PCMI Lecture #2

Metrics: Lecture 1 of 2

Metrics are ubiquitous: metrics measure distance, similarity (dissimilarity)

Random examples:

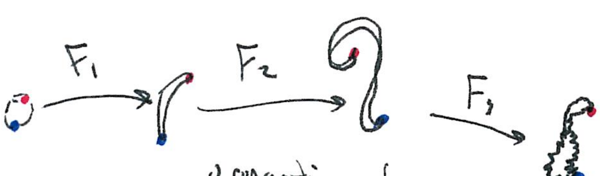
together

- how good is a prediction is measured by how close – metric used here – a match the prediction ~~with~~ to the reality.
- Image denoising: how good an image reconstruction is, is measured by how close the reconstruction is to the truth.
- The quality of a simulator is measured by how good the match is with the reality – this is just the first example restated.

together

together

- classification: patterns are classified into discrete classes & this is always done with the help of a metric. E.g. K-means clustering.
- Inference from scientific image data. This will depend on the extraction of information, with precise quantification of features by my impact. Metrics are used in the process of extraction and to measure how close the extracted features are to the real features in test cases.
- quantification of uncertainty: how big is the uncertainty? What is important and what is not is encoded into the metric.
- reconstruction from partial data: how does the uncertainty propagate through the inverse problem? How good is the reconstruction? Given a particular method, how bad can a reconstruction be? All these are measured with metrics.



propagation of uncertainty

As mathematicians we are used to define metrics:

Metric

① $\rho(x,y) \geq 0$ $=$ iff $x = y$

② $\rho(x,y) = \rho(y,x)$

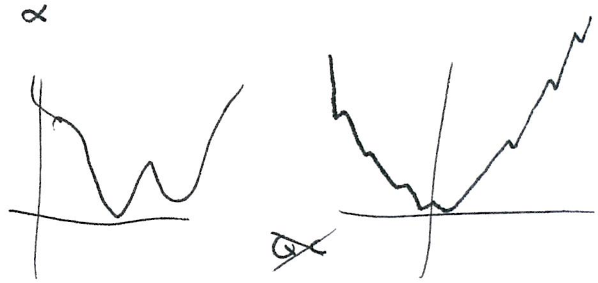
③ $\rho(x,y) + \rho(y,z) \geq \rho(x,z)$

$\Big\}$ $\{X, \rho\}$ metric space

Often, it is enough to have $\rho(x,\cdot) : \mathbb{R}^n \to \mathbb{R}$ simply be coercive and positive, or coercive and quasi-convex and positive [illegible]

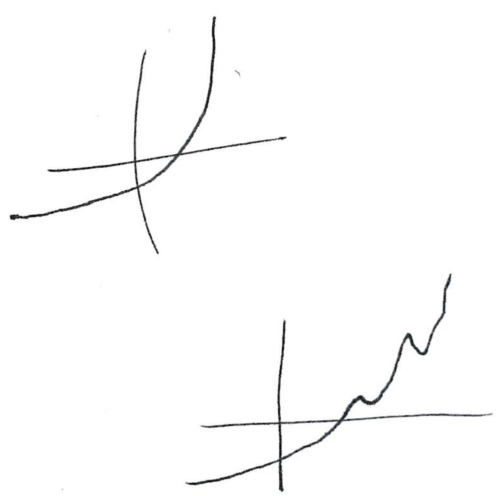
quasi-convex: $\rho^{-1}(-\infty, \alpha)$ is convex $\forall\ \alpha$

QC



~~QC~~

Coercive



non-coercive

Connection between probabalistic things and metrics

regularization parameter (some like to put in front of BV, ε) or noise parameter.

Famous Image functional: ROF (Rudin, Osher, Fatemi)

$$u^* \equiv \underset{u}{\text{argmin}}\, F(u) \equiv \underbrace{\int_\Omega |\nabla u|\, dx}_{\substack{\text{smoothness} \\ \text{BV seminorm}}} + \lambda \underbrace{\int_\Omega |u-d|^2 dx}_{\substack{\text{Data fidelity} \\ (L^2)^2}}$$

The probabalistic approach says:

choose measurement model $p(d|u)$ (choose here)
choose prior model $p(u)$

maximize $p(u|d) = \dfrac{p(d|u)\,p(u)}{p(d)}$ } Bayes

$\Downarrow$ constant

min $-\log p(u|d) = -\log p(d|u) - \log p(u) + \log p(d)$

choose:
$p(d|u) = e^{-\lambda \int |u-d|^2 dx}$ (Gaussian)
$p(u) = e^{-\int |\nabla u| dx}$
normalization ... ignore it

$\Rightarrow$ ROF!

So the regularization is just priors in a different form.

(one can argue about choosing to maximize the posterior, but it is often not a first step ... "many times" it is good enough)
(I will argue later that one should at least understand the shape at the peak, that knows how much variance there is in the family of possible solutions)

Robust statistics will use something like $e^{-\lambda \int |u-d| dx}$ to rob the outliers of their undue power. This leads to

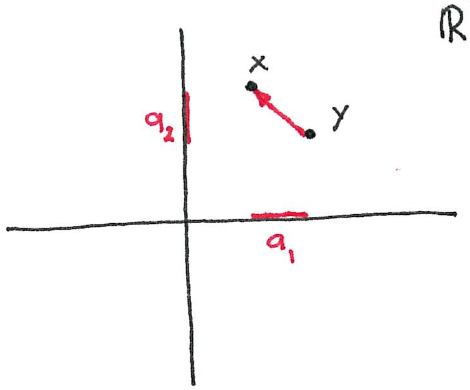
$$F(u) = \int_\Omega |\nabla u|\, dx + \lambda \int_\Omega |u-d|\, dx$$

L^1TV, TVL^1, Chan-Esedoglu which we study more later

Note: do this! $\alpha \Rightarrow p$

Metrics: A side trip, a detour.... off on a tangent

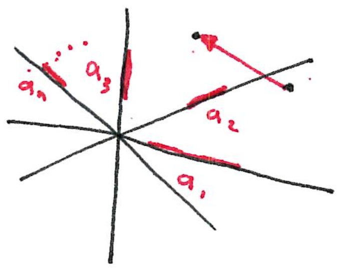
1. $\mathbb{R}^2$



$$\rho(x,y) = \begin{cases} a_1 + a_2 \\ \text{or} \\ (a_1^2 + a_2^2)^{1/2} \\ \text{or} \\ (a_1^{\alpha} + a_2^{\alpha})^{1/\alpha} \quad (\alpha \to p) \end{cases} \qquad \ell^{\alpha} \qquad 1 \le \alpha \le \infty \quad (\alpha \to p)$$

projections and $\gamma^{\alpha\, p}$

2. $\mathbb{R}^n$



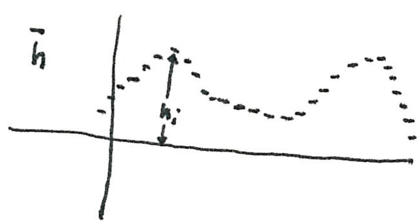
$$\rho(x,y) = \|x-y\|_{\alpha} = \left(\sum_{i=1}^{n} a_i^{\alpha}\right)^{1/\alpha}$$

projections and γ^{α}

3. $\mathbb{R}^{\infty}$

$\Downarrow$ approximate with

$\mathbb{R}^N \quad N \gg 1$

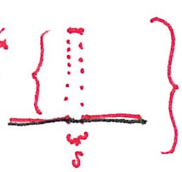


$$= \sum_{i=1}^{N} (\delta^{\frac{1}{\alpha}} h_i)\, b_i \quad \Rightarrow \text{length of } h \text{ is given by}$$

$$\left(\sum_{i=1}^{N} (\delta^{\frac{1}{\alpha}} h_i)^{\alpha}\right)^{\frac{1}{\alpha}}$$

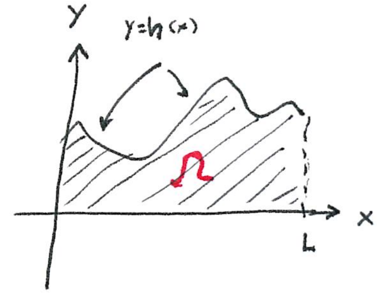
$$= \left(\sum_{i=1}^{N} \delta h_i^{\alpha}\right)^{\frac{1}{\alpha}}$$

$$= \left(\int \bar{h}^{\alpha}\, dx\right)^{\frac{1}{\alpha}}$$

$\left(\frac{1}{\delta}\right)^{\frac{1}{\alpha}}$  $b_i \qquad \int b_i^{\alpha}\, dx = 1$

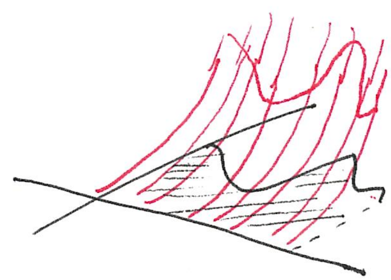
④ considering $\int h^\alpha dx$ more carefully:

$$\int_0^L h^\alpha dx = \iint_\Omega \alpha y^{\alpha-1}\, dy\, dx$$



$$= \iint_\Omega \alpha y^{\alpha-1}\, dx\, dy$$

$$= \int \alpha y^{\alpha-1} \ell(y)\, dy \qquad \ell(y) = M(x \mid f(x) \geq y)$$



$$\| h \|_\alpha = \left(\int \alpha y^{\alpha-1} \ell(y)\, dy \right)^{1/\alpha}$$

$$\| h \|_{f(y)} = \left(\int f(y)\, \ell(y)\, dy \right)^{?} \qquad f(y) \geq 0 \ \forall y$$

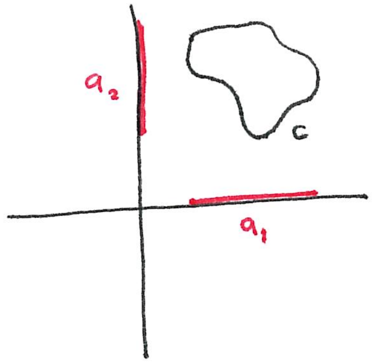
$$\underbrace{\| h \|_{f(x,y)} = \iint_\Omega f(x,y)\, dx\, dy}_{\text{generalized } L^p} \qquad \Omega \equiv \{(x,y) \mid 0 \leq y \leq f(x),\ 0 \leq x \leq L\}$$

In particular ~~diagonal~~ weighted metrics are a special case

$$L_g^\alpha = \left(\int g(x) h^\alpha \right)^{1/\alpha} \quad \ldots \text{ lots to explore}$$

(notice $(\quad)^?$ is important since $\| \cdot \|_2^2$ leads to paths of length 0 !!)

⑤ looking back at projections of $\gamma^{\cancel{x}\,r}$ and $\mathbb{R}^2$



ℓ^1 length of $c = 2a_1 + 2a_2$

$= 2(a_1 + a_2)$

$=$ sum of projections counting multiplicities

$$= \int \Pi_\theta \, d\eta(\theta)$$

η is concentrated on

$\theta = 0 \quad \theta = 90^\circ = \frac{\pi}{2}$

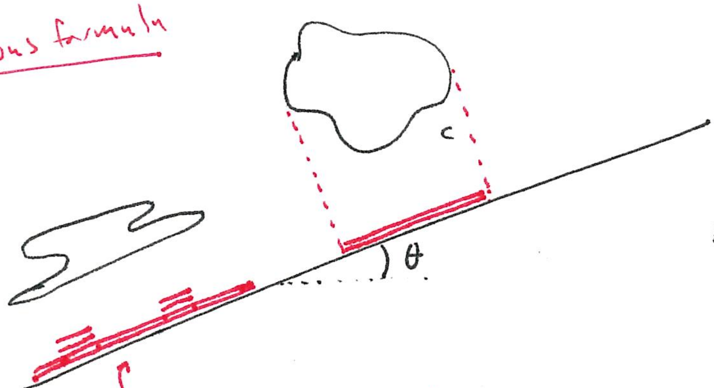
$$\int d\eta(\theta) = 1 + 1 = 2$$

<u>But</u> ... we usually want to know the Euclidean length of C.

(Notice that / & □ have equal lengths)

(Notice that ℓ^1 length is <u>not</u> rotationally invariant)

<u>Croftons formula</u>



$\Pi_\theta =$ sum of projected lengths

$=$ projection <u>counting</u> multiplicity

Euclidean length of C

$= \ell^2(C)$ ⊗

$$= \int_0^T \|\dot{c}(t)\|_2 \, dt$$

$$= \frac{1}{2} \int_0^\pi \Pi_\theta(C) \, d\theta$$

to get this look at the circle

(6) Finally one can play around in $\mathbb{R}^n$

$$\rho(x,y) = g\left(f_1(|x_1-y_1|), f_2(|x_2-y_2|), \ldots, f_n(|x_n-y_n|)\right)$$

... what must g & f satisfy to get coercivity, quasi-convexity ...

End of detour (and we didn't even get to distances like Gromov-Hausdorff Ü)

Certainly, the L^p metrics are used a great deal, but there is no reason we should stick only with them. But even the L^p used creatively give us much to think about.

Metrics are typically chosen as a compromise between what models the situation well and what leads to ease of computation or calculation.

$\boxed{L^1, L^2, L^\infty}$

It is often useful to consider the easy case L^2 as well as the the boundary cases, L^1 and L^∞.

$f: \mathbb{R}^n \to \mathbb{R}^m$ $\quad n = 1, 2, 3, 4 \quad$ $m = 1$ (grayscale) or sometimes 3 (color) or even say 1000 (hyperspectral)

We will concentrate on $m = 1$ in this course. (i.e. row & column vectors coming from "same space")

using the usual identification of $\mathbb{R}^n$ with its dual space

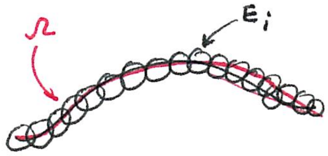
$$\nabla f: \mathbb{R}^n \to \mathbb{R}^n$$

$$\int |f|^p \, dx, \quad \int |\nabla f|^p \, dx \quad : \quad \|f\|_p^p, \ \|\nabla f\|_p^p$$

$$\int |f-d|^p \, dx, \quad \int_\Omega 1 \, dx, \quad \int \chi_\Omega \, dx \quad : \quad \|f-d\|_p^p, \ \mathcal{L}^n(\Omega), \ \mathcal{L}^n(\Omega)$$

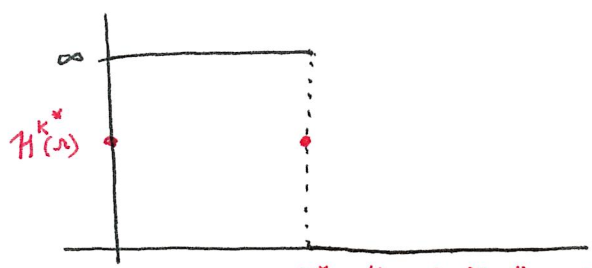
What if Ω is a curve in $\mathbb{R}^2$? Answer: $\mathcal{L}^n(\Omega) = 0$

$\boxed{\mathcal{H}^k}$



$$\Omega \subset \bigcup_{E_i \in \mathcal{F}} E_i$$

$$\mathcal{H}^k \equiv \sup_{\delta > 0} \left(\inf_{\substack{\mathcal{F} \\ \Omega \subset \bigcup E_i \\ E_i \in \mathcal{F} \\ \mathrm{diam}(E_i) \le \delta}} \left(\sum_i \alpha(k) \left(\frac{\mathrm{diam}(E_i)}{2} \right)^k \right) \right)$$



k^* = Hausdorff dimension of Ω

note that $0 \le \mathcal{H}^{k^*}(\Omega) \le \infty$

Examples of uses of these metrics

(Lectures 5, 9 & 12)

ROF: $$F(u) \equiv \int_\Omega |\nabla u|\, dx + \lambda \int_\Omega |u-d|^2\, dx$$

L^1TV: $$F(u) \equiv \int_\Omega |\nabla u|\, dx + \lambda \int_\Omega |u-d|\, dx$$

M-S: $$F(u,\Gamma) \equiv \int_{\Omega \setminus \Gamma} |\nabla u|^2\, dx + \lambda \int_\Omega |u-d|^2\, dx + \mu \int_\Gamma d\mathcal{H}^1$$

$$= \int_{\Omega \setminus \Gamma} |\nabla u|^2\, dx + \lambda \int_\Omega |u-d|^2\, dx + \mathcal{H}^1(\Gamma)$$

Since images naturally include discontinuities, we have to know what $\int |\nabla u|\, dx$ means for images. We will return to this in detail later, but we look at these three in a bit of detail now, especially from the perspective of what these different metrics imply for minimizers.