Case-study:
Non-rigid Registration

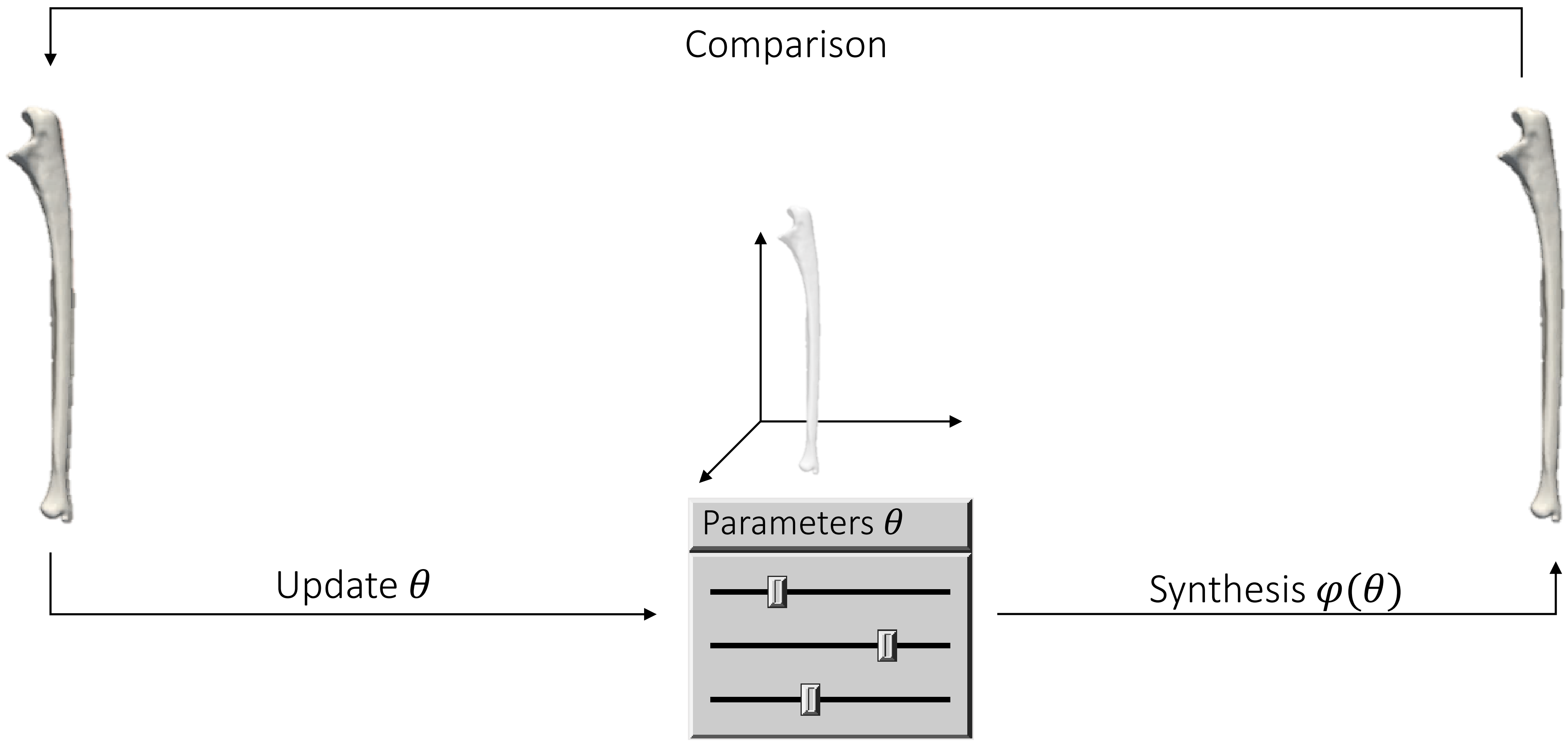
Marcel Lüthi

Graphics and Vision Research Group
Department of Mathematics and Computer Science
University of Basel

Outline

- Reminder: Analysis by Synthesis
- Non-rigid registration: The basic formulation
- A selection of useful likelihood functions
- A selection of useful priors
- Optimization

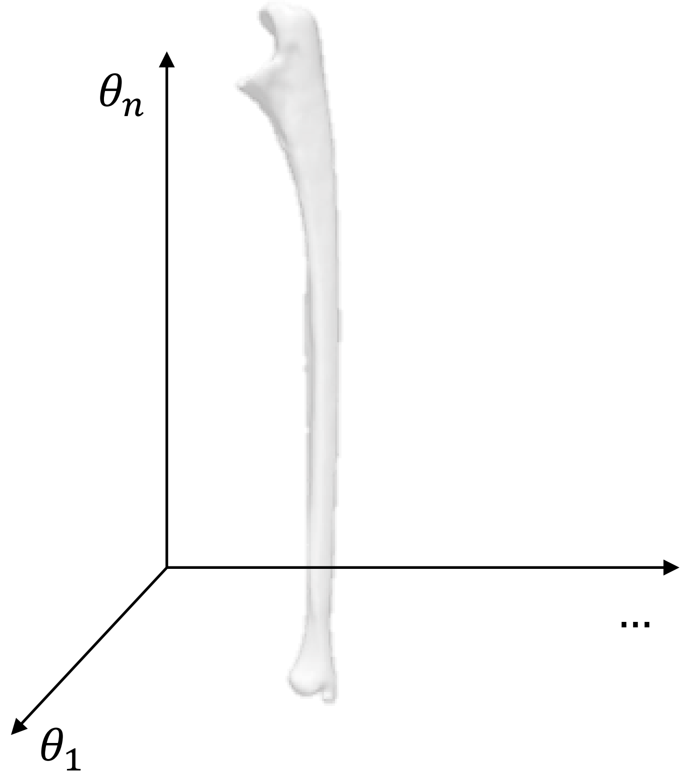
Conceptual Basis: Analysis by synthesis



Analysis by synthesis in 5 simple steps

1. Define a parametric model

- a representation of the world
- State of the world is determined by parameters $\theta = (\theta_1, \dots, \theta_n)$

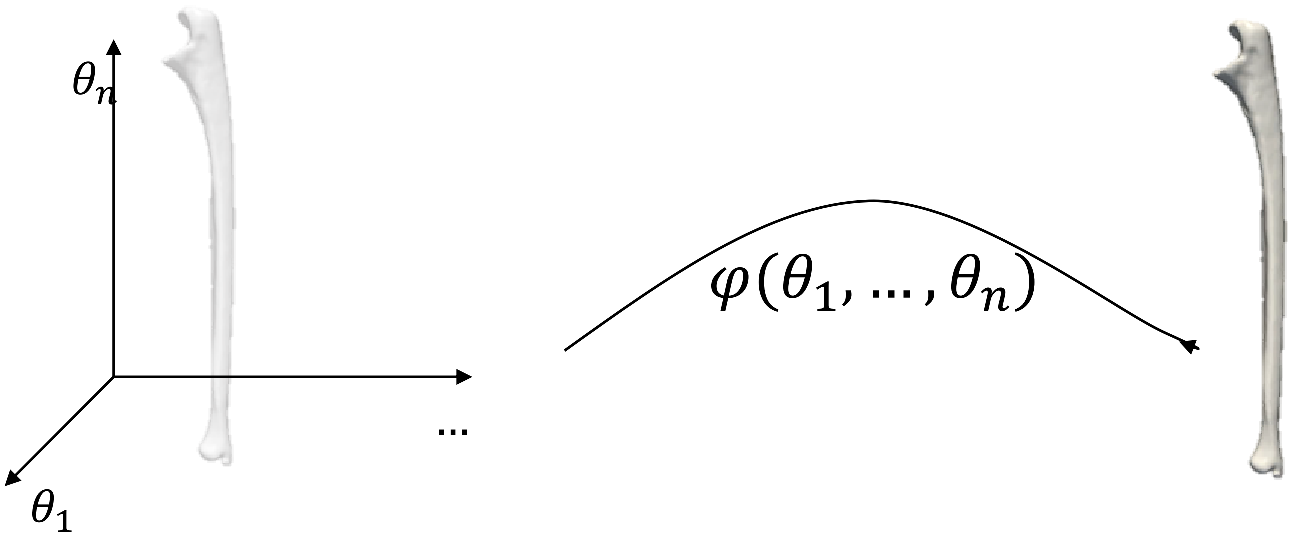


Defines what part of the world our model can explain.

Analysis by synthesis in 5 simple steps

2. Define a synthesis function $\varphi(\theta_1, \dots, \theta_n)$

- **generates/synthesize** the data given the "state of the world"
- φ can be deterministic or stochastic

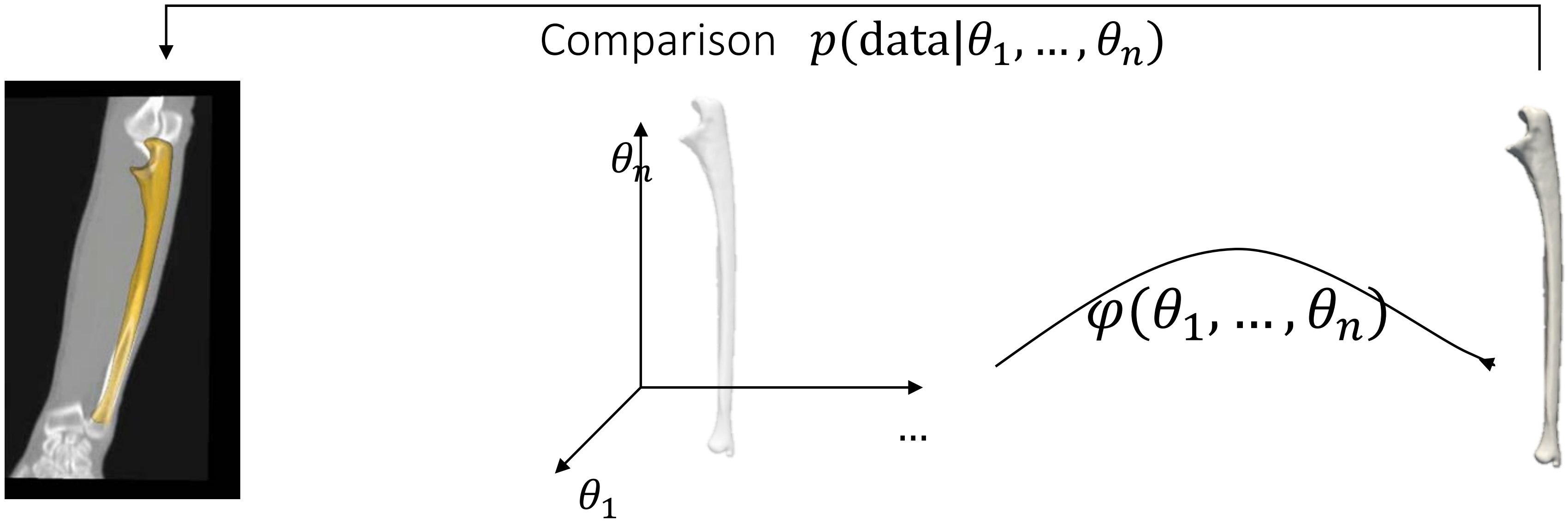


The model's view of the world.

Analysis by synthesis in 5 simple steps

3. Define likelihood function:

- Define a probabilistic model $p(\text{data}|\theta_1, \dots, \theta_n)$
- Includes stochastic factors on the data, such as noise



How well does the generated model's view correspond to the data?

Bayesian inference

We have: $P(data|\theta_1, \ldots, \theta_n)$
We want: $P(\theta_1, \ldots, \theta_n|data)$

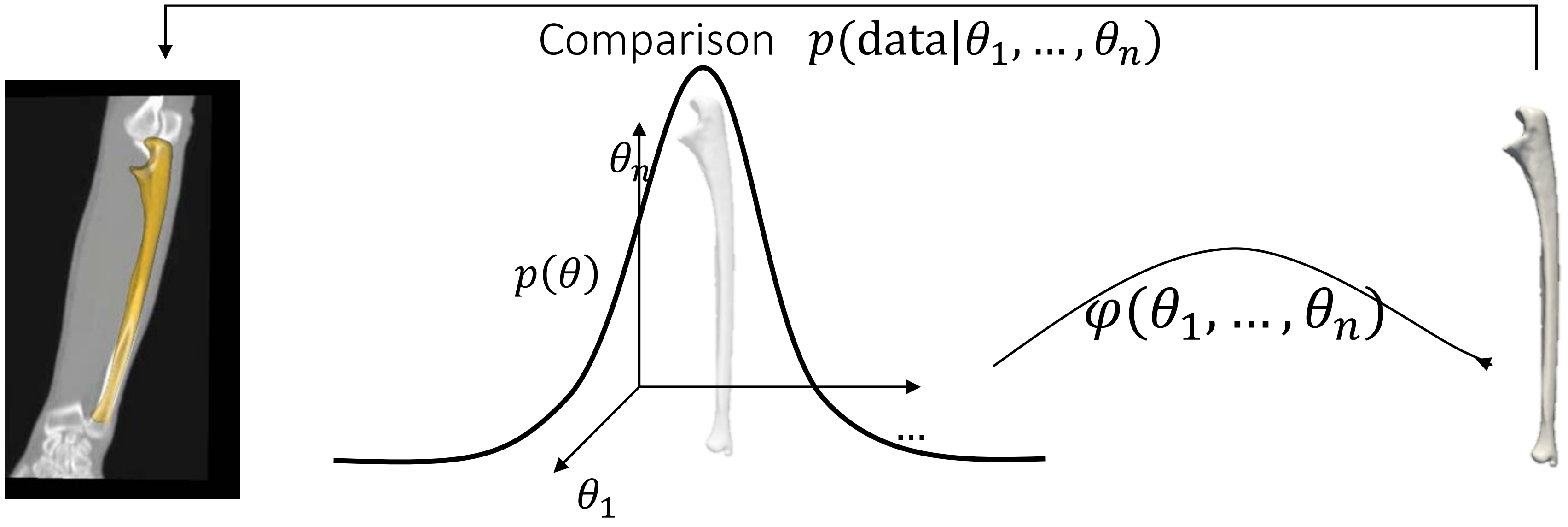
Bayes rule:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Lets us compute from $p(D|\theta)$ *its "inverse"* $p(\theta|D)$

Analysis by synthesis in 5 simple steps

4. Define prior distribution: $p(\theta) = p(\theta_1, \ldots, \theta_n)$
   - Our believe about the “state of the world”

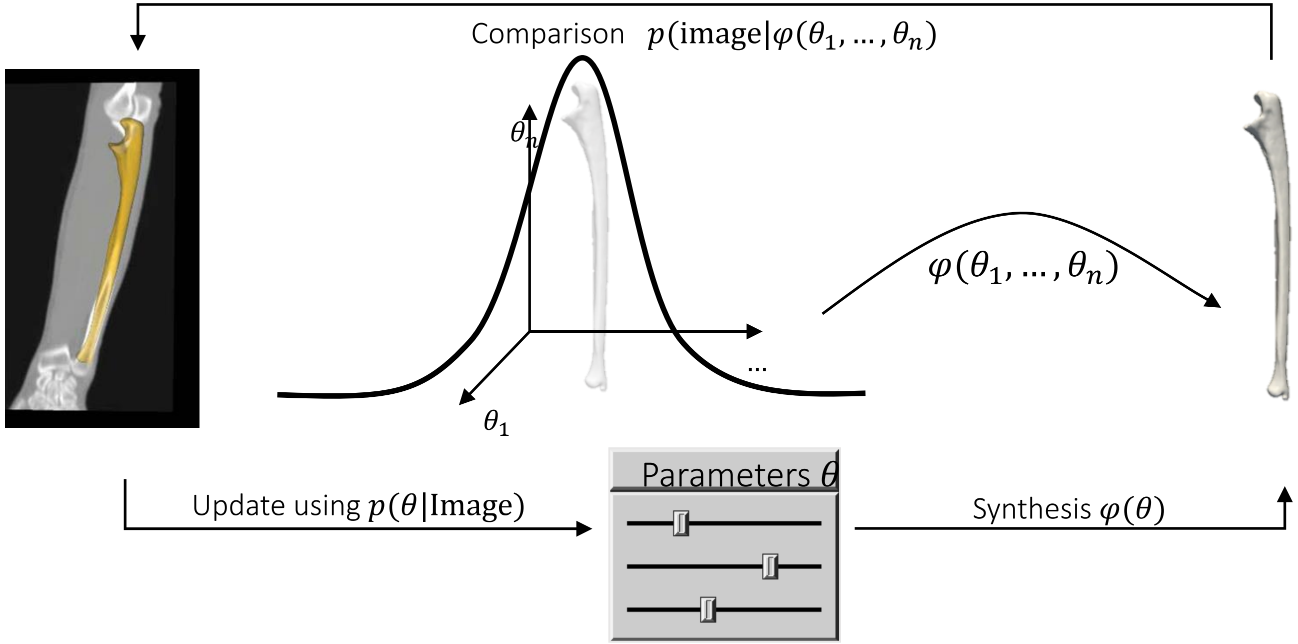


Makes it possible to invert mapping $p(data|\theta_1, \ldots, \theta_n)$

Analysis by synthesis in 5 simple steps

5. Inference

$$p(\theta_1, \dots, \theta_n | \text{data}) = \frac{p(\theta_1, \dots, \theta_n) p(\text{data} | \theta_1, \dots, \theta_n)}{p(\text{data})}$$



Estimates (fits) parameter of the model based on the data

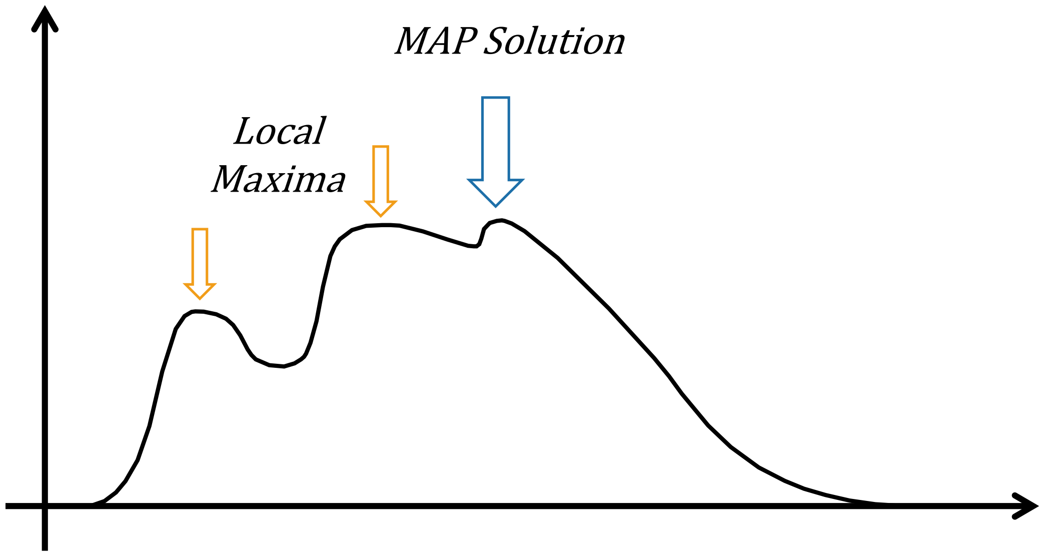
Analysis by synthesis in 5 simple steps

5. Possibility 1: Find MAP-solution:

$$\underset{\theta_1,\ldots,\theta_n}{\arg\max}\, p(\theta_1, \ldots, \theta_n|\text{data}) = \underset{\theta_1,\ldots,\theta_n}{\arg\max} \frac{p(\theta_1, \ldots, \theta_n)p(\text{data}|\theta_1, \ldots, \theta_n)}{p(\text{data})}$$

- Usually based on gradient-descent
- May miss good solutions

Today's topic

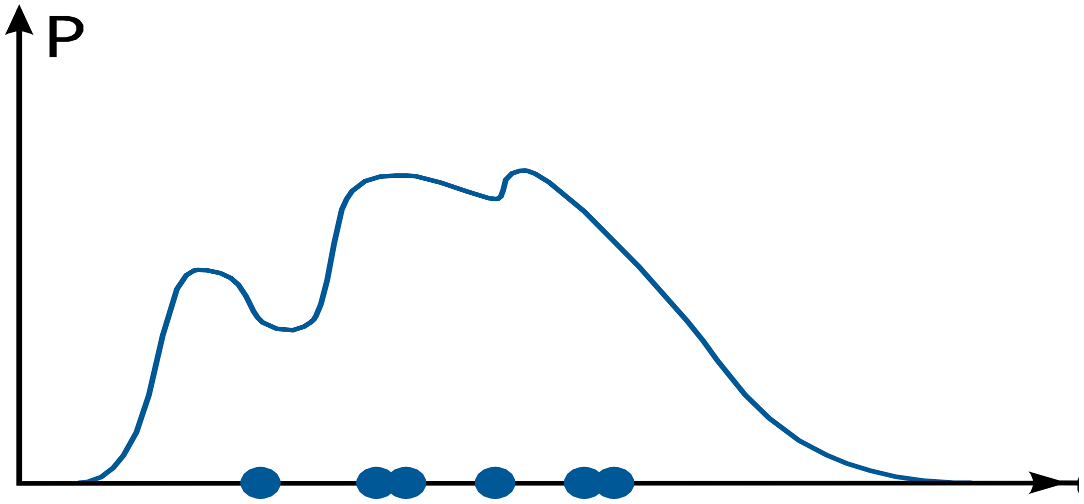


Analysis by synthesis in 5 simple steps

5. Possibility 2: Find full posterior distribution:

$$p(\theta_1, \dots, \theta_n | \text{data}) = \frac{p(\theta_1, \dots, \theta_n) p(\text{data} | \theta_1, \dots, \theta_n)}{p(\text{data})}$$

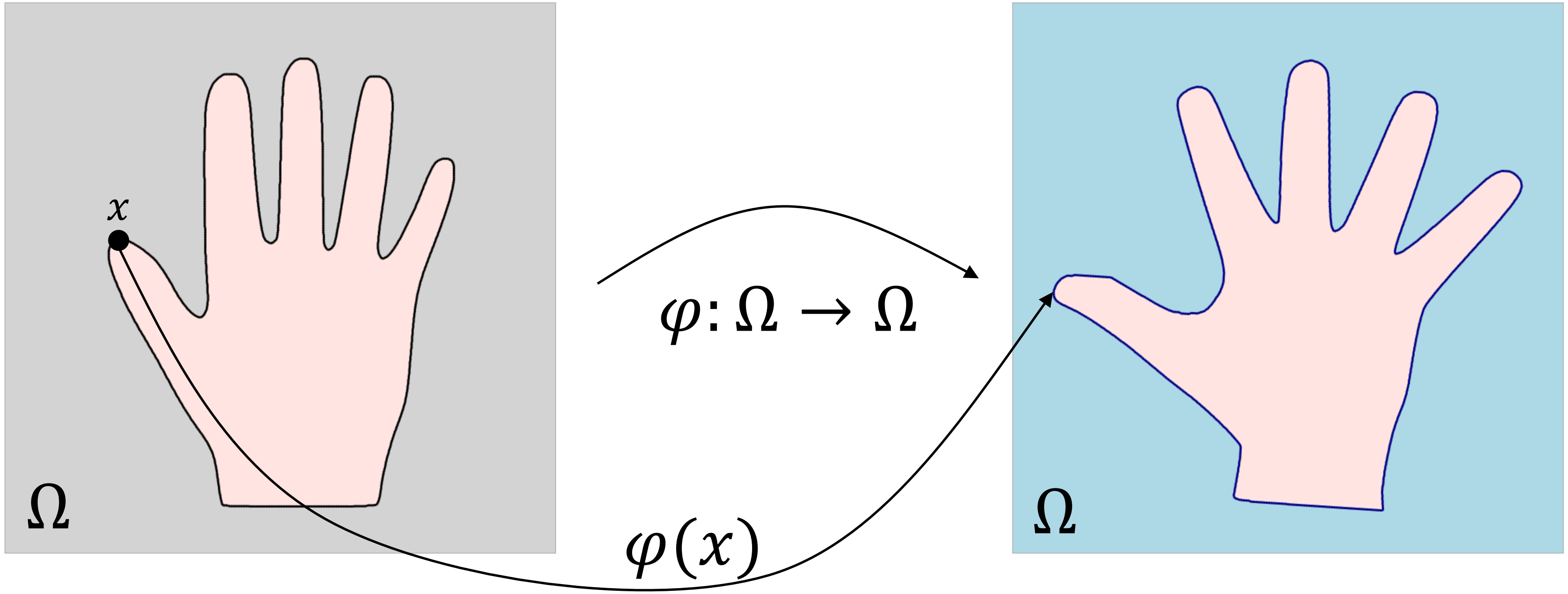
- Obtain samples from the distribution
- Based on Markov Chain Monte Carlo methods



Topic of next the next weeks.

Non-rigid registration – basic formulation

The registration problem



Reference:

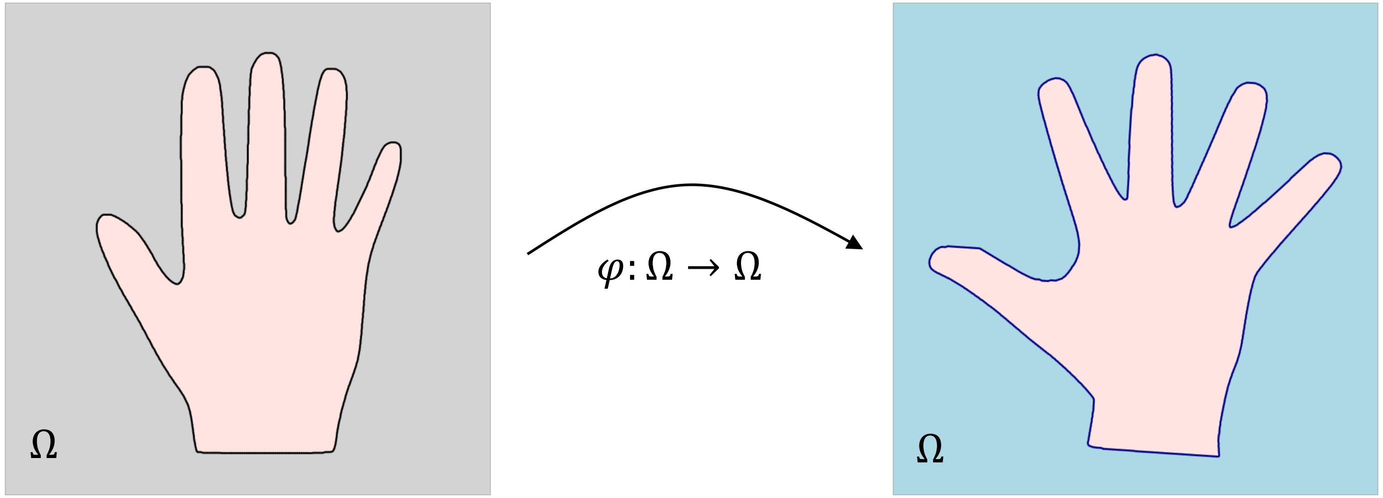
$I_R: \Omega \rightarrow \mathbb{R}$

Target:

$I_T: \Omega \rightarrow \mathbb{R}$

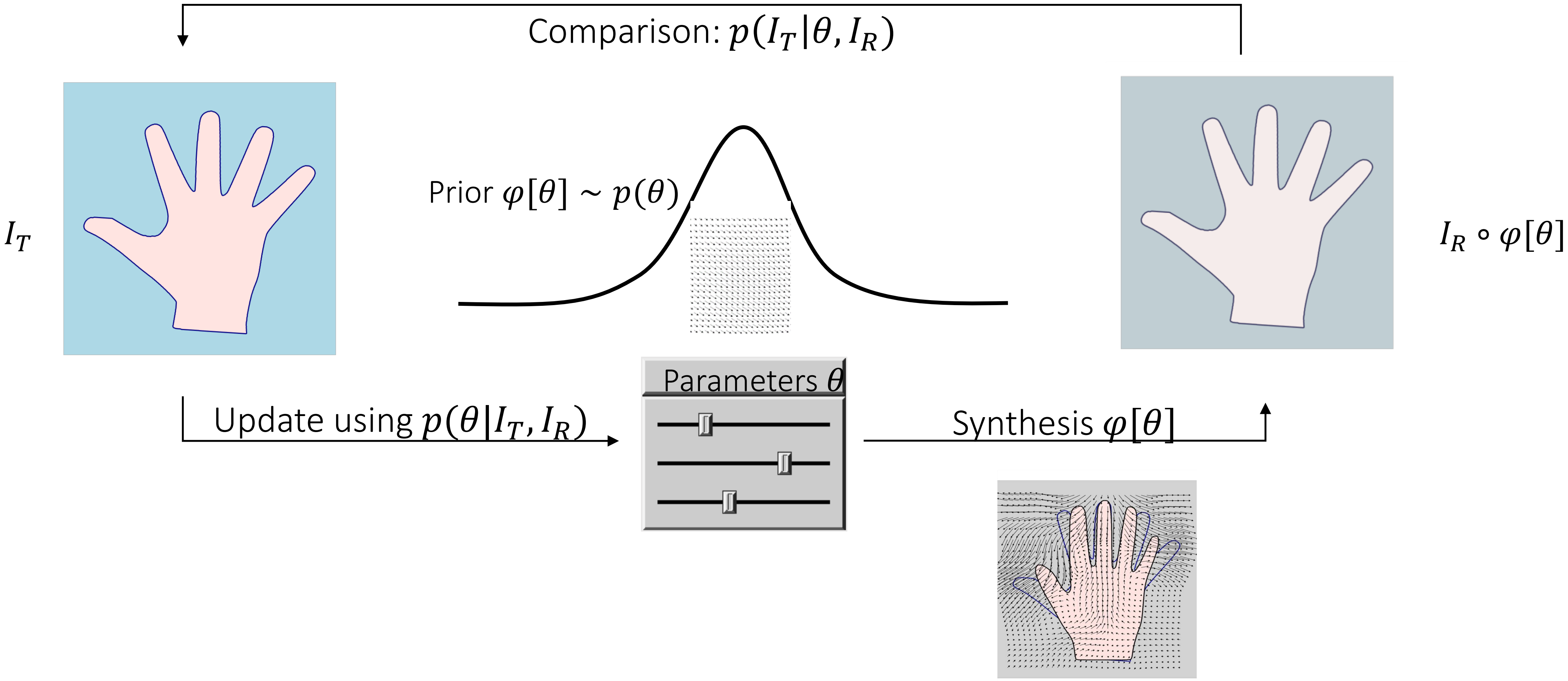
Why is it important?

- Do automatic measurements
- Compare shapes
  - Statistics
  - Build statistical models
- Transfer labels and annotations
  - Atlas based segmentation



Maybe the most important problem in computer vision and medical image analysis

Registration as analysis by synthesis



The registration problem

MAP-Estimation

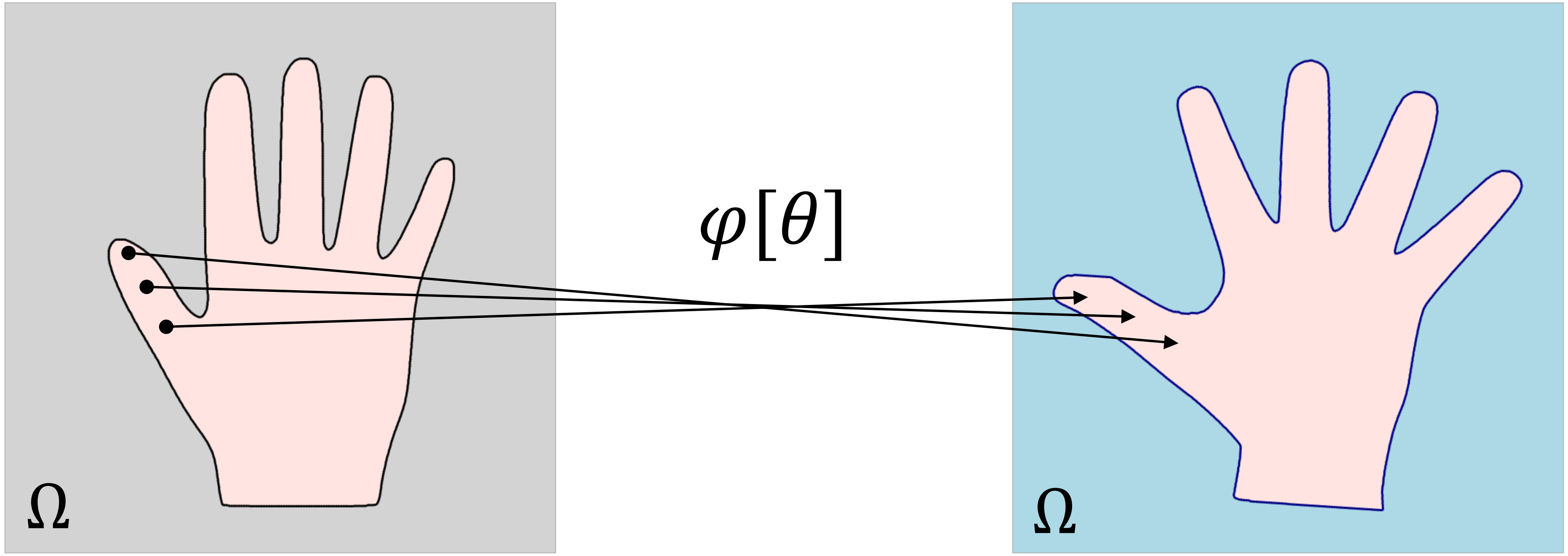
$$\theta^* = \arg\max_{\theta} p(\theta | I_T, I_R) = \arg\max_{\theta} p(\theta) p(I_T | \theta, I_R)$$

Mapping $\varphi[\theta^*]$ is trade-off that

- how well does the mapping explain the target image (likelihood function)
- matches the prior assumptions (prior distribution)

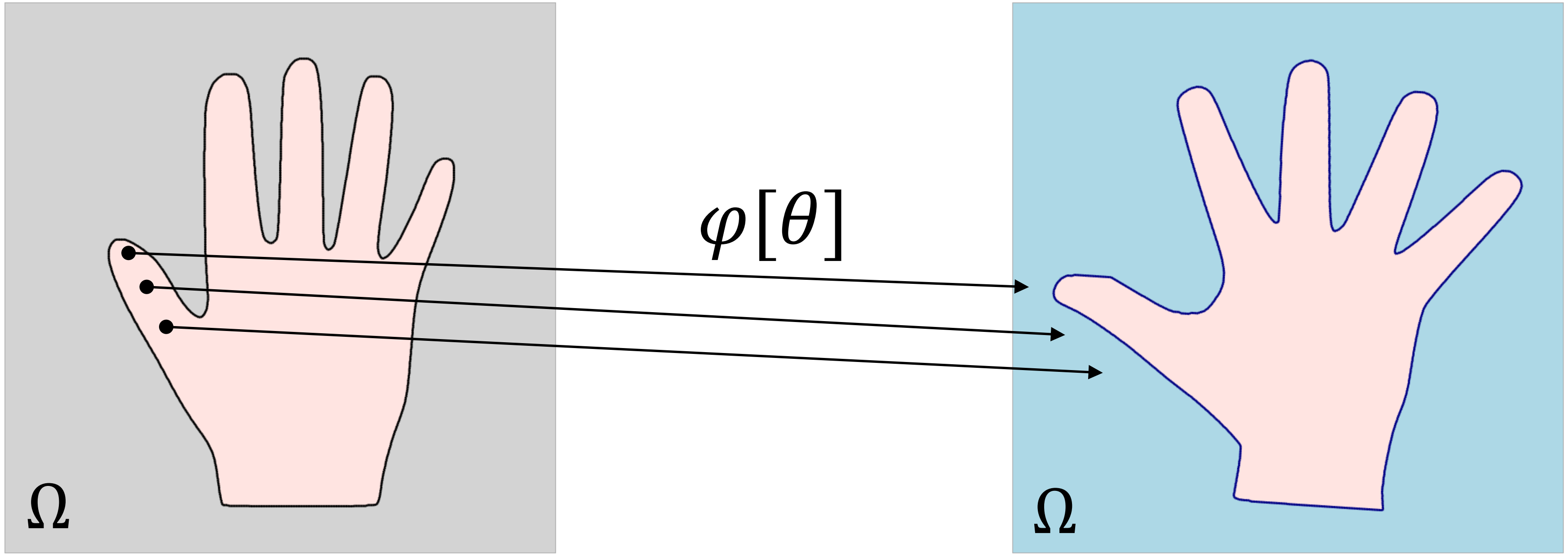
The registration problem

$$\theta^* = \arg\max_\theta p(\theta|I_T, I_R) = \arg\max_\theta p(\theta)p(I_T|\theta, I_R)$$



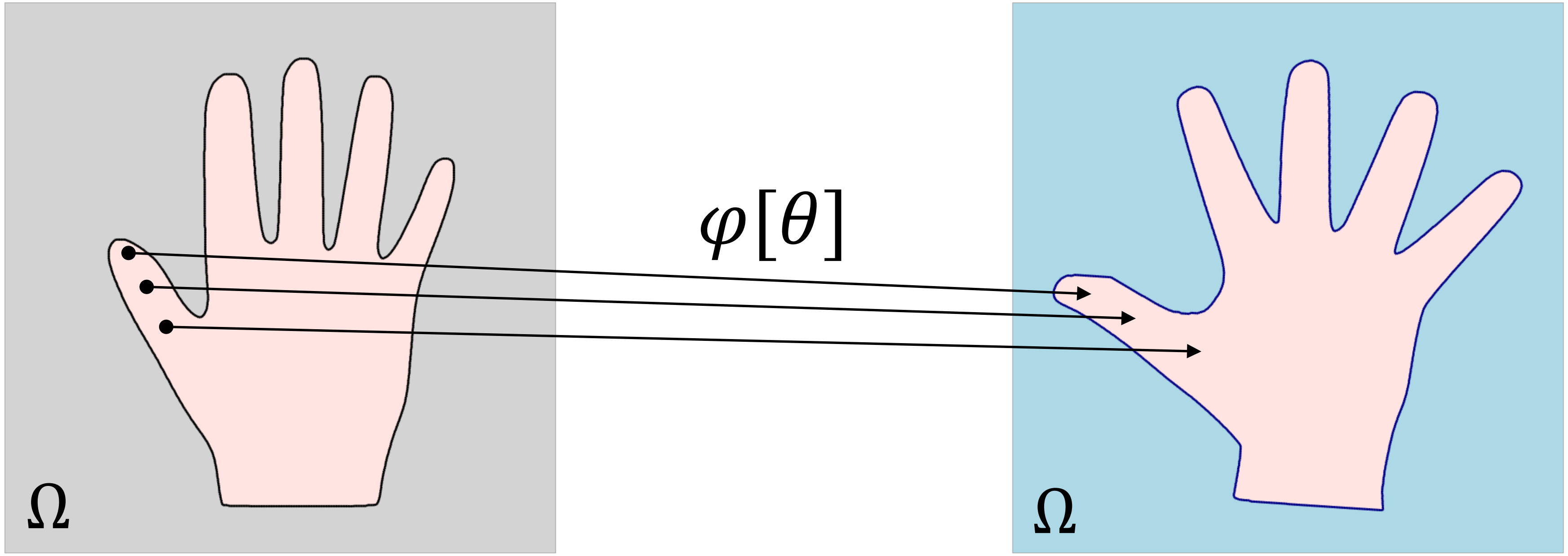
The registration problem

$$\theta^* = \arg\max_{\theta} p(\theta | I_T, I_R) = \arg\max_{\theta} p(\theta) p(I_T | \theta, I_R)$$



The registration problem

$$\theta^* = \arg\max_\theta p(\theta | I_T, I_R) = \arg\max_\theta p(\theta) p(I_T | \theta, I_R)$$



The registration problem

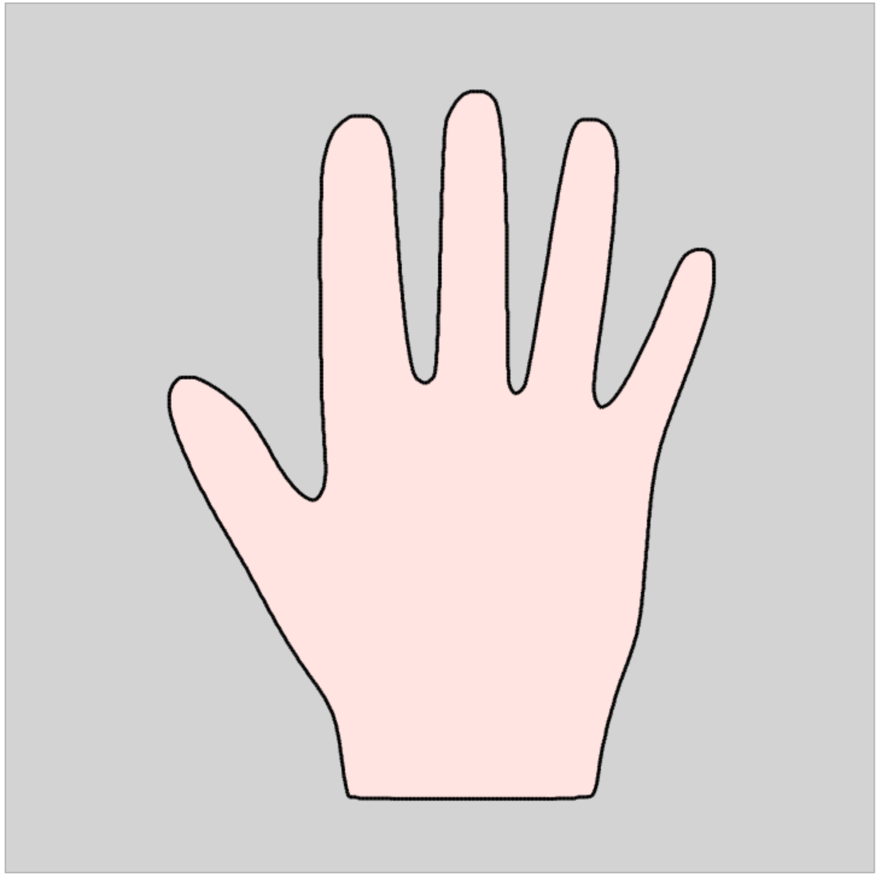
Probabilistic formulation

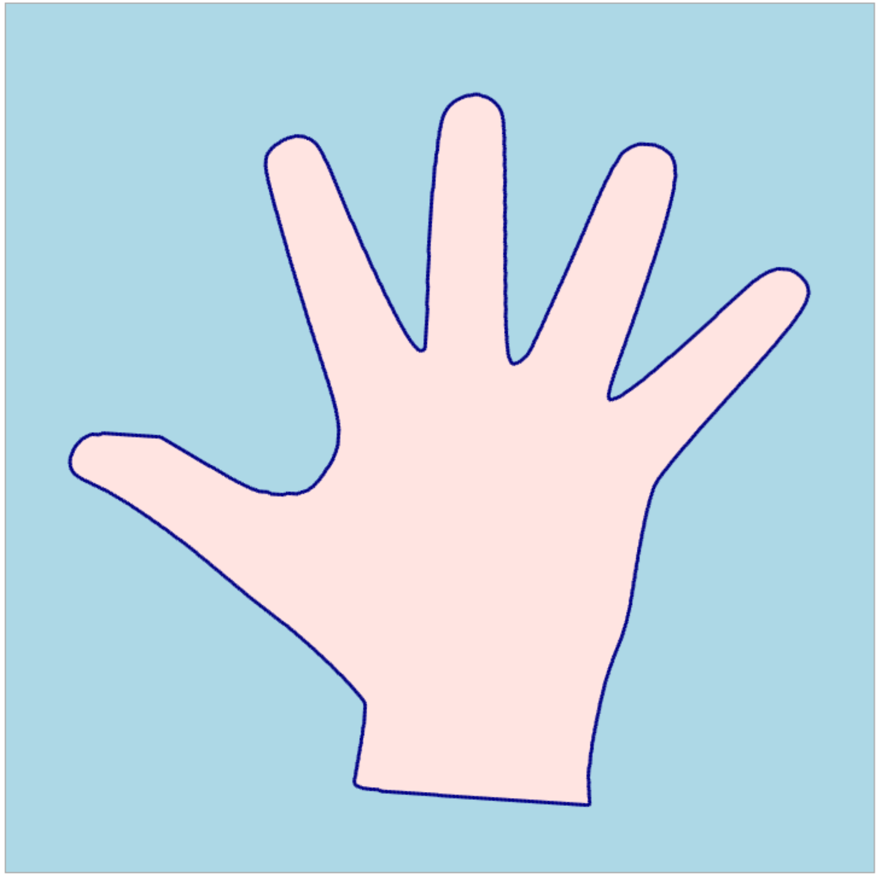
$$\varphi^* = \arg\max_{\varphi} p(\varphi|I_T, I_R) = \arg\max_{\varphi} p(\varphi)p(I_T|\varphi, I_R)$$

Main questions:

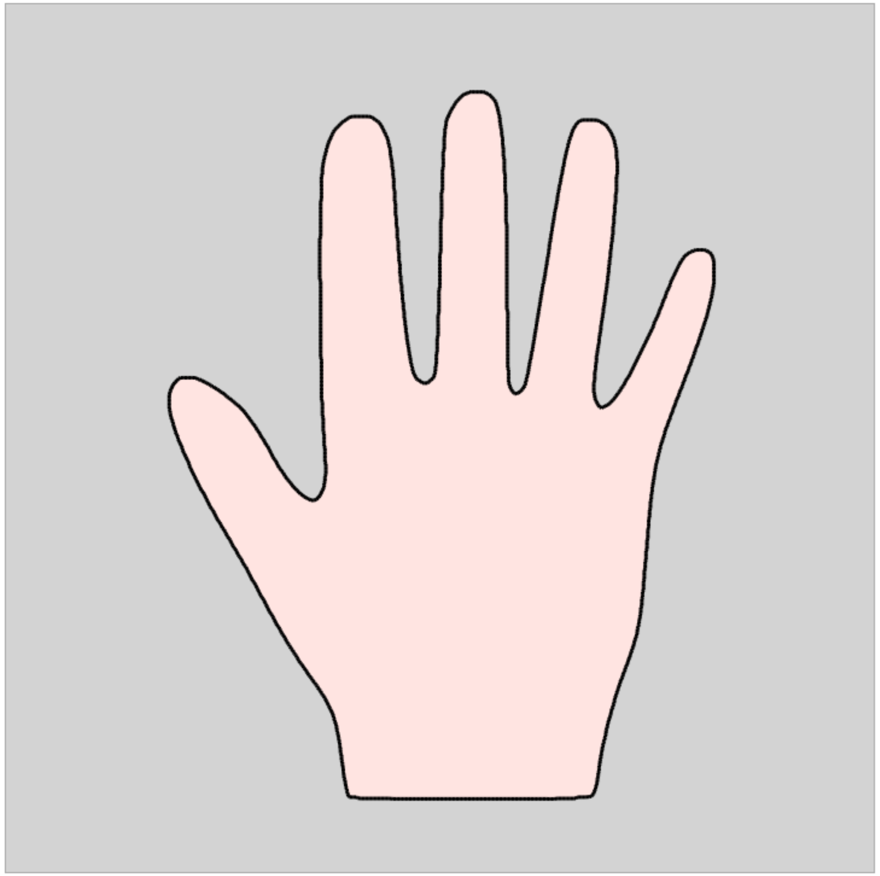
- How do we represent the mapping?
- How do we define the prior?
- What is the likelihood function?
- How can we solve the optimization problem?

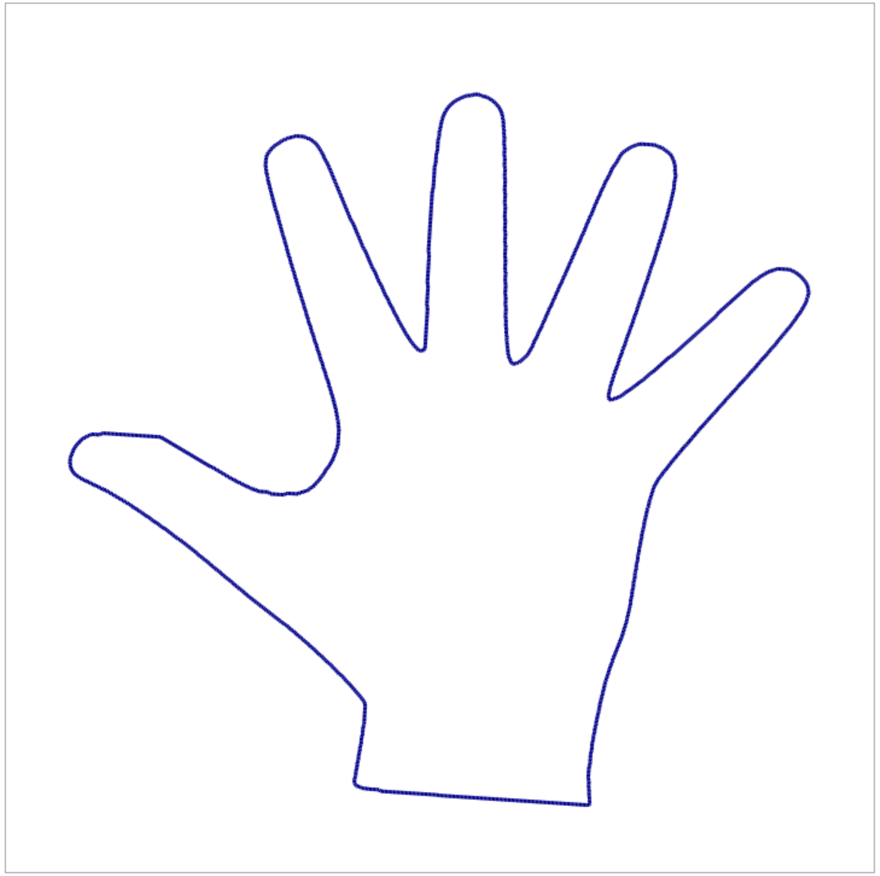
Representation of the mapping φ



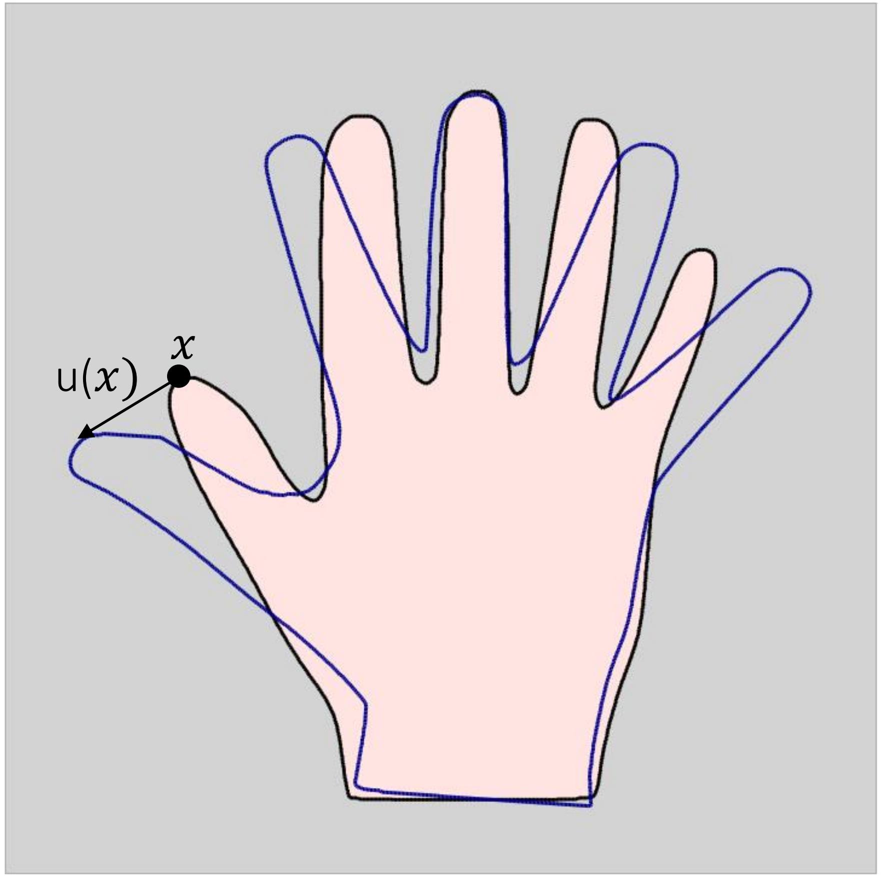


Representation of the mapping φ





Representation of the mapping φ

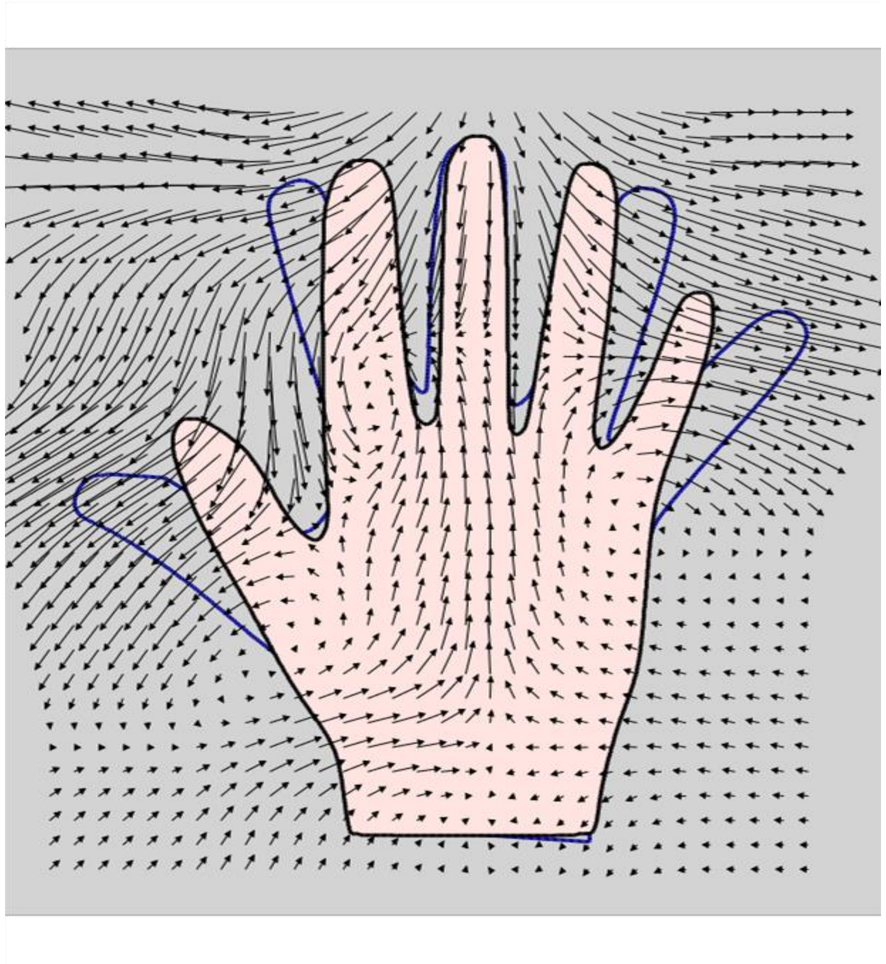


Assumption:
Images are rigidly aligned

- Mapping can be represented as a displacement vector field:

$$\varphi(x) = x + u(x)$$
$$u : \Omega \to \mathbb{R}^d$$

Representation of the mapping φ



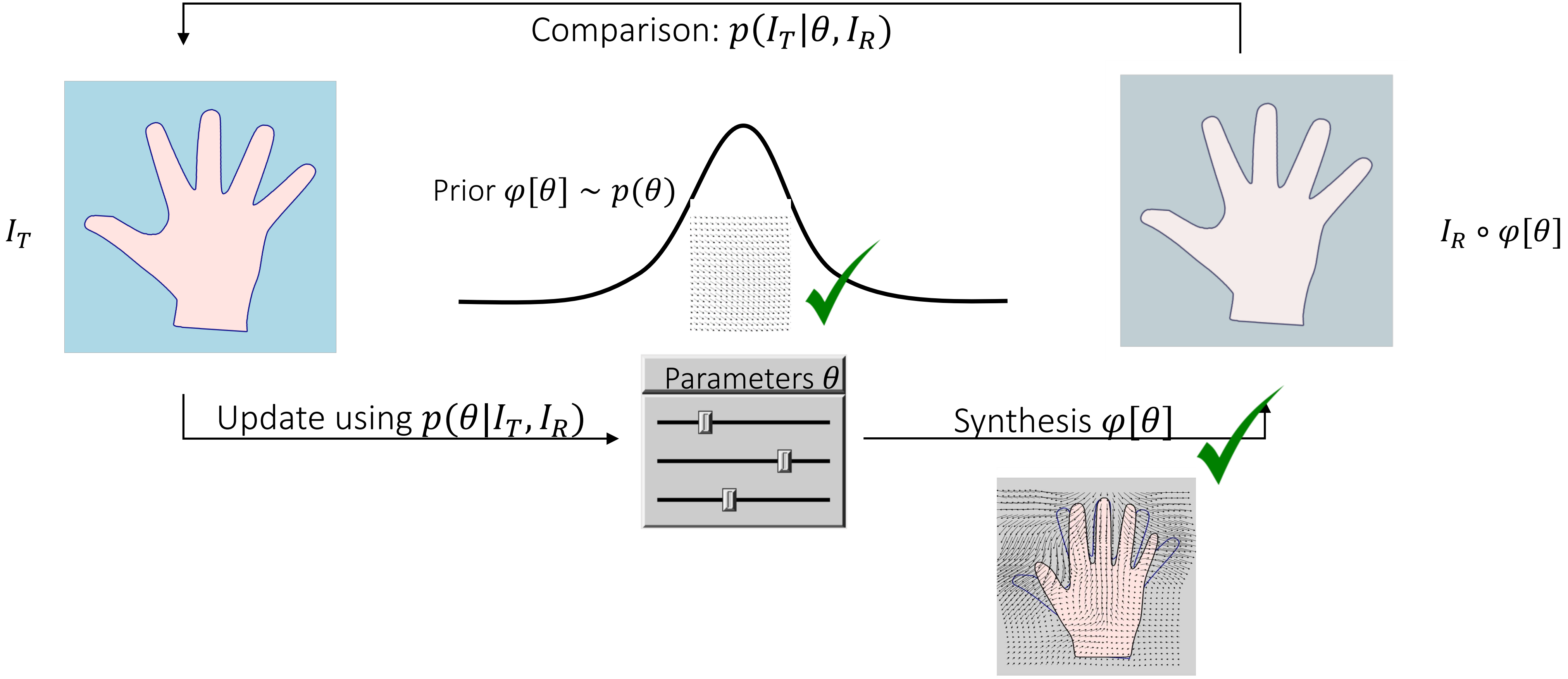
Assumption:
Images are rigidly aligned

- Mapping can be represented as a displacement vector field:

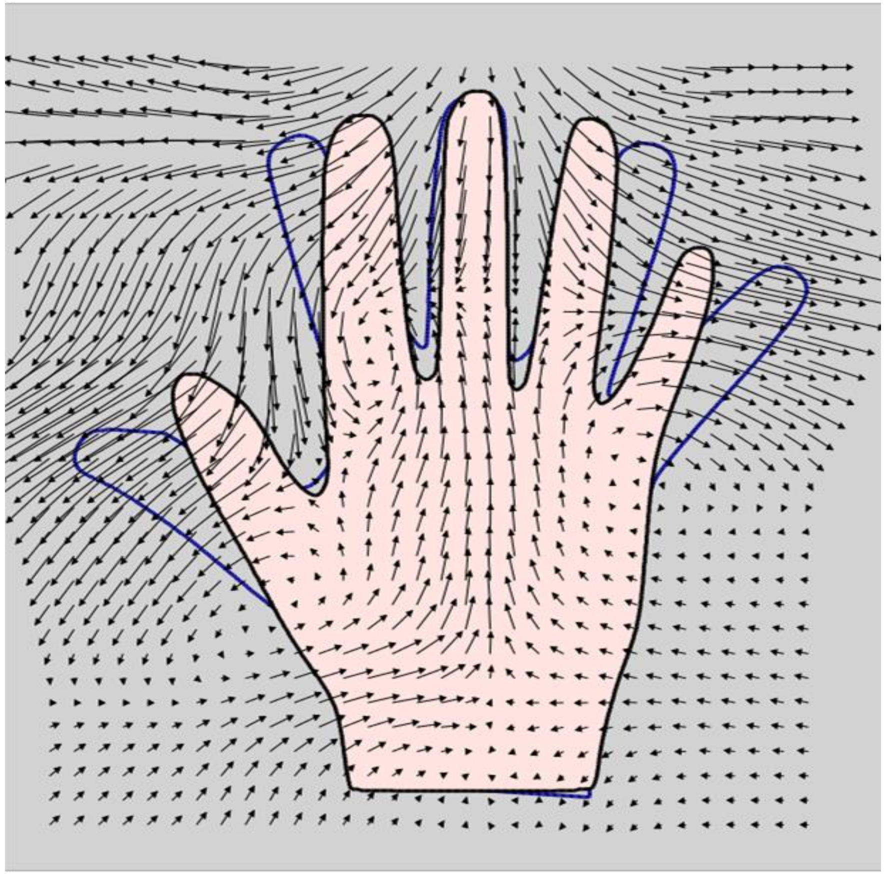
$$\varphi(x) = x + u(x)$$
$$u : \Omega \to \mathbb{R}^d$$

Observation:
Knowledge of u and I_R allows us to synthesize target image I_T

Registration as analysis by synthesis



Priors



Define the Gaussian process

$$u \sim GP(\mu, k)$$

with mean function

$$\mu: \Omega \rightarrow \mathbb{R}^2$$

and covariance function

$$k: \Omega \times \Omega \rightarrow \mathbb{R}^{2\times 2}\ .$$

Example prior: Smooth 2D deformations

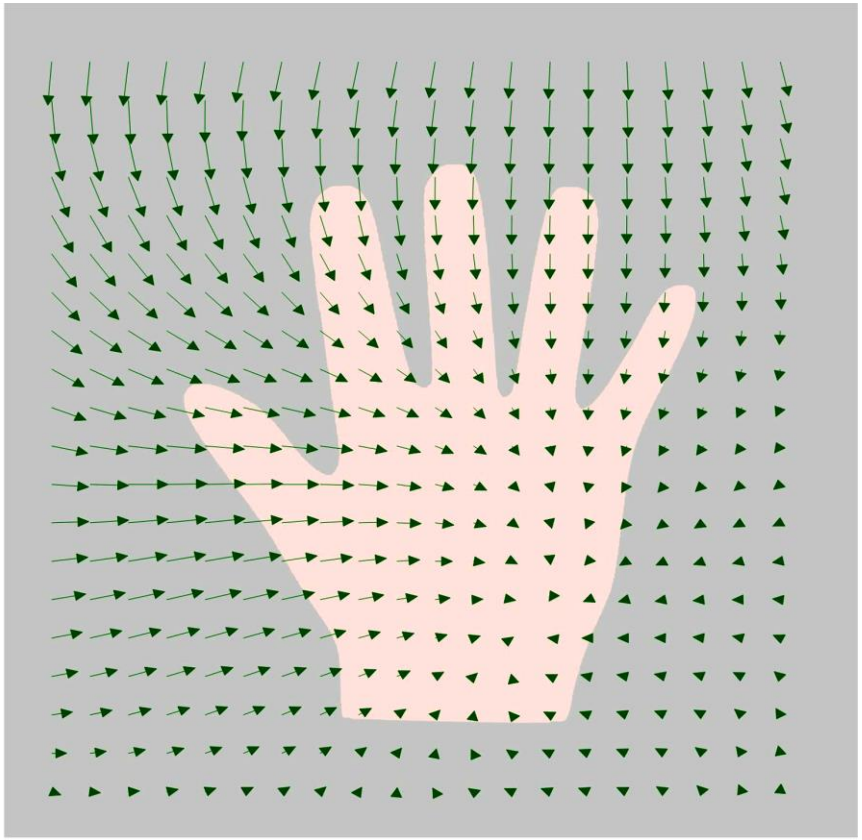
Zero mean:

$$\mu(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Squared exponential covariance function (Gaussian kernel)

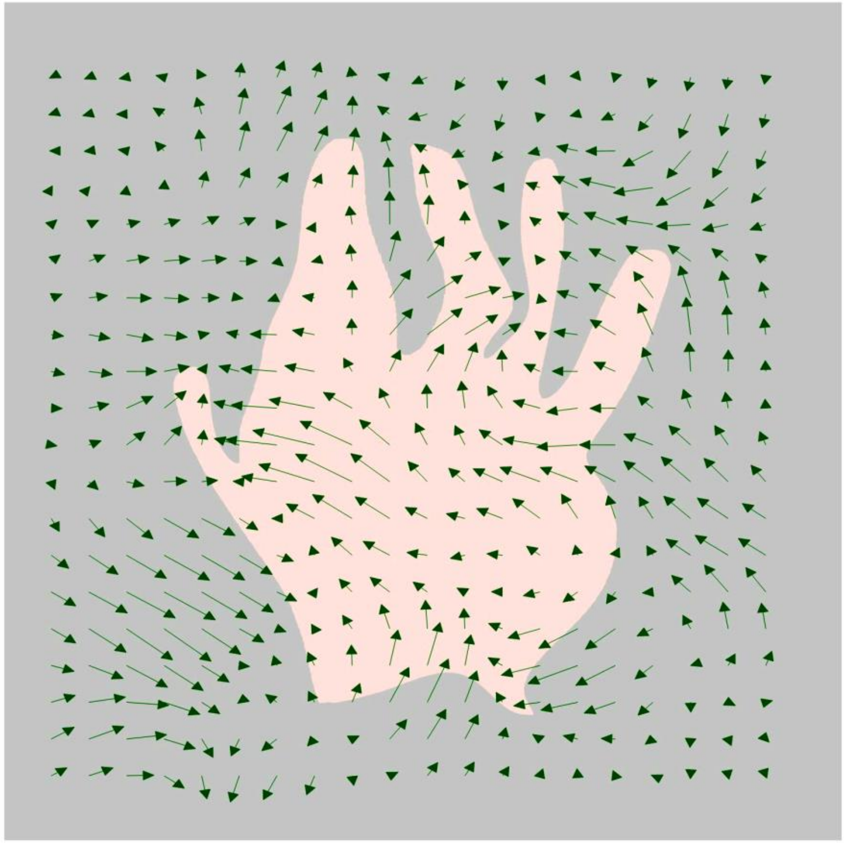
$$k(x, x') = \begin{pmatrix} s_1 \exp\left(-\frac{\|x - x'\|^2}{\sigma_1^2}\right) & 0 \\ 0 & s_2 \exp\left(-\frac{\|x - x'\|^2}{\sigma_2^2}\right) \end{pmatrix}$$

Example prior: Smooth 2D deformations



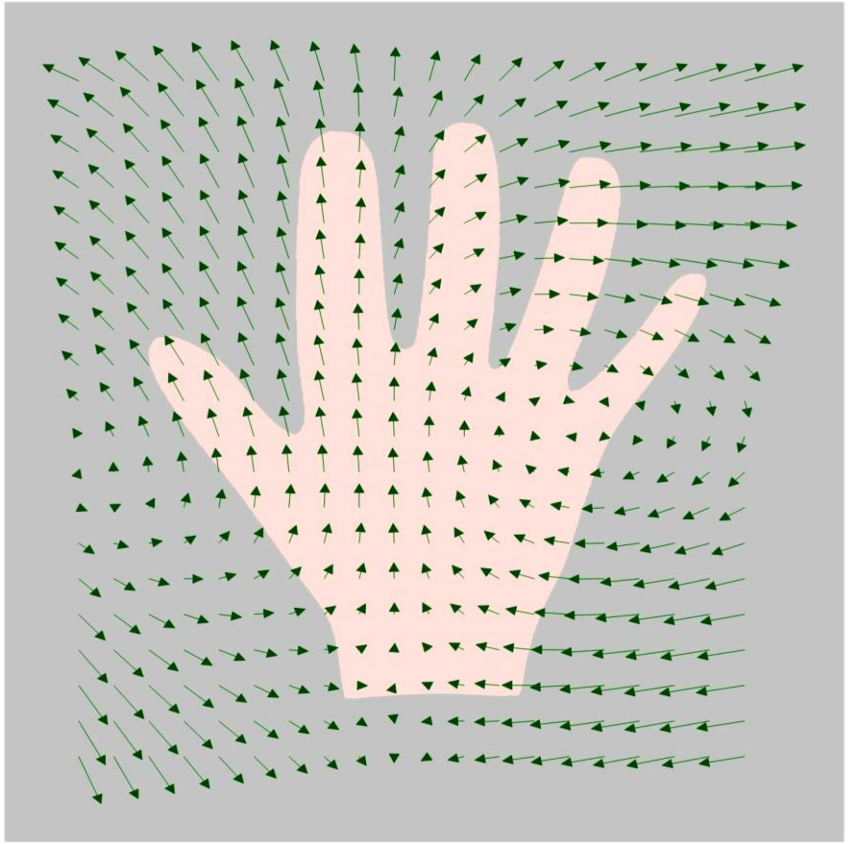
$s_1 = s_2$ small, $\sigma_1 = \sigma_2$ large

Example prior: Smooth 2D deformations



$s_1 = s_2$ small, $\sigma_1 = \sigma_2$ small

Example prior: Smooth 2D deformations



$s_1 = s_2$ large, $\sigma_1 = \sigma_2$ large

Parametric representation of Gaussian process

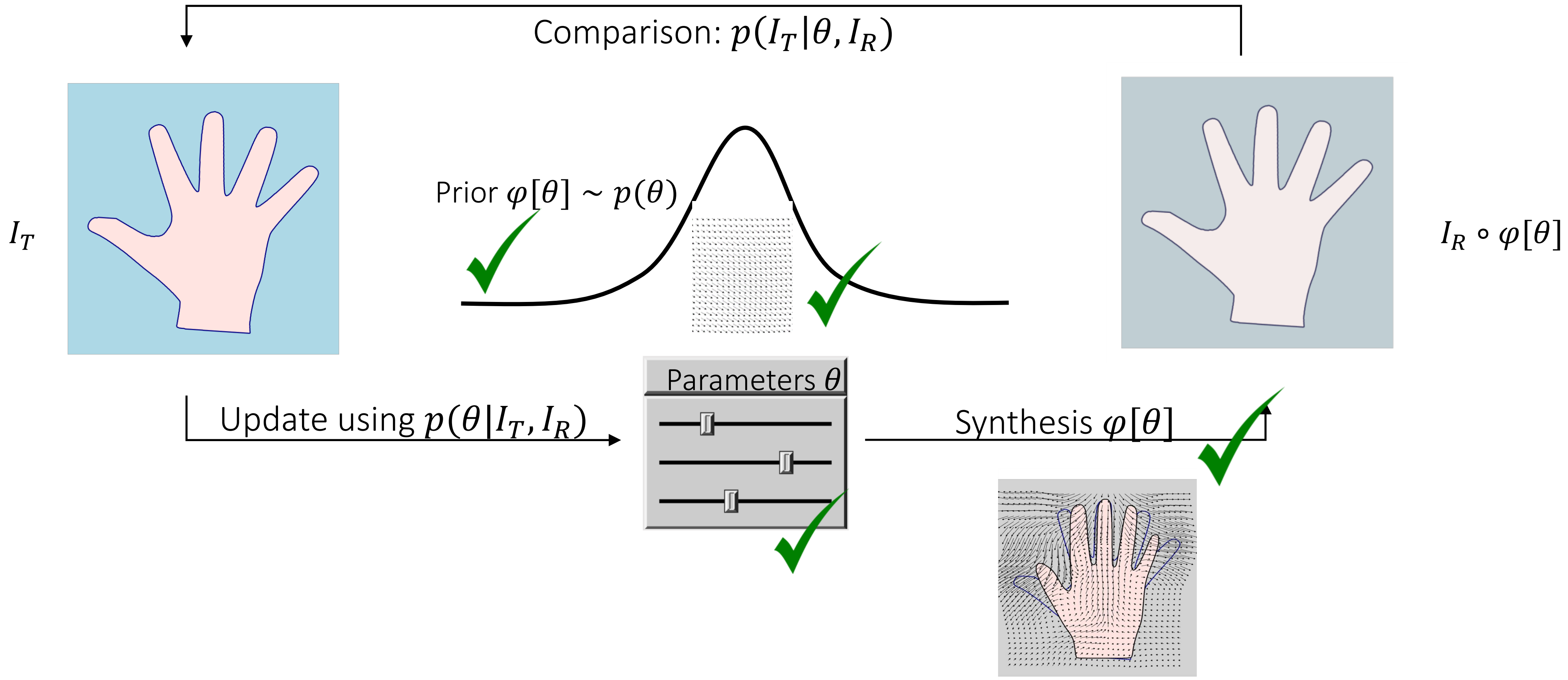
Represent $GP(\mu, k)$ using only the first r components of its KL-Expansion

$$u = \mu + \sum_{i=1}^{r} \alpha_i \sqrt{\lambda_i}\, \phi_i, \qquad \alpha_i \sim N(0,1)$$

- We have a finite, parametric representation of the process.
- We know the pdf for a deformation u

$$p(u[\alpha]) = p(\alpha) = \prod_{i=1}^{r} \frac{1}{\sqrt{2\pi}} \exp(-\alpha_i^2/2) = \frac{1}{Z} \exp(-\frac{1}{2}\|\alpha\|^2)$$

Registration as analysis by synthesis

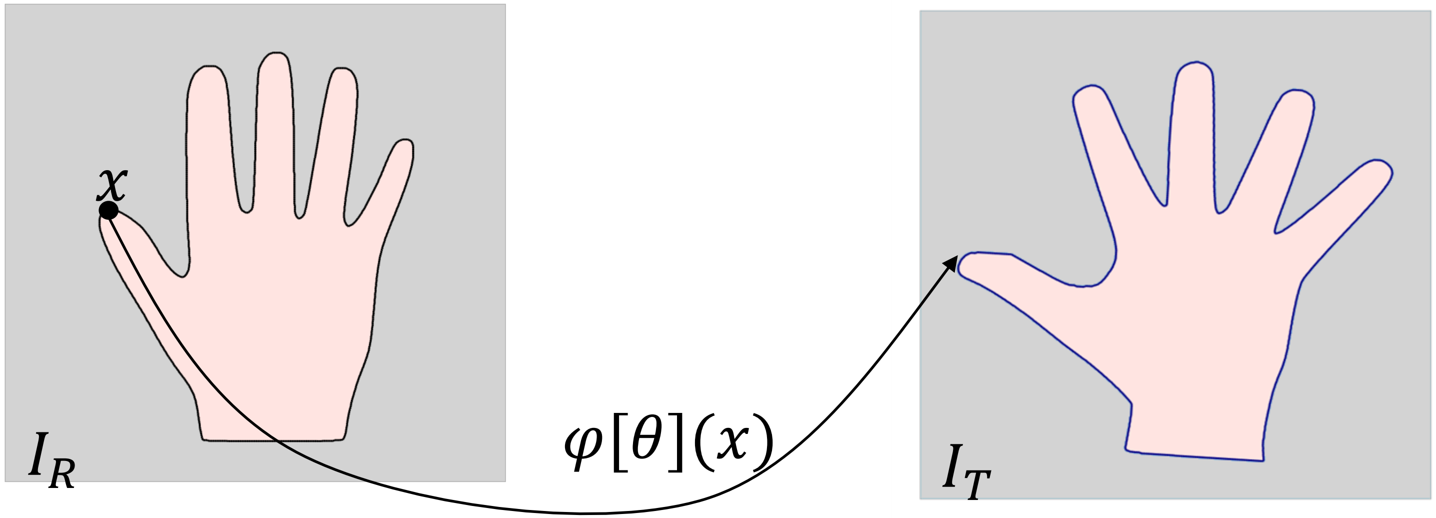


Likelihood function: Image registration

Images are similar when the intensities match

Assumptions:

- Corresponding points have the same image intensity (up to i.i.d. noise)



$$p(I_T(\varphi[\theta](x))|I_R,\theta,x) \sim N(I_R(x),\sigma^2)$$

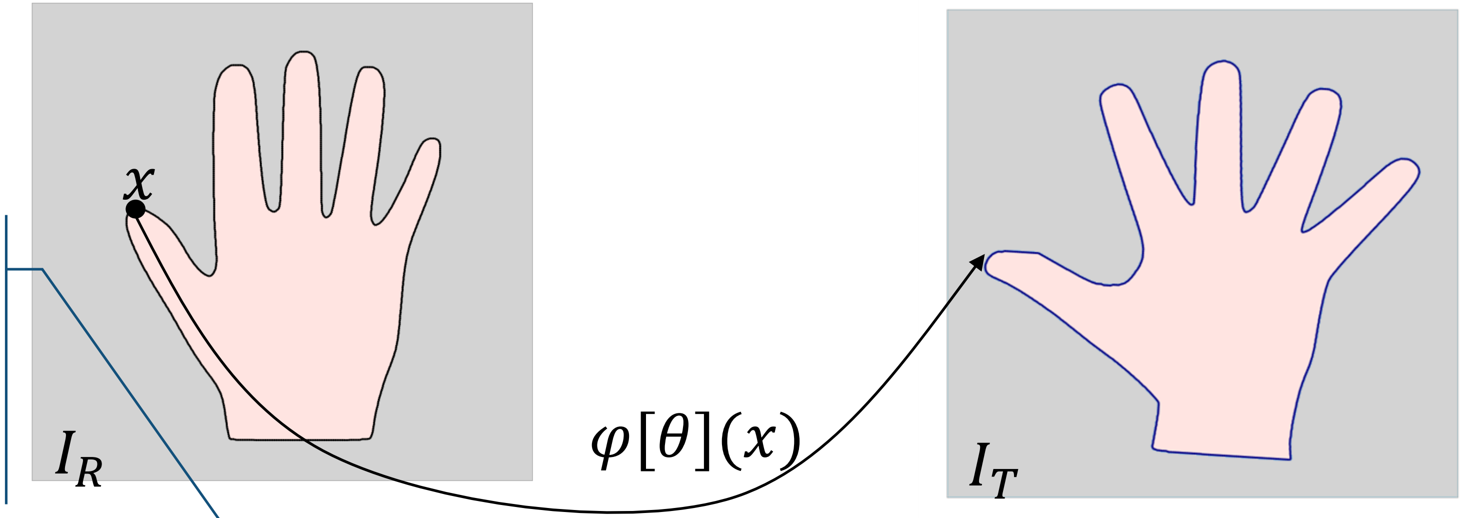
Likelihood function: Image registration

Images are similar when the intensities match

Assumptions:

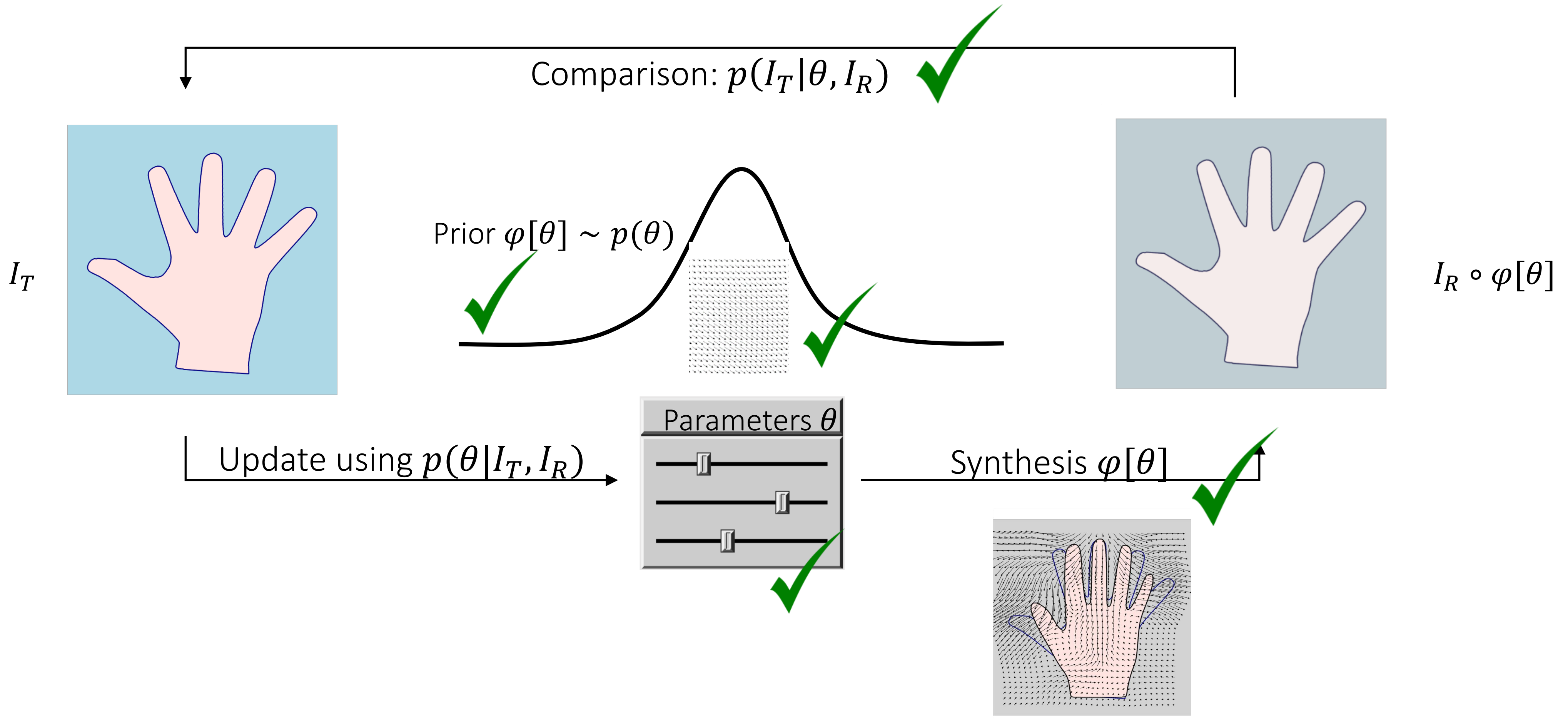
- Corresponding points have the same image intensity (up to i.i.d. noise)

Image term outside mapping function. Makes problem really difficult



$$p(I_T|I_R,\theta) = \prod_{x\in\Omega} p(I_T(\varphi[\theta](x))|I_R,\theta,x) = \prod_{x\in\Omega} \frac{1}{Z}\exp\left(-\frac{(I_T(\varphi(x)) - I_R(x))^2}{\sigma^2}\right)$$

Registration as analysis by synthesis



Registration problem

$$\theta^* = \arg\max_\theta p(\varphi[\theta]) p(I_T | \varphi[\theta], I_R)$$

$$= \arg\max_\theta \frac{1}{Z_1} \exp\left(-\frac{1}{2}\|\theta\|^2\right) \frac{1}{Z_2} \prod_x \exp\left(-\frac{\left(I_T(\varphi[\theta](x)) - I_R(x)\right)^2}{\sigma^2}\right)$$

- Parametric problem, since:

$$\varphi[\theta](x) = x + \mu(x) + \sum_{i=1}^{r} \theta_i \sqrt{\lambda_i}\, \phi_i(x)$$

- Can be optimized using gradient descent

Variational formulation

$$\arg\max_{\theta} \frac{1}{Z_1} \exp\left(-\frac{1}{2}\|\theta\|^2\right) \frac{1}{Z_2} \prod_{x} \exp\left(-\frac{\left(I_T\left(\varphi[\theta](x)\right) - I_R(x)\right)^2}{\sigma^2}\right)$$

$$= \arg\max_{\theta} \; \ln \frac{1}{Z_1} \exp\left(-\frac{1}{2}\|\theta\|^2\right) + \ln \frac{1}{Z_2} \prod_{x} \exp\left(-\frac{\left(I_T\left(\varphi[\theta](x)\right) - I_R(x)\right)^2}{\sigma^2}\right)$$

$$= \arg\max_{\theta} \; \ln \frac{1}{Z_1} - \frac{1}{2}\|\theta\|^2 + \ln \frac{1}{Z_2} - \sum_{x \in \Omega} \frac{\left(I_T\left(\varphi[\theta](x)\right) - I_R(x)\right)^2}{\sigma^2}$$

$$= \arg\min_{\theta} \sum_{x \in \Omega} \frac{\left(I_T\left(\varphi[\theta](x)\right) - I_R(x)\right)^2}{\sigma^2} + \frac{1}{2}\|\theta\|^2$$

Image metric

Regularizer

The registration problem

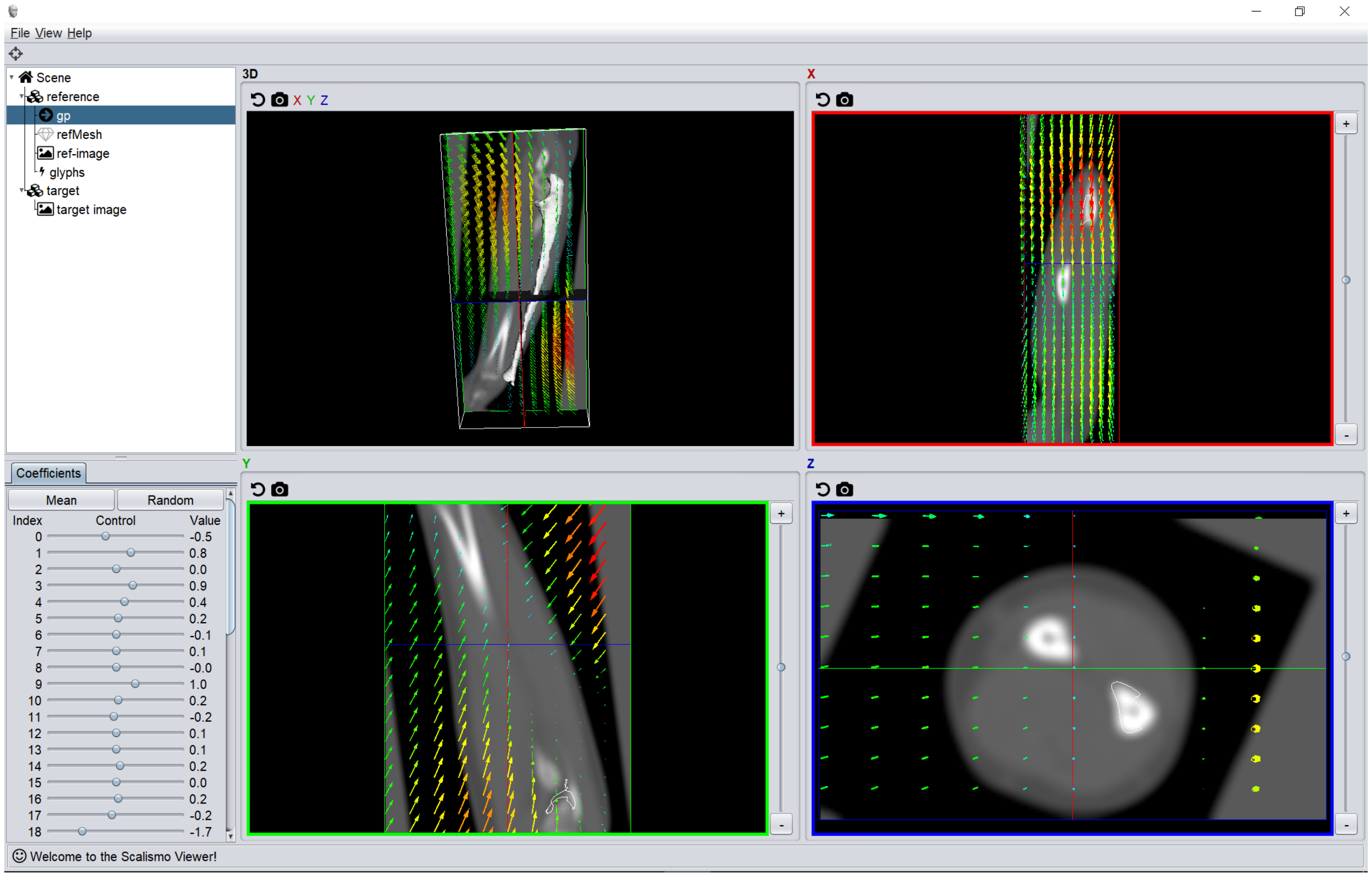
Probabilistic formulation

$$\theta^* = \arg\min_{\theta} -\ln\big(p(I_T|I_R, \varphi[\theta])\big) - \ln p(\varphi[\theta])$$

Variational formulation

$$\theta^* = \arg\min_{\theta} D[I_T, I_R, \varphi[\theta]] + \lambda R[\varphi[\theta]]$$

GP-Registration in Scalismo



A selection of useful likelihood functions

Landmark likelihood

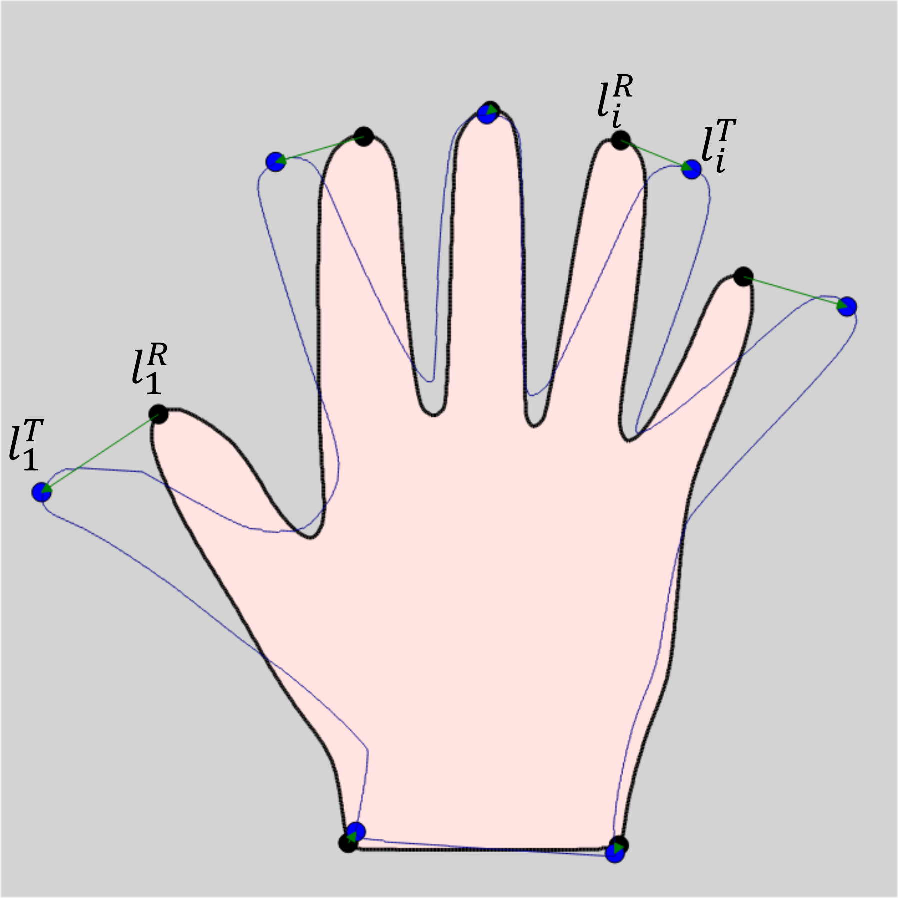
For one landmark pair (l_R, l_T):

$$p(l_T|\theta, l_R) = N(\varphi[\theta](l_R), I_{2x2}\sigma^2)$$

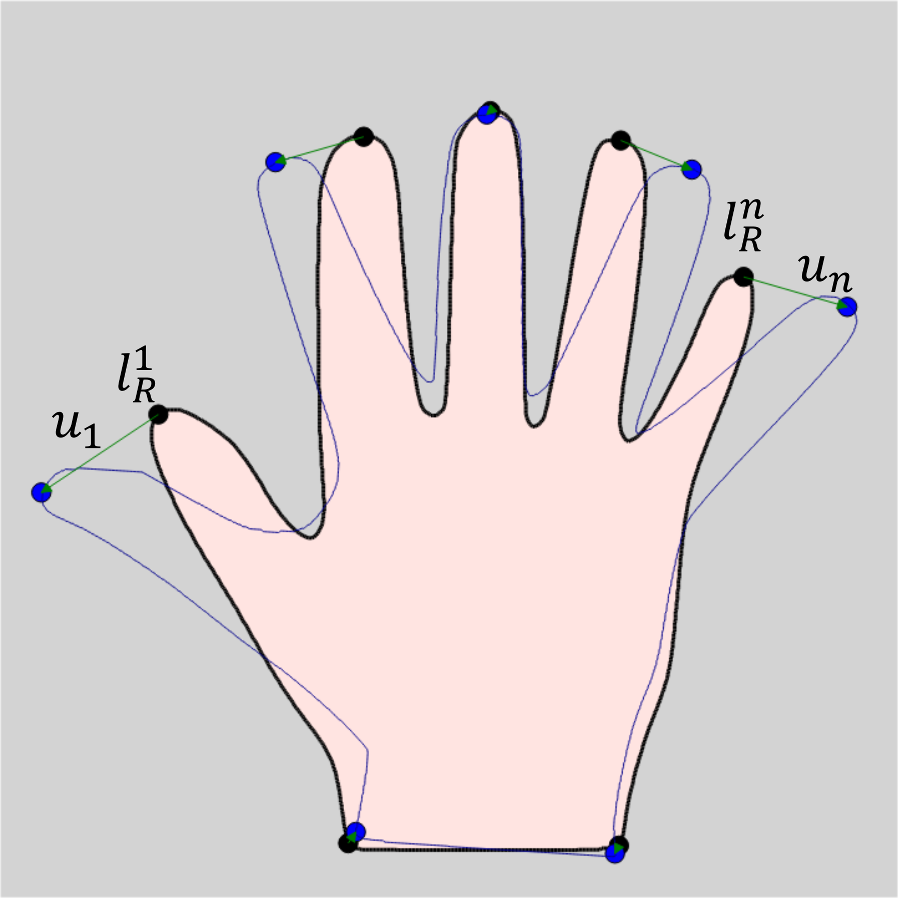
For many landmarks

$$L = ((l_R^1, l_T^1), \dots, (l_R^n, l_T^n))$$

$$p(l_1^T, \dots, l_n^T \middle| \theta, l_R^1, \dots, l_R^n) = \prod_i N(\varphi[\theta](l_R), I_{2x2}\sigma^2)$$



Landmark registration using GP Regression



Given:

- Gaussian process: $u \sim GP(\mu, k)$
- Observations: $\{(l_i^R, \tilde{u}_i), i = 1, \dots, n\}$

Assume:

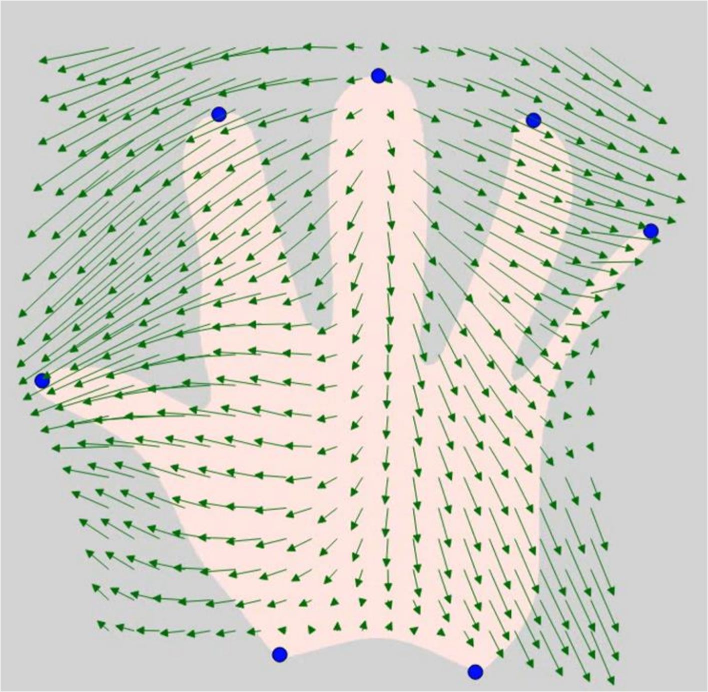
$$\tilde{u}_i = u(l_i) + \epsilon \text{ with } \epsilon \sim N(0, \sigma^2 I_{2\times 2}).$$

Goal:

- Find posterior distribution

$$u \mid l_1^R, \dots, l_n^R, \tilde{u}_1, \dots, \tilde{u}_n$$

Gaussian process regression



The posterior

$$u \mid l_1^R, \ldots, l_n^R, \tilde{u}_1, \ldots, \tilde{u}_n$$

is a Gaussian process

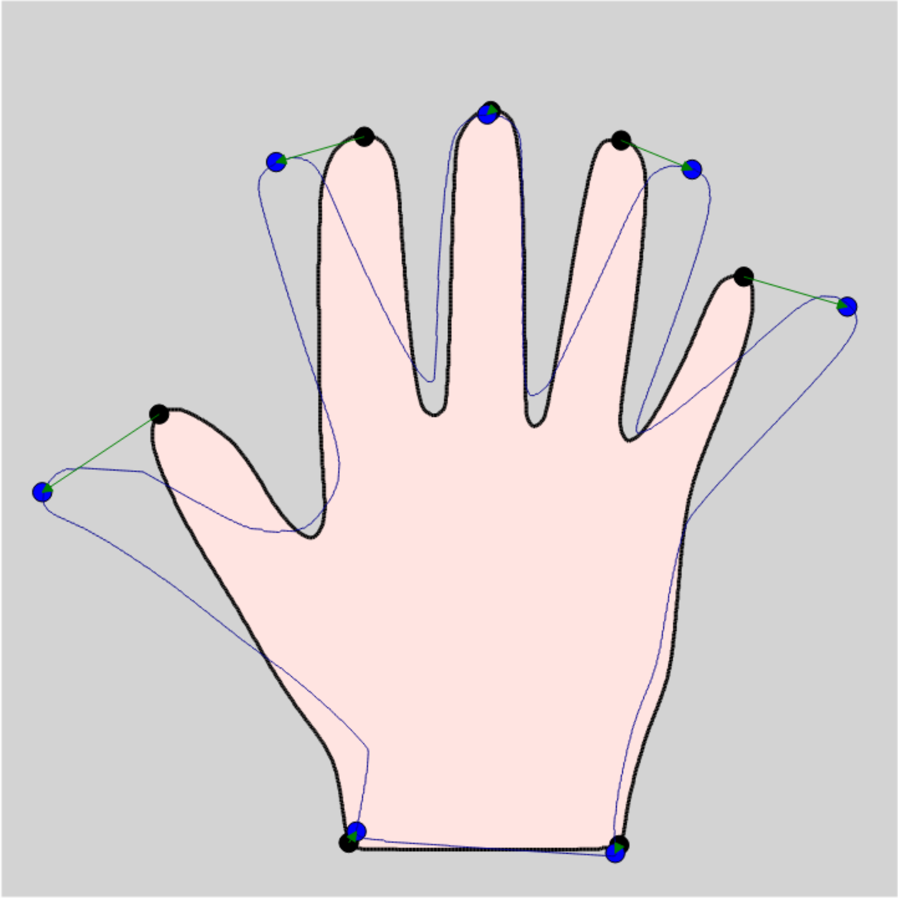
$$GP(\mu_p, k_p)$$

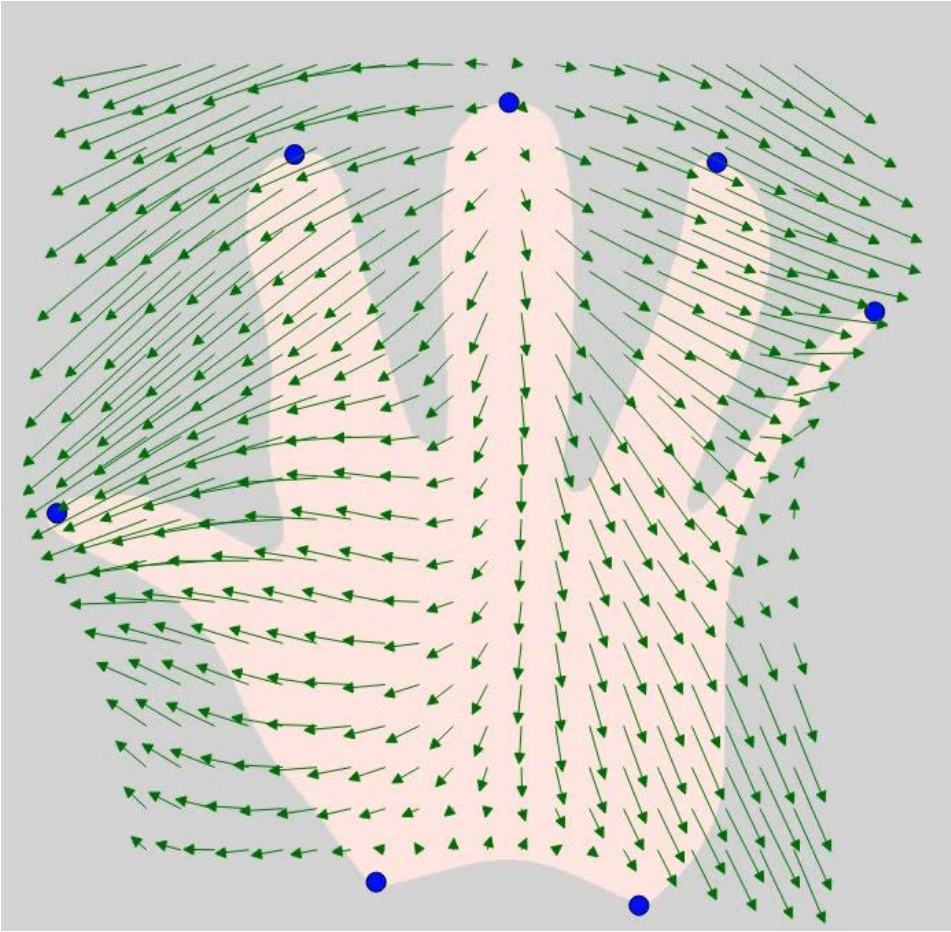
Its parameters are known analytically.

$$\mu_p(x) = \mu(x) + K(x, Y)(K(Y, Y) + \sigma^2 I_{2n\times 2n})^{-1}(\tilde{\boldsymbol{u}} - \mu(Y))$$

$$k_p(x, x') = k(x, x') - K(x, Y)\,(K(Y, Y) + \sigma^2 I_{2n\times 2n})^{-1} K(Y, x')$$

Landmark registration using GP Regression



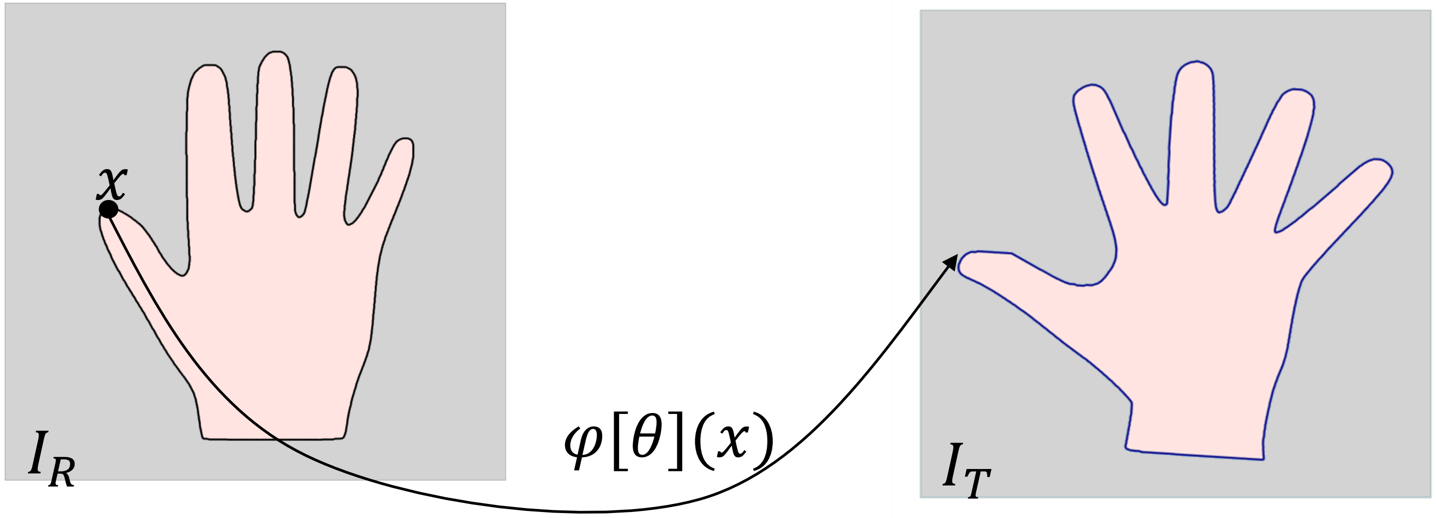


Likelihood function: Image registration

Images are similar when the intensities match

Assumptions:

- Corresponding points have the same image intensity (up to i.i.d. noise)



$$p(I_T(\varphi[\theta](x))|I_R, \theta, x) \sim N(I_R(x), \sigma^2)$$

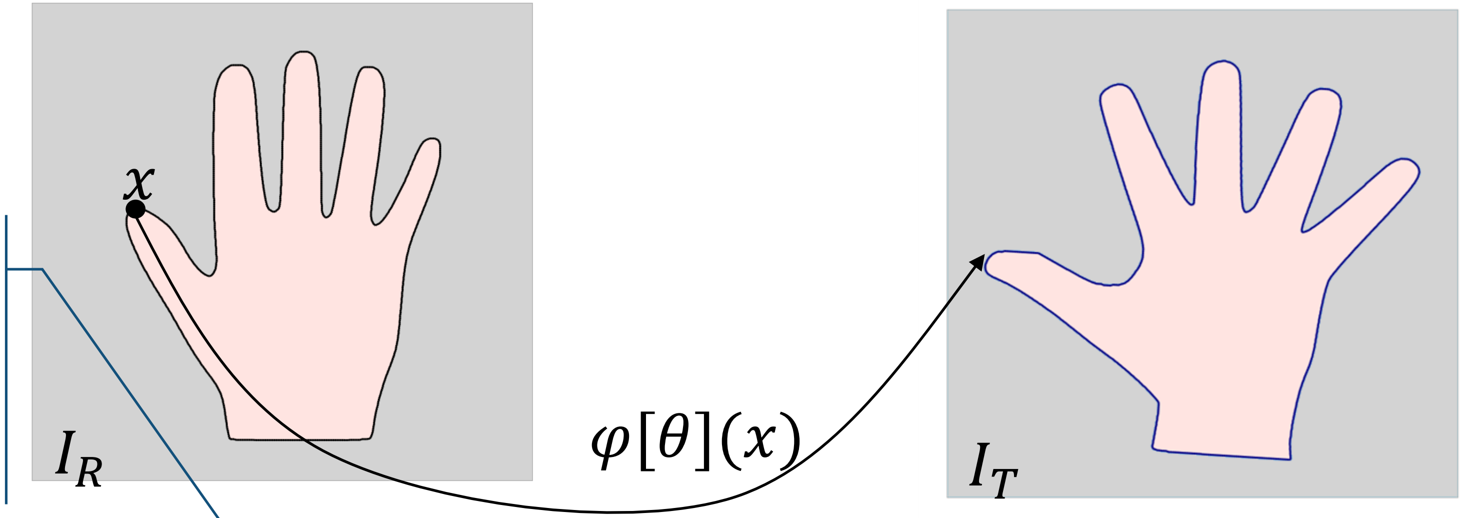
Likelihood function: Image registration

Images are similar when the intensities match

Assumptions:

- Corresponding points have the same image intensity (up to i.i.d. noise)

Image term outside mapping function. Makes problem really difficult



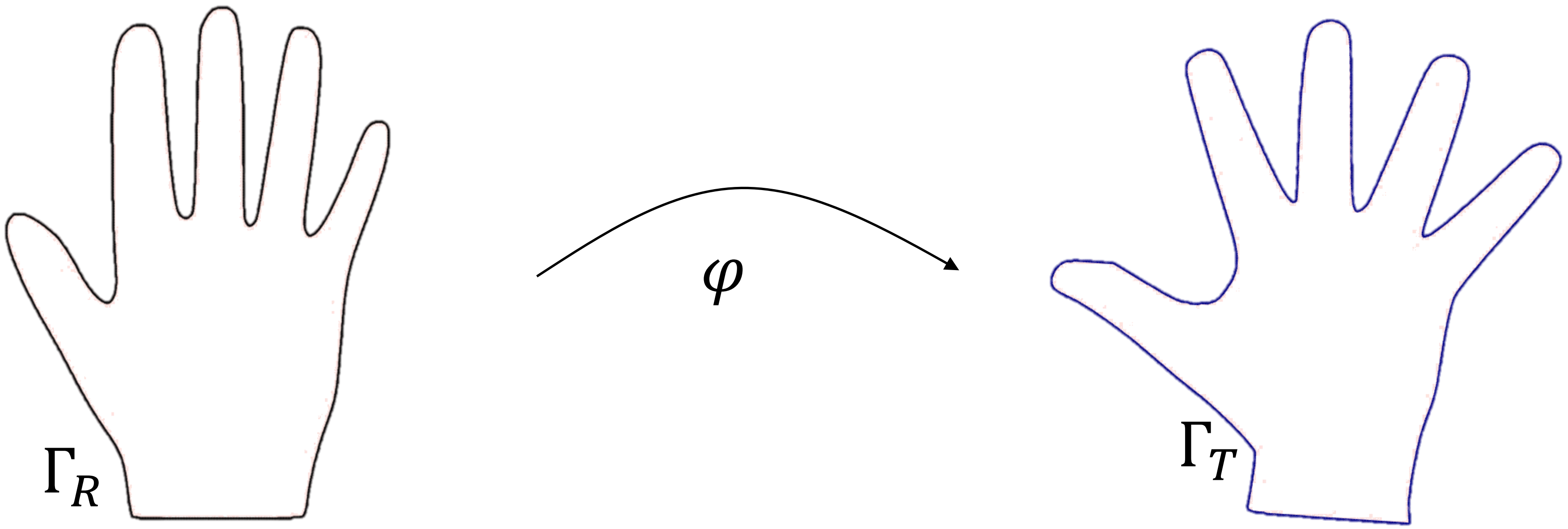
$$p(I_T|I_R,\theta) = \prod_{x\in\Omega} p(I_T(\varphi[\theta](x))|I_R,\theta,x) = \prod_{x\in\Omega} \frac{1}{Z}\exp\left(-\frac{\left(I_T\big(\varphi(x)\big) - I_R(x)\right)^2}{\sigma^2}\right)$$

Image vs. Landmark registration

- Landmark registration is easy
  - All components are Gaussian
  - Closed form solution using Gaussian process regression

- Image registration is hard
  - Image destroys Gaussian assumption
  - Likelihood function is not Gaussian
  - Problem with many local minima

What about surface registration?



Reference (surface):
Γ_R

Target (surface):
Γ_T

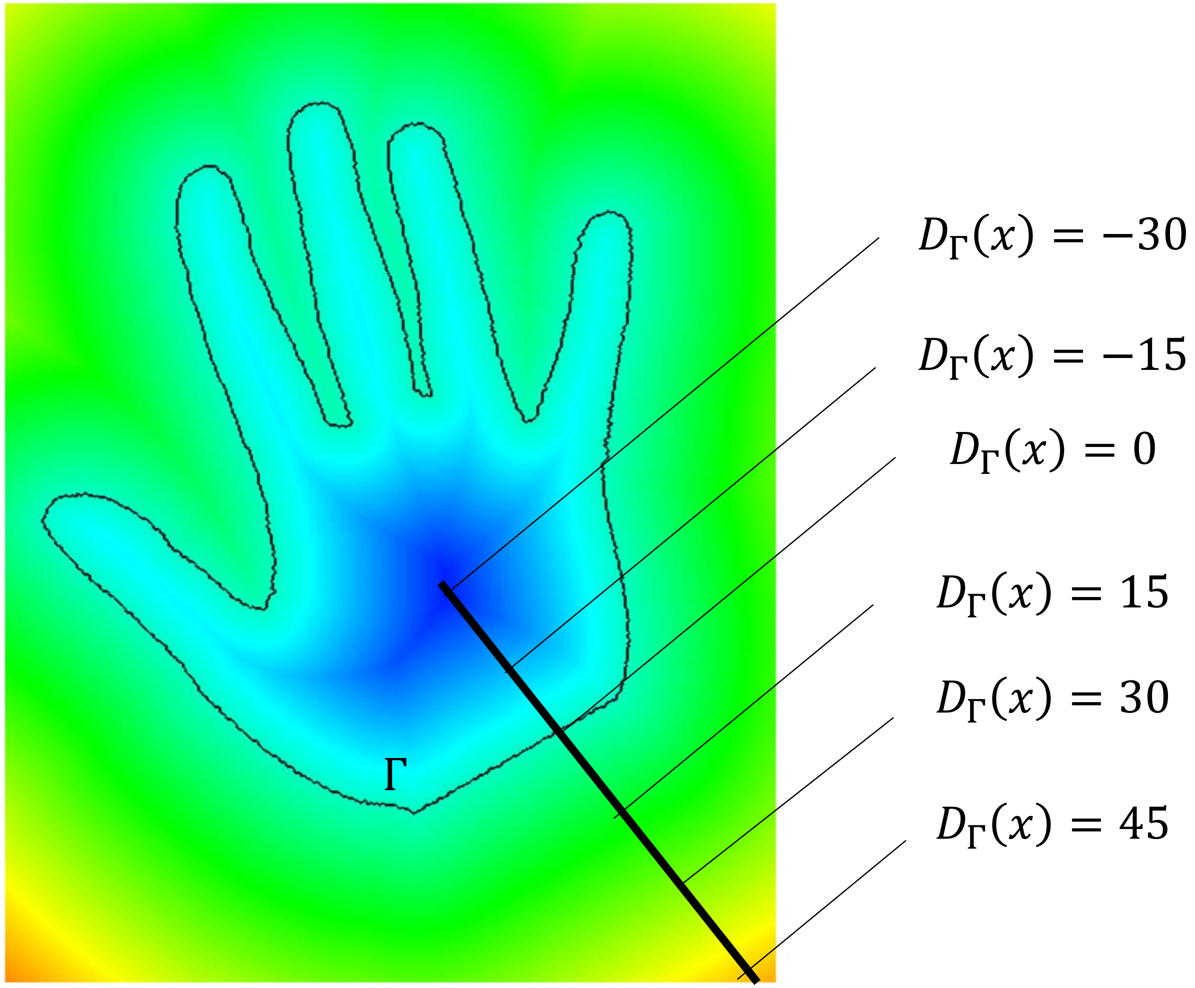
A trick: Implicit definition of a surface

- Surface Γ can be represented as the zero level set of a distance function defined as

$$D_\Gamma(x) = \|\text{ClosestPoint}_\Gamma(x) - x\|$$

with

$$\text{ClosestPoint}_\Gamma(x) = \arg\min_{x' \in \Gamma} \|x - x'\|$$

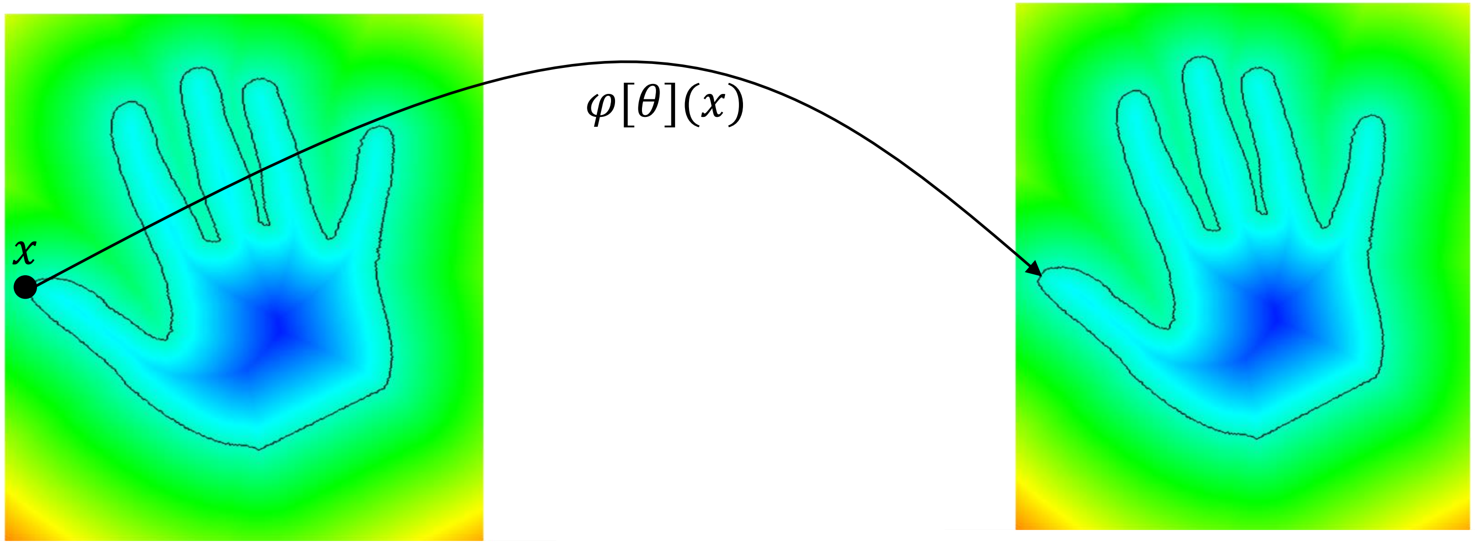


Likelihood function: Surface registration

Surface registration becomes image registration of distance images:

$$p(D_T(\varphi[\theta](x))|\theta, D_R, x) \sim N(D_R(x), \sigma^2)$$

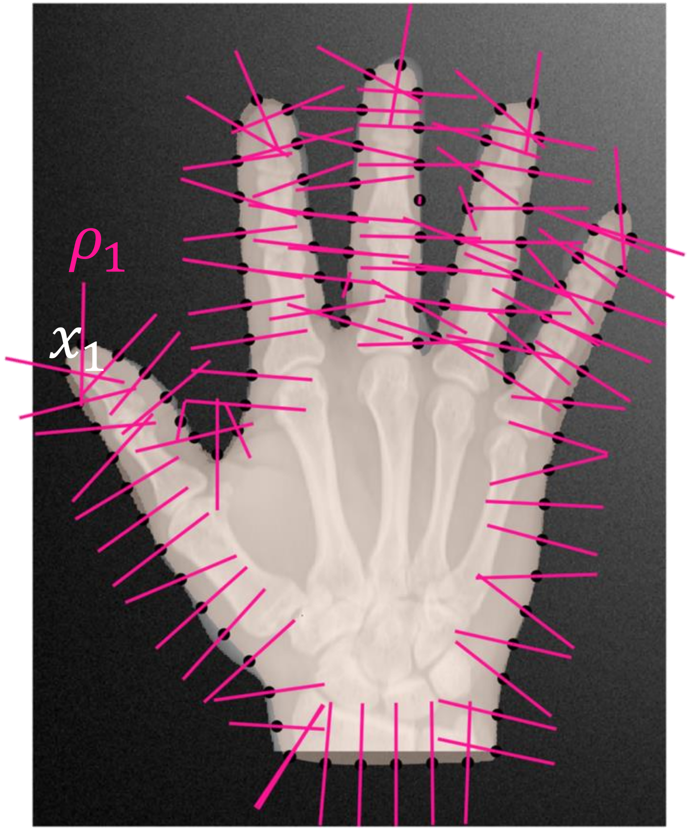
- Most likely solution: Points with same distances are mapped to each other
- σ^2 has now geometric interpretation



Reference $D_R: \Omega_R \rightarrow \mathbb{R}$ Target $D_T : \Omega_T \rightarrow \mathbb{R}$

Likelihood function: Active shape models

Shape is well matched if environment around profile points is likely under trained model.



- ASMs model each profile $\rho(x_i)$ as a normal distribution

$$p(\rho(x_i)) = N(\mu_i, \Sigma_i)$$

Extracts profile (feature) from image

- Single profile point x_i:

$$p(\rho(\varphi[\theta](x_i))|\theta, x_i) = N(\mu_i, \Sigma_i)$$

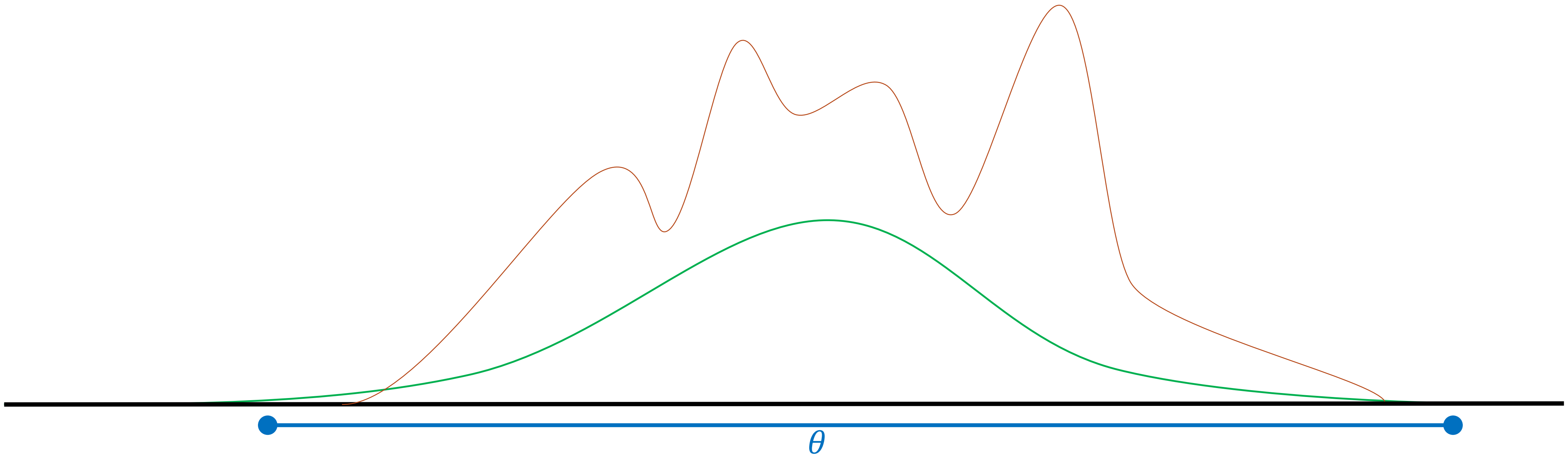
- Image likelihood:

$$p(\rho(\varphi[\theta](x))|\theta, \Gamma_R) = \prod_i N(\mu_i, \Sigma_i)$$

A selection of useful GP priors

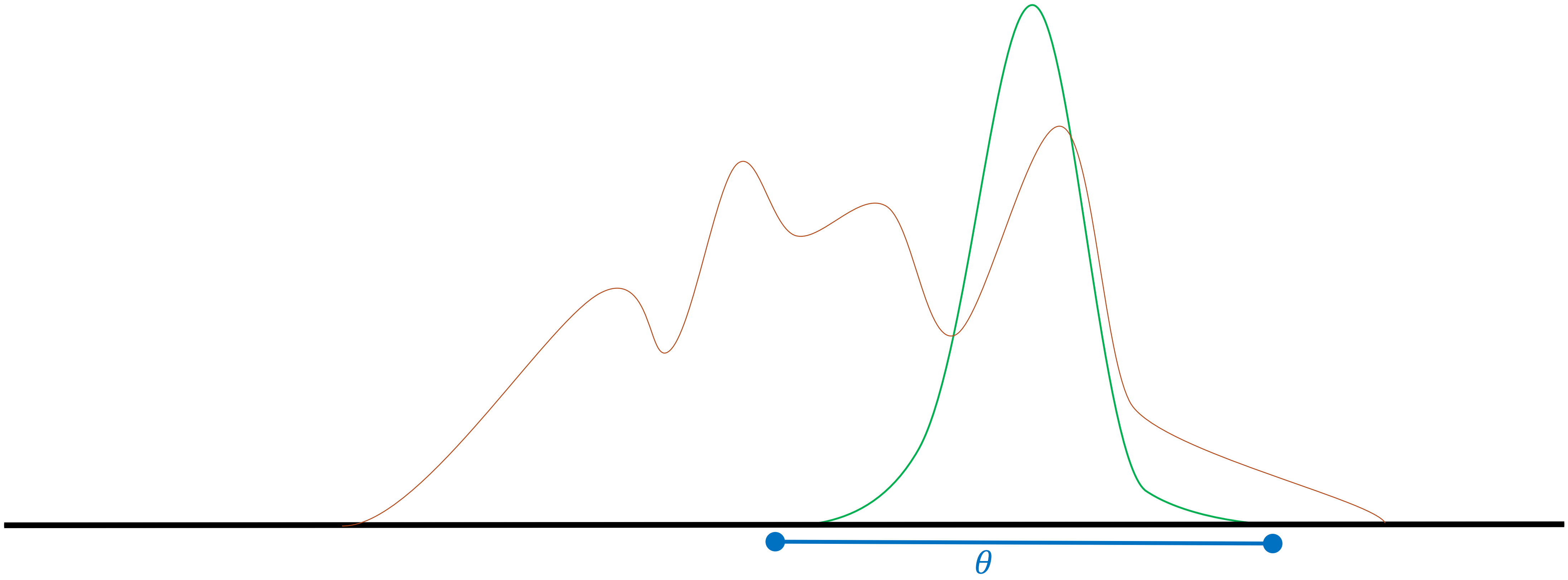
Why are priors interesting?

$$\theta^* = \arg\max_{\theta} p(\varphi[\theta])p(I_T|I_R,\varphi[\theta])$$



Why are priors interesting?

$$\theta^* = \arg\max_{\theta} p(\varphi[\theta])p(I_T|I_R, \varphi[\theta])$$



Constructing s.p.d. kernels

1. $k(x, x') = f(x)\, f(x')^T, f: X \rightarrow \mathbb{R}^d$
2. $k(x, x') = \alpha k_1(x, x'), \alpha \in \mathbb{R}_+$ (scaling)
3. $k(x, x') = B^T k_1(x, x') B, \quad \mathrm{B} \in \mathbb{R}^{r \times d}$ (lifting)
4. $k(x, x') = k_1(x, x') + k_2(x, x')$ (or relationship)
5. $k(x, x') = k_1(x, x') \cdot k_2(x, x')$ (and relationship)

Multi-scale kernels

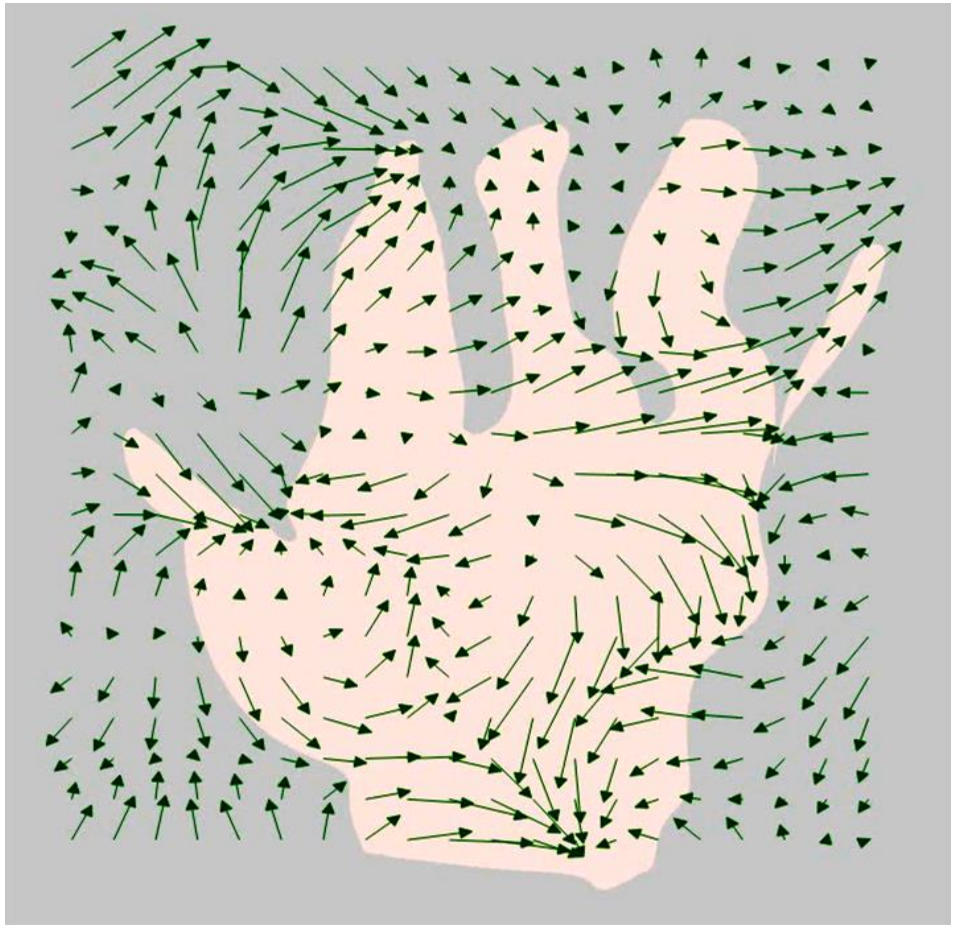
Add kernels that act on different scales:

$$k(x, x') = \sum_{i=0}^{n} \sum_{k \in \mathbb{Z}^d} \beta(2^{-i}x - k)\beta(2^{-i}x' - k)$$

- Wavelet like multiscale representation

Opfer, Roland. "Multiscale kernels."
Advances in computational mathematics 25.4 (2006): 357-380.

Multi-scale kernel



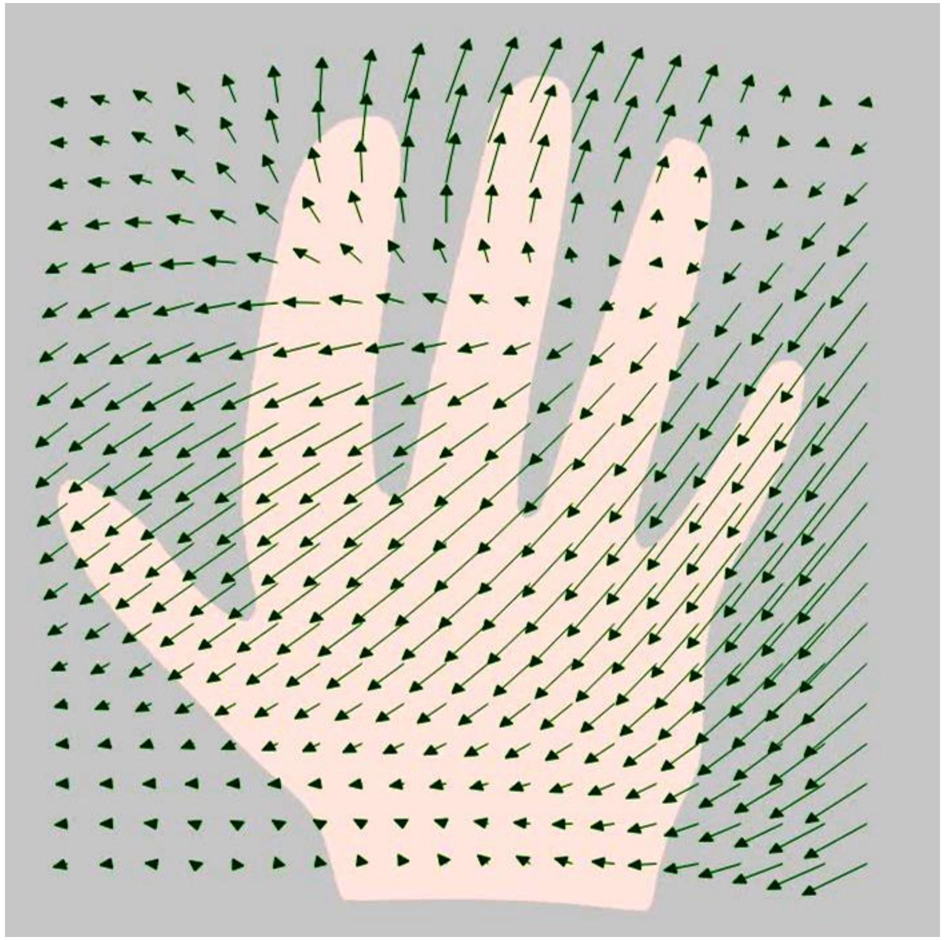
Anisotropic priors

Scale deformations differently in each direction

$$\mathrm{k}(x, x') = R^T \begin{pmatrix} \sqrt{s_1} & 0 \\ 0 & \sqrt{s_2} \end{pmatrix} k(x, x') \begin{pmatrix} \sqrt{s_1} & 0 \\ 0 & \sqrt{s_2} \end{pmatrix} R$$

- R is a rotation matrix
- k is scalar valued
- s_1, s_2 scaling factors

Anisotropic priors

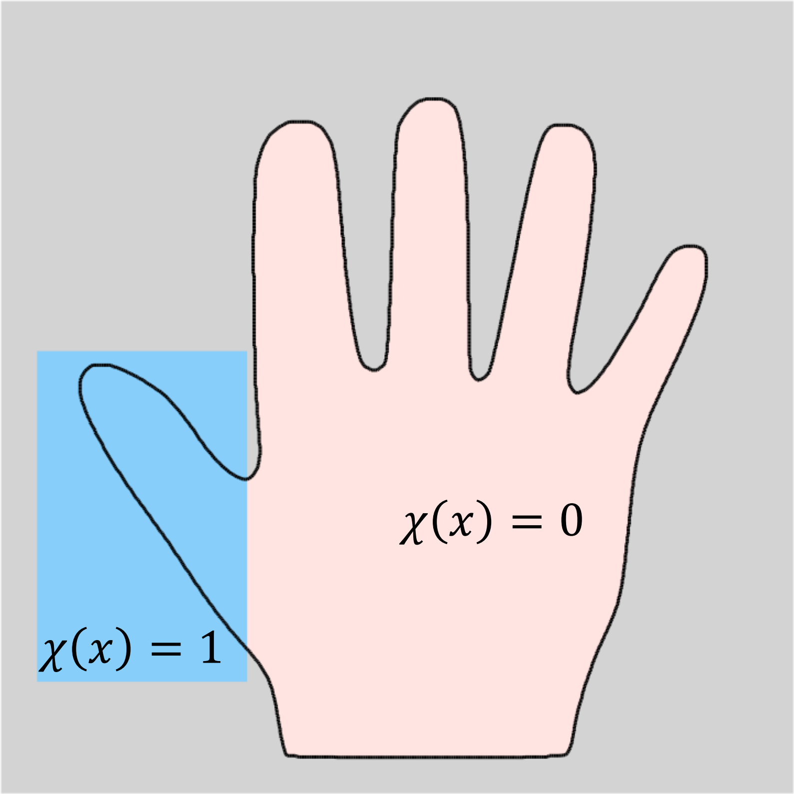


Spatially-varying priors

Use different models for different regions

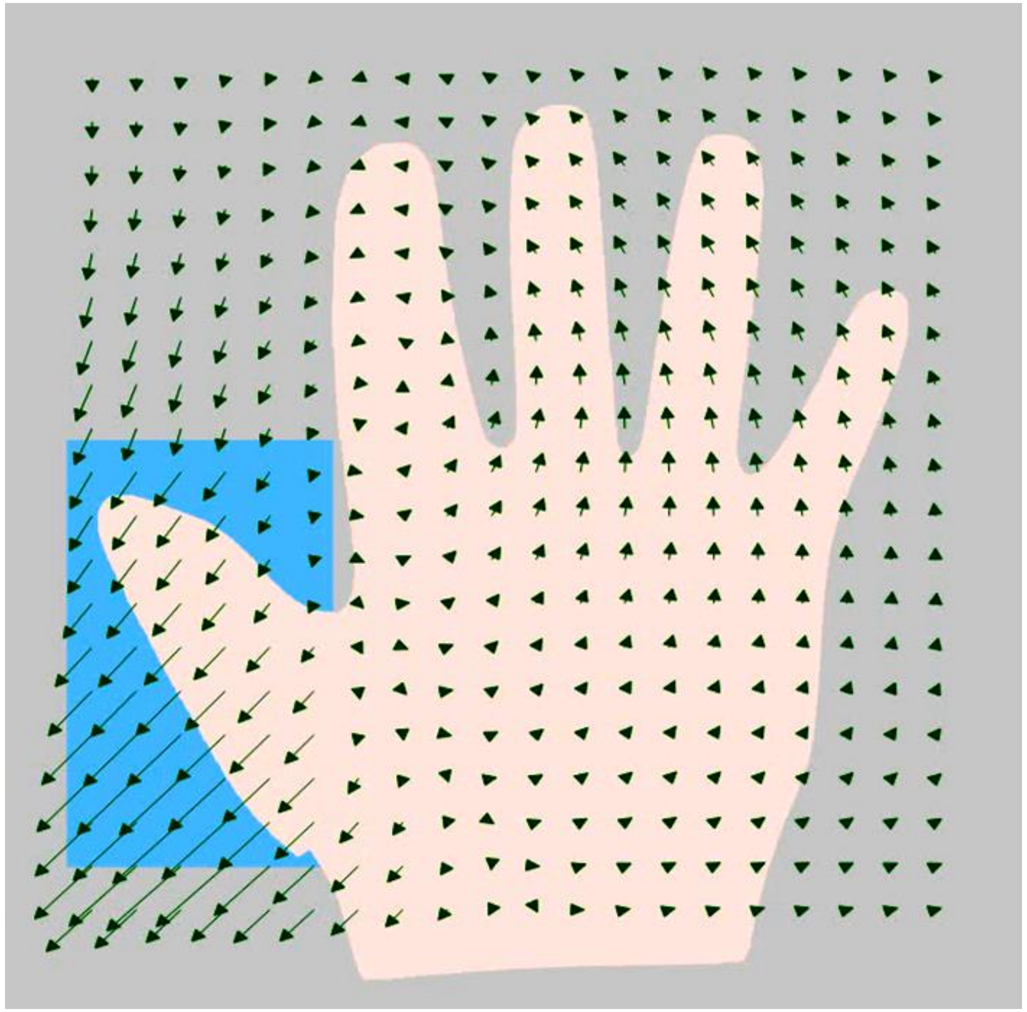
$$k(x,x') = \chi(x)\chi(x')k_1(x,x') + \big(1-\chi(x)\big)(1-\chi(x'))\,k_2(x,x')$$

$$\chi(x) = \begin{cases} 1 & \text{if } x \in \text{thumb region} \\ 0 & \text{otherwise} \end{cases}$$



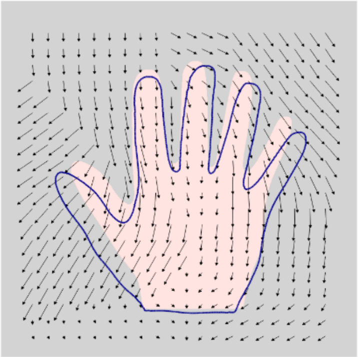
Freiman, Moti, Stephan D. Voss, and Simon K. Warfield. "Demons registration with local affine adaptive regularization: application to registration of abdominal structures." *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011.

Spatially-varying priors

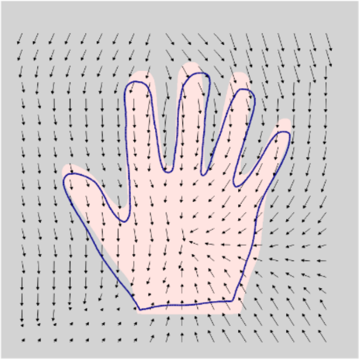


Statistical deformation models

Estimate mean and covariance function from data:

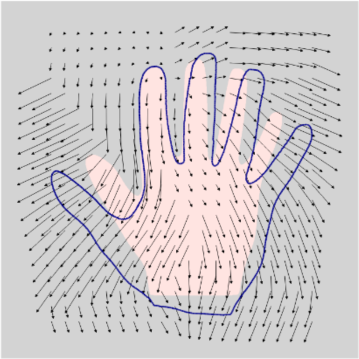


$u^1 : \Omega \rightarrow \mathbb{R}^2$



$u^2 : \Omega \rightarrow \mathbb{R}^2$

…

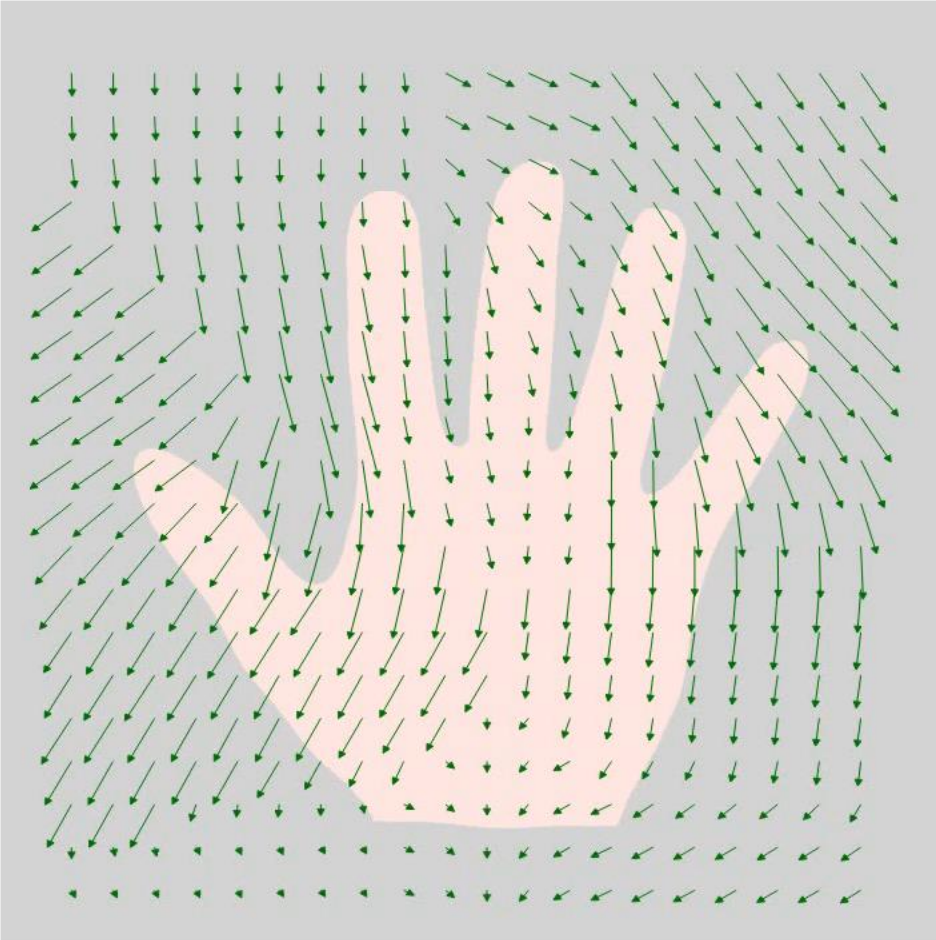


$u^n : \Omega \rightarrow \mathbb{R}^2$

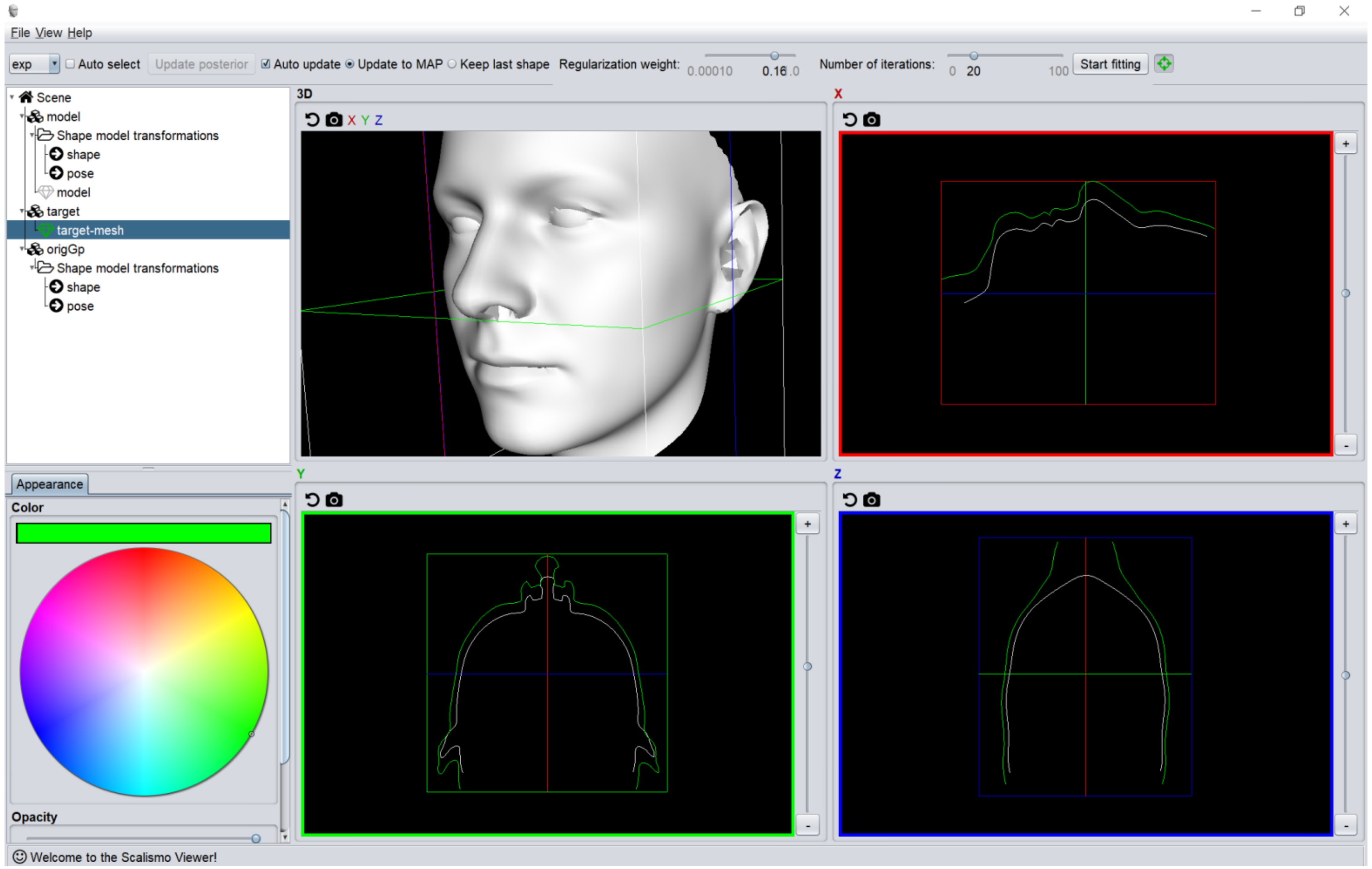
$$\mu(x) = \overline{u}(x) = \frac{1}{n}\sum_{i-1}^{n} u^i\,(x)$$

$$k_{SM}(x, x') = \frac{1}{n-1}\sum_{i}^{n}(u^i(x) - \overline{u}(x))\big(u^i(x') - \overline{u}(x')\big)^T$$

Example 5: Statistical deformation models



Demo: Priors and interactive registration

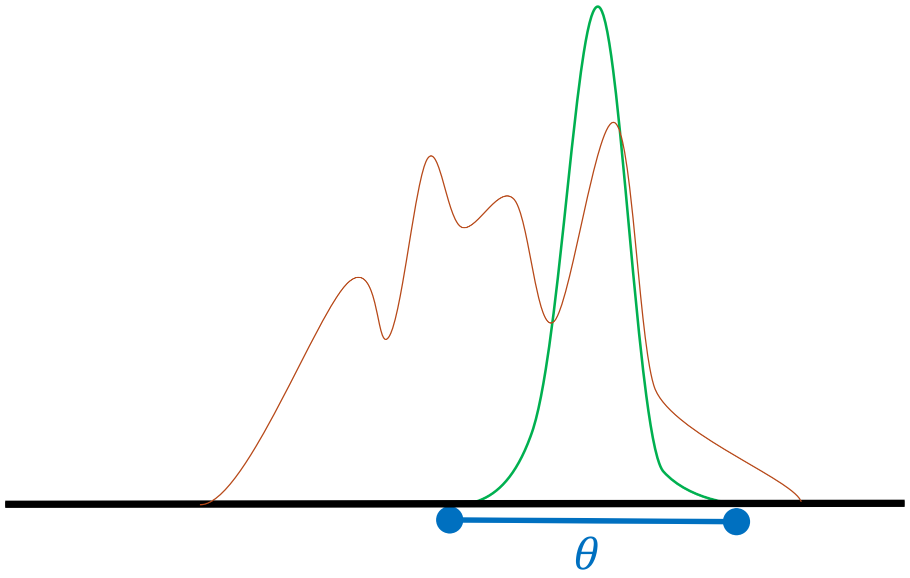


Optimization

The optimization problem

$$\theta^* = \arg\max_{\theta} p(\varphi[\theta])p(I_T|I_R \circ \varphi[\theta])$$

- The final problem is a difficult optimization problem
- Possibly many local minima
- Non-linearity due to image term
  - Not possible to avoid it
- Flexible models makes things worse



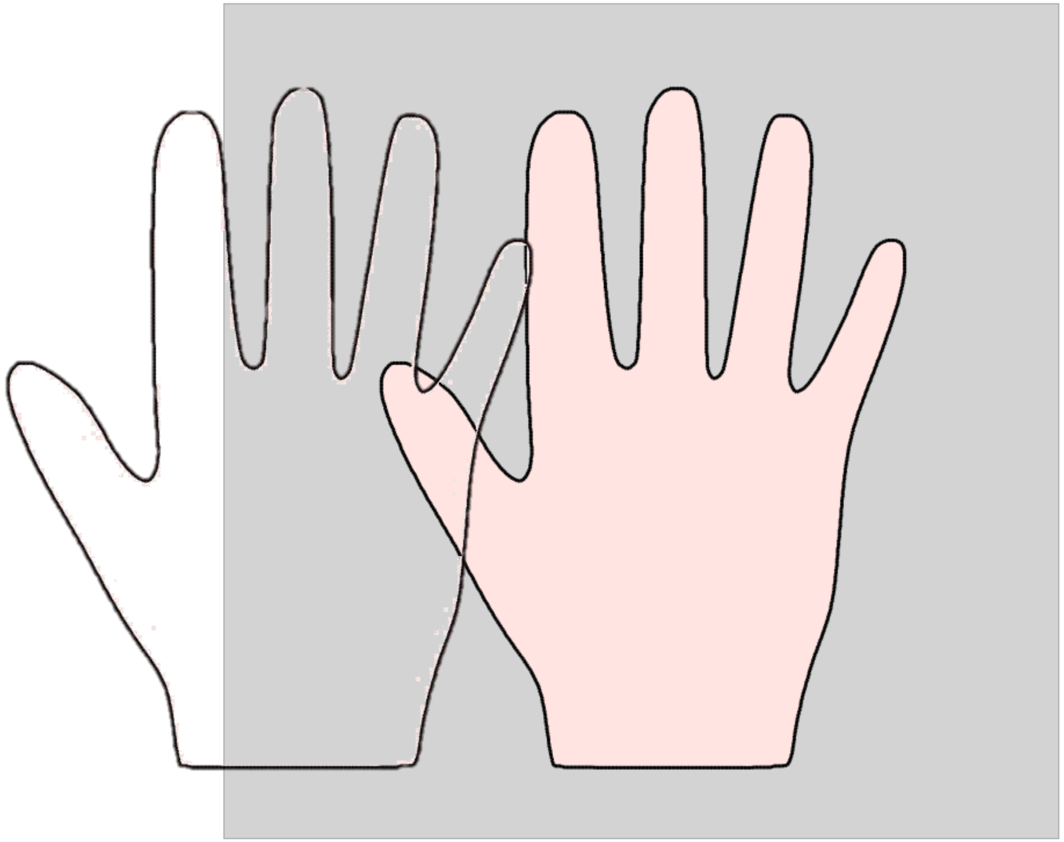
Local minima

- Rigid Transformation
  - Minima due to structure of object

  Possible approach: Multi-resolution

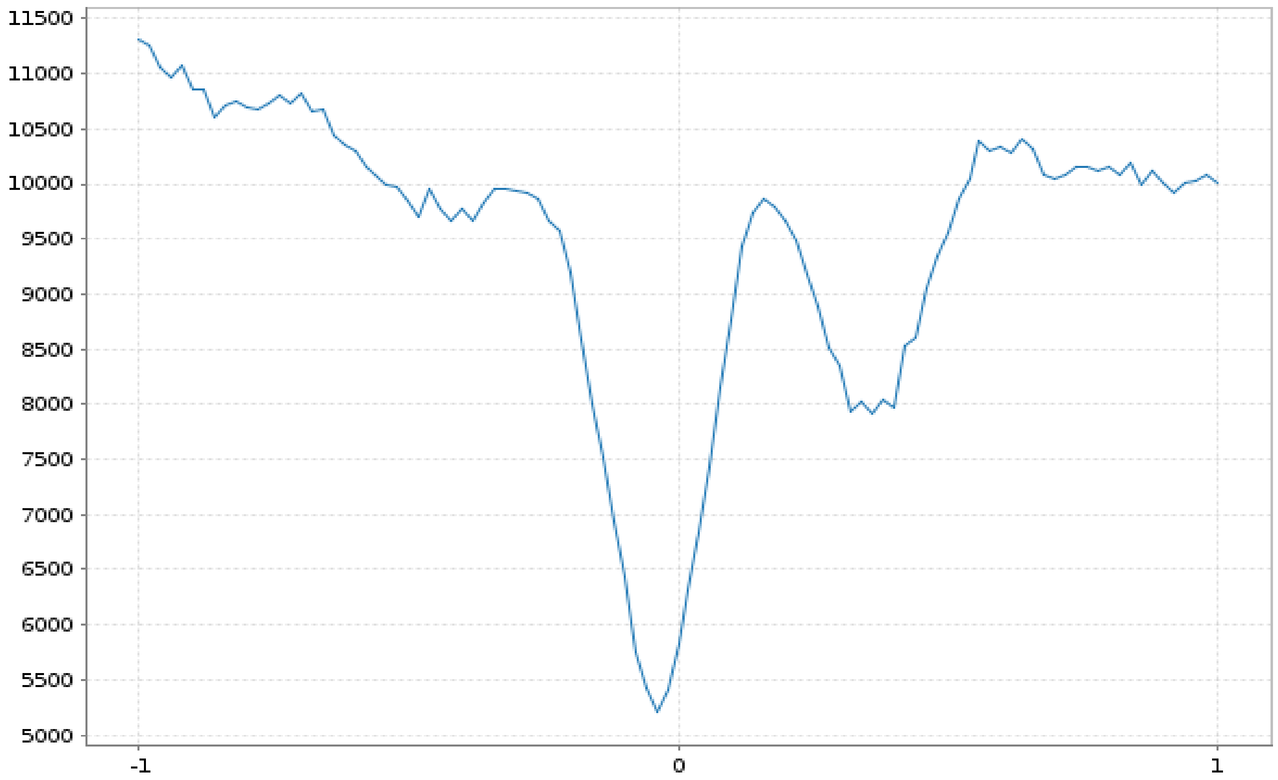
- Non-rigid Transformation
  - Minima appear/dissappear when shape changes

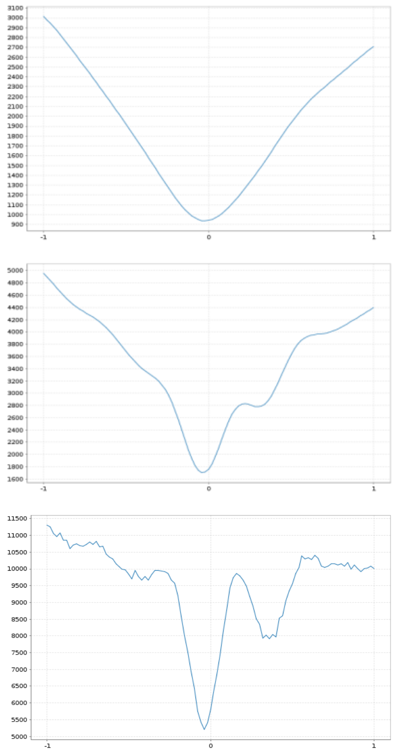
  Possible approach: Multi-scale models, regularization



Multi resolution

Idea: Solve optimzation problem for a sequence of smoothed out objects.





Implementation

- Smooth the input shapes
- For images, achieved by Gaussian blurring

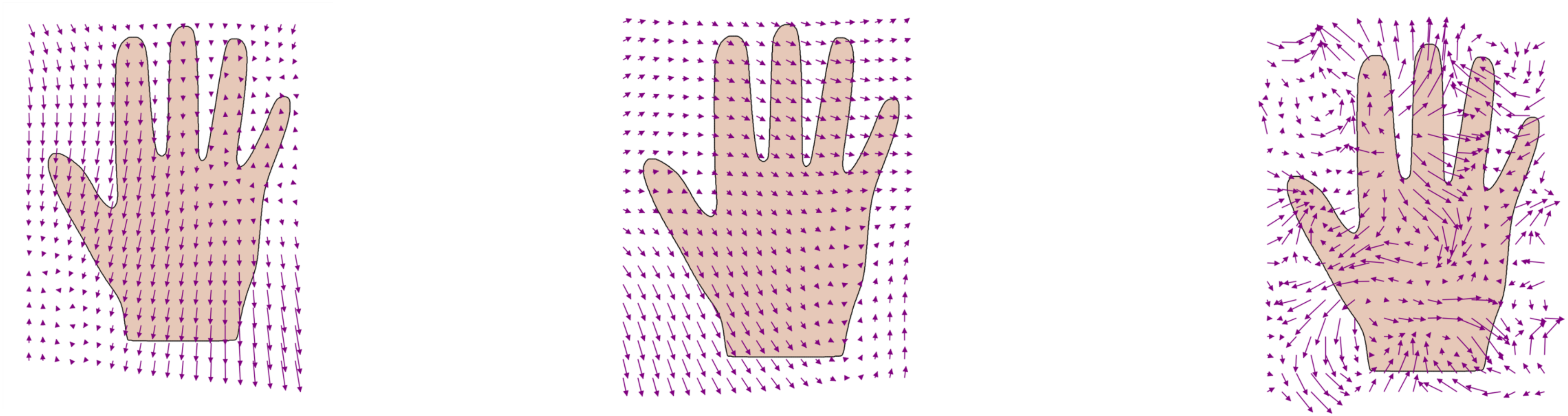
Initial registration → Final registration

Almost no local minima
No-details

Many local minima
All-details

Multi-scale / Regularization

Idea: Solve optimzation problem for a sequence of increasingly detailed deformations



Initial registration

Only large, smooth deformations
Large regularization value

Final registration

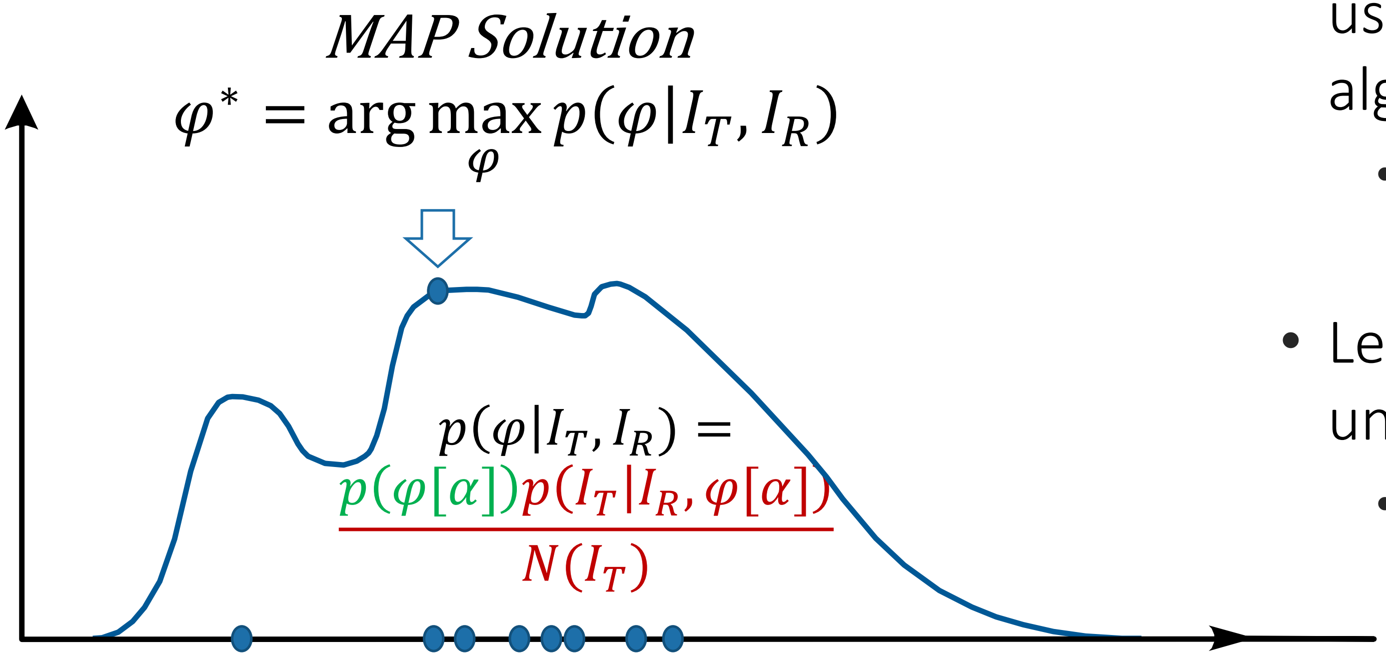
Allow detailed deformations
Almost no regularization

Doing the registration

Strategies:

- Gradient-based registration
  - Compute gradient and use local optimization methods
  - Quasi-Newton schemes , SGD, …
- Gradient free registration
  - Use global optimization method directly on cost function
  - Examples: Simulated annealing, Particle Swarm, …
- ICP-based methods
  - Assume correspondence and solve in each iteration analytic problem
  - Examples: Non-rigid ICP, Active Shape models, CPD

Model-fitting using Markov Chain Monte Carlo



- Can obtain full posterior distribution using the Metropolis Hastings algorithm
  - Needs only point-wise evaluation of unnormlized posterior
- Leads to principled way to integrate unreliable bottom up methods
  - Automatically detected landmarks