PCMI Lecture # 3

Metrics : Lecture 2 of 2

Note: we can (and do) use these functionals on images that are not corrupted by noise. This leads us to image decompositions in the case of ROF and L^1TV (CE) and to a decomposition and segmentation in the case of M-S.

- $\int_\Omega |\nabla u|\, dx$ for u with discontinuities $\Rightarrow$ Functions of Bounded Variation and the representation of ∇u as Radon measure + vector (direction) field σ.

  I. e. $\nabla u = \vec{\sigma}\, d\mu$, μ Radon

- $\int_\Omega |\nabla u|^2 dx$ does not work on discontinuities so, in M-S, we exclude the jump set Γ from the integral.

- $\mathcal{H}^1(\Gamma)$ is the length of the jump set Γ

$\boxed{\text{Discussion}}$

- There is another, older model

$$F(u) \equiv \int |\nabla u|^2 + \lambda \int |u-d|^2$$

but this strongly discourages discontinuities. The reason why this model is tempts is that it is all L^2... nice analysis!

The Euler-Lagrange for this model gives

$$-\Delta u + \lambda(u-d) = 0$$

at a minimizer.

- Here is how that goes:

$$F(u+\alpha v) = \int (\nabla u + \alpha \nabla v)\cdot(\nabla u + \alpha \nabla v)\, dx + \lambda \int (u+\alpha v - d)(u + \alpha v - d)\, dx$$

$$= \int (\nabla u \cdot \nabla u + 2\alpha \nabla u \cdot \nabla v + \alpha^2 \nabla v \cdot \nabla v)\, dx$$

$$+ \lambda \int u^2 + d^2 + \alpha^2 v^2 + 2\alpha u v - 2ud - 2\alpha v d \; dx$$

$$\left.\frac{d}{d\alpha}\right|_{\alpha=0} \Rightarrow 2\int \nabla u \cdot \nabla v + 2\lambda \int uv - vd\, dx = 2 \int [-\Delta u + \lambda(u-d)]\, v\, dx$$

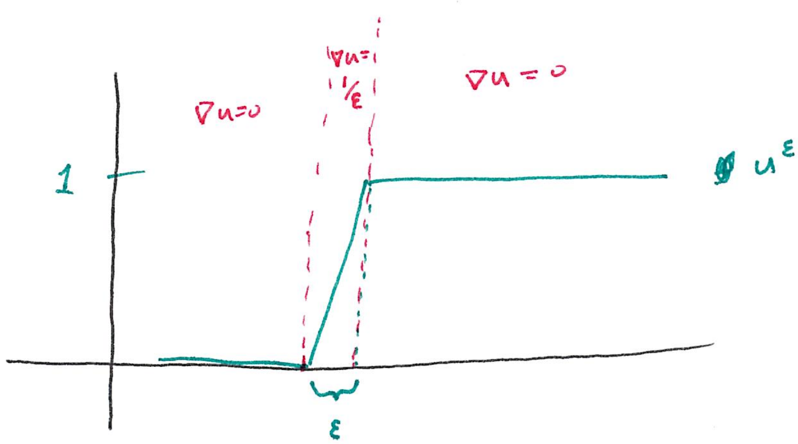
setting this $= 0$ $\rightarrow$ $\boxed{-\Delta u + \lambda(u-d) = 0}$

<u>so</u>: This model naturally leads to heat equation like behavior
$\Rightarrow$ smooths at discontinuities.

- If we measure smoothness via

$$E(u) \equiv \int |\nabla u|\, dx$$

instead we do <u>not</u> bias ourselves against edges, something which is critical for its use in image analysis.



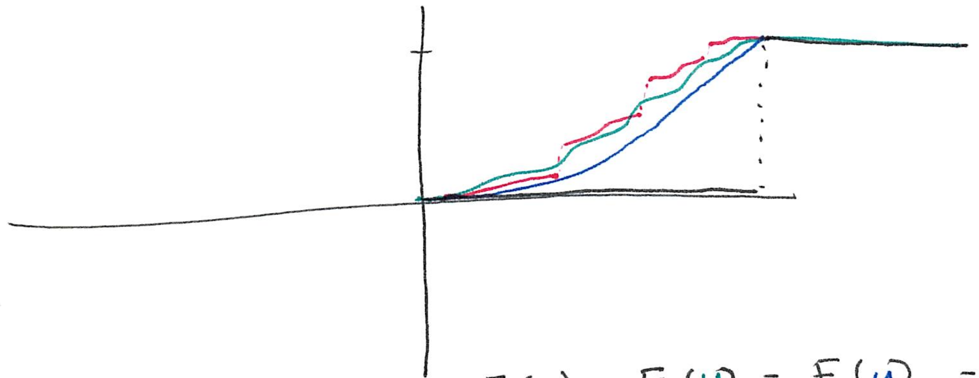
$$E_\alpha(u^\varepsilon) \equiv \int |\nabla u^\varepsilon|^\alpha dx = \left(\frac{1}{\varepsilon}\right)^\alpha \cdot \varepsilon = \varepsilon^{1-\alpha}$$

as $\varepsilon \rightarrow 0$ we get

$\alpha < 1$ $\quad E_\alpha(u^\varepsilon) \rightarrow 0$

$\alpha = 1$ $\quad E_\alpha(u^\varepsilon) = 1 \;\; \forall \varepsilon$

$\alpha > 1$ $\quad E_\alpha(u^\varepsilon) \rightarrow \infty$



$E(u) = E(u) = E(u) = E(u)$

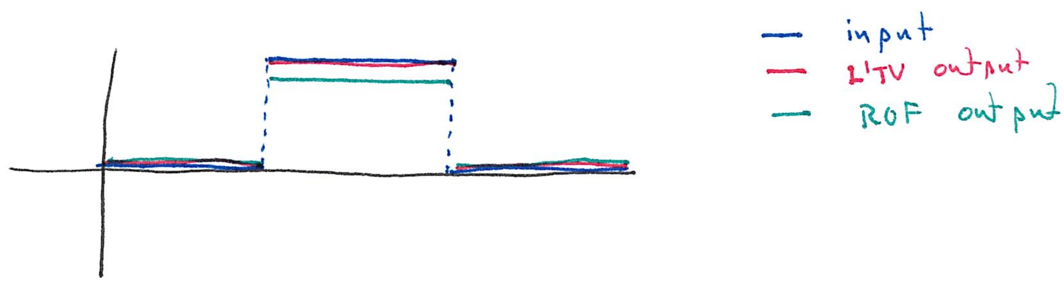
- what about the difference between

$$DF_2(u) \equiv \int |u-d|^2\, dx$$

$$DF_1(u) \equiv \int |u-d|\, dx \qquad ?$$

$DF_2(u)$ allows small perturbations to data at small quadratic cost whereas
$DF_1(u)$ is linear in the error.

This leads to (for example) contrast preservation with DF_1 and not DF_2.



- Decompositions:

Minimizers u decompose the data into

$$d = \{u\} + \{d-u\}$$

In the case of L^1TV or CE, this turns out to be the ~~key~~ important flat norm decomposition.

Starting with the work of Y. Meyer, there has been quite a bit of work on decompositions. Very briefly (we will return to this later) we look for minimizers u of

$$\min_{u} F(\alpha, u, f) = \alpha E_1(u) + (1-\alpha) E_2(f)$$

$$d = u + f$$

$$0 \leq \alpha \leq 1$$

Examples are the cartoon/texture decompositions of Vese, Le, et al

E_1 & E_2 are norms in some carefully chosen function space.

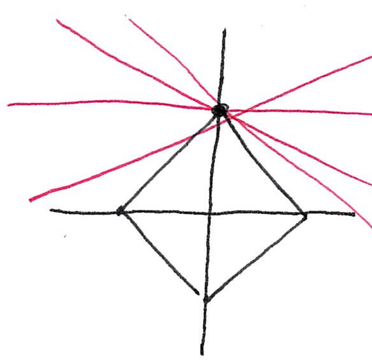
Example: Y. Meyer

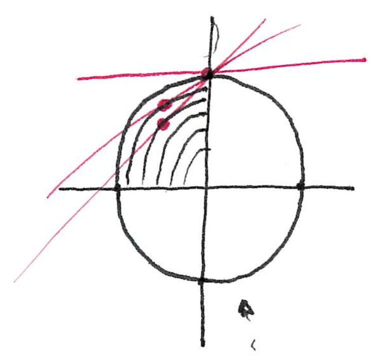
$$E_1(u) = \int |\nabla u|$$

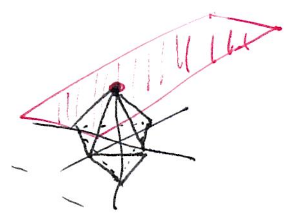
$$E_2(f) = \|f\|_* = \inf\{\|\vec{g}\|_\infty \mid f = \nabla \cdot \vec{g},\ \vec{g} \in L^\infty\}$$

Two Examples illustrating the difference between L^1 & L^2.

$\boxed{\text{Compressive Sensing}}$







$Ax = b$, A measurement Matrix $K \times M$

b measurements K measurements

x sparse signal n non-zero elements

$x^* \equiv \min \|x\|_1$ subject to $\boxed{Ax = b}$ typically works $K \geq 3n$

$x^* \equiv \min \|x\|_2$ subject to $\boxed{Ax = b}$ Essentially never works

p-laplacian

$$E(u) = \int_\Omega |\nabla u|^p \, dx$$

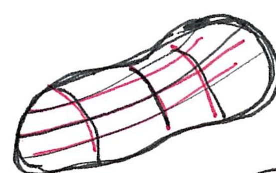
$$\Downarrow$$

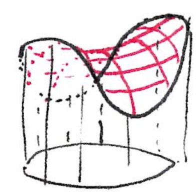
$$EL \Rightarrow -p \, \nabla \cdot \left(|\nabla u|^{p-2} \nabla u \right) \underbrace{= 0}_{\text{stationary condition}}$$

$p = 2 \longrightarrow -\Delta u = 0$ or $\Delta u = 0$

$p = 1 \longrightarrow \nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) = 0 \longrightarrow$ minimal surface equation i.e. level sets are stationary, the mean curvature is zero.

$p = 2$ Harmonic (2-Harmonic)

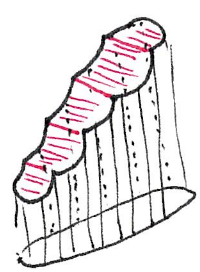




$\Delta u = 0$ on Ω

$u = f$ on $\partial\Omega$

generates smooth solutions on u, irregardless of regularity of f.



$p = 1$ 1-Harmonic

$\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) = 0$ on Ω

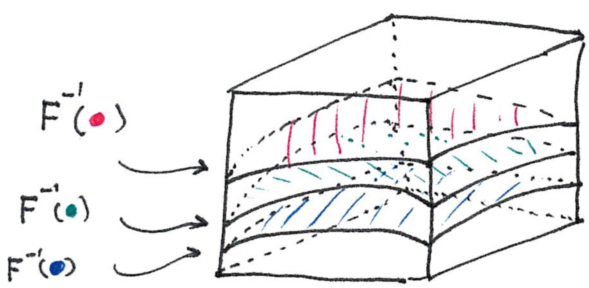
$\Downarrow$

level sets are straight lines.

u in Ω has regularity of f on $\partial\Omega$.

Merit Function Approach to Metric Generation

$x \in \mathbb{R}^n$ $\qquad$ $n > m$ $\qquad$ $y \in \mathbb{R}^m$



F

merit function

...

we make inferences based on $y = F(x)$ and we have some metric M_y on $\mathbb{R}^m$ that tells us what are large variations and not.

We can pull back the metric M to $\mathbb{R}^n$

$$M_x \equiv DF^T M_y DF$$

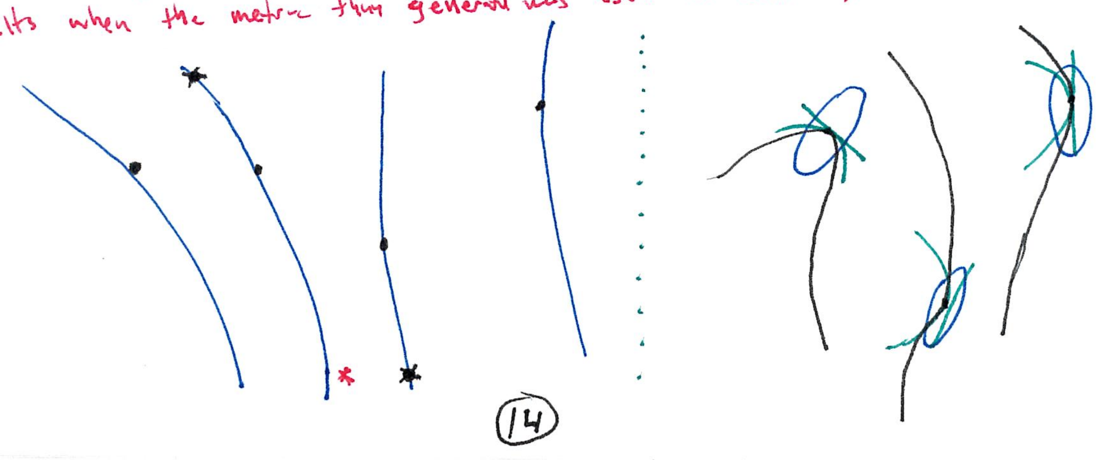
and regularize (if we need to) to get

$$M_x^{\varepsilon} \equiv DF^T \cdot M \cdot DF + \varepsilon I$$

where, as $\varepsilon \to 0$, we recover M_x.

This leads to a singular metric field (nonnegative, symmetric), and this leads to interesting questions (on the pure end of things).

In the case of the face recognition problem, we didn't even know how to compute F (that is the whole problem) but we knew something about the structure of the level sets of F. We used this to build a metric in $\mathbb{R}^n$ that was adapted to this knowledge (differences along $F^{-1}(y)$ were not ~~counted~~ counted as important as differences along the orthogonal directions). This gave us better face recognition results when the metric thus generated was used to classify an unknown face.

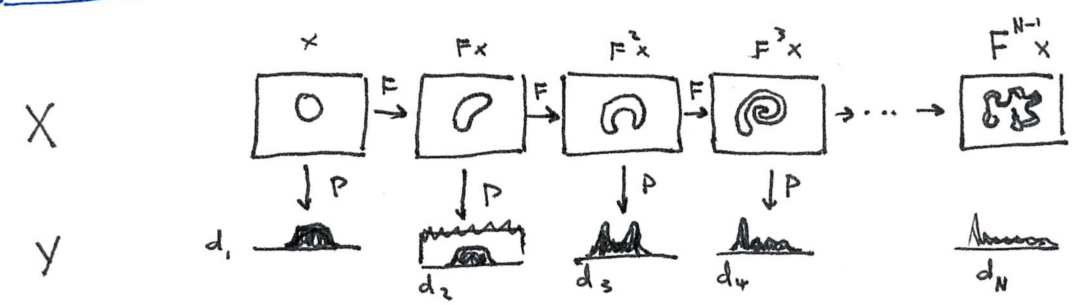


comments on the merit function approach

- in general, we want the metric to pay attention to what is important and ignore what is not. Alternatively, one can say that the metric encodes the importance of variations... for better or worse (worse if the metric is not chosen carefully!)
- This approach is not always practical, but it always is theoretically informative.
- It has not been very well explored - anybody interested in collaborating on it?
- The basic idea of pulling back metrics / uncertainty balls is fundamental to progress on uncertainty propagation.

The idea of pulling back metrics shows up (or at least lurks) all over the place.

"Curve Evolve"



$$x^* \equiv \underset{x}{\operatorname{argmin}} \sum_i \| d_i - PF^{i-1}x \|$$

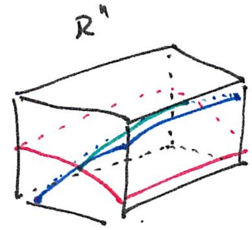
we really want the minimum to be zero and the size of the ~~singular~~ minimizing set to be a single point.

Geometric Reason why we expect N big enough to yield true x

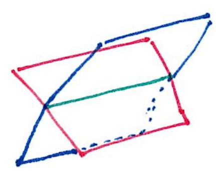
Answer: Transverse intersections are generic.

Another ("Equivalent") Answer: Dimension of the null space of ${}_m[A]^n$ $n \geq m$ is $n-m$ for typical matrices.

Suppose Y is m dimensional and X is n dimensional $n > m$. Then the dimension of $P^{-1}(y)$ is typically $n-m$. The dimension of $(PF)^{-1}(y)$ is also typically $n-m$, and is a different subset of X



• = • ∩ •



dim − dim + dim = n

⇓

dim = dim + dim − n

⇒ intersection of two sets of dimension $n-m$ will typically have dimension $n-2m$

...

after k intersections we will have an intersection of $n-km$. Now just choose $k > \frac{n}{m}$. This is the N we need.

In the linear case

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} P \end{bmatrix} \end{array}$$

$$\downarrow$$

$$\begin{array}{cc} & n \\ 2m & \begin{bmatrix} P \\ \hline PF \end{bmatrix} \end{array}$$

$$\downarrow \quad n$$

$$\begin{array}{cc} 3m & \begin{bmatrix} P \\ \hline PF \\ \hline PF^2 \end{bmatrix} \end{array}$$

$$\downarrow$$

$$\vdots$$

$\downarrow$ dim of null set decreases in steps of m

We are choosing x based on a metric we pull back. That is we evolve x in X according to

$$\sum_{i=1}^{N} \| d_i - PF^{i-1}x \|$$

and this will give us a unique solution only if N is big enough (and we are not unlucky).

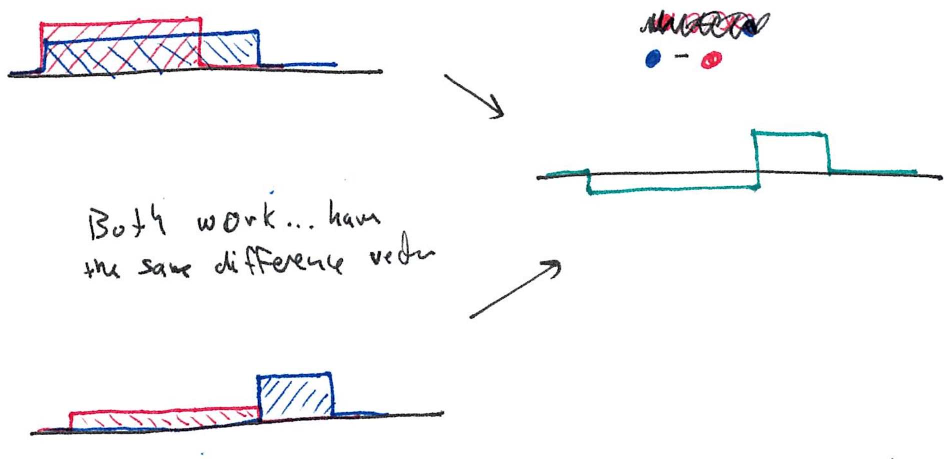
The pullback of the (Euclidean) metric on Y is simply

$$M_x = \sum_{i=1}^{N} DF^{i-1\,T} P^T P \, DF^{i-1}$$

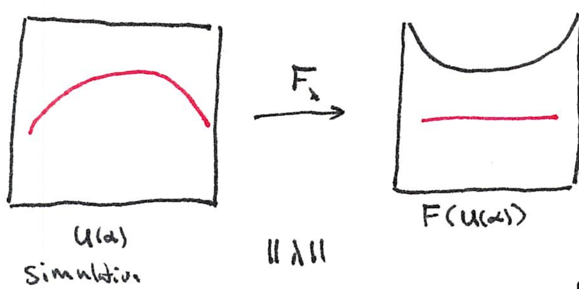
This is an example where we typically would not compute M_x.

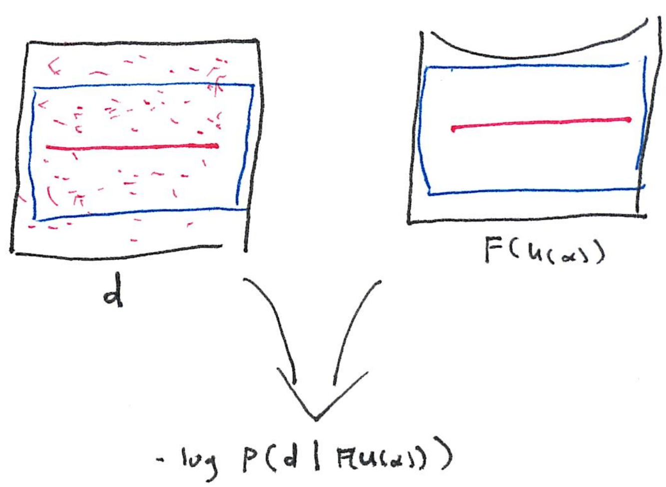
TSWARP

norms are not always what you need when measuring distances.



A norm would say the distance between these shapes on the top is the same as the distance between the shapes on the bottom. Physically this would easily not make sense.





distance between u & d given by

$$\rho(U(\alpha), d) \equiv \|\lambda\| - \mu \log P(d \mid F(u(\alpha)))$$

parameter