PCMI Lecture #7

Classification Modulo Invariance (CMODI) and friends

In this lecture, I will describe some work I did with collaborators on invariant recognition.

detection = finding all the faces or boats or cars or trees in an image

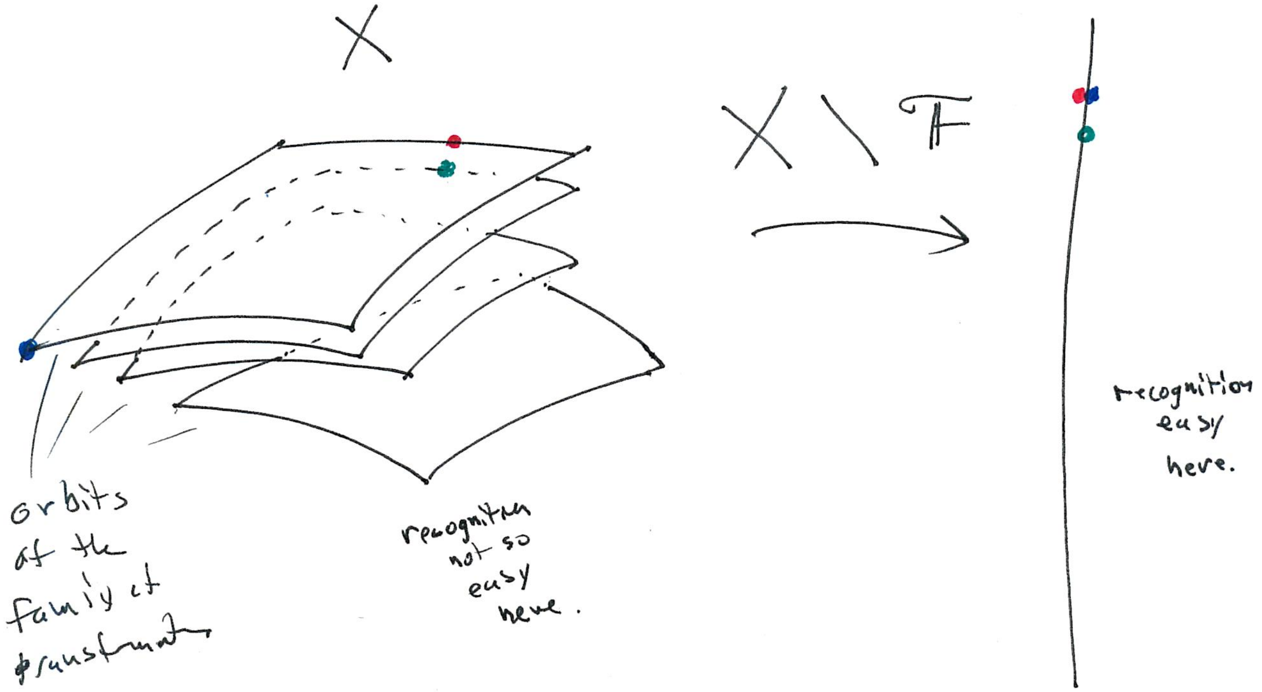
recognition = finding the face of John Doe III in an image

invariant recognition = finding John Doe III in the image irregardless of what pose he has, what lighting conditions prevail, what partial occlusions prevail.

Another Example: suppose you are in a big bay and want to be able to detect "bad" underwater vehicles (e.g. torpedos) by monitoring audio signals recorded using hydrophones. Suppose further that you have a library of all possible "vehicles" that you will hear using your sonar equipment. The reason you cannot simply do correlations to id. signal sources is the fact that you have - in addition to noise - doppler effects. Building algorithms that automatically factor out this effect is is what invariant recognition is all about.

Key Idea: work in the quotient space

Suppose we have a family of transformations on the signal space X mapping the ideal signal to a family of different signals which are nevertheless equivalent to the original signal. Example: for any face in an image, all other images with the same face scaled, translated, rotated are different images yet are images of the same face.



i.e. $\bigcup_{f \in \mathbb{F}} f(x^*)$

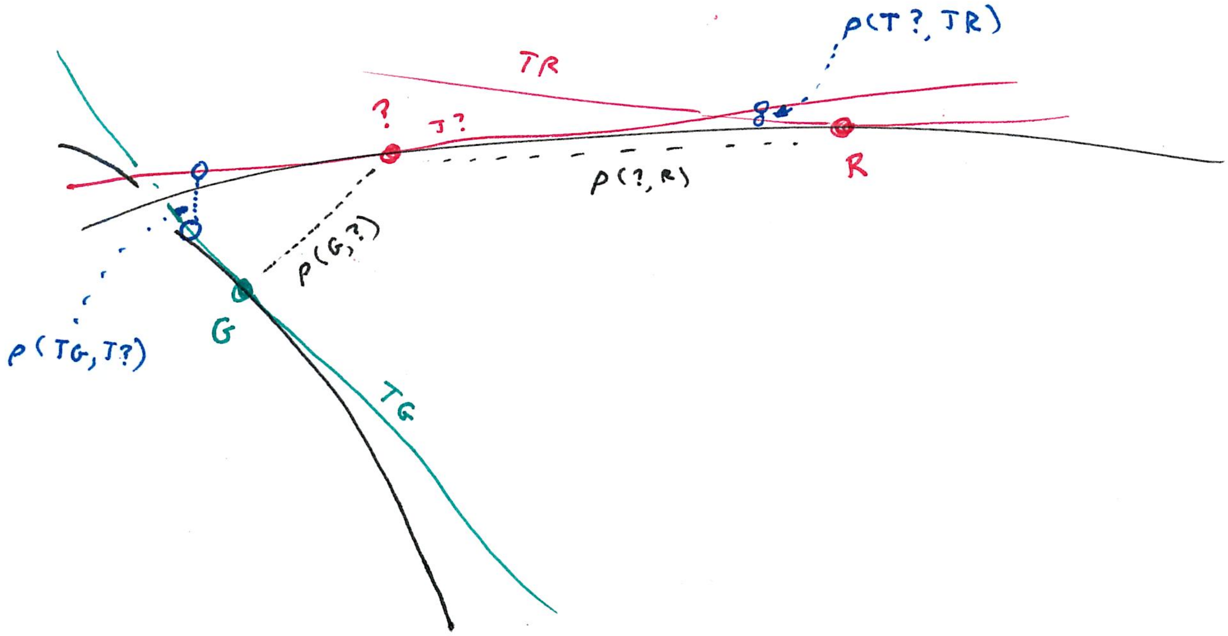
we get a different orbit or sheet for each distinct face, signal, object...

This is the ideal, but usually not computationally feasible.

Principled hack: modify the metric used in X in a spatially varying way to ignore those directions tangent to the orbits, invariant sets.

Simard, LeCun, Denker, Victorri 1998

They used the tangents at the signals and then measured distances between tangent spaces.



In X, the distance between ? and R exceeds the distance between G and ? tempting you to label ? as another G. i.e.

$$\rho(G,?) < \rho(?,R)$$

But, if we instead approximate the invariant sets with the tangent spaces TG, T?, TR, then we can compute the distances between those objects and get the correct classification

$$\rho(TG,T?) > \rho(T?,TR).$$

Simard et al. used this idea to get better results on handwritten digit recognition problems.

(one can imagine various regularizations of this e.g. using segments to control errors introduced by {tangent space ≠ invariant set} errors.)

Fraser, Hengartner, Vixie, Wohlberg 2003:

we thought we could improve recognition of faces (and we were using faces as an example application, not as a final goal) by improving on how we approximated the invariant ~~your~~ sets.

Here is what we did, in some detail:

we considered the relatively simple case of recognition of faces when there might be scaling, translation, and rotation generating differences between otherwise identical face images

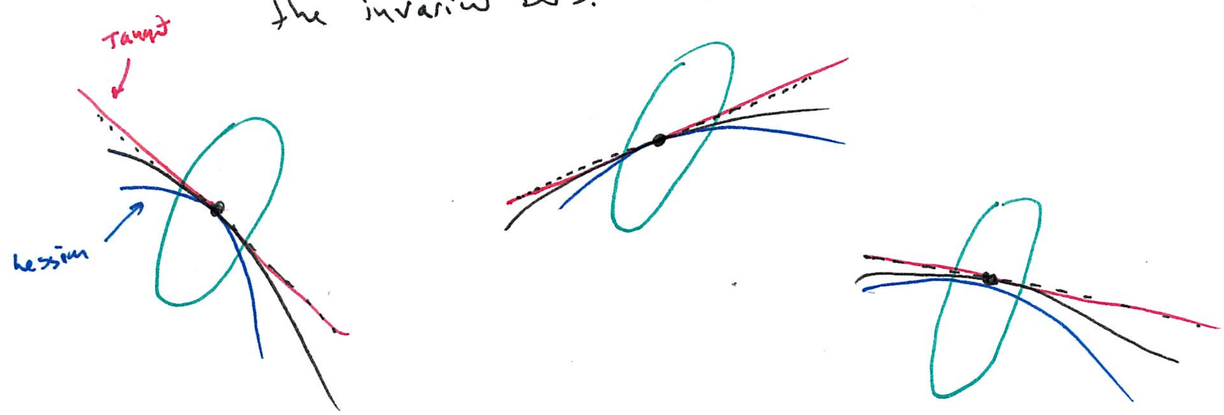
we generated the "within class" covariance matrix using all the labeled ~~[illegible]~~ training faces images. This gave us an estimate of what kind of variations are important and what variations ~~[illegible]~~ were not important

$$C_w$$

The Mahalanobis distance thus generated

$$\sqrt{(x-\bar{x}_i)^T C_w^{-1} (x-\bar{x}_i)}$$

was modified using both First and Second order derivative information from the invariant sets.



This was done by letting $x - \bar{x}_i$ be longer in the directions contained in the tangent space to the invariant set ... but not too long, since then errors arising from the fact that {invariant set $\neq$ second order approx} would overwhelm you.

$$C_i = V_i C_{\theta,i} V_i^T + C_W$$

$\Rightarrow$ new Mahalanobis distance

$$\sqrt{(x - \bar{x}_i)^T C_i^{-1} (x - \bar{x}_i)}$$

comment: If $C_{\theta,i} = \alpha I$, then we can - by choosing large α, simply make deviations in the tangent directions V_i have zero length or as small as you like.

We made a careful argument to derive a $C_{\theta,i}$ that depended on the second order Hessian tensor H.

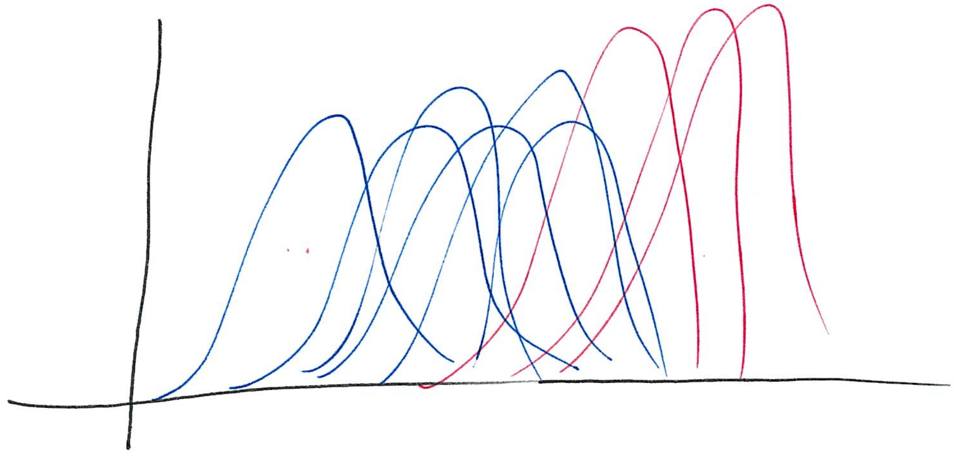
$$C_{\theta,i} \simeq \alpha H^{-1}$$

... but there are more details.

Essentially: we generated a second order tensor that weighted directions that C_W <u>did not</u> care about less heavily.

What kind of results did we get?

They were quite good. The method was compared in a very strictly controlled way to 13 other algorithms in the CSU database (at the time of the work).



These curves are the montecarlo rank 1 recognition results:

$$\{ 1 \text{ target}, \ 160 \text{ gallery} \} \times 10{,}000$$

sort distances to 160 then

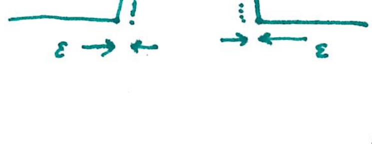
recognition rate

... do this a bunch of times and estimate Gaussian ... plot it.

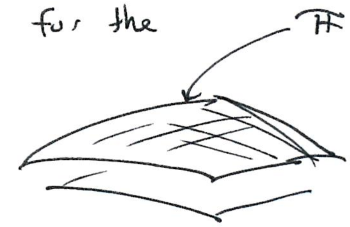
Some details: In order to compute derivatives, you need to smooth the images. This is because

The presence of edges in images leads to infinite curvature in the invariant set.

$K \sim \frac{1}{\varepsilon}$



This approach suggests another: Compute geodesic's for the metric generated by



$$C_\alpha \equiv C_W + \alpha V V^T \qquad \left(\sqrt{x^T C_\alpha^{-1} x}\right)$$

where α is large.

<u>or</u> ...

Defining $V^\perp$ = orthogonal complement of the basis for $T\mathbb{F}$, V

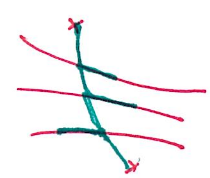
define the Reimann metric M_α

$$M_\alpha \equiv (V^\perp)(V^\perp)^T + \alpha I$$

where α small means we are ignoring the directions $v \in V$.

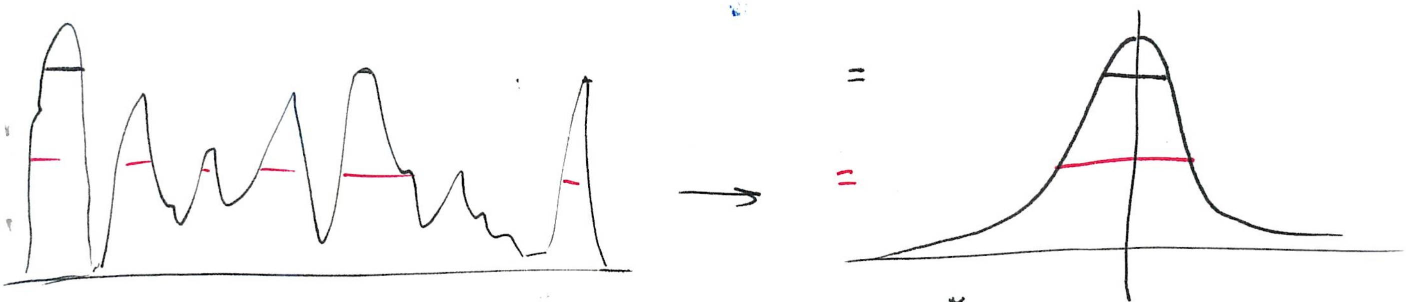
For $\alpha = 0$ we have a singular metric.

This is an idea worth pursuing, but which fell by the wayside in our work.



* one certainly <u>does not</u> have uniqueness for geodesics
* in an unbounded domain it is easy to concoct cases without existence of geodesics
* it seems there are some interesting questions to look at
* this should not be confused with what is already known as singular Reimannian geometry: allows geodesics that are tangent to some distribution, i.e., some subspace of the tangent space at each point.

A very simple example: connected to highly nontrivial analysis



f $\qquad$ f^* Symmetric decreasing Rearrangement

L^p norms only depend on the symmetric decreasing rearrangement of a function.

$$\|f\|_p = \|f^*\|_p$$

$$\|\nabla f\|_p \geq \|\nabla f^*\|_p$$

We used the SDR map $\mathcal{R}$ to solve an image based problem ... match the x-ray driven jetting data to simulations. It worked.

Interesting Tidbit: Almgren & Lieb (1989) showed that $\mathcal{R}$ is not continuous in $W^{1,p}$ when we are in $\mathbb{R}^n$ $n > 1$. This is linked to Coarea Irregular functions..

* $G_f(y) \equiv \int \chi_{\{f > y\}} \chi_{\{\nabla f = 0\}} \, d\mathcal{L}^n$ if $\frac{d}{dy} G_f(y)$ is singular wrt $\mathcal{L}^1$ on $\mathbb{R}$ then f is coarea regular

* if f is not coarea regular it is coarea irregular.