Metrics and Regularizations in Image Analysis

Kevin R. Vixie
WSU-Mathematics
Center for Geometric Analysis and Data
Analysis+Data group
and
Blue Space Labs

A course given when I was in the
Theoretical Division, T-7
Los Alamos National Laboratory

July 19-21 2005

Outline

An brief schedule for this tutorial is as follows:

Tuesday Introduction to metrics and regularization: basic concepts and connection to statistical modeling

Wednesday Metrics: examples, data fidelity terms, warping, and face recognition

Thursday Regularization: examples, denoising and total variation based methods, and geometric analysis

Disclaimer: this tutorial is not an unbiased review of metrics and regularization for image analysis. I present my own views and perspectives while at the same time presenting a range of ideas, many of which are important for the work of our team at LANL.

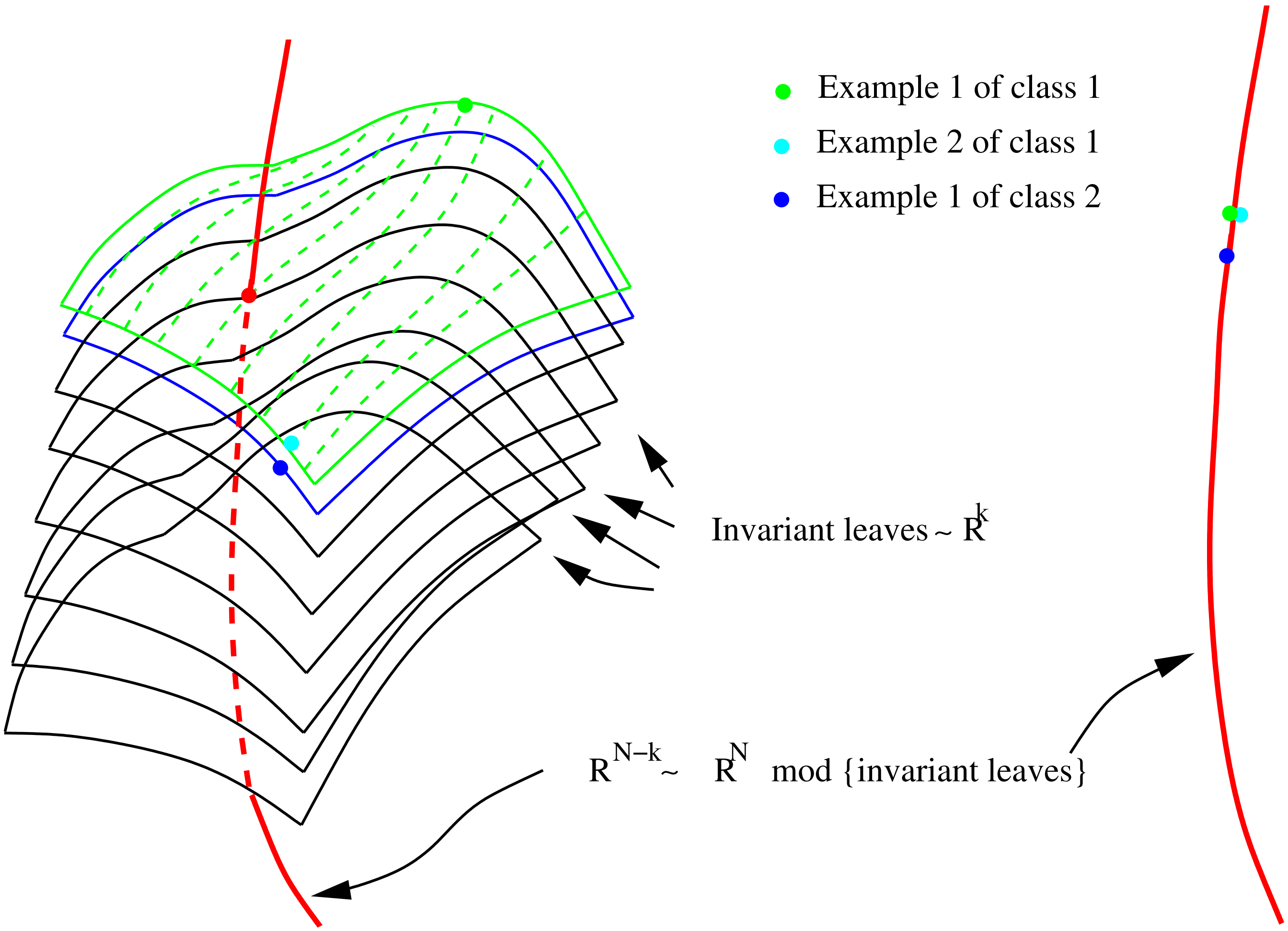
Acknowledgements: Members of the DDMA team at Los Alamos and connected academic departments – UCLA, Duke, Courant, Georgetown, Rice and Clarkson as well as visitors to the DDMA team at Los Alamos.

Information extraction from image comparisons

Find a particular person:

Metrics

Metrics should measure what matters, ignore what doesn't



The "viewgraph" norm or metric

Standard practice: the "viewgraph" norm – how similar are two images to the human eye.

Difficulties with this: Which human? Why should what we notice be important? How do you quantify the degree to which things are different? How much are biases helping or hurting the determination?

Forcing the difference measure to be formulated concretely allows biases to be examined and varied or modified.

Regularization

Image analysis problems are inverse problems $\Rightarrow$ regularization is needed

Inpainting: fill in missing pieces of images.
Denoising: reverse the degradation due to noise.
Segmentation: find the objects and their boundaries in an image.
Detection: find persons, houses, and stars.
Recognition: find a particular person or house or star.

Since it is almost always the case that many "true" images could give rise to the degraded, noisy measured image, we need to pick a solution from the many candidates.

Regularization helps us pick a solution

Example: L^2TV functional for denoising

An example functional we will examine more carefully on Thursday is the ROF or L^2TV functional which is minimized to obtain denoised reconstructions of degraded images:

$$\min_u F(u) = \int |\nabla u| dx + \lambda \int |d - u|^2 dx \tag{1}$$

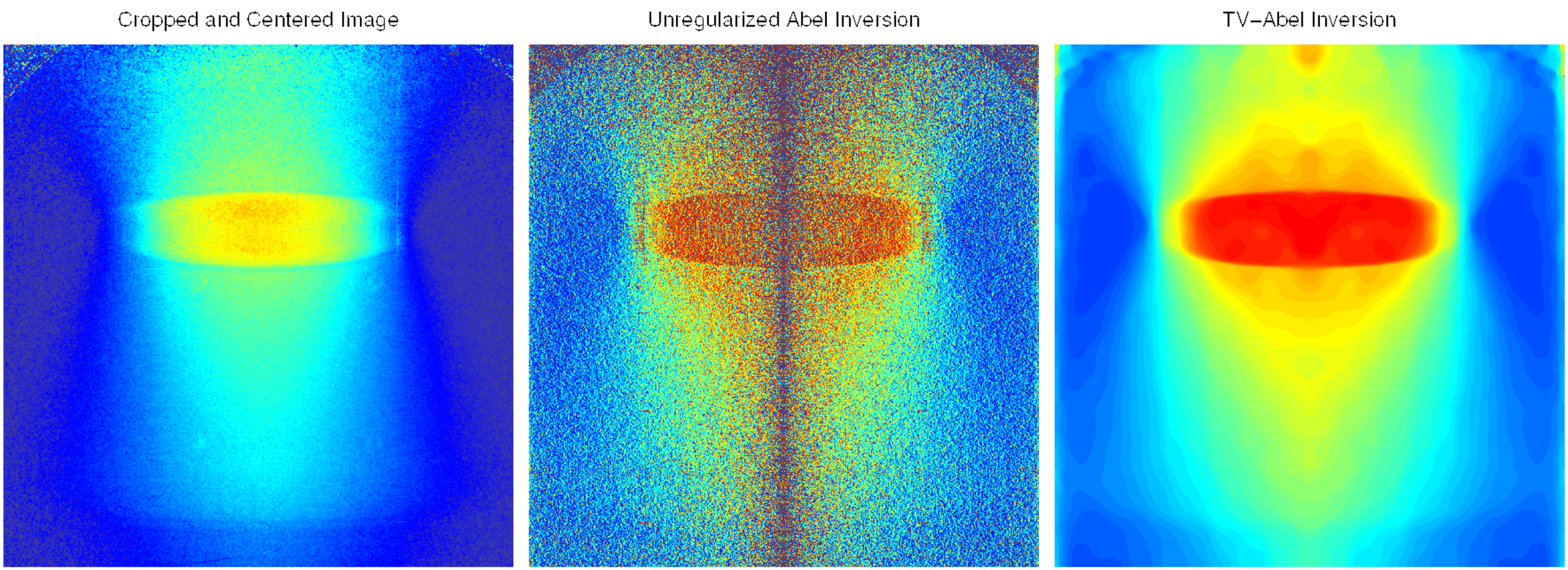
Metric: $\int |d - u|^2 dx$
Regularization: $\int |\nabla u| dx$
What about λ?: Lagrange multiplier, a balance between the regularization and the fidelity of the solution to the measured data. (Has important geometric implications as we will see on Thursday).

Example: L^2TV functional for denoising

Here is an example with a slight twist:

$$\min_u F(u) = \int |\nabla u| dx + \lambda \int |Pu - d|^2 dx \tag{2}$$

where the P is the measurement operator (Abel projection).



Metrics: from a variety of sources

The term metric has at least two uses: one mathematical and rigorous and another use that is more suggestive than precise.

Let's represent images by $u, v, w \in U$.

Precise: A function $\rho(u, v)$ such that

1 $\rho(u, v) \geq 0$ with equality if and only if $u = v$,
2 $\rho(u, v) = \rho(v, u)$, and
3 $\rho(u, v) \leq \rho(u, w) + \rho(w, v)$ for all $u, v, w \in U$

Suggestive: $f(u, v)$ such that (usually) $f(u, v) > 0$ if $u \neq v$ and there is a sense that $f(u, v)$ measures similarity between u and v. (Notice the the ambiguity.) We do not require (or necessarily have) symmetry and/or the triangle inequality.

Metrics: mathematically speaking

Examples of metrics in the mathematical sense:

1 x and y in $\mathbb{R}^n$: $\rho(x,y) = ||x - y||_p = ((x_1 - y_1)^p + (x_2 - y_2)^p + ... + (x_2 - y_2)^p)^{\frac{1}{p}}$. Common choices for p = 1,2 and ∞.

2 f and g functions from $\mathbb{R}^n$ to $\mathbb{R}$: $\left(\int |f-g|^p dx\right)^{\frac{1}{p}}$

3 x and y points on a manifold M: $\rho(x,y) \equiv$ distance of shortest path from x to y that lies on the manifold. To be more precise, we need a Riemannian metric field that we can integrate to get geodesics ($\equiv$ shortest paths).

4 Sobolev norms: example, $||f-g||_{H^1} \equiv \left(||f-g||_2^2 + ||\nabla f - \nabla g||_2^2\right)^{\frac{1}{2}}$

Examples of metrics in the suggestive sense:

1 f,g probability densities: relative entropy, not symmetric, no triangle inequality.

2 f,g functions: TV seminorm $\int |\nabla f - \nabla g| dx$, 0 $\not\to$ equality.

3 $f,g \in F$ functions: choose some $h : F \to \mathbb{R}$. Now consider the "metric" obtained by $|h(f) - h(g)|$. Common scientific data approach.

We will look at metrics for image comparisons in much more detail tomorrow.

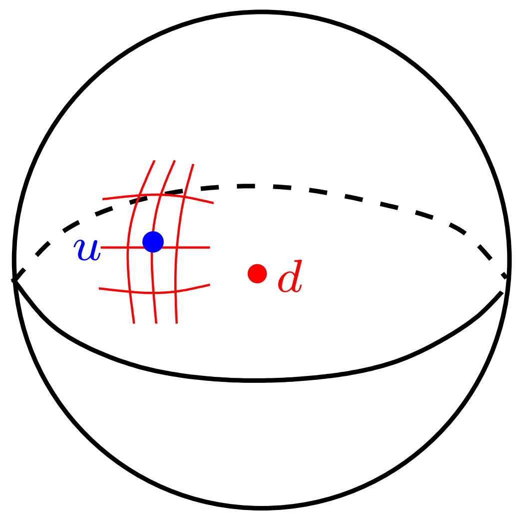
What is regularization?

I will answer using an example:

Suppose a measured image d is equal to some true image u corrupted by noise η, $d = u + \eta$ and η is Gaussian: $\eta \sim C\exp(-\eta^t \cdot \eta / 2\sigma^2)$.

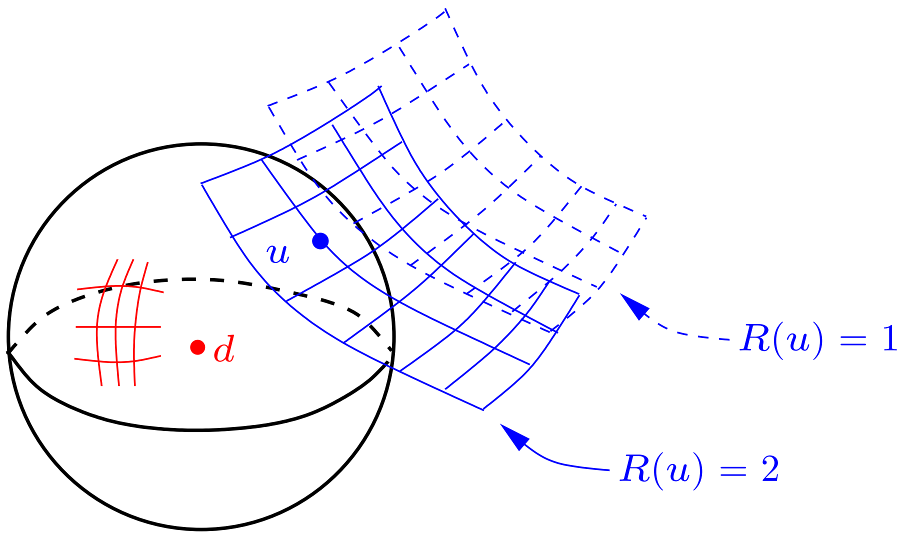
Task: use d and knowledge of the noise level σ to recover u.

$\boldsymbol{Any}$ point in U on $S(d, \sigma)$, the sphere centered at d of radius σ, is a perfectly good "reconstruction".



What is regularization?

But we always know more: the solution is smooth or continuous or has no high frequency components or is piecewise polynomial, etc. Suppose we can quantify this: the smaller $R(u)$ is, the more *regular* u is. Level sets of R (frequently) pick a unique *best* u.



We invert by balancing *fidelity* to the data, measured by $\rho(u, d)$, against solution *regularity*, measured by a *regularization term* $R(u)$.

$\min_u F(u) \equiv R(u) + \lambda\rho(u, d) =$ (for example) $\int |\nabla u| dx + \lambda \int |u - d|^2 dx$.

What is regularization?

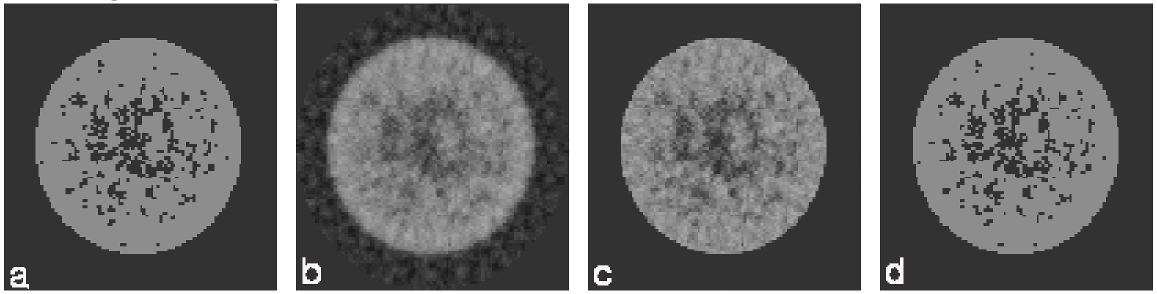
Another example: regularized tomographic inversions. Linear subspaces of solutions match the measured data.

$$\min_u F(u) = \int |\nabla u| dx + \lambda \int |Pu - d|^2 dx \tag{3}$$

Regularization chooses $u \in N + u^*$, where $N \equiv \{u : Pu = 0\}$ is the subspace of *null* solutions and u^* is any solution to $Pu = d$.

Above we regularize be minimizing some energy $R(u)$. We can also regularize by restriction to some lower dimensional subset or submanifold

Example: choosing images restricted to the subset of binary images.



(More details in Asaki's talk later.)

Regularization: a look ahead

We have seen the L^2TV or ROF functional above, an older one seems very similar at first sight:

$$\min_u F(u) = \int |\nabla u|^2 dx + \lambda \int |d - u|^2 dx \tag{4}$$

In this case $|\nabla u|^2$ replaces $|\nabla u|^1$ in the regularization term – and we lose edges: see Asaki's talk latter this afternoon.

Regularization and Metrics: probabilistic connection

The last example suggests a connection between priors and regularization.

$$
\begin{aligned}
p(u|f) &\sim p(f|u)p(u) \\
\arg\max_u p(u|f) &= \arg\max_u p(f|u)p(u) \\
\arg\min_u \{-\log(p(u|f))\} &= \arg\min_u \{-\log(p(f|u)) - \log(p(u))\}
\end{aligned}
$$

Supposing that (for example) $p(f|u) \sim e^{-\int \lambda |u-f|^2 dx}$ and $p(u) \sim e^{-\int |\nabla u| dx}$, we get:

$$
\arg\min_u \left\{ \int |\nabla u| dx + \lambda \int |u-f|^2 dx \right\}.
$$

More generally, with $p(f|u) \sim e^{-\lambda \rho(u,f)}$ and $p(u) \sim e^{-E(u,\nabla u)}$, we get:

$$
\arg\min_u \{E(u, \nabla u) + \lambda \rho(u,f)\}.
$$

Regularization and Metrics: probabilistic connection

Regularization can be viewed as

Method of stabilizing inverse problems: simplest point of view. But we understand why this is reasonable by seeing regularization as

Enforcement of prior: choosing the solution that best fits what we know about the solution;

Dimension reduction: reducing the dimension of the object we are trying to infer or estimate from the data;

It should now be clear that while regularization can be viewed as a stabilization procedure for inverse problems, it is justified by the fact that this is how we can insert prior information into the solution.

Mathematical background

For the rest of this talk, I will introduce mathematical concepts important to the topics of this tutorial.

Some of these concepts will be used only rather lightly in the tutorial, but are very useful for understanding the topics we will discuss and will, for example, be assumed of readers of research papers in this area.

Perspective: I am (unapologetically) highly biased towards things geometric and geometric-analytic, so I try to "see" what motivates and leads us towards solutions and understanding. I will attempt now to pass to you a few of these intuitive tools.

References: I recommend Evans' PDE text, Evans' and Gariepy's monograph, Mattila's book on sets and measures, Folland's graduate analysis text as well as Ekeland and Temam, Zeidler's books, and Dacorogna's new text "Introduction to the Calculus of Variations".

Properties of minimizers

So: you concoct a functional the incorporates measured data and a prior model and end up with $\min_u F(u) \equiv R(u) + \rho(d,u)$: what now?

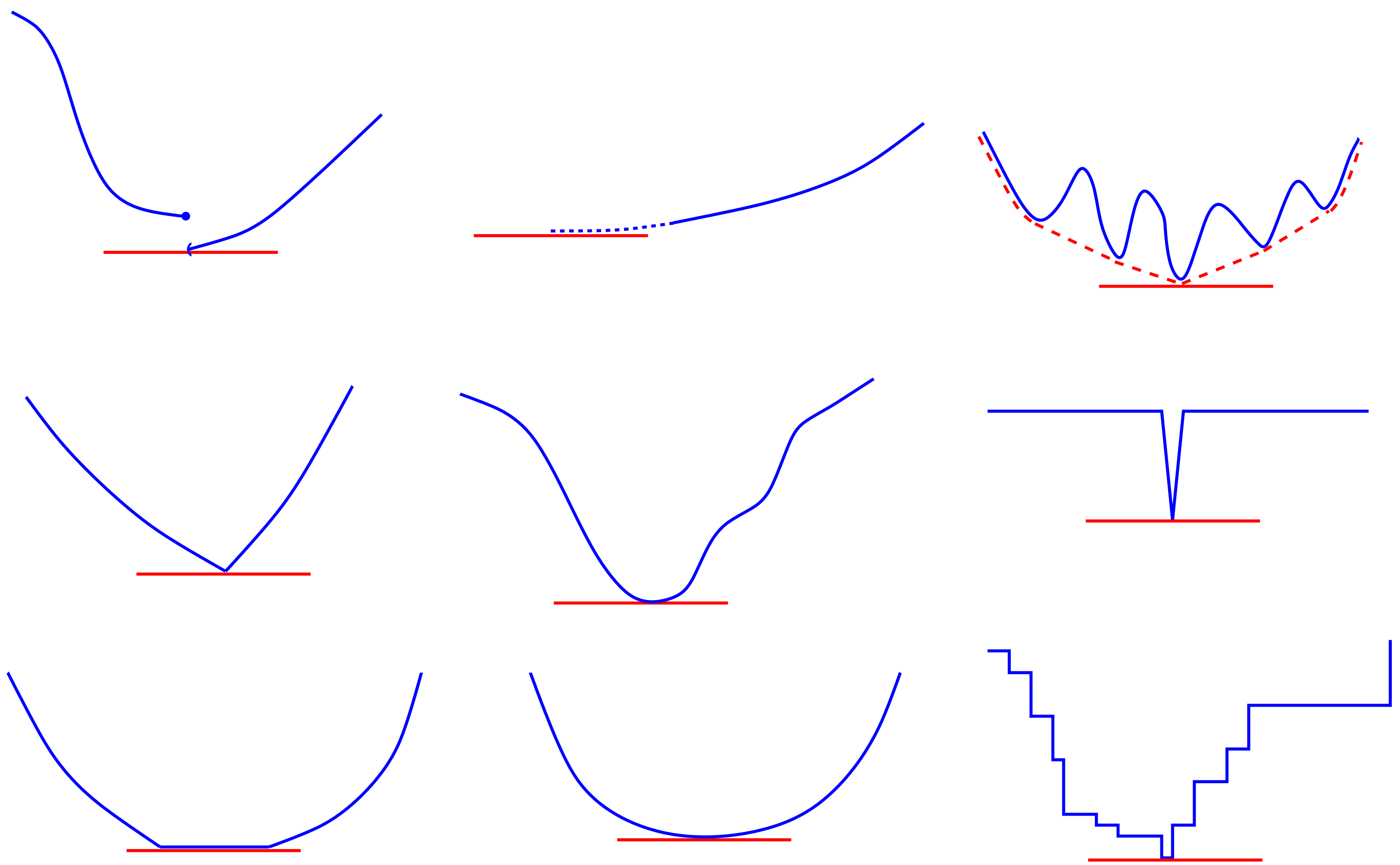
Mathematicians can't help themselves: they must ask (no matter how silly this seems to some physicists)

- E "Is there a minimizer" (existence)
- U "How many minimizers" (uniqueness)
- R "How nice is the minimizer(s)" (regularity)

and if that mathematician has an "applied" bent,

- C "Can I construct a convergent algorithm to find the minimizer" (computation)
- S "How stable is my solution to perturbations of the data?" (stability*)
- G "Can I characterize exact solution properties in terms of useful solution properties?" (geometry)

Looking for minimizers: one-dimensional examples



Looking for minimizers: existence

The direct method of the calculus of variations:

Bounded Below: f is bounded below; $f^* \equiv \inf_u f(u)$
f Coercive: $L(a_0) \equiv \{u | f(u) \leq a_0\}$ bounded for some $a_0 > f^*$
f lower semicontinuous: $L(a) \equiv \{u | f(u) \leq a\}$ is closed for all $a_0 \geq a > f^*$.
f has a minimizer: here is a proof.

* $L(a_0) \supset L(a_1) \supset L(a_2) \supset \ldots$ is a nested sequence of compact sets.
* $L(f^*) = \bigcap_i L(a_i) \neq \emptyset$.
* QED

f has a minimizer: another (equivalent proof).

* minimizing sequence exists; $f(u_i) \to f^*$
* the level set $L(a_0)$ is compact
* $\Rightarrow u^* \equiv \lim_k u_{i_k}$; a subsequence converges.
* $f(u^*) \leq \liminf_k f(u_{i_k}) = f^*$; from lower semicontinuity
* $f(u^*) = f^* \Rightarrow u^*$ is a minimizer

Mathematical background

Caution: important issue glossed over – compactness is non-trivial infinite dimensional spaces (e.g. function spaces).

Finite Dimensions: compact = closed and bounded.

Infinite dimensions: closed unit ball not compact.

Weak topologies: Changing to weak topologies can give us compactness ... interesting details here.

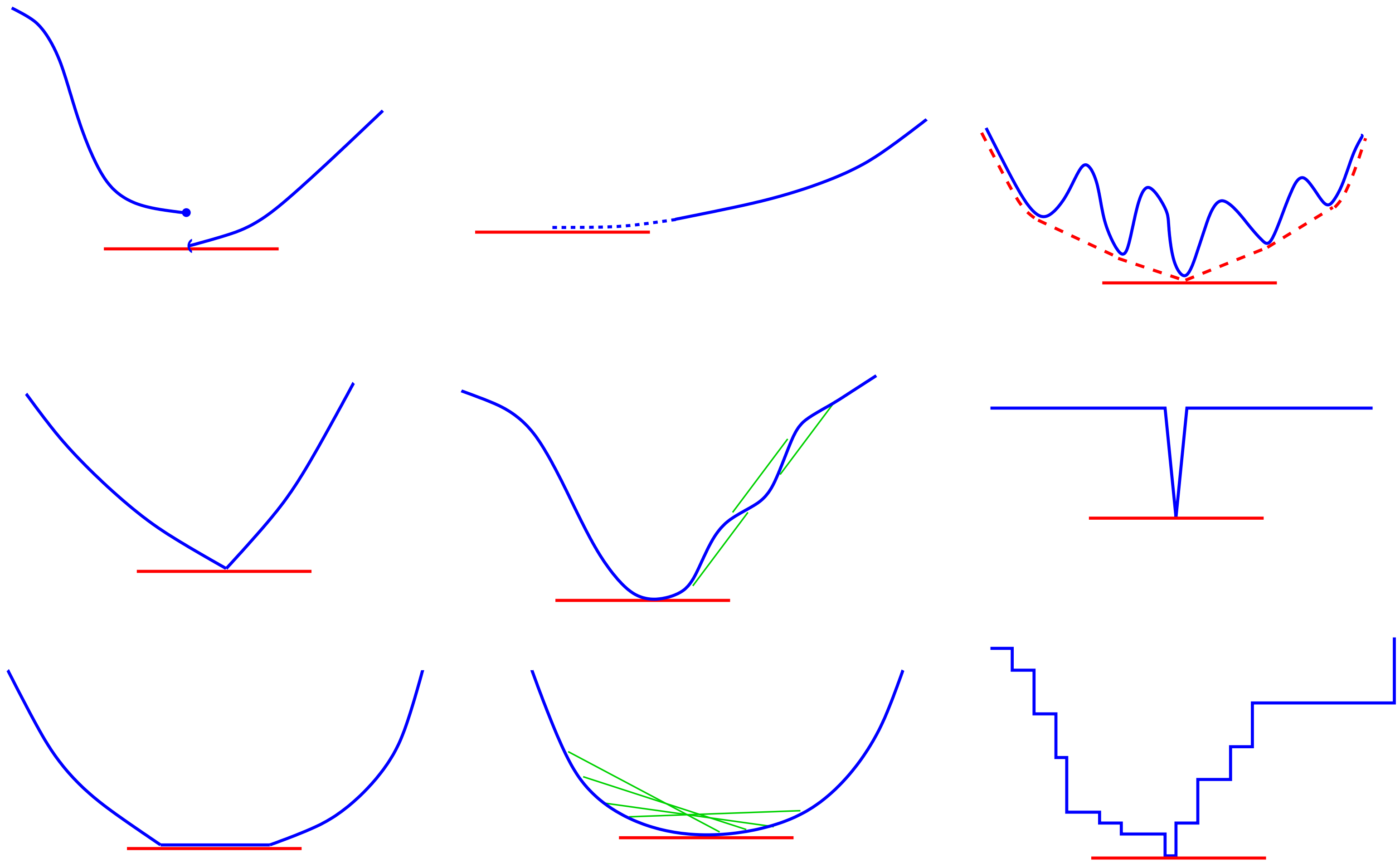
Fix: Typically, the existence proofs above will use weak convergence and compactness.

Weak convergence: define $F_g(f) \equiv \int fg dx$

$$f_i \rightharpoonup f^* \text{ if } F_g(f_i) \rightarrow F_g(f^*) \tag{5}$$

for all g in some set or space (details here!).

Looking for minimizers: convexity and uniqueness



How nice is the minimizer? characterization and image representation

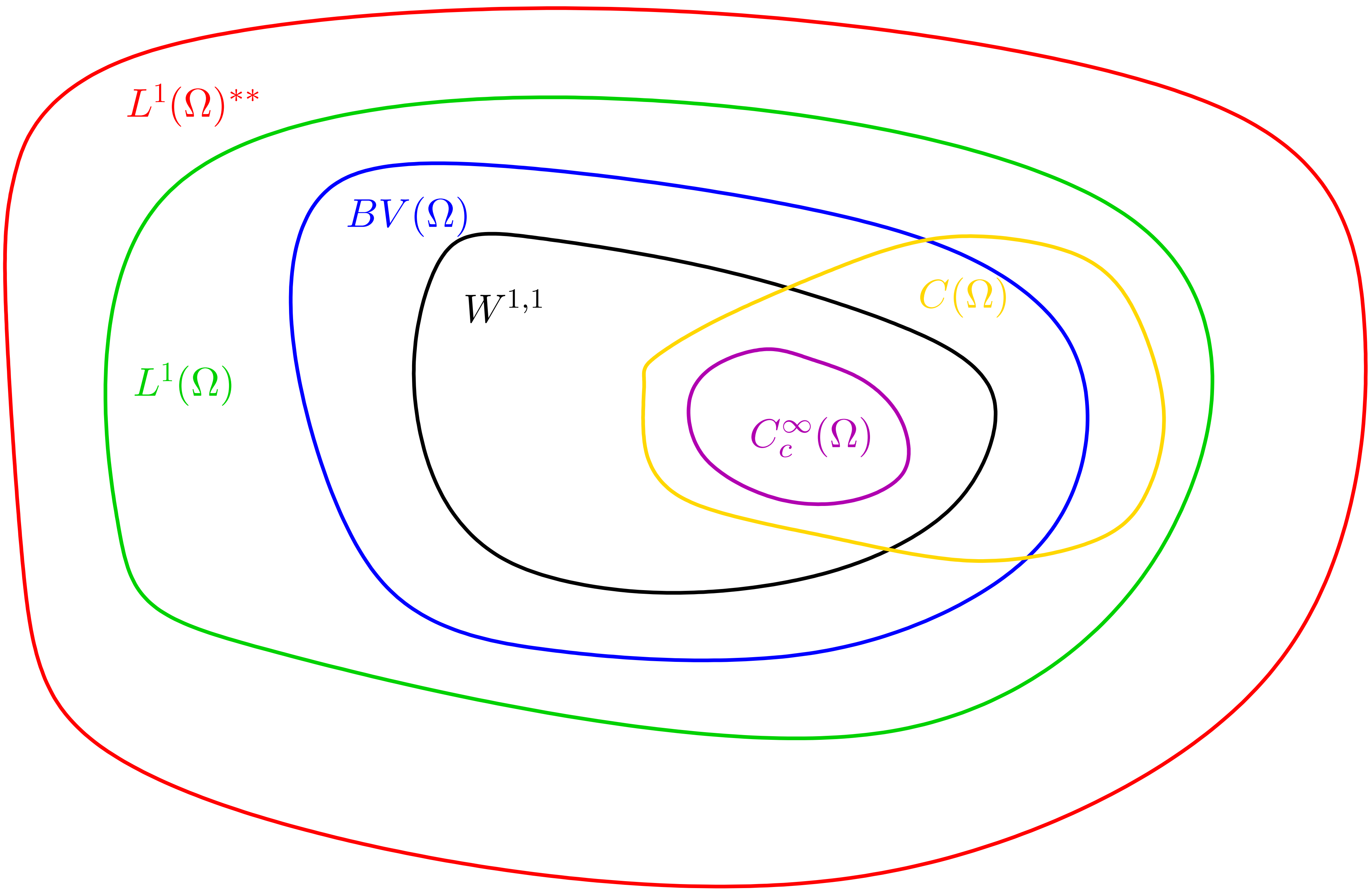
Note: not unreasonably, how nice a solution is usually referred to as solution *regularity*. This is related to but not completely identical with the *regularization* that we have discussed today and is treated at length on Thursday.

Nice makes sense only for high dimensional u; for u that are functions or discretized functions.

* How smooth is the minimizer?
* Do the level sets have bounded curvature?
* Is the minimizer really just a sensibly denoised version of the original data?
* Is the minimizer equal to the true image?
* What artifacts has the regularization introduced into the minimizer?
* Can we find or construct exact, non-trivial data-minimizer pairs?

In short, we want to know how nice minimizers are and be able to characterize them as closely as possible. Now a picture of function spaces ...

Mathematical background



Mathematical background

Flowing towards minimizers: derivatives and Euler-Lagrange equations

A first approach to finding minimizers is often essentially undergraduate calculus: we descend the gradient.

$$u^t = -DF(u). \tag{6}$$

To do this we need to understand DF(u) when $u \in U$, an infinite dimensional space. (There are details we gloss over: e.g. when is $DF(u) \in U$?)

So, what does $DF(u)$ mean? Answer: exactly what you expect!

* Derivatives are simply local linear approximations.
* $F(u+\delta) = F(u) + DF(u)(\delta) + o(|\delta|)$.
* Next page: we calculate DF for $F(u) = \int |\nabla u|^2 dx + \lambda \int |d-u|^2 dx$

Note: Convexity (when we have it) means the gradient is reliable (though possibly very slow) and leads to a global minimizer.

Mathematical background

Calculate $DF(u)\delta \approx F(u+\delta) - F(u)$ for $F(u) = \int |\nabla u|^2 + \lambda \int |u-d|^2$:

$$F(u+\delta) = \int \nabla u \cdot \nabla u + 2\nabla u \cdot \nabla \delta + \nabla \delta \cdot \nabla \delta + \lambda \int u^2 + d^2 + \delta^2 - 2ud + 2u\delta - 2\delta d$$

$$F(u) = \int \nabla u \cdot \nabla u + \lambda \int u^2 + d^2 - 2ud$$

subtracting:

$$F(u+\delta) - F(u) = \int 2\nabla u \cdot \nabla \delta + \nabla \delta \cdot \nabla \delta + \lambda \int 2u\delta - 2d\delta + \delta^2$$

Ignoring second order terms (we assume that δ and $\nabla \delta$ are "very small") we get:

$$F(u+\delta) - F(u) = \int 2\nabla u \cdot \nabla \delta + \lambda \int 2u\delta - 2d\delta$$

Integrating the first term by parts and combining we get:

$$F(u+\delta) - F(u) = \int 2\left(-\nabla \cdot \nabla u + \lambda 2(u-d)\right)\delta.$$

SO: $DF(u)\delta = \int 2\left(-\Delta u + \lambda 2(u-d)\right)\delta = < 2\left(-\Delta u + \lambda 2(u-d)\right), \delta >$

Mathematical background

Representing Images: functions, sets and measures

Radon Measures: are nice measures, include well behaved singular measures. Examples: point masses (Dirac delta "functions") and other measures concentrated on lower dimensional subsets of the plane.

BV functions: a natural choice for image representation

* $f \in BV(\Omega)$ if $f \in L^1(\Omega)$ and $\int \nabla f \cdot \vec{\phi} dx = \int \vec{\phi} \cdot d\nu$ where ν is a vector valued Radon measure with finite mass.
* $f \in BV(\Omega)$ may have discontinuities!
* we can do analysis with BV functions.

Multiscale representations: Examples abound from previous lectures. I will briefly mention another tomorrow.

Currents and varifolds: currents (think of them as generalized manifolds) have the advantage of allowing wild behavior to be handled with "ease" analytically. Their disadvantage is that this branch of geometric measure theory takes some concentration to master. (example: Federer's tome.) There are close connections to BV functions.

Mathematical background

Level sets: can be used to do very nice things with sets of codimension k where k is a very small integer like 1. Huge amount of work here. Very interesting from every angle (theoretical, computational, usefulness for applications).

Radon Measures:

Define: $\mu_f(A) \equiv \int_A f dx$ for any positive f in L^1. Then: $f dx = d\mu_f$
What if we generalize allowable f's? Answer: Radon measures happen. (catch: we only know how to generalize μ)

BV functions:

Typical gradient fields: $d\vec{F}(x) = \vec{\sigma}(x) f(x)$, $\vec{\sigma}(x)$ a unit vector field, $f(x)$ the magnitude $|d\vec{F}(x)|$.

Define the gradient measure $\mu_{dF}(A) \equiv \int_A \vec{\sigma}(x) f(x) dx$.

A BV function is one whose gradient measure can be written $\int_A \vec{\sigma}(x) d\mu$ where μ is a Radon measure.

Mathematical background

Geometric structure: Derivatives are local linear approximations

$F : u \in \mathbb{R}^n \to \mathbb{R}$: zoom into a function whose derivative is not identically 0. It looks like a tipped plane.

$F : u \in \mathbb{R}^n \to \mathbb{R}^m$: zoom into a function with full rank derivative. The derivative tells us everything.

Technicalities for infinite dimensions: Even linear maps become nontrivial(!) in infinite dimensional spaces, but you can get a very long ways by leveraging a solid understanding of the case $F : \mathbb{R}^n \to \mathbb{R}^m$ and the corresponding DF.

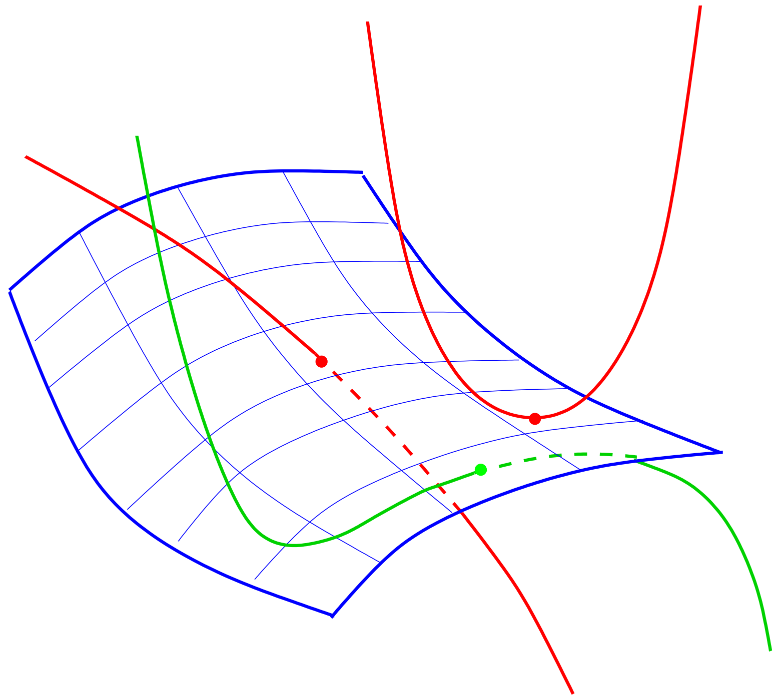
What can local linear approximations tell us?

Inverse function theorem: if DF is invertible, then locally, so is F.

Implicit function theorem: if DF has a null space of dimension k, then F has level sets of dimension k.

Transverse intersections: $\dim(A \cap B) = \dim A + \dim B - \dim W, A\&B \subset W$

Mathematical background



What do we need higher order approximations for? Answer: understanding singularities.

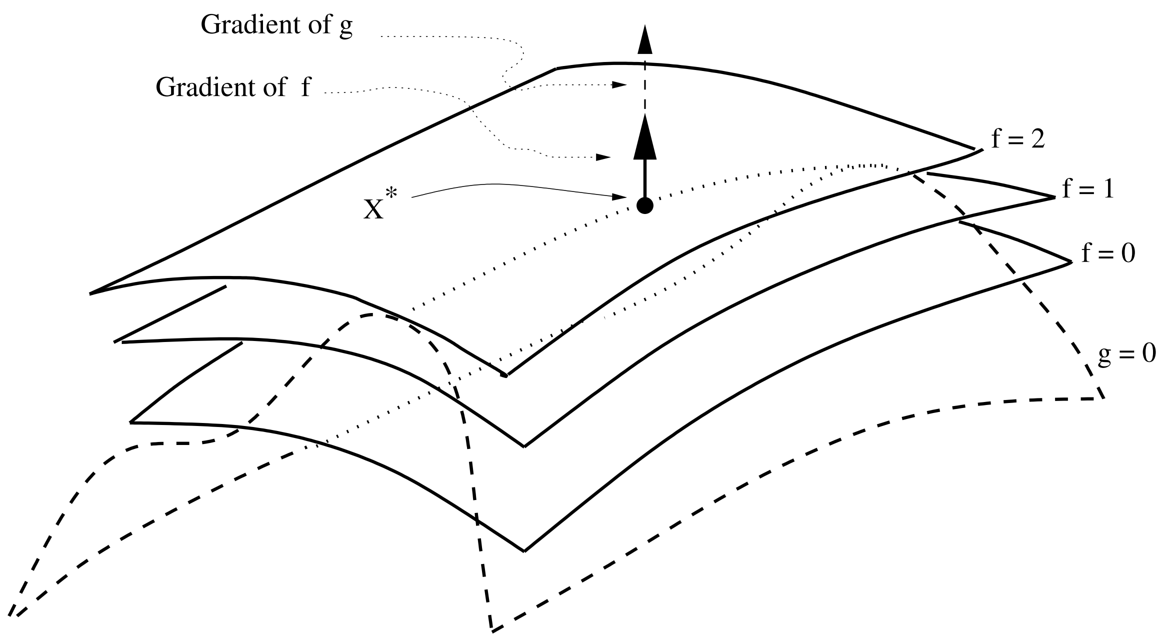
Example: second order variations at singularities of the derivative can tell us a great deal. We will see an example Thursday in the derivations of non-trivial, exact minimizers.

The Lagrange Multiplier Picture

Problem Maximize $f(x)$ subject to $g(x) = c$.

Multiplier method Find stationary points of $L(x, \lambda) \equiv f(x) + \lambda g(x)$

Do ... solve $DL(x, \lambda) = Df(x) + \lambda Dg(x) = 0$



Tomorrow: Metrics

1. Brief review
2. Examples of metrics
3. Warping metrics
4. Metrics that ignore unimportant differences: Classification mod Invariance (CMODI)
5. New directions

See you tomorrow!

Outline for Wednesday

Metrics: examples, data fidelity terms, warping, and face recognition

1. Brief review from yesterday
2. Examples of metrics
3. Warping metrics
4. Metrics that ignore unimportant differences: Classification mod Invariance (CMODI).
5. New directions

Central motivation: principled comparisons of image data that measure what is important.

Starting point: viewgraph "norm", very little development of metrics which incorporate domain knowledge.

Philosophy: think about what we would like to do ideally, use this to inform whatever approximations of this ideal get implemented.

Metrics: data fidelity, likelihood, and image comparisons

Recall: metrics enter into variational image analysis methods as *data fidelity* terms with a connection to maximum likelihood estimation.

Example:

$$\min_u F(u) = \int |\nabla u| dx + \lambda \int |d - u|^2 dx \tag{7}$$

This data fidelity can be viewed as the negative log-likelihood of the Gaussian likelihood of data d given truth (or model) u.

Data fidelity terms with "projection" operators – $||Pu - d||$.

$$\min_u F(u) = \int |\nabla u| dx + \lambda \int |Pu - d|^2 dx \tag{8}$$

There is often a significant null or effective null space associated with P. Typically, the measurements alone are not enough to infer the value u should have – $P^{-1}(u)$, even when it exists, is not what we want to compute.

Metrics: data fidelity, likelihood, and image comparisons

Image comparisons for scientific purposes require metrics implicitly "knowing" which differences are important and which are not.

What is important depends on how the data is being used:

Validation Complex simulation codes predict a few scalar quantities. Then if we predict $y \in \mathbb{R}^3$ using states in $\mathbb{R}^{1000000}$ then there will be a codimension 3 $(= 999,997$ dimensional) submanifold ALL mapping to the same output.

Recognition Rotation, scaling, translation, lighting changes all produce different images given a fixed subject. These sets of images produced highly curved submanifolds that we want the metrics to effectively factor out.

Generally Even in cases in which there are no null directions, there should be a ranking in the importance of different perturbations $\rightarrow$ weighted metrics.

Example: L^2

The L^2 metric (norm) has lots of nice properties.

In the ROF example we use the L^2 norm squared:

$$\min_u F(u) = \int |\nabla u| dx + \lambda \underbrace{\int |d-u|^2 dx}_{(L^2)^2} \tag{9}$$

Properties:

This norm gives us a Hilbert space: which is the nicest kind of normed vector space.

Our intuition works: due to the fact that we have an inner product.

Derivative of $(L^2)^2$ is simple: leading to nice derivations

$(L^2)^2$ is strictly convex: leading to unique solutions.

$(L^2)^2$ can be seen as motivated by Gaussian perturbations: what most are comfortable assuming (even when they shouldn't be!)

Example: L^1

Using an L^1 data fidelity leads to very interesting geometric properties: we will discuss this in detail tomorrow.

$$\min_u F(u) = \int |\nabla u| dx + \lambda \int |d-u|^1 dx \tag{10}$$

Properties:

This norm is more robust to outliers than $(L^2)^2$: This has practical advantages as we have seen from Asaki's talk.

L^1 **is not strictly convex:** we can (and do) get nonuniqueness.

L^1 **is a boundary case in the $1 \leq p \leq \infty$ continuum:** interesting things often happen at boundaries.

There is no inner product giving rise to L^1: and correspondingly, intuition for this space is not immediate.

Examples: L^∞

What about $p = \infty$, the other boundary case for the L^p norms?

$$\min_u F(u) = \int |\nabla u| dx + \lambda \sup_x |u - d|? \tag{11}$$

Not robust to outliers: recall Peter Jones' β numbers and the reasons for considering the L^2 versions that were studied carefully in Gilad Lerman's dissertation.

Derivatives are difficult: in fact they are singular and require regularization for computation purposes. (If $u \geq 0$, u is continuous, $u^* = \sup u$ and $A \equiv \{x | u(x) = u^*\}$ then the directional derivative in the direction of v is $\sup_A v(x)$.)

There are interesting connections: tomorrow we will briefly touch on the case where we use take the L^∞ norm of the gradient term (regularization). This leads to very interesting work on absolutely minimizing functions.

Actually: , in the case of binary images, the above equation can be solved by inspection $\rightarrow$ exercise)!

Examples: Poisson

New work: since it is not unreasonable that noise could be Poisson since many measurements are based on *counting* the data fidelity term derived by taking the negative log-likelihood can be considered.

$$\min_u F(u) = \int |\nabla u| dx + \lambda \int (u - f \log u) dx \tag{12}$$

Poisson distribution: $p(k) = \frac{e^{-\mu}\mu^k}{k!}$; mean and variance are both equal to μ

CCD cameras and scintillators: are measuring particle counts. Therefore, Poisson statistics are appropriate.

Examples: relative entropy

an interesting and very useful non-metric distance is the relative entropy distance also known as the Kulback-Liebler distance. In particular, this distance dominates information theory. A beautiful reference for those interested in more details is the text by Thomas and Cover.

Definition: Given two pdf's $\rho_1(x)$ and $\rho_2(x)$ $D(\rho_1||\rho_2) \equiv \sum_x \rho_1(x) \log(\frac{\rho_1(x)}{\rho_2(x)})$.

Not Symmetric: Generally, $D(\rho_1||\rho_2) \neq D(\rho_2||\rho_1)$.

Triangle inequality: not satisfied!

Large deviation theory: The theory of types (see chapter 12 of Cover and Thomas)exploits relative entropy to obtain results on large deviations.

Opinion: Information theory still has much to offer image analysis – a good place to start would be with relative entropy distances between image derived pdf's.

Examples: Meyer's suggestion

Y. Meyer suggested that to better separate the cartoon and texture components of an image the following function be used:

$$\min_u F(u) = \int |\nabla u| dx + \lambda ||d - u||_* \tag{13}$$

where $||w||_*$ is given by

$$||w||_* = \inf_{g_1, g_2 \in A} \sup_x (g_1^2(x) + g_2^2(x))^{\frac{1}{2}} \tag{14}$$

for $A \equiv (g_1, g_2)$ such that $g_1, g_2 \in L^\infty$ and $w = \nabla \cdot (g_1, g_2)$.

The set of all generalized functions that can be written $w = \nabla \cdot (g_1, g_2)$ with $g_1, g_2 \in L^\infty$ is a Banach space with the norm $|| \cdot ||_*$ given above and is in some sense the dual space of BV.

Reference: Y. Meyer's monograph, "Oscillating Patterns in Image Processing and Nonlinear Evolution Equations", AMS 2002

Examples: Vese and Osher's implementation

Vese and Osher implemented a version of Meyer's suggestion of using something close to the dual space of BV for the data fidelity and obtained:

$$\inf_{u,\vec{G}} F(u,\vec{G}) = \int |\nabla u| dx + \lambda \int |f - u - \nabla \cdot \vec{G}|^2 dx + \gamma || \sqrt{\vec{G} \cdot \vec{G}} ||_p \tag{15}$$

where $\vec{G} = (g_1, g_2)$ and $g_1, g_2 \in L^\infty$.

The parameters λ and γ control the balance between terms. The idea is to let $p \to \infty$. In practice $p \approx 1$ works quite well.

Advantages: over a strictly faithful implementation of Meyer's ideas are the fact that you can actually compute this functional and minimize it.

New work: continues to improve the cartoon texture separation

Results: show an improvement in the cartoon-texture separation ... results from the Vese-Osher paper are shown on the next slide.

Examples: Vese and Osher's implementation

original thumb print

ROF separation

Vese-Osher Separation

Examples: Warping metrics

In a bit we will look in more detail at a couple of warping metrics. Here I simply list a few of the methods that have been proposed.

Miller's Computational Anatomy: In this approach two images are connected by paths in image space whose length is calculated by integrating the Sobolev norm of the increments along the path. Then the minimal length path is found (geodesic). This length is defined to be the distance between the images.

Monge-Kantorovich Warping: The classic Monge Optimal transport problem looks for the cheapest way to move a pile of dirt in one spot to a hole with the same volume in another spot. Essentially we can view any two images (after normalization) as mass distributions. The optimal transport of one to another is the MK warping distance. We will say more about this today.

Curvature Warping: Here we warp the domain around and penalize for curvature by having having something like $|\kappa|^2$ (where κ is the curvature) in the cost functional. The cost of the cheapest warp between two images is then the distance between the two images.

Elastic Warping: Same as curvature warping except we use elastic energies to construct the cost function.

Example: invariant recognition metrics

The idea of constructing metrics which ignore differences between objects or images which are in the same orbit of some group or set of transformations is certainly not new.

Difficulty: This is computationally very difficult for those problems for which it cannot be done analytically.

The Idea is compelling: It is the right thing to do. Find "all" the invariant manifolds or sets using a "large" amount of data and methods that Laurence Saul talked about and then construct a quotient map. Given new data, use the quotient map to project the new data on the quotient space. Invariant recognition is then easy.

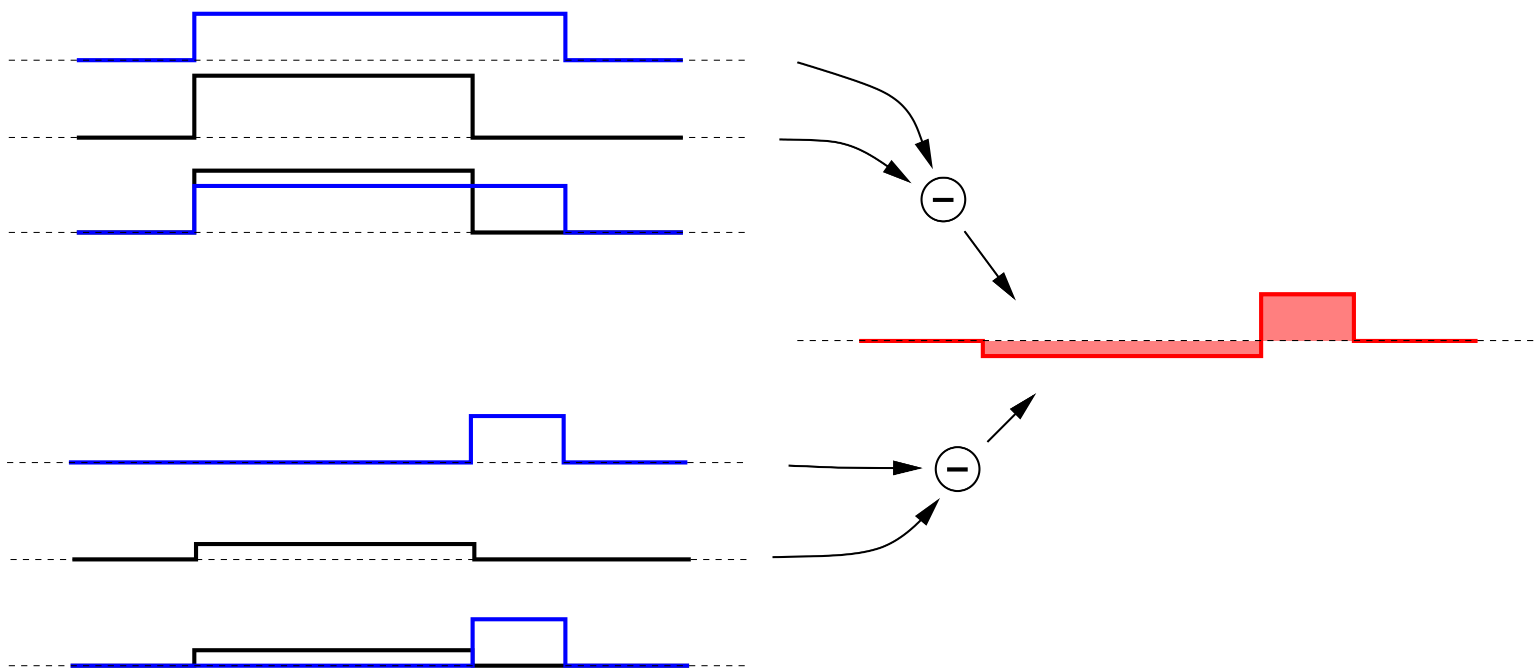
Practically: You must do some sequence of approximations and hacks.

One example is examined below: What we call classification mod invariance (CMODI) was constructed to factor our invariances for which we know the infinitesimal generators (the tangent space.

We now look at warping and CMODI – tomorrow we will look at L^1 and L^2 metrics in conjunction with TV regularization.

Why Warp?

Typical metrics are *norms* of differences, $\rho(u, f) = ||u - f||$. This can have undesirable effects.

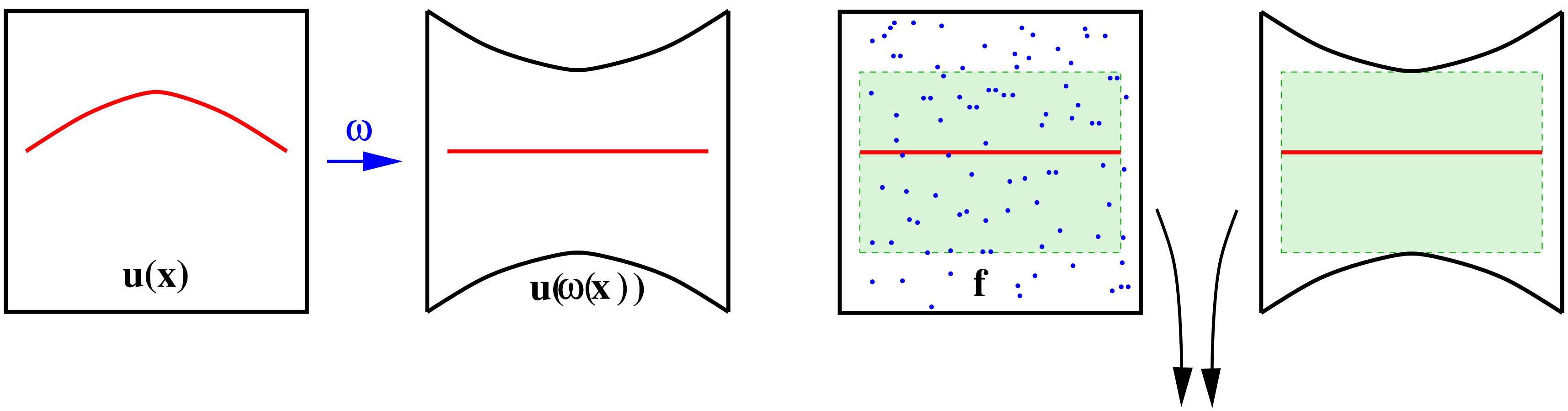


We want metrics which in effect split differences nonlinearly and weight the factors differently.

Why Warp?

Idea: Warp domain to match image u and f. Use a natural stochastic term $\log(p_n())$ to measure remaining difference.

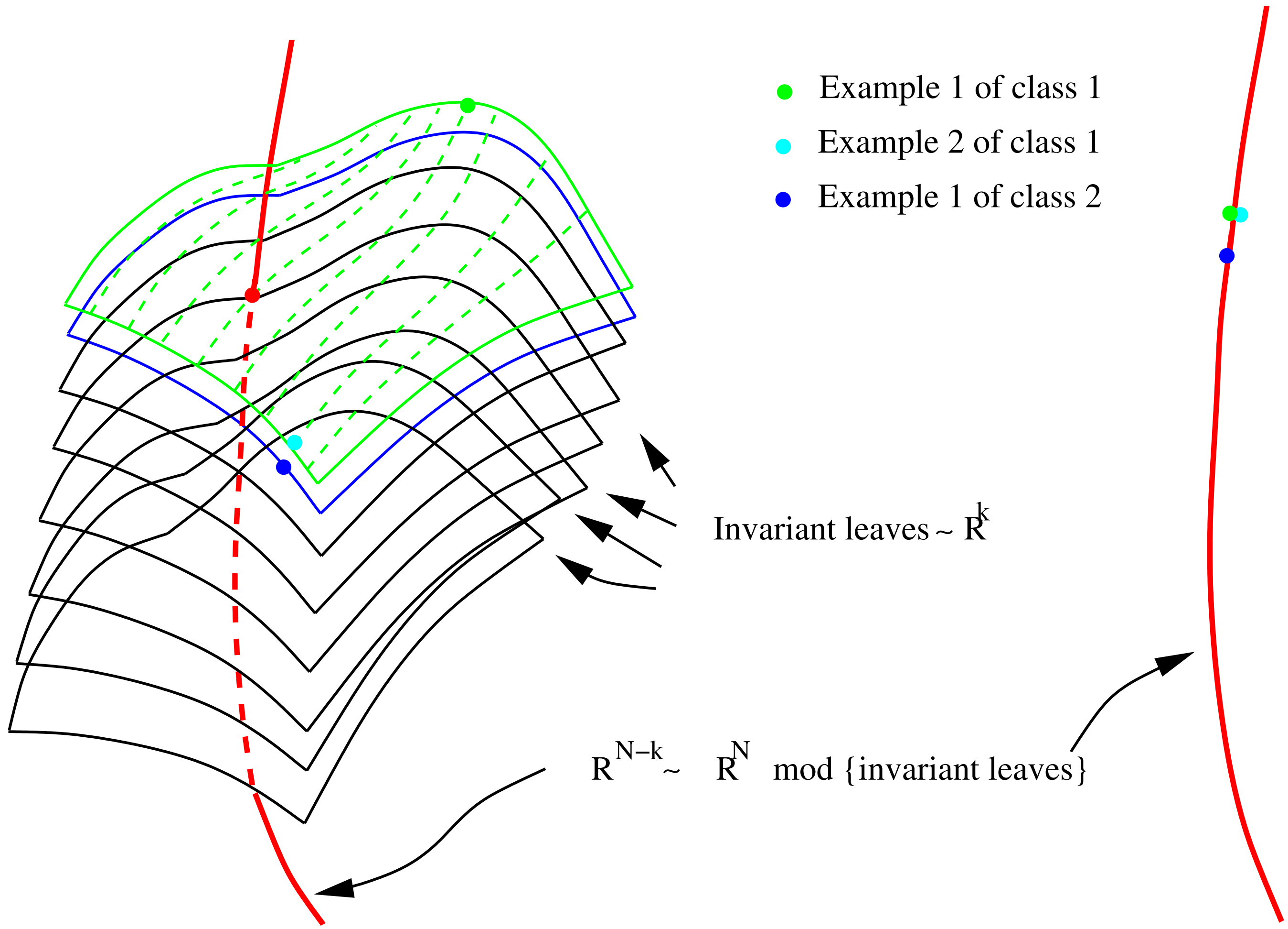
$$\rho(u(x), f) = E(\omega) - \alpha \log(p_n(f|u(\omega(x)))$$



$$\rho(\mathbf{u}, \mathbf{f}) = \mathbf{E}(\omega(\mathbf{x})) - \alpha \log(\mathbf{p}_n(\mathbf{f} \mid \mathbf{u}(\omega(\mathbf{x}))))$$

(on green comparison region)

Quotients: Metrics that ignore unimportant differences.



Warps and Quotients: Splitting Differences

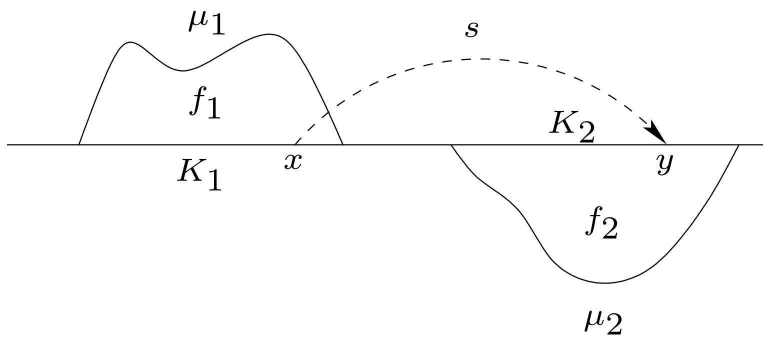
We want metrics which care more about some directions than others or at least measure them differently:

Warps: Metrics which split the differences nonlinearly and measure components differently

Quotients: Metrics which split the differences nonlinearly ignore one component.

Monge-Kantorovich (MK) Warping: the idea

How to move the pile to the hole?



The mapping s should be injective, map K_1 to K_2, and satisfy the pullback condition:

- $\mu_2 = \mu_1 \circ s^{-1}$; or
- $\int_{K_1} (h \circ s) f_1 dx = \int_{K_2} h f_2 dx$, for all $h \in C(K_2)$; or
- $f_1 = (f_2 \circ s)|Ds|$.

MK Warping: New method and results

New results obtained by Chartrand et al. use insights from convex analysis to obtain the gradient of the cost functional directly.

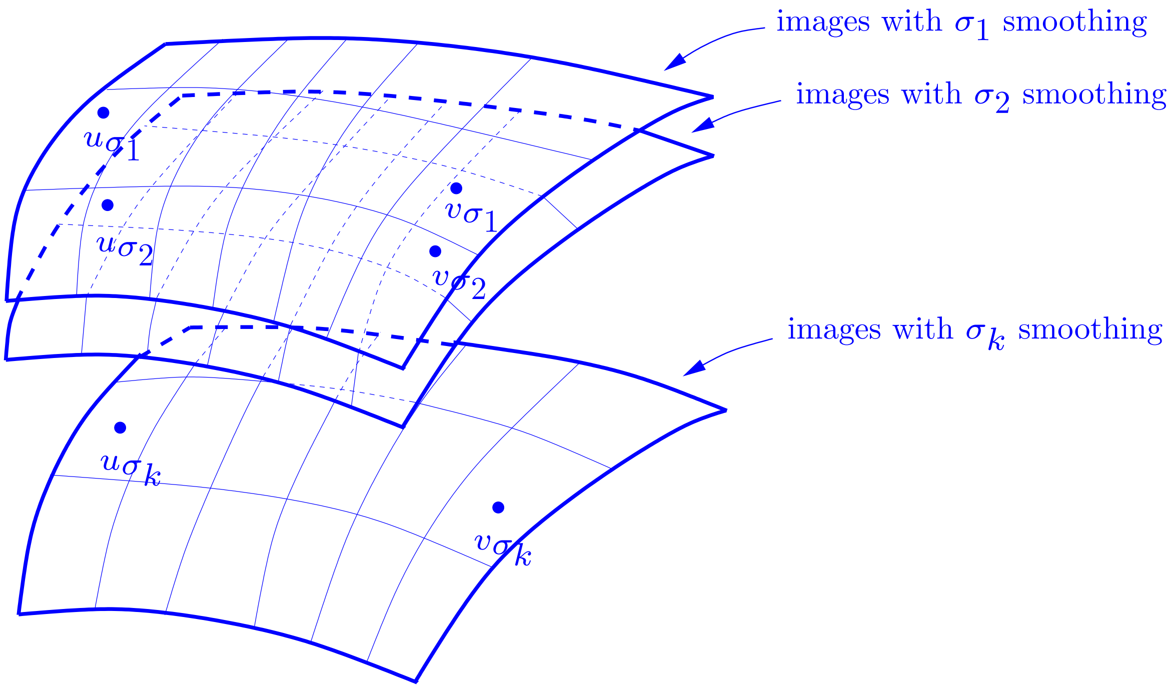
Kantorovich: relaxed and dualized the problem:

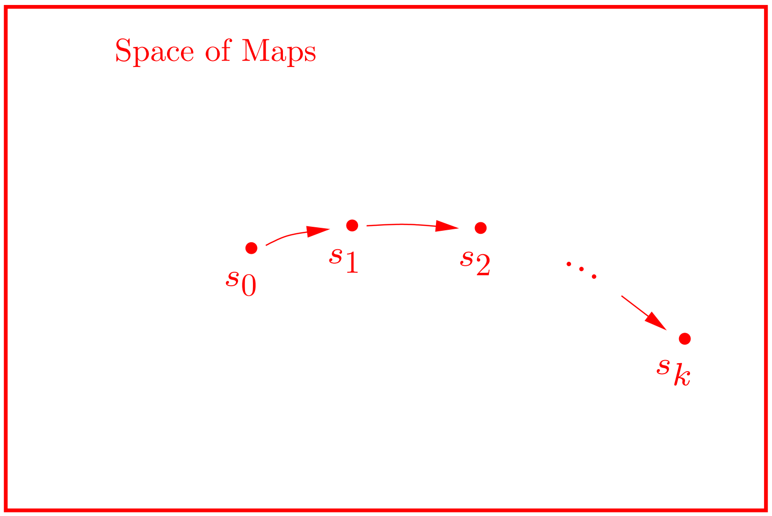
After some manipulation: insights from convex analysis allow the calculation of the derivative of the functional.

Amazingly: in the process the problem is converted from a constrained to an unconstrained problem!

MK Warping: New method and results

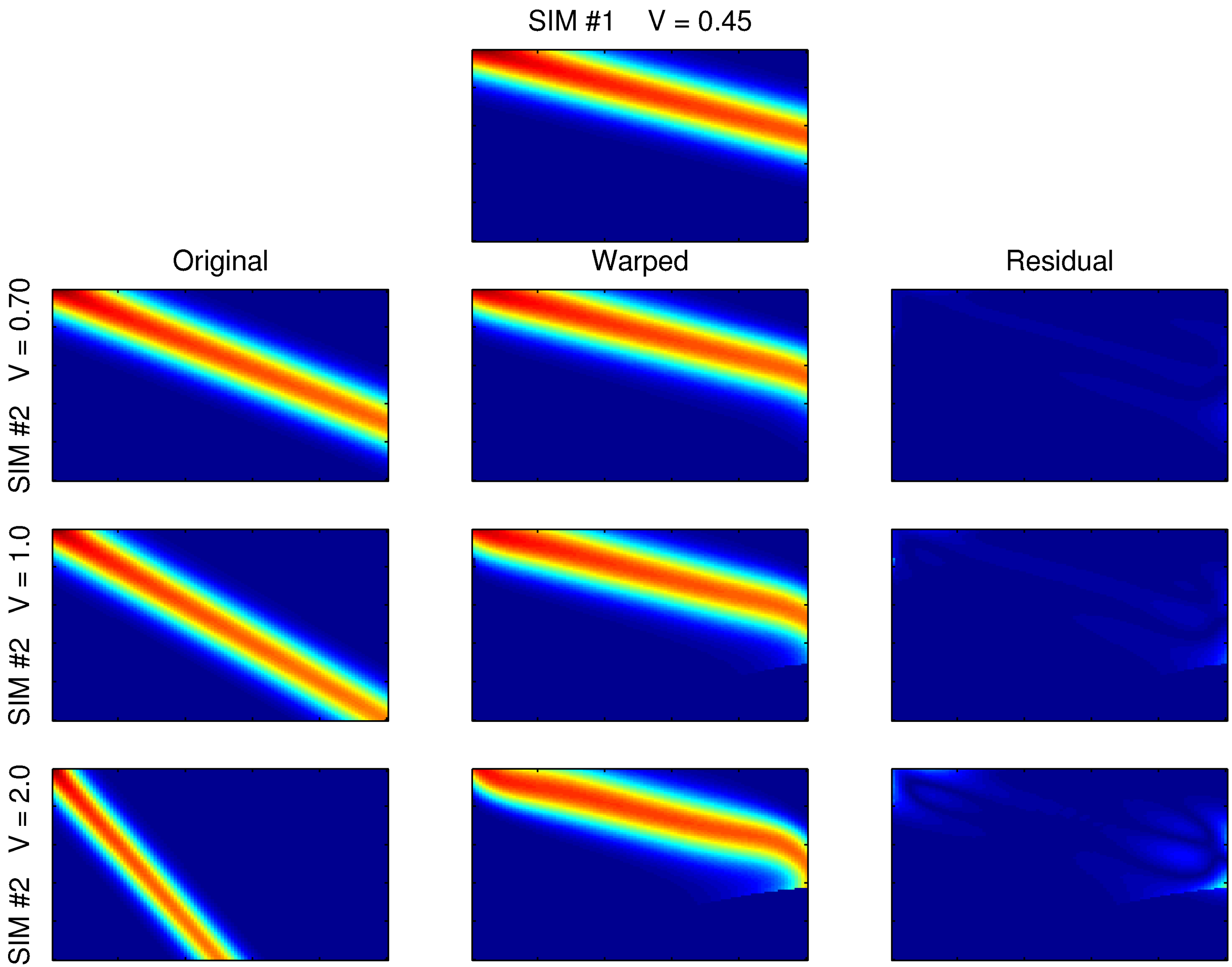
Schematic illustrating the algorithm: The idea is that we start with some map s_0 and evolve to find s_1 transporting u_{σ_1} to v_{σ_1}. Now move to a smaller scale σ_2 and, starting our evolution in S at s_1, find s_2 transporting u_{σ_2} to v_{σ_2}. Repeat this until we reach a fine enough resolution.



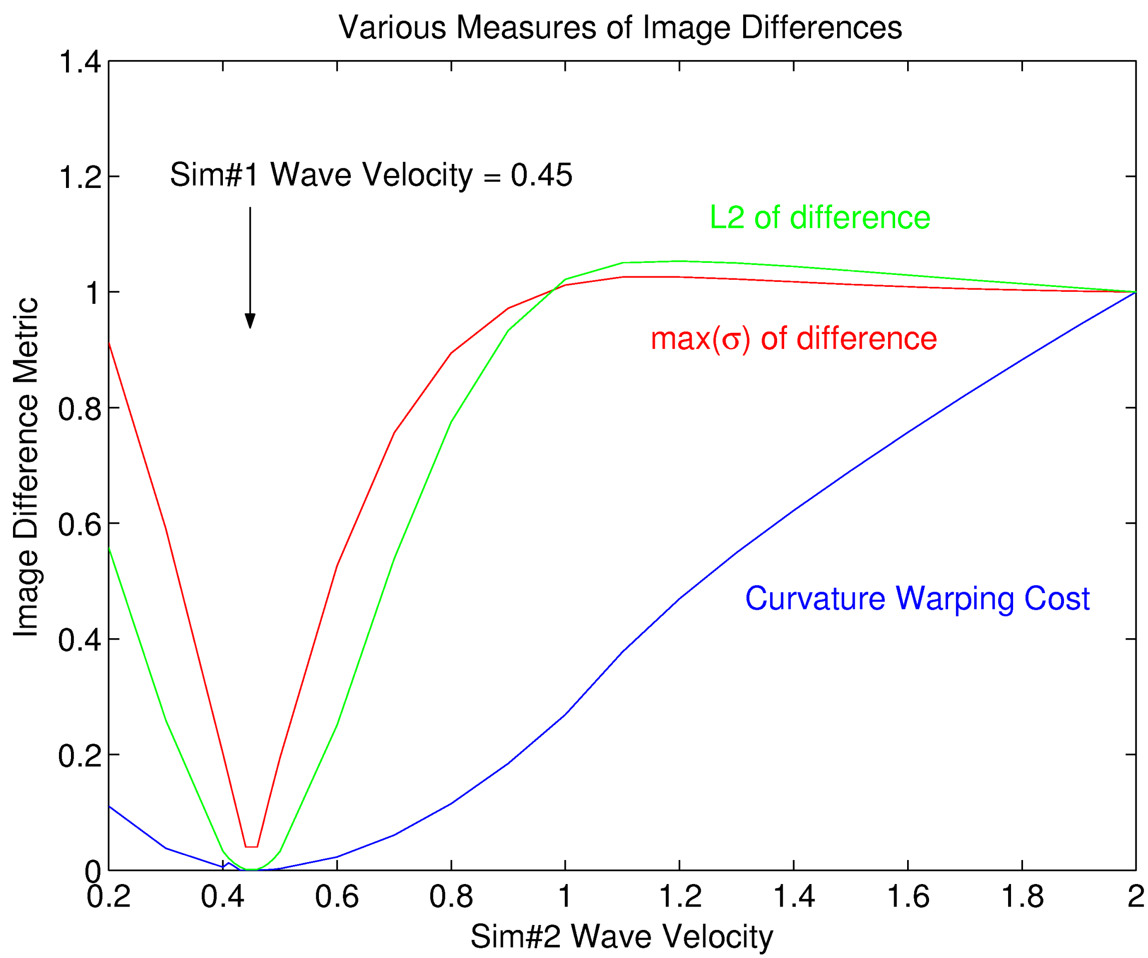


$$\phi_t = -(v - u(\nabla\phi)\det(D(\nabla\phi))) \text{ and } s = \nabla\phi.$$

Curvature warping example

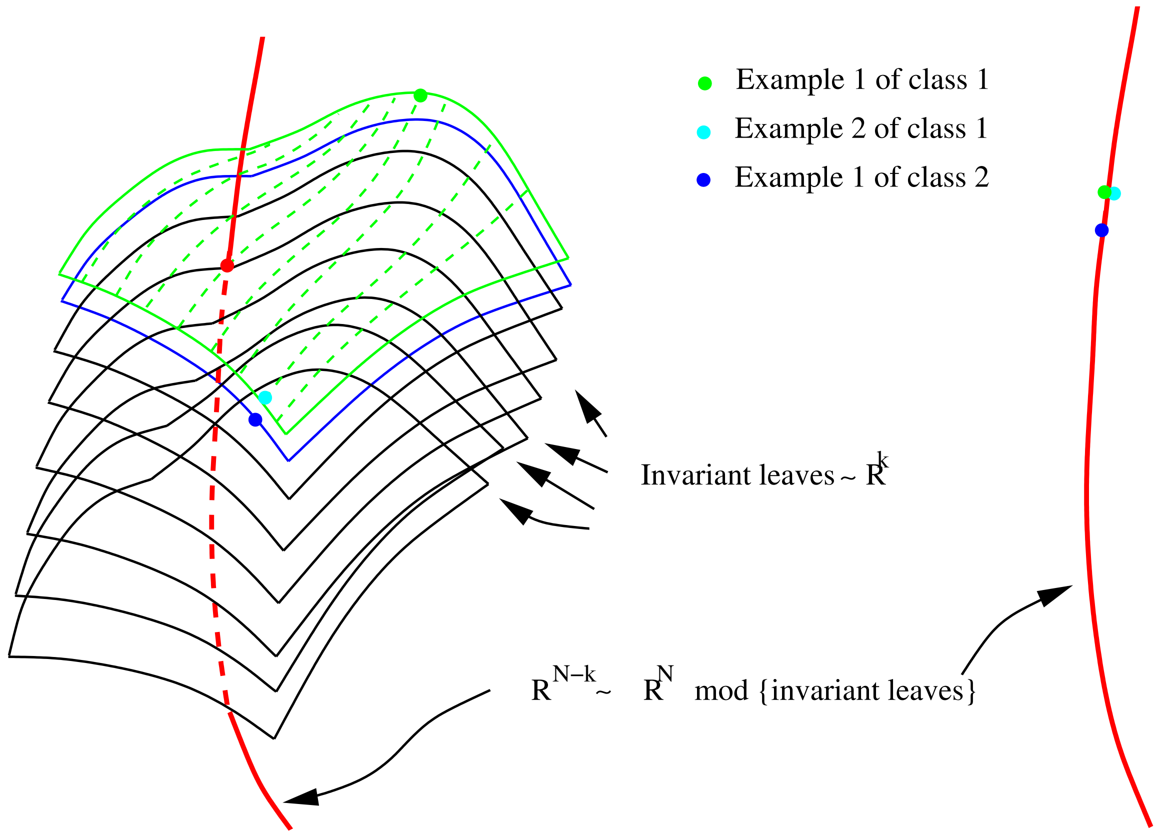


Curvature warping example



CMODI: ignoring unimportant things

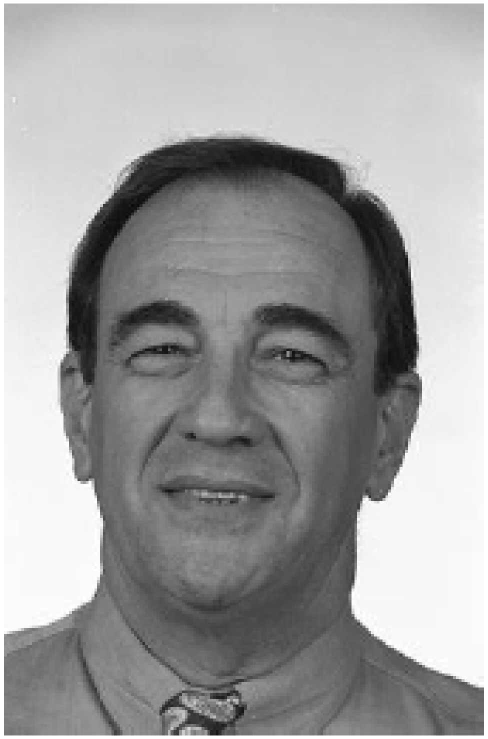
(for the last time(!), quotients)

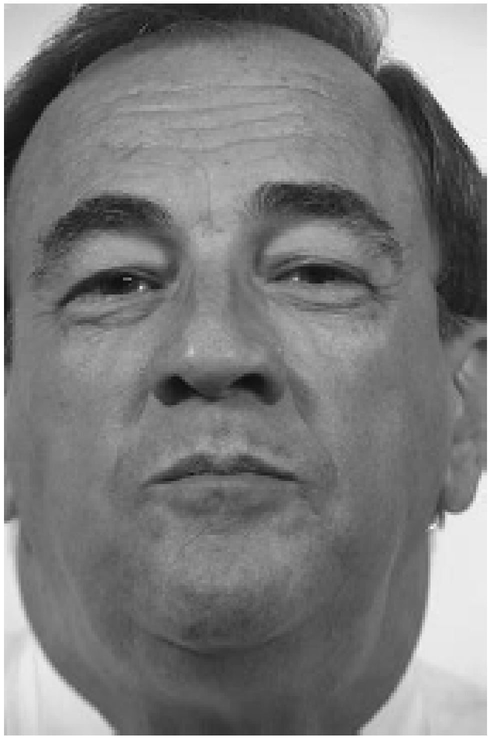


What we would like to do: factor out the orbits of transformations to which we desire invariance.

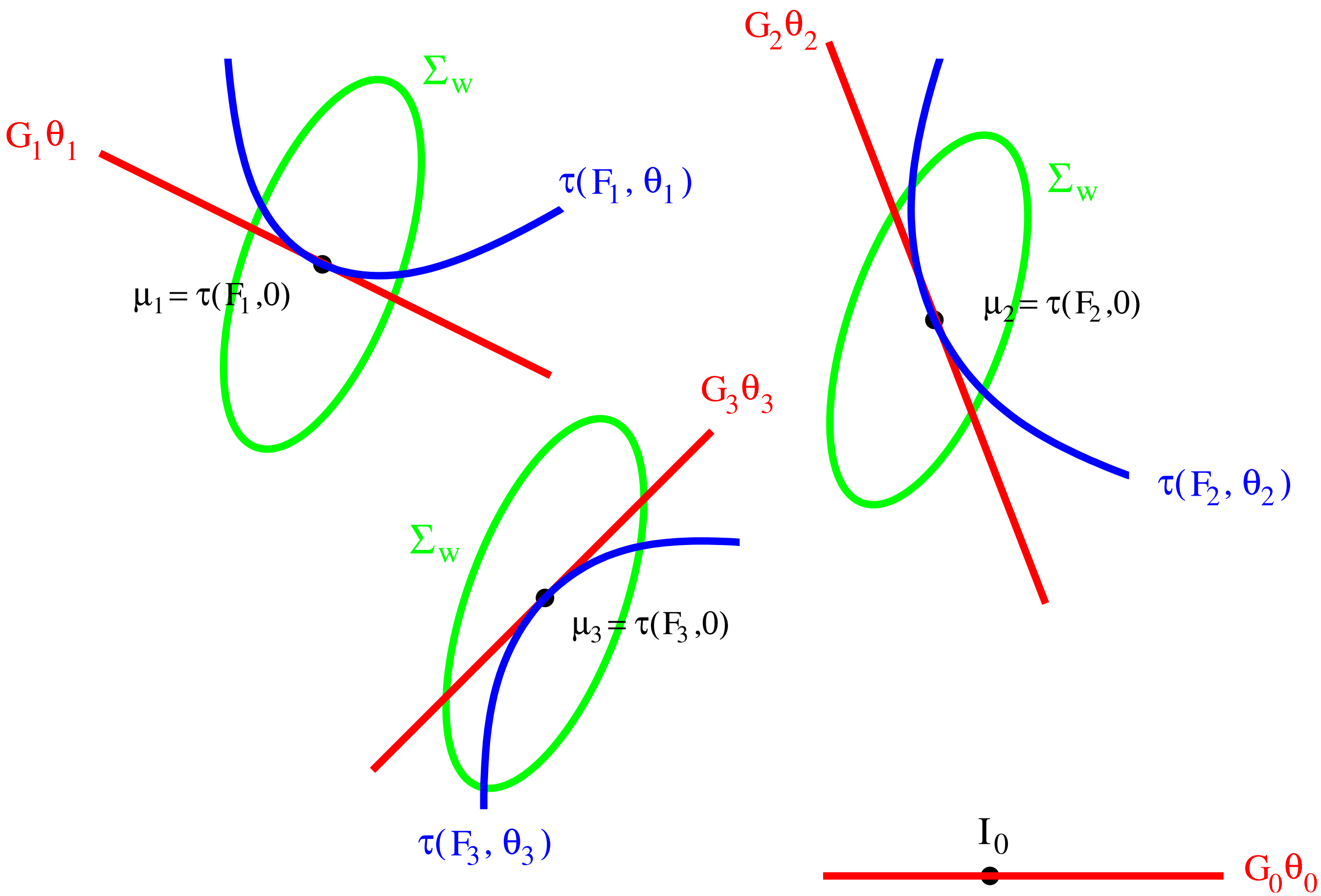
CMODI: ignoring unimportant things

Test case: find images of the same person.





CMODI: ignoring unimportant things



How we get conditional measures which approximate the quotient space metrics.

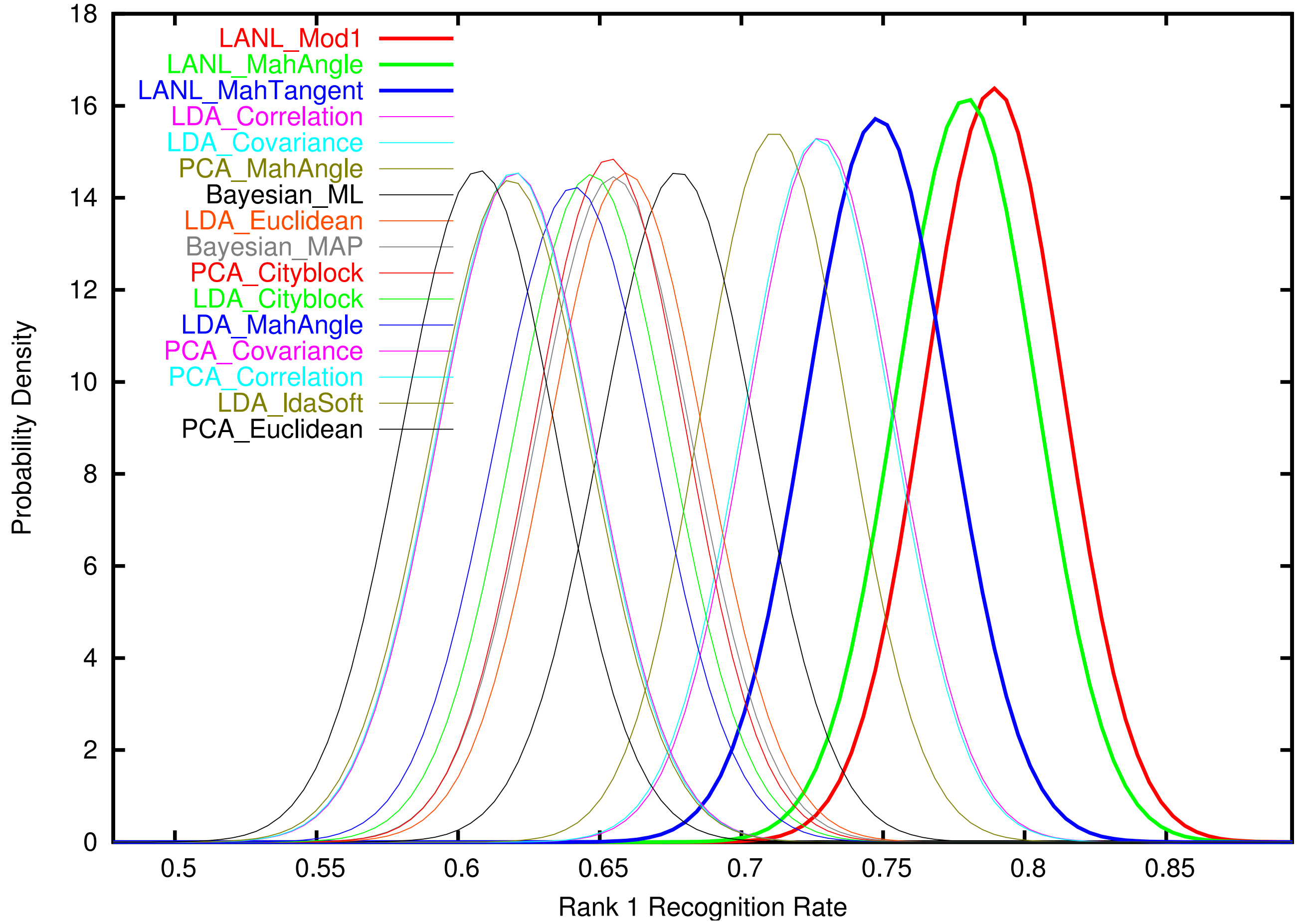
CMODI: ignoring unimportant things

Example of shifts

Image convolved with kernel	First order approximation to shift	Second order approximation to shift

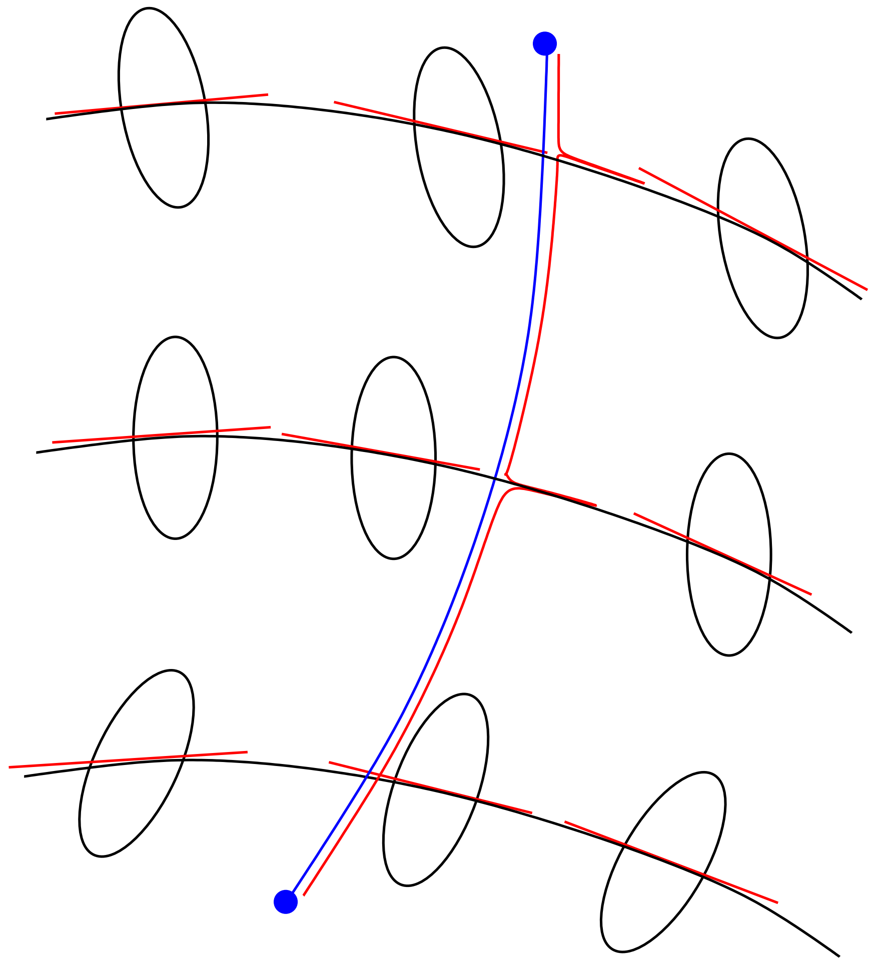
CMODI: ignoring unimportant things

Results:



Our results (the three rightmost curves) are quite good!

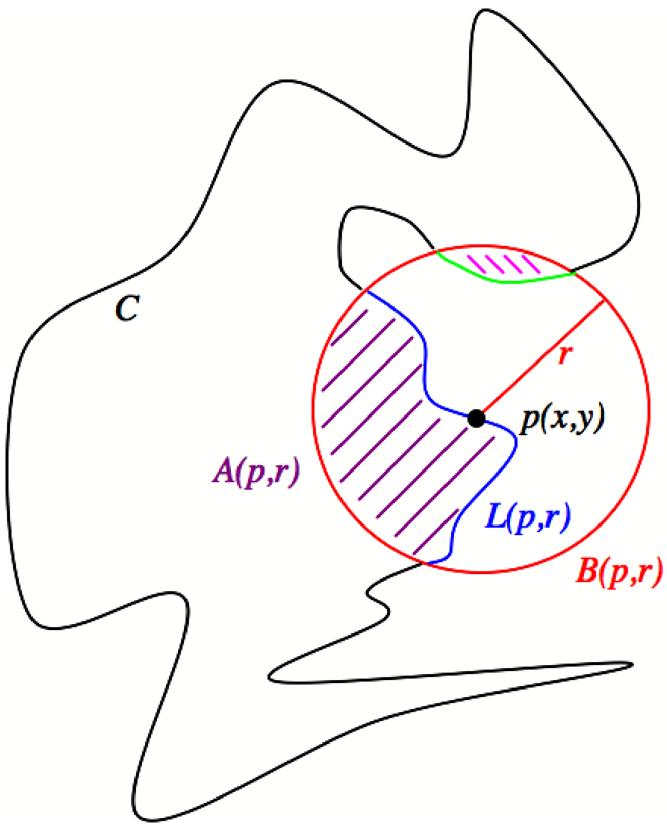
New directions: Geodesics and singular metric fields

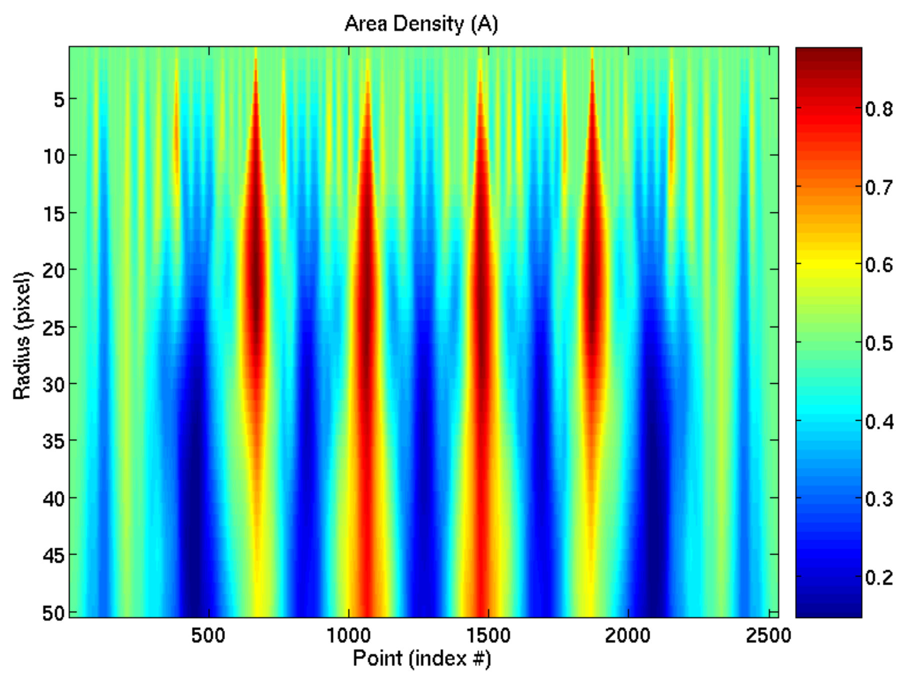


Replace $M \to PP^TMPP^T + \alpha I$ then let $\alpha \to 0$. (P is a projection annihilating ignored directions.)

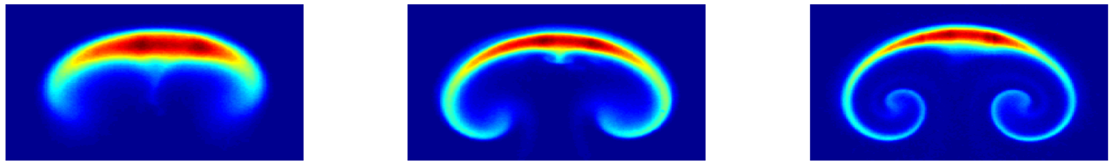
New directions: Multiscale densities

Details on metrics and densities

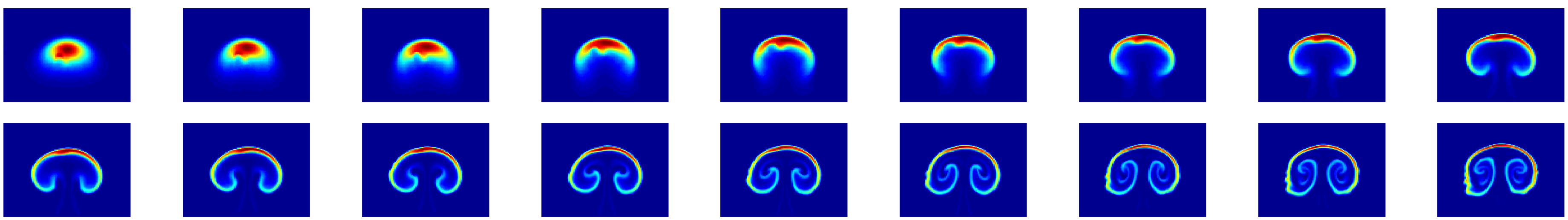




New directions: Multiscale densities



Three experimentally measured images.



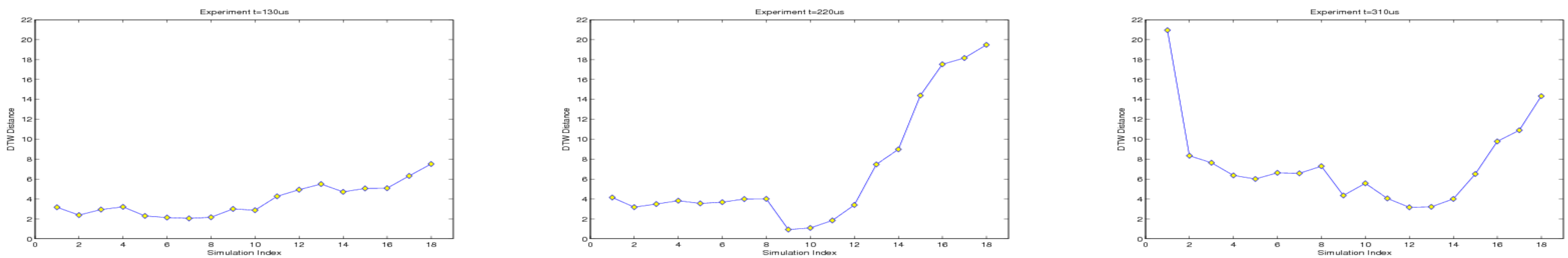
18 simulated images.

New directions: Multiscale densities

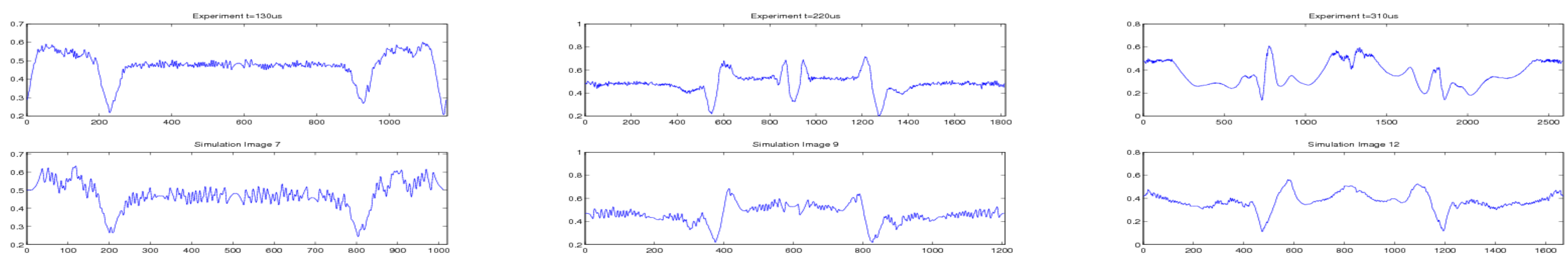
Sequence distances: are computed for the density sequences. These distances are based on *dynamic time warping*. (In the same family of algorithms as the Smith-Waterman and related methods for sequence similarity in the context of genetic sequences.)

Minimizing the distances: The simulation best matching a given experimental measurement is determined by minimizing distances. Preliminary results are very good!

New directions: Multiscale densities



Sequence similarity of each experiment with every simulation.



Density sequences for each simulation and the closest simulation.

A look ahead: L^1TV and L^2TV

We will look at regularization tomorrow. In particular we will study TV regularization and how the regularization couples with the data fidelity – L^1 and L^2 in particular.

1. L^1TV and L^2TV
2. Review of regularization.
3. Examples
4. More mathematical background
5. p-Laplacian and H^1TV and L^2TV
6. results using L^2TV
7. L^1TV, L^2TV and geometric analysis
8. Summary
9. References

Outline for Thursday

Today we will finish this tutorial series with a closer look at regularization and the interesting geometric/analytic aspects of TV regularized functionals and their minimizers.

1 L^1TV and L^2TV
2 Review of regularization.
3 Examples
4 More mathematical background
5 p-Laplacian and H^1TV and L^2TV
6 results using L^2TV
7 L^1TV, L^2TV and geometric analysis
8 Summary
9 References

Perspective: geometric/analytic insights provide the power needed for creating and understanding the best image analysis methods.

My Goal: to motivate you to learn more.

L^1TV and L^2TV: Our focus for the day

Today we will study quite closely the following two functionals paying special attention to the TV regularization.

$$\min_u F(u) = \int |\nabla u| dx + \lambda \int |u - d|^2 dx \tag{16}$$

$$\left(\min_u F(u) = \int |\nabla u| dx + \lambda \int |Pu - d|^2 dx \right)$$

$$\min_u F(u) = \int |\nabla u| dx + \lambda \int |u - d|^1 dx \tag{17}$$

But before we do this, I will review regularization and list a few examples.

Review of Regularization: the concept and pictures

Regularization can be viewed as

Method of stabilizing inverse problems: simplest point of view. But we understand why this is reasonable by seeing regularization as

Enforcement of prior: choosing the solution that best fits what we know about the solution;

Dimension reduction: reducing the dimension of the object we are trying to infer or estimate from the data;

While regularization can be viewed as a stabilization procedure for inverse problems, it is justified by the fact that this is how we can insert prior information into the solution.

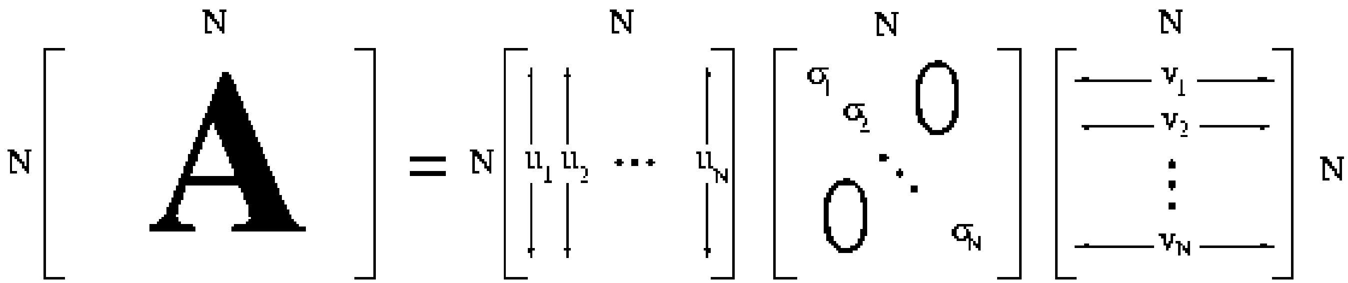
Illustration: The need for regularization in inverse problems can be seen by looking at a very simple linear prototype: measure $\tilde{b} = b + \eta$; now find x given $Ax = b$ and the fact that η is noise.

We finish the recap of regularization by looking closely at this problem.

Review of Regularization: the concept and pictures

Linear model: Given $Ax = b$ and $\tilde{b} = b + \eta$, find x. Assume for simplicity that A is an $N \times N$ invertible matrix.

Recall the SVD: $A = U\Sigma V^t$ where Σ is a diagonal matrix with ordered, non-negative entries and U and V are $N \times N$ orthogonal matrices.



Express b and η in terms of the u_i's: Suppose that $b = \alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_N u_N$ and $\eta = \beta_1 u_1 + \beta_2 u_2 + ... + \beta_N u_N$.

Note: $A : v_i \rightarrow \sigma_i u_i$

Therefore: $A^{-1} : \alpha u_i \rightarrow \frac{\alpha}{\sigma_i} v_i$ and $A^{-1} : (\alpha + \epsilon) u_i \rightarrow \frac{(\alpha+\epsilon)}{\sigma_i} v_i$

In other words: An ϵu_i perturbation of b leads to an $\frac{\epsilon}{\sigma_i} v_i$ perturbation of x.

Conclusion: Small singular values (all interesting inverse problems are afflicted with them!) induce instabilities into the inverse problem that must be dealt with using *regularization*.

Examples

Examples of regularization terms:

$\int |u|^p dx$ **:** used, for example, in the tomographic reconstruction technique known as ART for algebraic reconstruction technique (ART).

$\int |\nabla u|^p dx$ **:** The Euler-Lagrange equations $\rightarrow$ p-Laplacian.

$\int |\nabla u|^{p(|\nabla u|)} dx$ This term proposed by Blomgren, Chan, Mulet and Wong, has recently been shown to have a minimizer by myself and other members of the DDMA team. The actual functional we consider is $\int |\nabla u|^{p(|\nabla G_\delta \star u|)} dx + \lambda \int |u - d| dx$. Due to the L^1 data fidelity term, there is no uniqueness. We typically choose $p(w)$ to be a decreasing function of w such that $p(0) \geq 2$, $p(w) = 1$ for $w \geq M > 0$,

$\int |\nabla u|^{h(x)} dx$ This is a simplified version of the regularization introduced by Blomgren et al. Recently Stacey Levine and collaborators have made some very nice progress. Here $h(x)$ is something like $p(|\nabla G_\delta \star d|)$, a smoothed version of the data d plugged into $p()$ above.

$d(u, A)^p$**:** $A =$ a submanifold learned from data. Choice of d is critical here.

Examples

SVD Projection: regularize by setting small singular values to zero and projection. I.E. project $\tilde{b}$ onto the span of the u_i whose corresponding singular values satisfies $\sigma_i \geq 1$. Now inversion will not magnify errors.

Mumford-Shah Segmentation: uses $|\nabla u|^2$ and curve length to regularize the solution to the segmentation problem.

$$\min_u F(u) = \int_{\Omega \backslash \Gamma} |\nabla u|^2 dx + \lambda H^1(\Gamma) + \gamma \int |u - d|^2 dx \tag{18}$$

Vese-Osher Texture separation: mentioned above for it's use of a novel data fidelity term (metric) uses total variation regularization.

$$\inf_{u, \vec{G}} F(u, \vec{G}) = \int |\nabla u| dx + \lambda \int |f - u - \nabla \cdot \vec{G}|^2 dx + \gamma || \sqrt{\vec{G} \cdot \vec{G}} ||_p \tag{19}$$

where $\vec{G} = (g_1, g_2)$ and $g_1, g_2 \in L^\infty$.

Examples

Blomgren et al. denoising: as mentioned above suggests using an adaptive exponent for $|\nabla u|$.

$$\int |\nabla u|^{p(|\nabla u|)} dx + \lambda \int |u-d|^2 dx \tag{20}$$

Esedoglu-Osher Anisotropic denoising: looks at

$$\int \phi(\nabla u) dx + \lambda \int |u-d|^2 dx \tag{21}$$

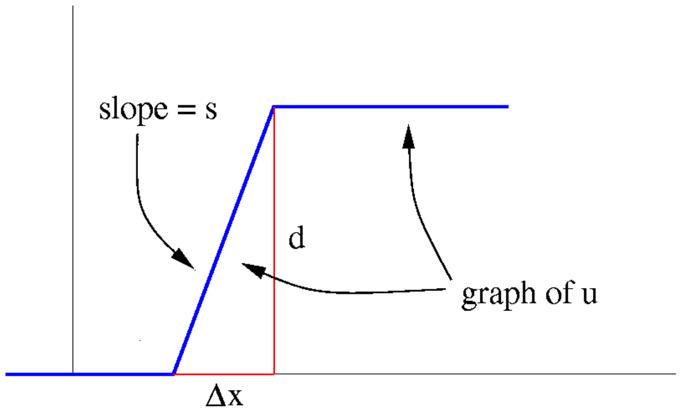
where $\phi(y)$ is positive and 1-homogeneous.

Now we look in detail at the L^2TV and L^1TV functionals and their minimizers. We begin by looking at properties of the TV seminorm and BV functions.

We begin with the question: Why choose $\int |\nabla u| dx$?

Background: TV functionals

Consider $F(u) \equiv \int |\nabla u|^p dx$



$$F(u) = s^p(\Delta x) = \frac{(s\Delta x)^p}{(\Delta x)^{p-1}} = d^p(\Delta x)^{1-p}$$

$(p > 1)$ $F(u) \underset{\Delta x \to 0}{\longrightarrow} \infty$ discontinuities are avoided: smooth u preferred,

$(p < 1)$ $F(u) \underset{\Delta x \to 0}{\longrightarrow} 0$ discontinuities cost nothing: step u preferred,

$(p = 1)$ $F(u) = d$ only jump magnitude "counts", no bias towards smooth or step.

Background: TV functionals

How do we manipulate and understand the TV seminorm rigorously?

Radon Measures:

Define: $\mu_f(A) \equiv \int_A f dx$ for any positive f in L^1.
Then: $f dx = d\mu_f$
What if we generalize allowable f's? Answer: Radon measures happen. (catch: we only know how to generalize μ)

BV functions:

Typical gradient fields: $d\vec{F}(x) = \vec{\sigma}(x) f(x)$, $\vec{\sigma}(x)$ a unit vector field, $f(x)$ the magnitude $|d\vec{F}(x)|$.

Define the gradient measure $\mu_{dF}(A) \equiv \int_A \vec{\sigma}(x) f(x) dx$.

A BV function is one whose gradient measure can be written $\int_A \vec{\sigma}(x) d\mu$ where μ is a Radon measure.

Background: TV functionals

Another approach to the definition is through a weak formulation:

If $u \in C^1(\Omega)$ *total variation of* $u = \int |\nabla u| dx$

If $u \in W^{1,1}$

$$\int |\nabla u| = \sup\{\textstyle\int \nabla u \cdot \vec{g} dx \text{ for } |\vec{g}| < 1, \vec{g} \in C_0^1(\Omega; \mathbb{R}^n)\} \tag{22}$$

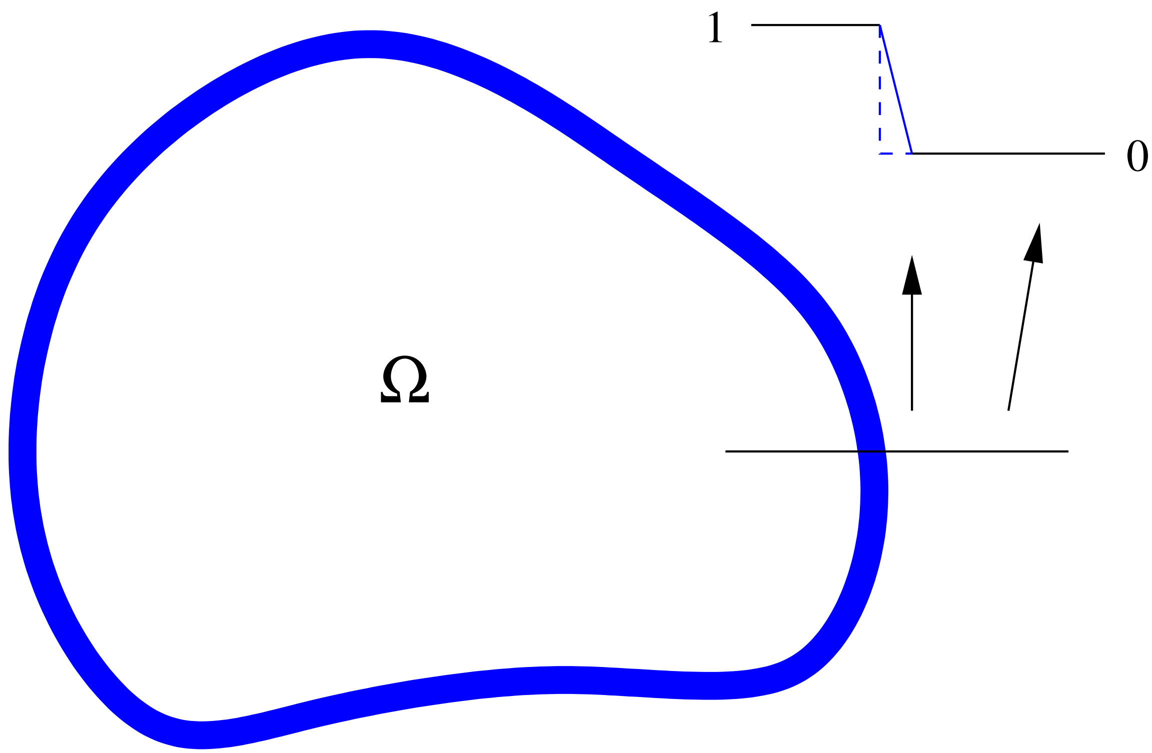
$$= \sup\{\textstyle\int u \text{div} \vec{g} dx \text{ for } |\vec{g}| < 1, \vec{g} \in C_0^1(\Omega; \mathbb{R}^n)\} \tag{23}$$

makes sense.

Finally for $u \in L^1(\Omega)$, we use the last equation to *define* $\int |\nabla u| dx$

Background: TV functionals

Suppose u is a characteristic function of a set Ω? Can we see what $\mathrm{TV}(u)$ will be?

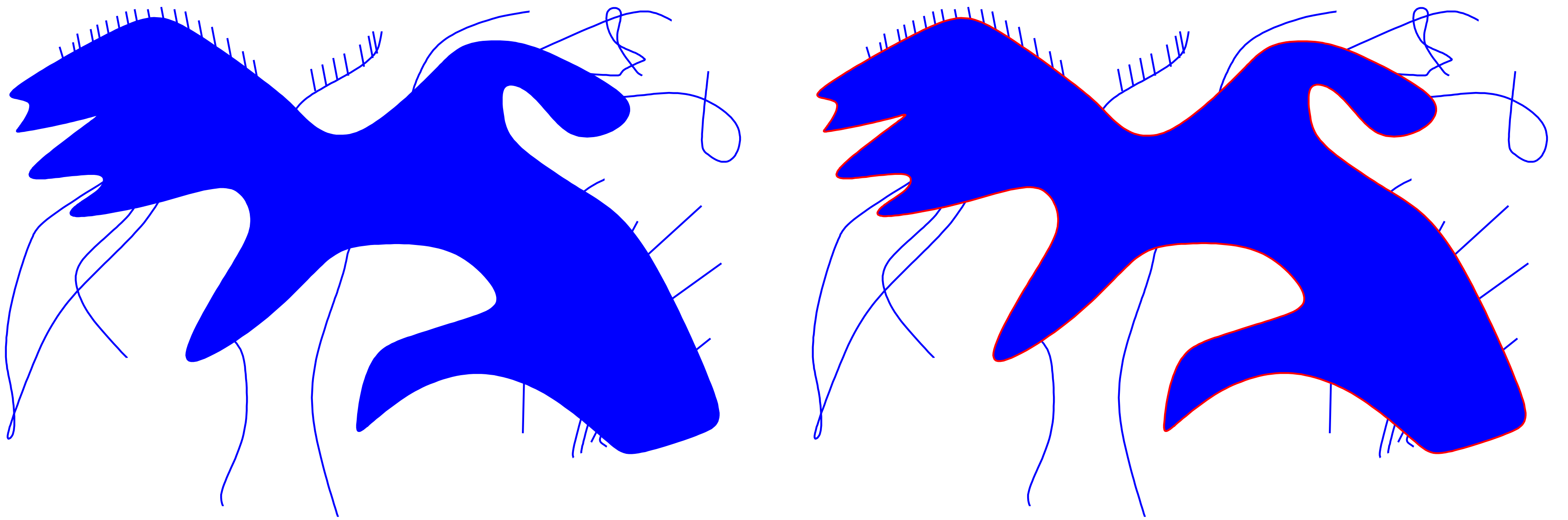


Using the figure of an approximate characteristic function, we can convince ourselves that $\mathrm{TV}(u)$ is simply the length of the boundary of the set Ω.

Background: TV functionals

Suppose Ω is really wild? The "argument" above depended on Ω being nice. What can we conclude about $\mathrm{TV}(\chi_\Omega)$ in this case?

Without out going into all the gory details, there is a set $\partial^*\Omega$ called *reduced boundary* of Ω that coincides with the boundary that test functions can see. $\mathrm{TV}(\chi_\Omega)$ picks up the boundary that integration against smooth test functions "sees".



An aside: ∞-Laplacian

Let's look a little more closely at the regularization term:

$$\int |\nabla u|^p dx. \tag{24}$$

Computing the derivative and setting it equal to zero we get the p-Laplacian:

$$\nabla \cdot \left(|\nabla u|^{p-2} \nabla u\right) = 0 \tag{25}$$

Checking for the case of $p = 2$ we see that indeed, in agreement with our calculation on Tuesday, $\int |\nabla u|^p \to \nabla \cdot (\nabla u) = \Delta u$.

This term works for $1 < p < \infty$. The cases $p = 1$ and $p = \infty$ being boundary cases, are quite interesting. Of course the first case, $p = 1$ is the case we are looking at in more detail today.

What about $p = \infty$?

An aside: ∞-Laplacian

The ∞-Laplacian is a very intriguing and challenging nonlinear PDE that still holds forth challenges to the analyst.

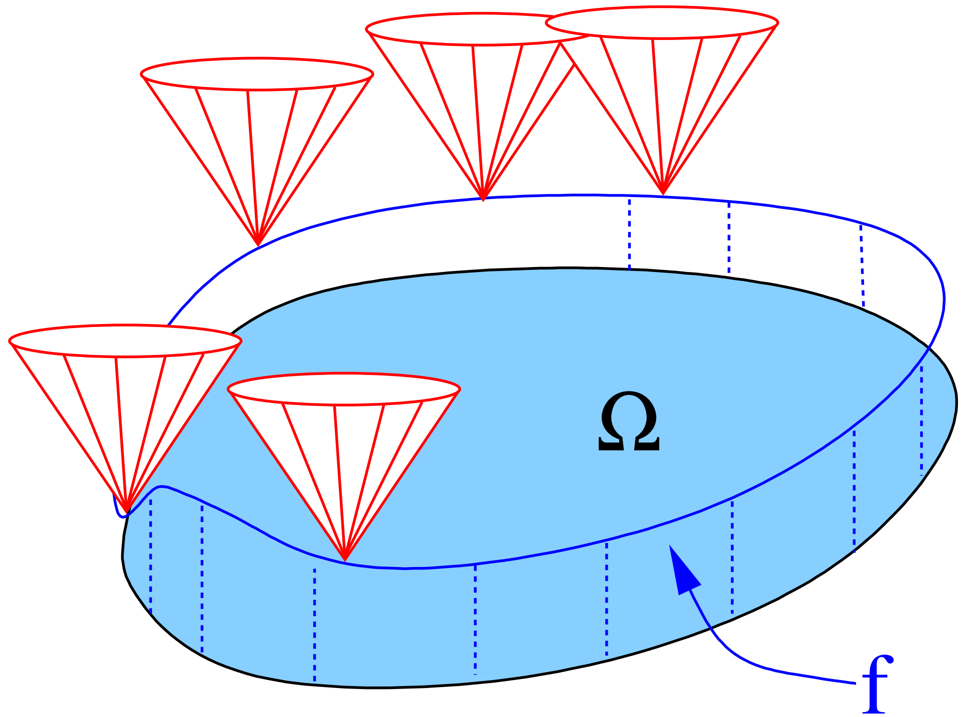
One way to interpret the limit as $p \to \infty$ of $\nabla \cdot \left(|\nabla u|^{p-2}\nabla u\right) = 0$ is to notice that:

$$\min_u \int_\Omega |\nabla u|^p dx = \min_u \left(\int_\Omega |\nabla u|^p dx\right)^{\frac{1}{p}} = \min_u ||\nabla u||_p \tag{26}$$

where of course we are specifying $u = f$ on $\partial\Omega$.

An *absolutely minimizing function* $f : \Omega \to \mathbb{R}$ is an optimal Lipschitz extension to the interior of Ω of the boundary data $f(\partial\Omega)$. Optimality has a sensible definition (see notes at the end of the tutorial). Absolutely minimizing functions can be viewed as the correct way of interpreting the solutions to the ∞-Laplacian.

An aside: ∞-Laplacian

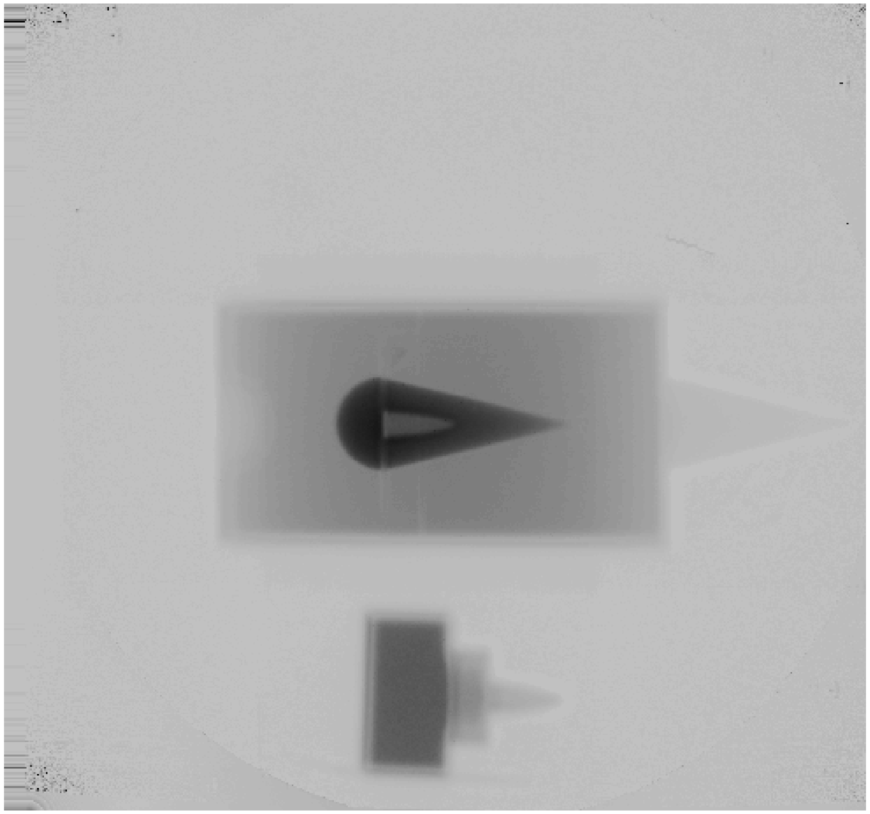


For more details see the article by Aronsson, Crandall and Juutinen in the October 2004 American Mathematical Society Bulletin and articles on L. Craig Evans website (Berkeley). Finally, it has been used to do inpainting by Caselles, Morel, and Sbert.

Cool fact: the $I_{\zeta\zeta}$ that Ron Kimmel talked about before lunch is the ∞-Laplacian! (Pointed out by Craig Evans after the talk.)

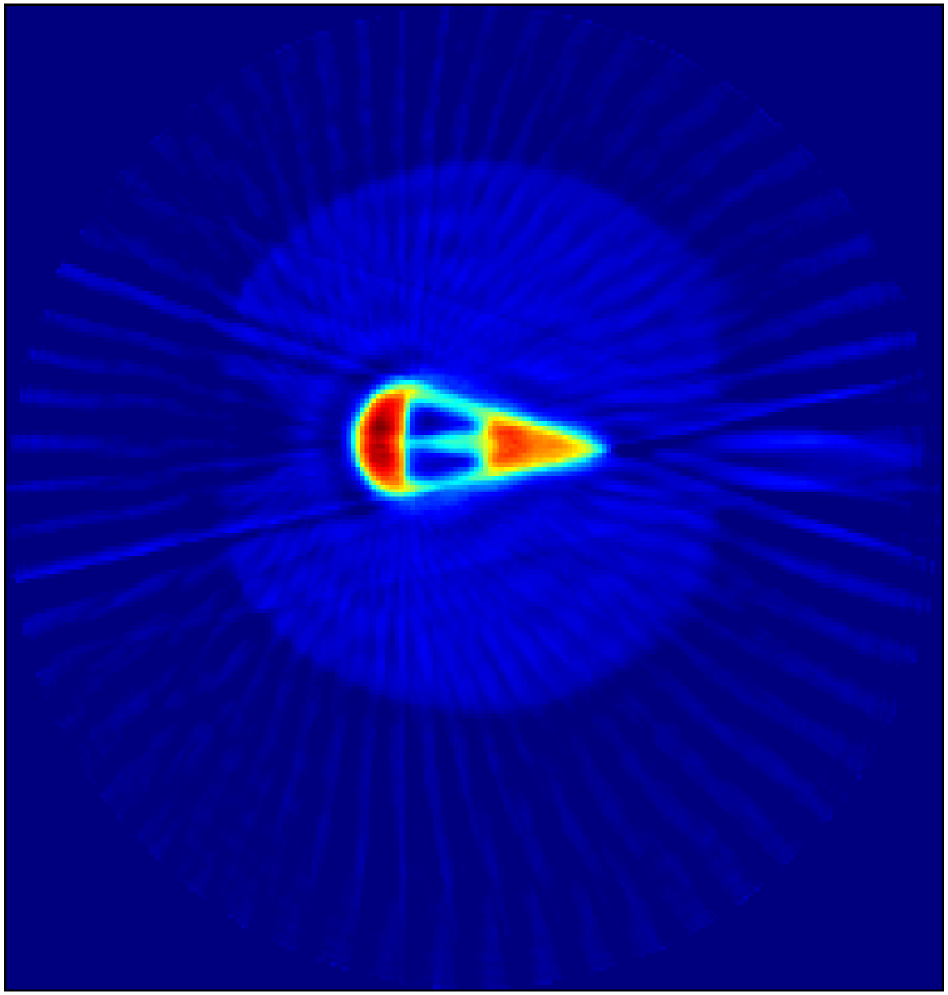
Example of L^2TV in action: BCO4

The multiple view test object: a proton radiograph from one of 30 viewing angles. The data was collected at the Los Alamos Neutron Science Center (LANSCE) in the proton radiography facility.



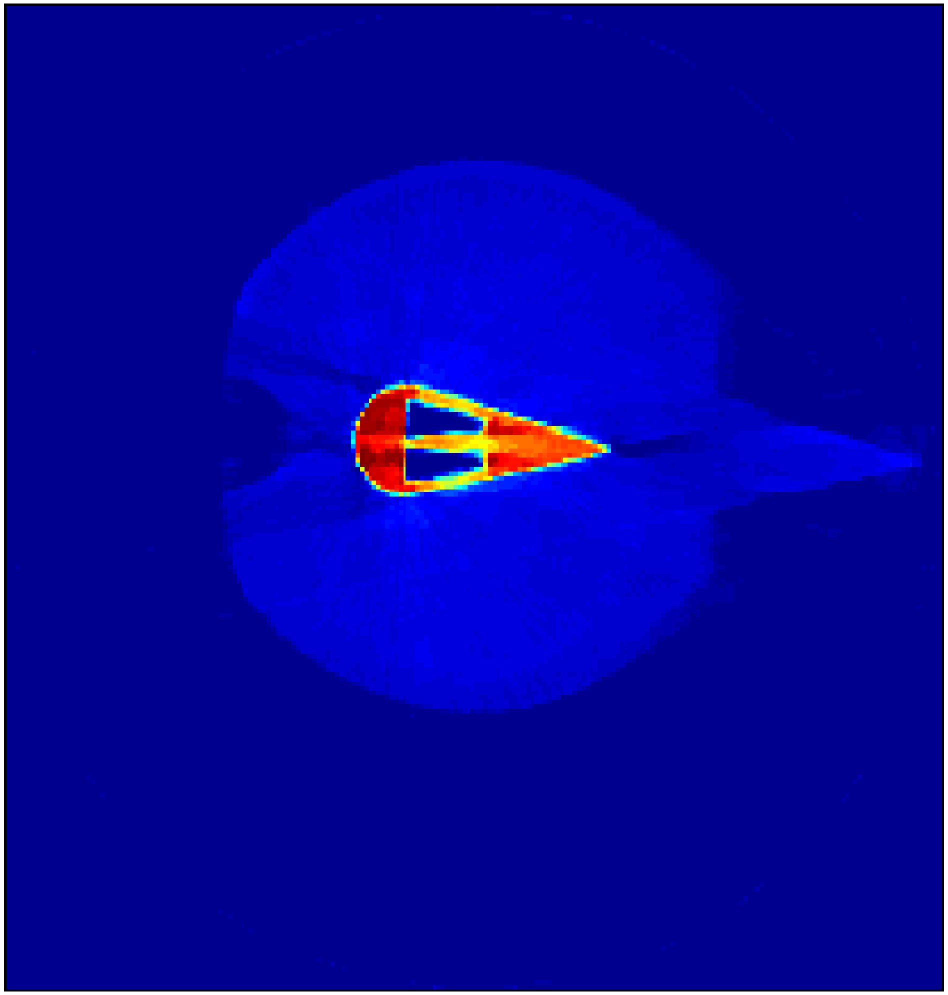
Example of L^2TV in action: BCO4

BCO4 reconstructed using SVD regularization



Example of L^2TV in action: BCO4

BCO4 reconstructed using TV regularization



Properties of L^1TV

Moving now to the TV regularized functional with L^1 data fidelity,

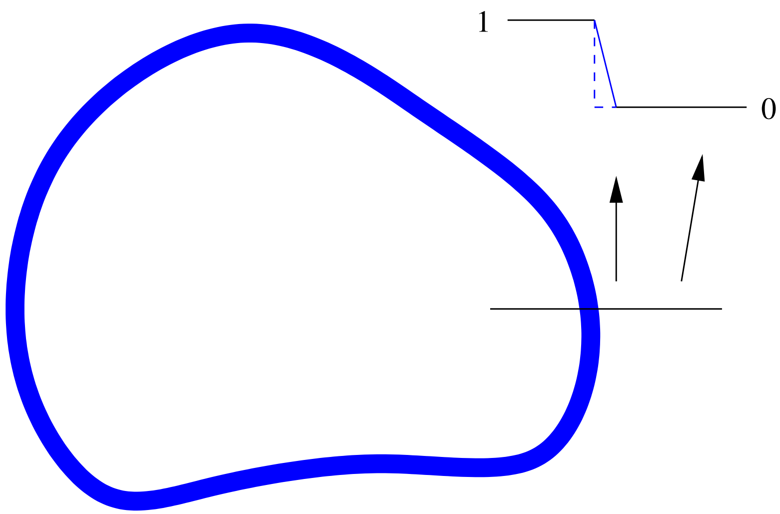
$$F(u) \equiv \int |\nabla u| dx + \lambda \int |u - d| dx \tag{27}$$

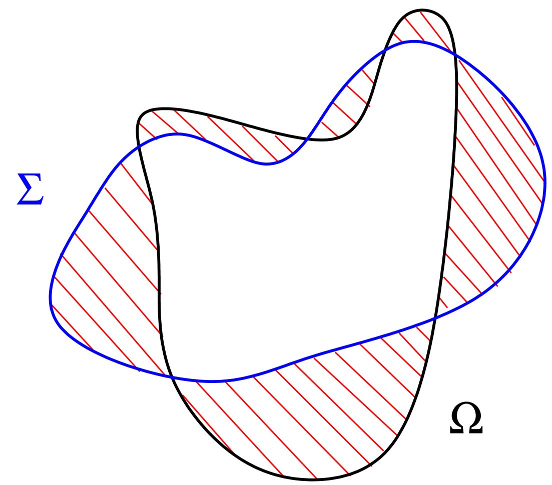
Not strictly convex: $F(u)$ is not strictly convex $\Rightarrow$ we do not have uniqueness!
Homogeneity: u is a minimizer for $d \to Cu$ is a minimizer for Cd
Existence: Since $TV(u)$ is lower semi-continuous in L^1, $F(u)$ is convex and coercive.

Properties of L^1TV

$u = \chi_\Sigma$ **and** $d = \chi_\Omega \Rightarrow F(\Sigma) \equiv F(\chi_\Sigma) = \textbf{Per}(\Sigma) + \lambda|\Sigma \Delta \Omega|$

- $u = \chi_\Sigma \to \int |\nabla u| dx =$ perimeter of Σ
- $u = \chi_\Sigma, d = \chi_\Omega \to \lambda \int |u - d| dx = \lambda \int |\chi_\Sigma - \chi_\Omega| dx = \lambda \text{Area}(\Sigma \triangle \Omega)$





Results for L^1TV

We now look at results appearing in papers by Chan and Esedoglu, Esedoglu and Vixie, and Allard.

$d = \chi_\Omega \Rightarrow u = \chi_\Sigma$, for some Σ, is a minimizer.

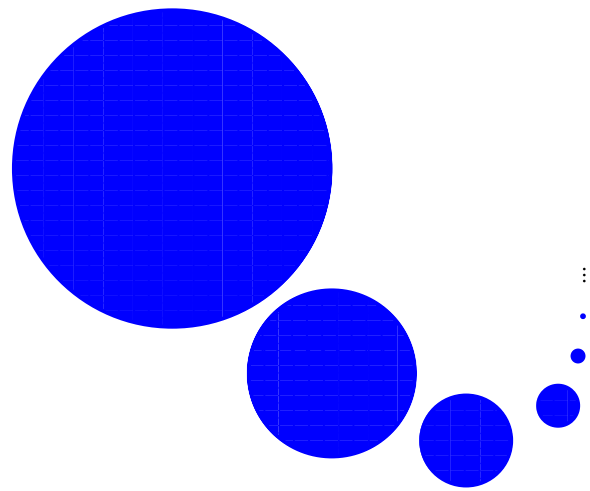
Non-convex minimizations of $F(\Sigma) = \mathrm{Per}(\Sigma) + \lambda|\Sigma\Delta\Omega|$ can be relaxed to $F(u) \equiv \int |\nabla u| dx + \lambda \int |u - \chi_\Omega| dx$ (convex). Now suppose – due to the non-uniqueness – u, the minimizer we find, is not a characteristic function? Use:

If u is any minimizer of $F_\lambda(u)$ then for almost all $\mu \in [0, 1]$, $\chi_{\{x:u>\mu\}}$ is also a minimizer of $F_\lambda(u)$,

to get a minimizer that is a characteristic function.

If $\Omega = B_{\frac{2}{\lambda}}$ then $u = \alpha \chi_{B_{\frac{2}{\lambda}}}$ is a minimizer for any $\alpha \in [0, 1]$.

One can therefore concoct Ω's whose solutions $\Sigma(\lambda)$ have, as $\lambda \to \infty$, an infinite number of non-uniqueness points ...

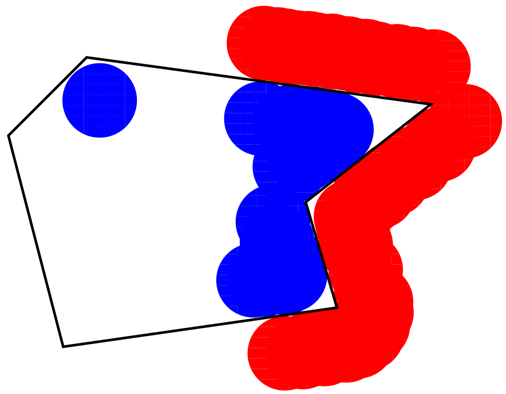


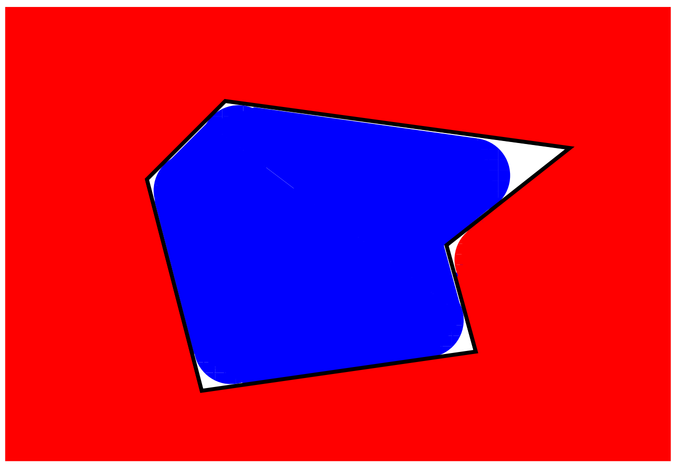
Results for L^1TV

$B_{\frac{2}{\lambda}} \subset \Omega \to \;\; B_{\frac{2}{\lambda}} \subset \Sigma$**:** Consequence – for the L^1TV functional, edges are perfectly preserved if they can be touched by $\frac{2}{\lambda}$ balls in and out.

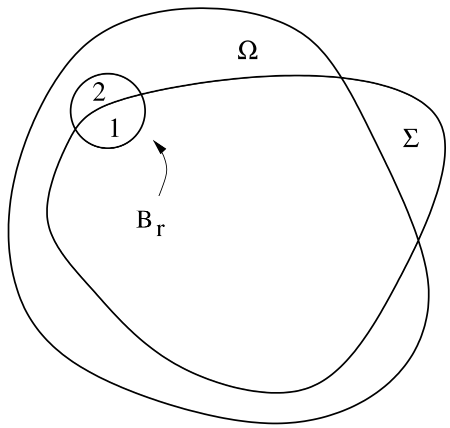
Theorem 1. *If $B_r \subset \Omega$ where $r \geq \frac{2}{\lambda}$, then $B_r \subset \Sigma$.*

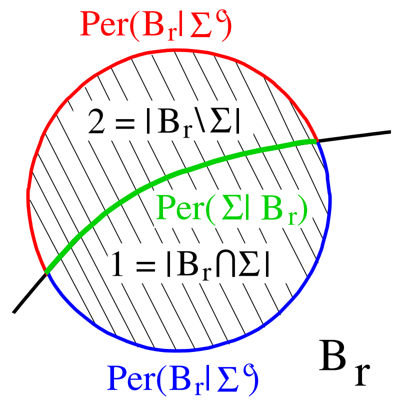
In particular, we can conclude that the boundary of Σ is in the envelope between inside and outside $\frac{2}{\lambda}$ balls.





Results for L^1TV





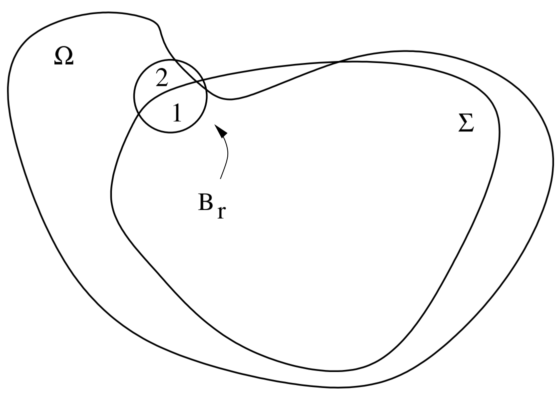
$$
\begin{aligned}
E(\Sigma \cup B_r) - E(\Sigma) &= (\mathsf{Per}(B_r) - \lambda|B_r|) + (\lambda|B_r \cap \Sigma| - \mathsf{Per}(B_r \cap \Sigma)) && (28)\\
&= \left(2\pi r - \frac{2}{R}\pi r^2\right) + \left(\frac{2}{R}\pi \rho^2 - 2\pi\rho^*\right) && (29)\\
&= 2\pi r(1 - \frac{r}{R}) + 2\pi\rho(\frac{\rho}{R} - \frac{\rho^*}{\rho}) && (30)
\end{aligned}
$$

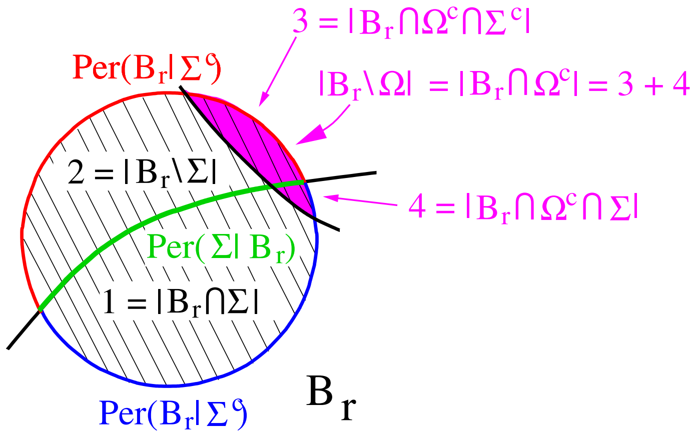
What about the case in which the $\frac{2}{\lambda}$ ball is *almost* included in Ω?

Results for L^1TV

Denoising shapes: almost $\frac{2}{\lambda}$ means slightly smaller included

Theorem 2: $B_{\frac{2}{\lambda}} \underset{\sim}{\subset} \Omega \rightarrow \; B_{\frac{2}{\lambda}-\epsilon} \subset \Sigma$





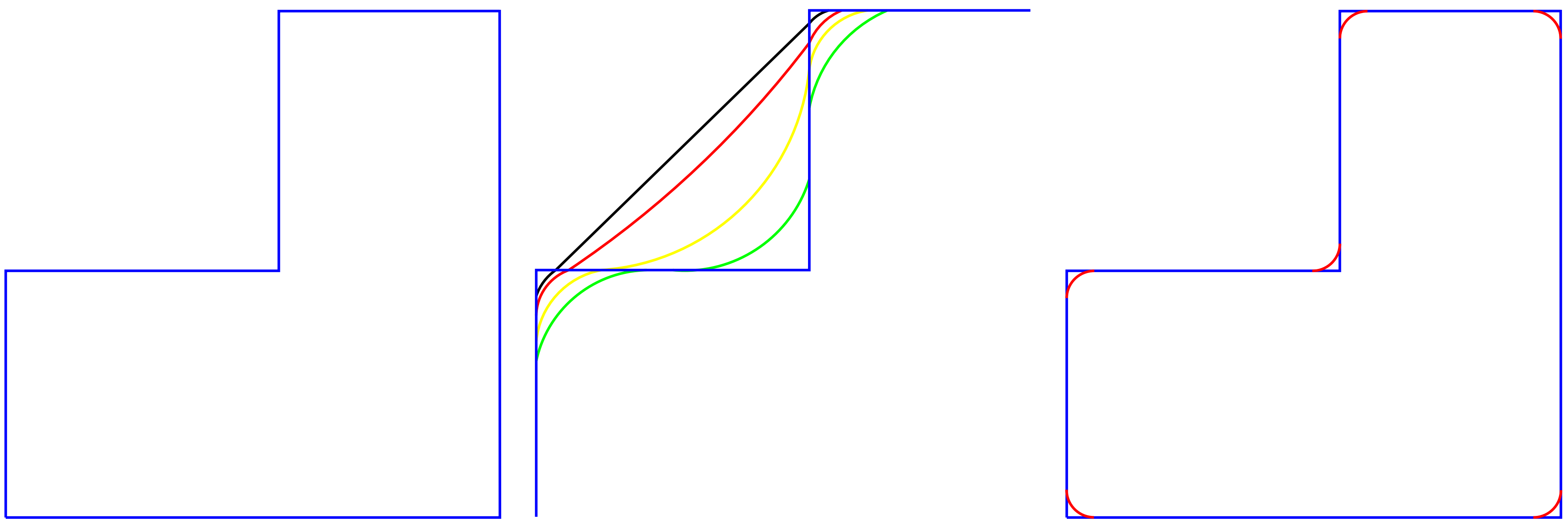
$$E(\Sigma \cup B_r) - E(\Sigma) \leq 2\pi r(1 - \frac{r}{R}) + 2\pi\rho(\frac{\rho}{R} - \frac{\rho^*}{\rho}) + 2\lambda|B_r \setminus \Omega| \qquad (31)$$

Results for L^1TV

Higher order results: Allard's formula from a first variation

$$H(x) = -\lambda\gamma'(y - d(x))\mathbf{n}_E(x) \text{ for } x \in M \tag{32}$$

where $E \equiv \{x | u(x) \geq y\}$ and $\gamma(w) = |w|$ or $|w|^2$.



Where does this handy curvature formula come from?

Results for L^1TV

Theorem: For any $y \in J$ and any $X \in \mathcal{X}(\Omega)$ we have

(I) $\int_\Omega (\frac{1}{\lambda} a_1 + b_1)(x, y)\, d||\partial[\{u \geq y\}]||x = 0$

and

(II) $\int_\Omega (\frac{1}{\lambda} a_2 + b_2)(x, y)\, d||\partial[\{u \geq y\}]||x \geq 0$

where

$P(x)$ equals orthogonal projection on $\{v \in \mathbb{R}^n : v \bullet \mathbf{n}_{\{u \geq y\}}(x) = 0\}$

$$
\begin{aligned}
t(x) &= P(x) \circ \partial X(x) \circ P(x) \\
h(x) &= P(x)^\perp \circ \partial X(x) \circ P(x) \\
a_1(x) &= \mathbf{trace}\, t(x) \\
a_2(x) &= (\mathbf{trace}\, t(x))^2 + \mathbf{trace}\, (h(x)^* \circ h(x) - t(x)^2) \\
l(x) &= \gamma'(y - d(x)) \\
k(x) &= \gamma''(y - d(x)) \\
v(x) &= \mathbf{div}\, X(x) X(x) - \partial X(x)(X(x)) \\
w(x) &= \nabla d(x) \bullet X(x) X(x) \\
b_1(x) &= l(x) X(x) \bullet \mathbf{n}_{\{u \geq y\}}(x) \\
b_2(x) &= \big(l(x) v(x) - k(x) w(x)\big) \bullet \mathbf{n}_{\{f \geq y\}}(x).
\end{aligned}
$$

The second variation limits the size of an arc of Σ boundary to $< \pi$ radians!

Summary of Tutorial

There is much more to say of course, but I will stop here. After all:

The secret to wearying consists in saying everything

Voltaire

Metrics: difference measures between images should be designed with the goal in mind. In particular what is needed are metrics which ignore unimportant differences. These metrics can be learned – at least in part – from data.

Regularization: Using what we know or are willing to assume – prior models – we can improve results of many image analysis methods. In fact, this regularization of solutions is critical for any sort of sensible answer in many cases.

Geometric Analysis: the power of analysis and geometric analysis for the purposes of innovation and insights in image analysis should not be underestimated.

Caution: Image analysis and processing $\neq$ mathematics. Anyone who wants to do the best work should be able to wear an engineer's or scientist's hat and be willing to sustain a connection to real data and real problems. The mathematics generated will be greatly enriched by such a perspective and practice.

References

1 Books on image analysis: “Mathematical Problems in Image Processing” (2002) by Aubert and Kornprobst; “Geometric Partial Differential Equations” (2001) by Sapiro; “Variational methods in Image Segmentation” (1995) by Morel and Solimini.

2 Books on geometric analysis: “Measure Theory and Fine Properties of Functions” (1999) by Evans and Gariepy; “Geometry of Sets and Measures in Euclidean Spaces” (1995) by Mattila; “Convex Analysis and Variational Problems” (1999) by Ekeland and Temam; “Introduction to the Calculus of Variations” (2004) Dacorogna; “Partial Differential Equations” (1998) Evans; Zeidler’s behemoth work on “Nonlinear Functional Analysis” – see especially vol. III.

3 Review Papers: “Inverse problems in image processing and image segmentation: some mathematical and numerical aspects” (2000) by Chambolle; “Variational PDE models in image processing” (2002) by Chan, Shen and Vese.

4 For more details on the tangent approximation CMODI method see: `http://ddma.lanl.gov/public/publications/fraser-2003-classification.shtml`.

5 Esedoglu and Chan, “Aspects of total variation regularized L^1 function approximation”, which can be downloaded from `http://www.math.ucla.edu/~esedoglu/papers.htm`. See references for previous works by Alliney and Nikolova.

6 Allard’s talk and preprint can be found here: `http://www.math.duke.edu/~wka/`.

7 Levine et al.’s papers on the $|\nabla u|^{P(|\nabla u|)}$ problem at: `http://www.mathcs.duq.edu/~sel/` (see Chen, Y., Levine, S. and Rao).

References

8 More information on sparse radiography methods and results at: http://ddma.lanl.gov/public/publications/. The Esedoglu and Vixie paper will also soon appear here as a preprint. Some may be interested in mildly unorthodox approach found in “Defensible Metrics and Merit Functions: Making Informative Comparisons of Computer Simulations and Experiments” (2004) by Vixie and Asaki.

9 For a large number of interesting and pertinent papers, see the CAM reports at http://www.math.ucla.edu/applied/cam/.

Notes: for online version of slides

1: These notes are meant to help those who are reading the slides and did not see the talks. In particular, I will add a few comments about details over the next week or so.

2: Page 7: On this slide, I refer to λ as a Lagrange multiplier. The reason for this is that some have considered this functional to arise from minimization of the TV seminorm under the constraint that the data fidelity residual equaled some fixed value. Otherwise, this parameter really simply a balance between the two terms.

3: Page 8: On the Abel projection operator, see papers on the TV Abel inversions at `http://ddma.lanl.gov/public/publications/`.

4: Page 15: In finite dimensions, there is no problem defining the probability densities referred to in the slide. In infinite dimensions we have to be much more careful.

Suggestive explanation: we know the volume of a ball with radius r in an n-dimensional Euclidean space is $V_n(r) \equiv \alpha(n)r^n$ where $\alpha(n)$ is the volume of the n-dimensional unit ball (1-ball). Now set $\alpha(\infty) = 1$ so that $V_\infty(1) = 1$. Taking a limit as $n \to \infty$, we expect

Notes: for online version of slides

$$V_\infty(r) = \begin{pmatrix} \infty & \text{if} & r > 1, \\ 1 & \text{if} & r = 1 \\ 0 & \text{if} & r < 1 \end{pmatrix} \tag{33}$$

This suggests that as $n \to \infty$, a uniform distribution will have all of it's mass or measure concentrated on the surface of the closed ball. While this is an accurate intuitive picture for the phenomena of *concentration of measure*, it also tells us that we will have a problem defining something as innocuous as the uniform distribution on the closed 1-ball.

In more detail: Let's try to define the uniform distribution on the closed 1-ball in a separable infinite dimensional Hilbert space. Let the probability of the closed ball of radius $0 < \gamma < \frac{1}{4}$ (γ-ball) be $\epsilon > 0$. Let $B_i = B((1-\gamma)e_i, \gamma) \equiv$ the γ-ball centered at $(1-\gamma)e_i$), where the e_i are the orthonormal unit vectors. The B_i are a countably infinite family of disjoint, closed γ-balls contained in the closed 1-ball. Since we want the measure of the 1-ball to equal 1, we need $\sum_i B_i \leq 1$. This is only possible if $\epsilon = 0$. But then, by using a different countable set of closed γ-balls to cover the closed 1-ball, we see that the measure of the closed 1-ball equals 0.

Notes: for online version of slides

5: Page 18: Stability here refers to theoretical stability. There are also separate issues of numerical stability.

6: Page 70: G_δ is a smooth symmetric mollifier with support of radius δ.

7: Page 79: Define $\text{Lip}_A(f)$ to be the Lipschitz constant for f in the set A:

$$\text{Lip}_A(f) = \sup_{x,y \in A \text{ and } x \neq y} \frac{|f(x) - f(y)|}{|x - y|}. \tag{34}$$

An *absolutely minimizing function* $f : \Omega \to \mathbb{R}$ is a function that

1: continuously extends the boundary data $f(\partial\Omega)$ to the interior of Ω without increasing the Lipschitz constant: $\text{Lip}_\Omega(f) = \text{Lip}_{\partial\Omega}(f)$ and

2: for any $V \subset\subset \Omega$, $\text{Lip}_V(f) = \text{Lip}_{\partial V}(f)$.

MORE: I will add more notes ...