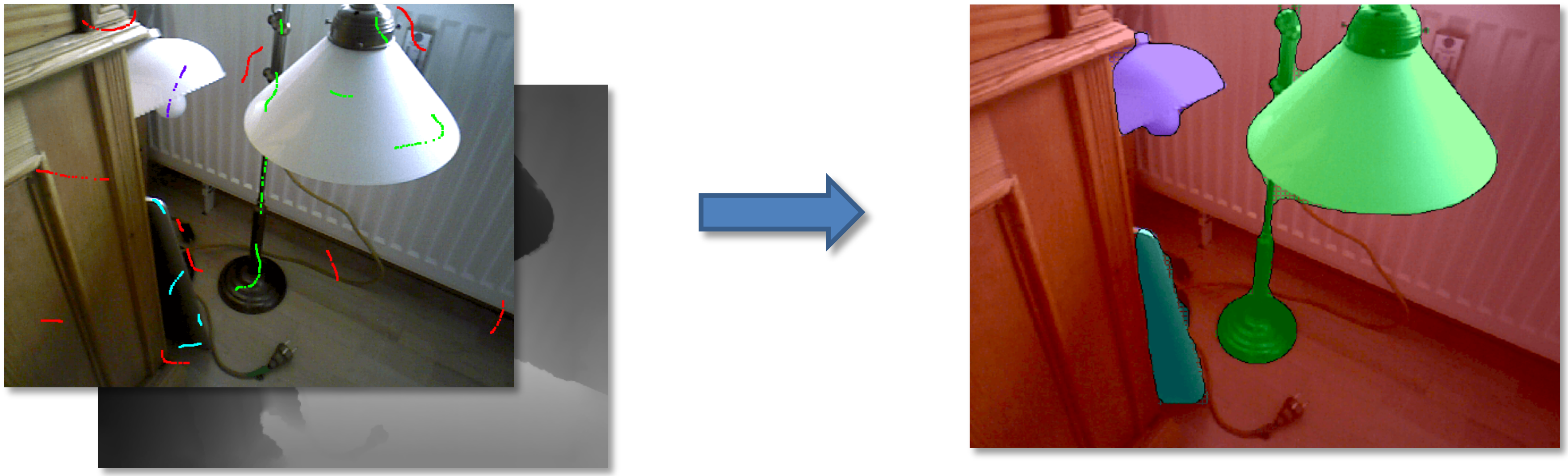


Cremers, Strekalovskiy '12
Lellmann, Strekalovskiy, Kötter, Cremers '13

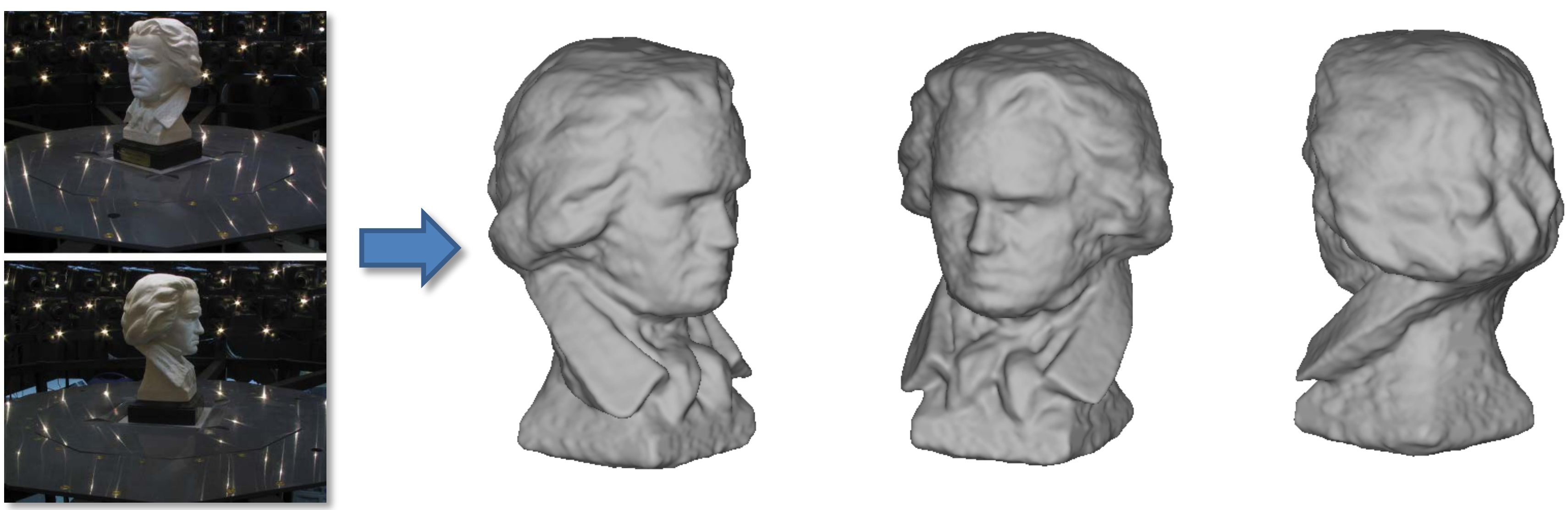
Manifold-valued data



Cremers, Strekalovskiy '12
Lellmann, Strekalovskiy, Kötter, Cremers '13



RGB-Depth Segmentation (Diebold et al. , SSVM '15)

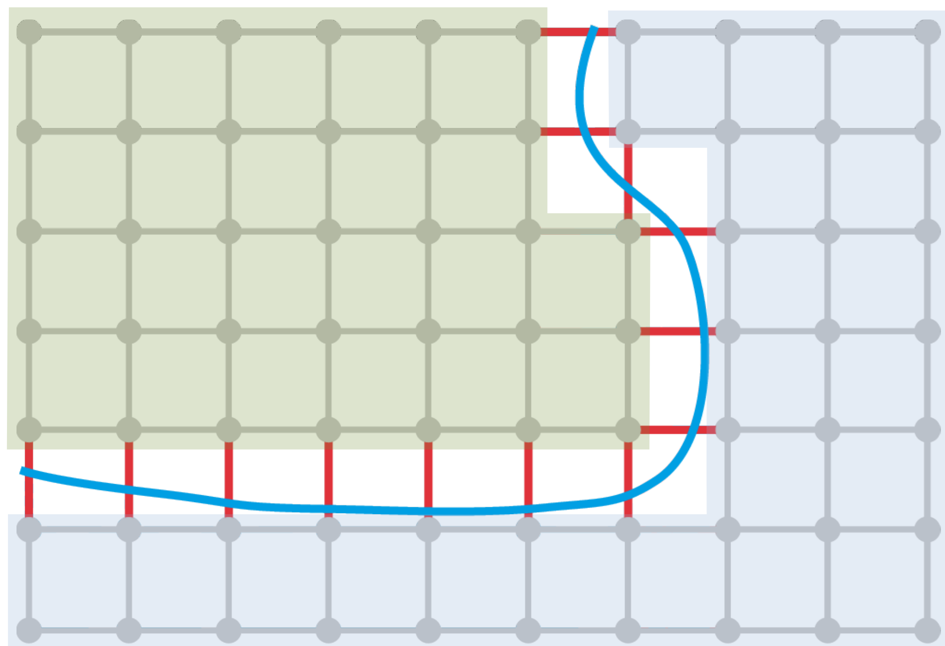


Kolev et al., Int. J. Comp. Vis. '09

Why so complicated?

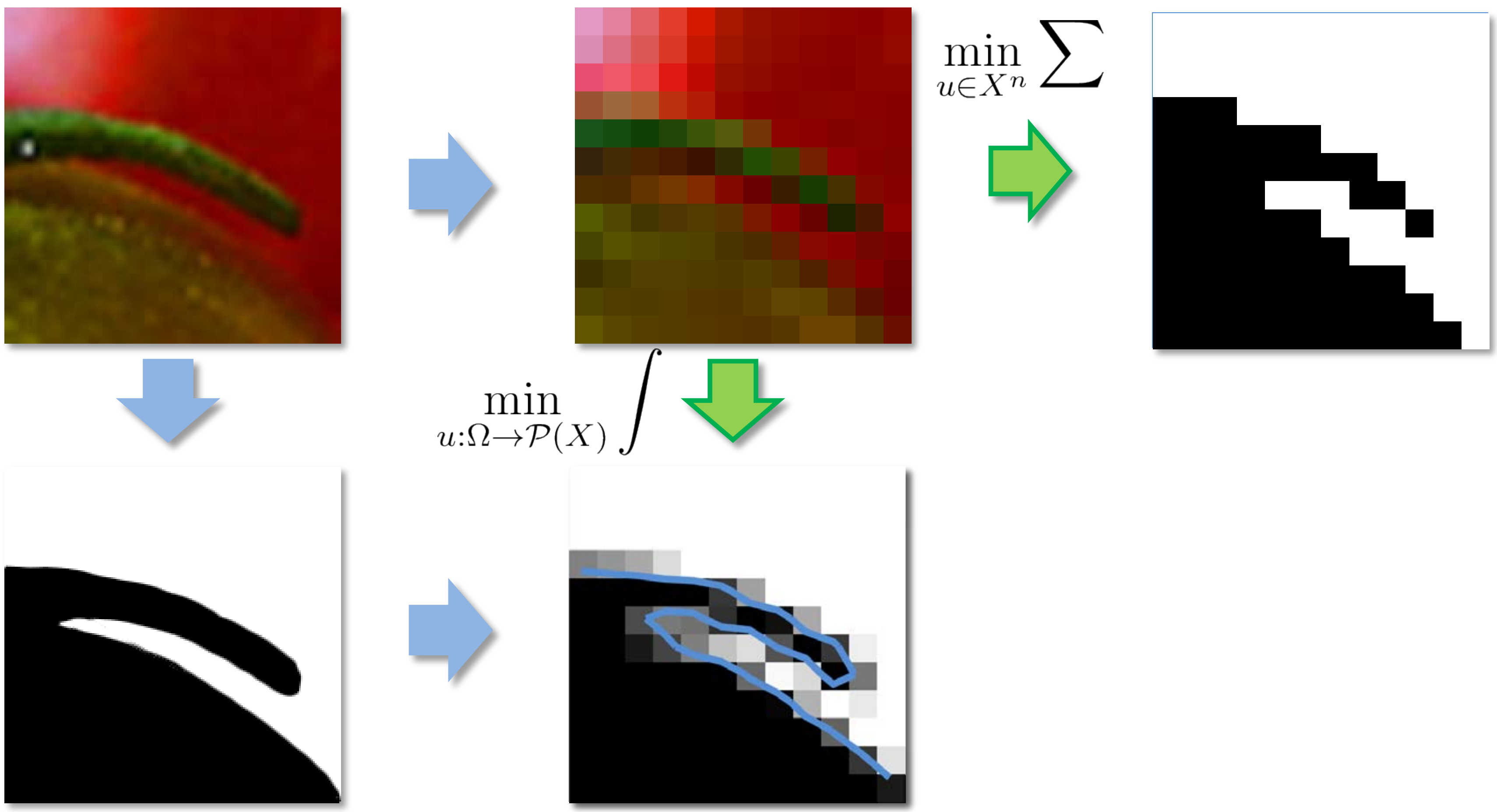
Combinatorial methods

Markov Random Fields, Graphical Models, Graph Partitioning,...



Solve 2-class case using min-cut/max-flow, n-class case using combinatorial solvers: integer program, branch and bound/cut, move making, commercial solvers

A continuous world



Take-home

- Variational methods are intuitive
- (True) non-smoothness is essential
- If it doesn't fit, think big!
- We live in a continuous world (and we actually need the hard math)

J.Lellmann@damtp.cam.ac.uk