graphics and vision gravis

University of Basel

Non-rigid Registration

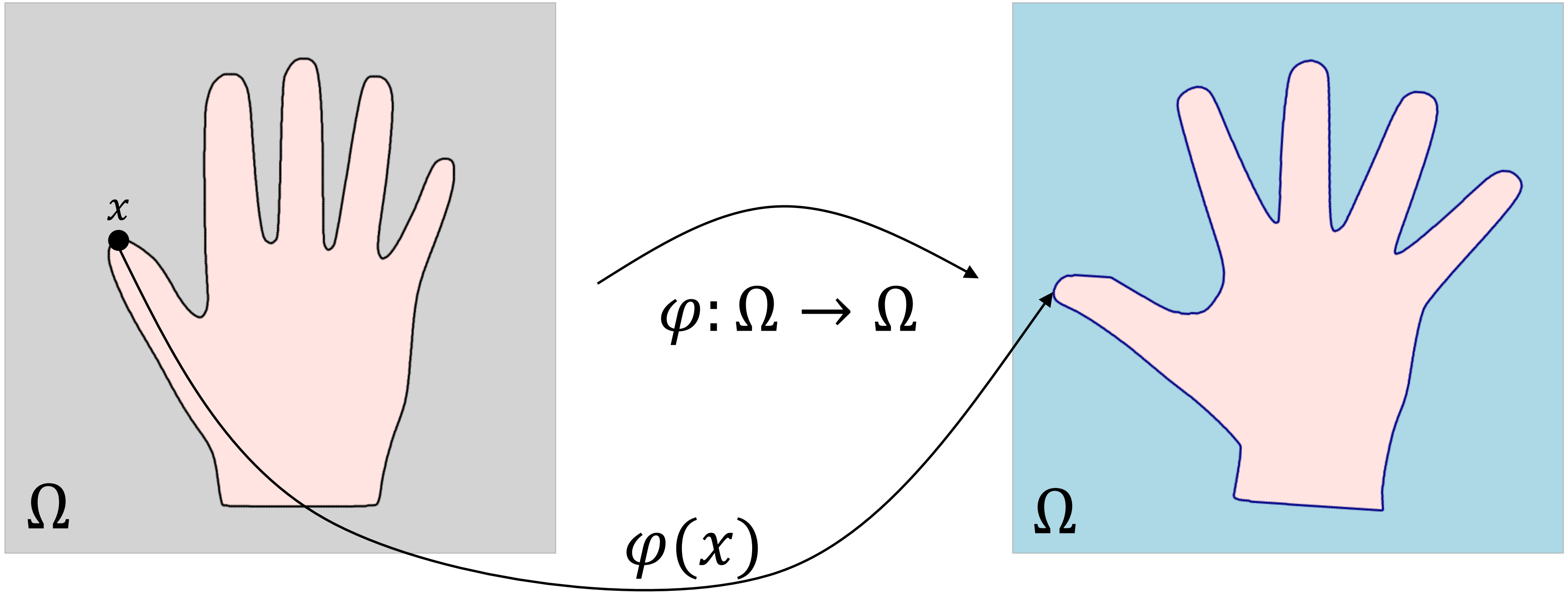
Marcel Lüthi

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

Outline

- Non-rigid registration: The basic formulation
  - Exercise: Parametric registration in Scalismo
- Advanced Priors
- Likelihood functions
  - Exercise: ASMs in Scalismo
- Optimization

The registration problem



Reference:

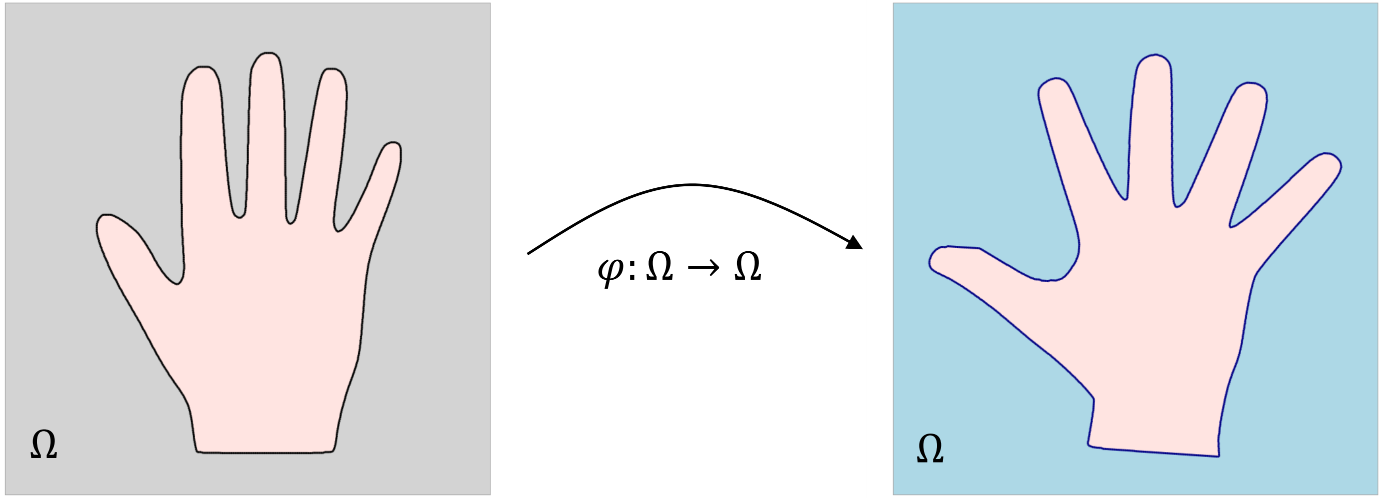
$I_R: \Omega \to \mathbb{R}$

Target:

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

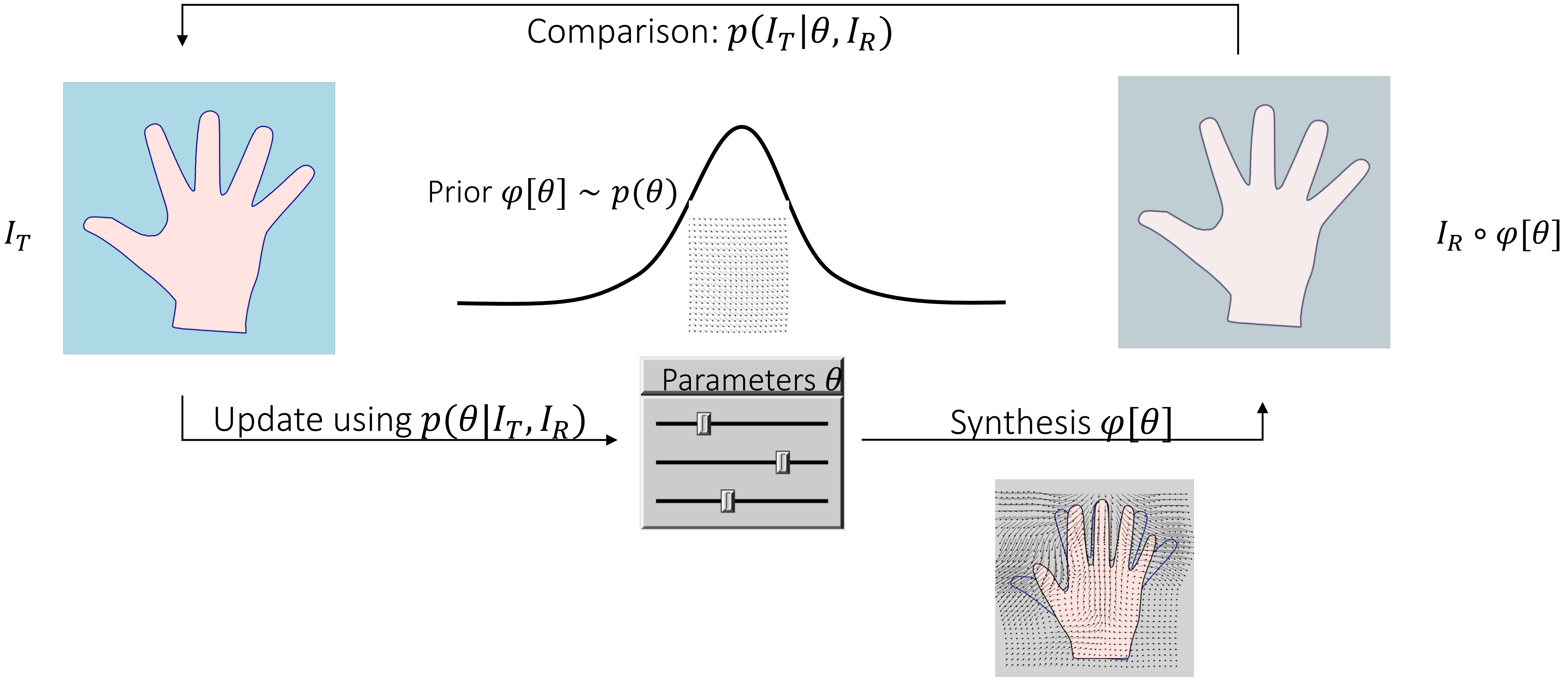
Why is it important?

- 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

Probabilistic formulation

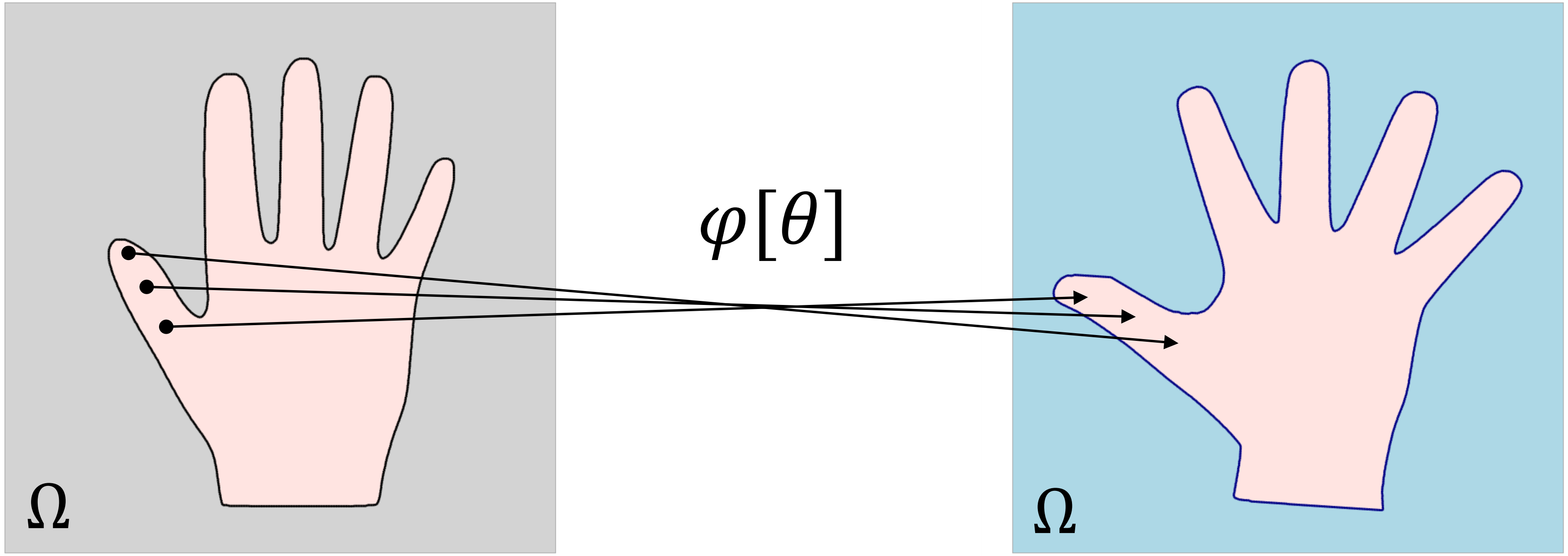
$$\theta^* = \arg\max_{\theta} p(\theta|I_T, I_R) = \arg\max_{\theta} p(\theta)p(I_T|\theta, I_R)$$

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

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

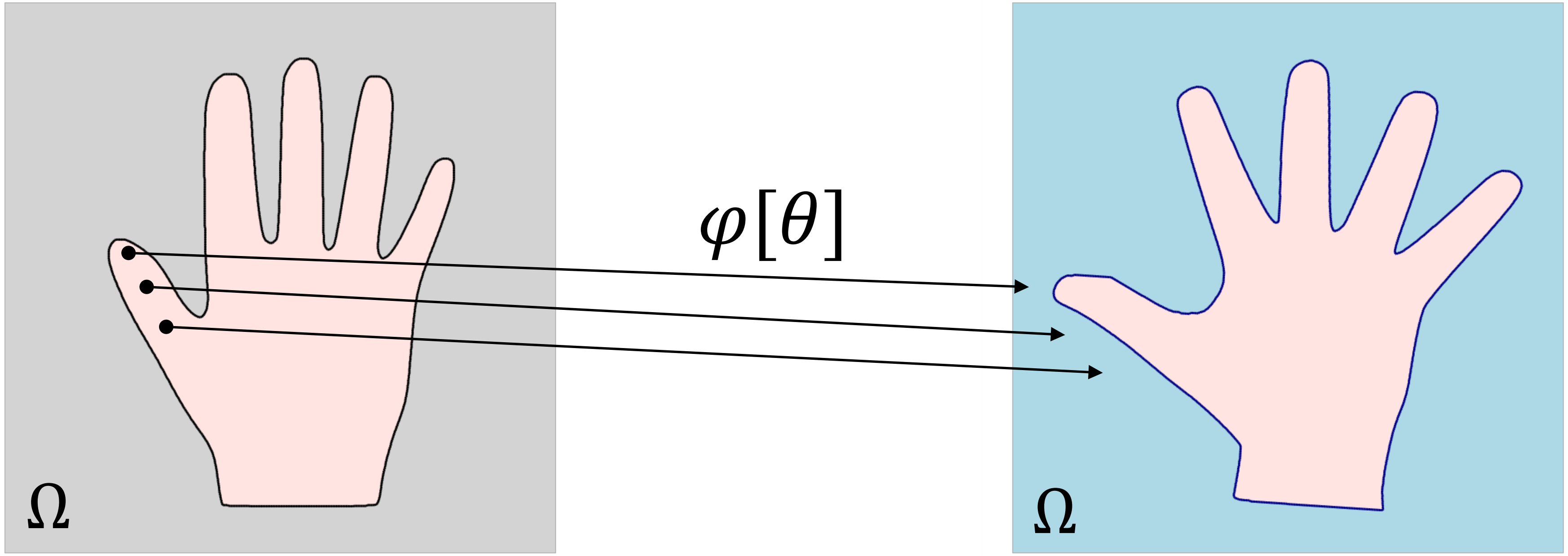
The registration problem

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



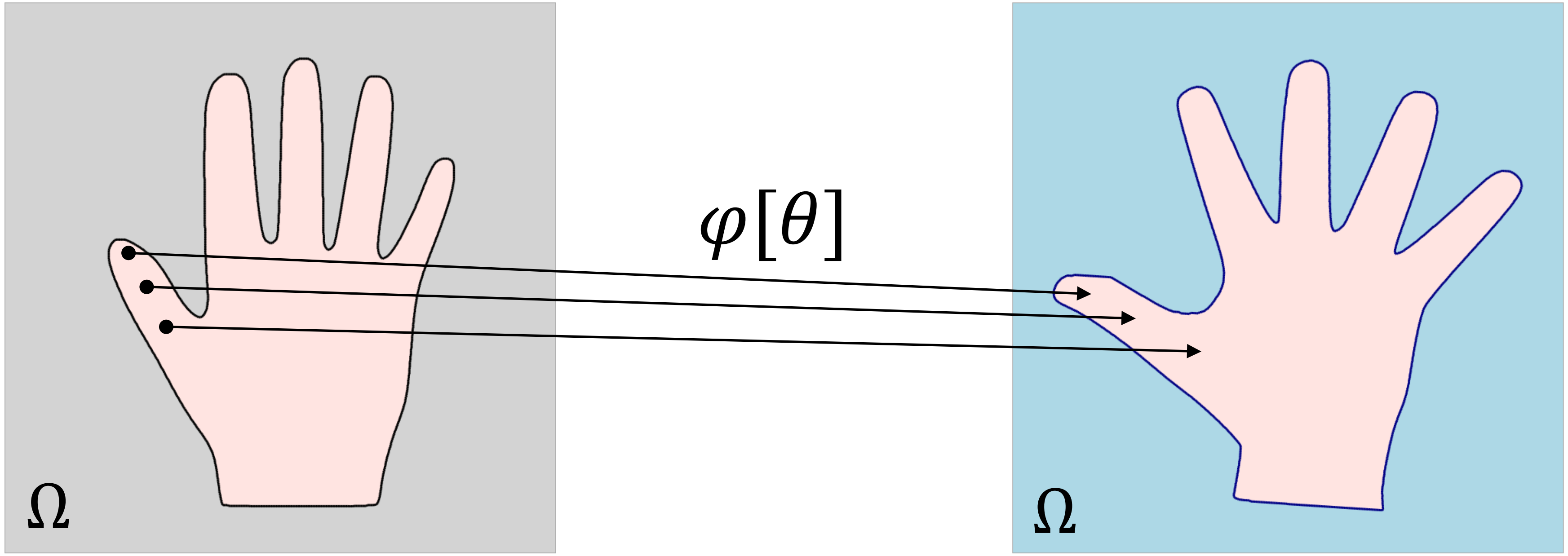
The registration problem

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



The registration problem

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



The registration problem

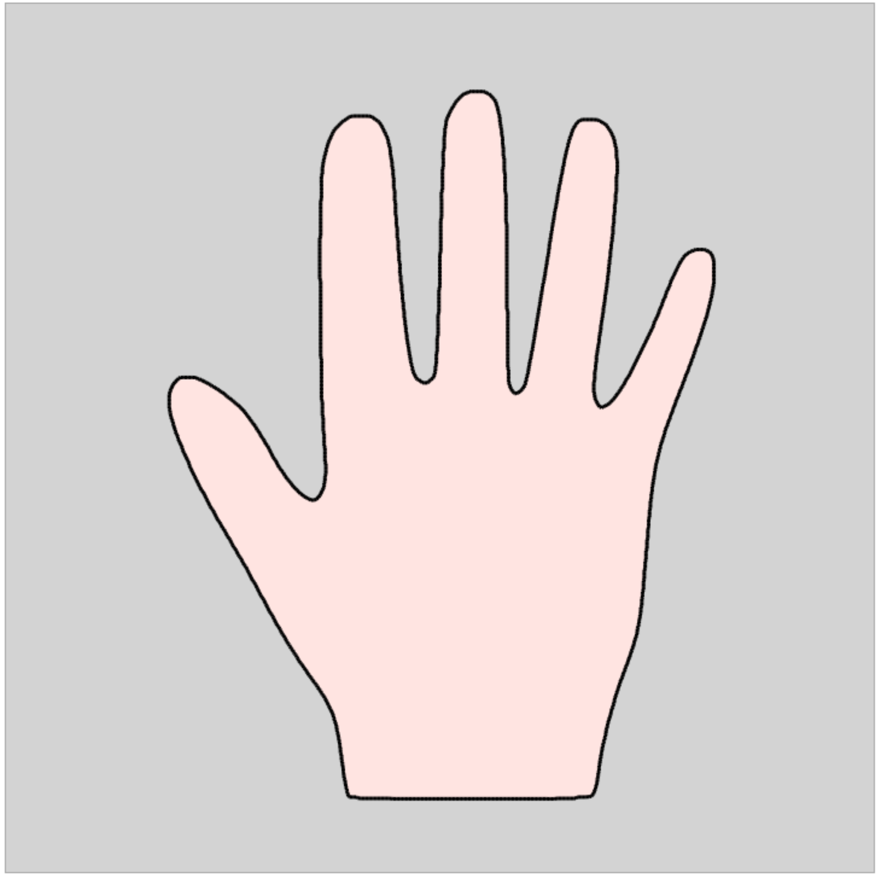
Probabilistic formulation

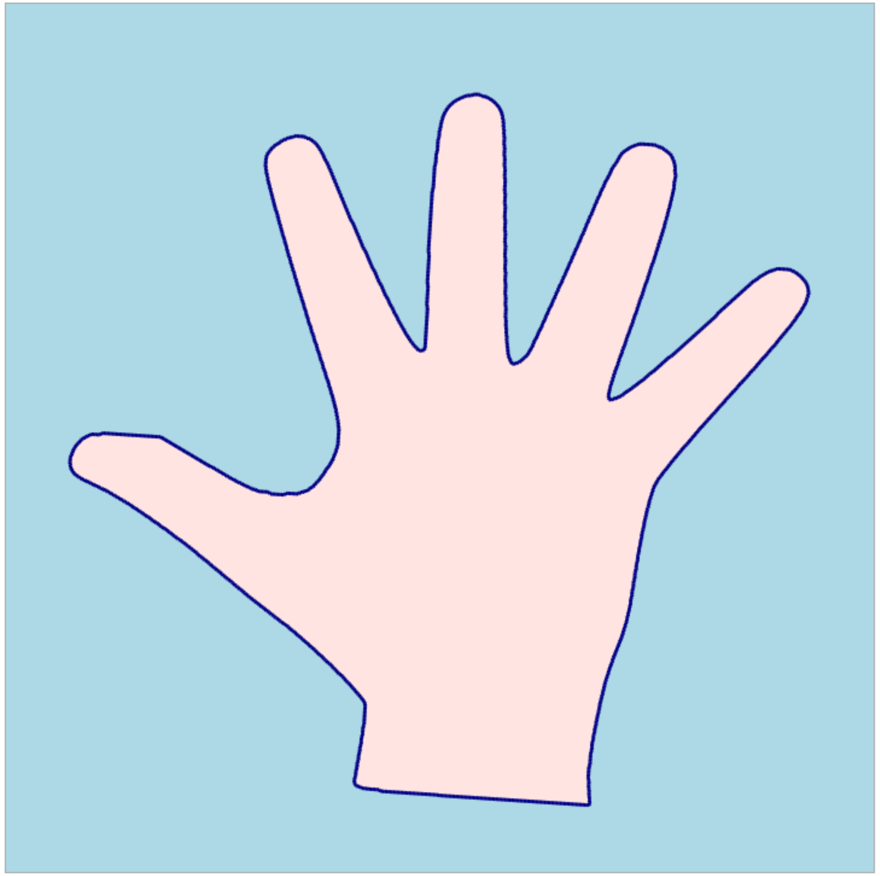
$$\varphi^* = \arg\max_{\varphi} p(\varphi | I_T, I_R) = \arg\max_{\varphi} p(\varphi) p(I_T | \varphi, I_R)$$

Main questions:

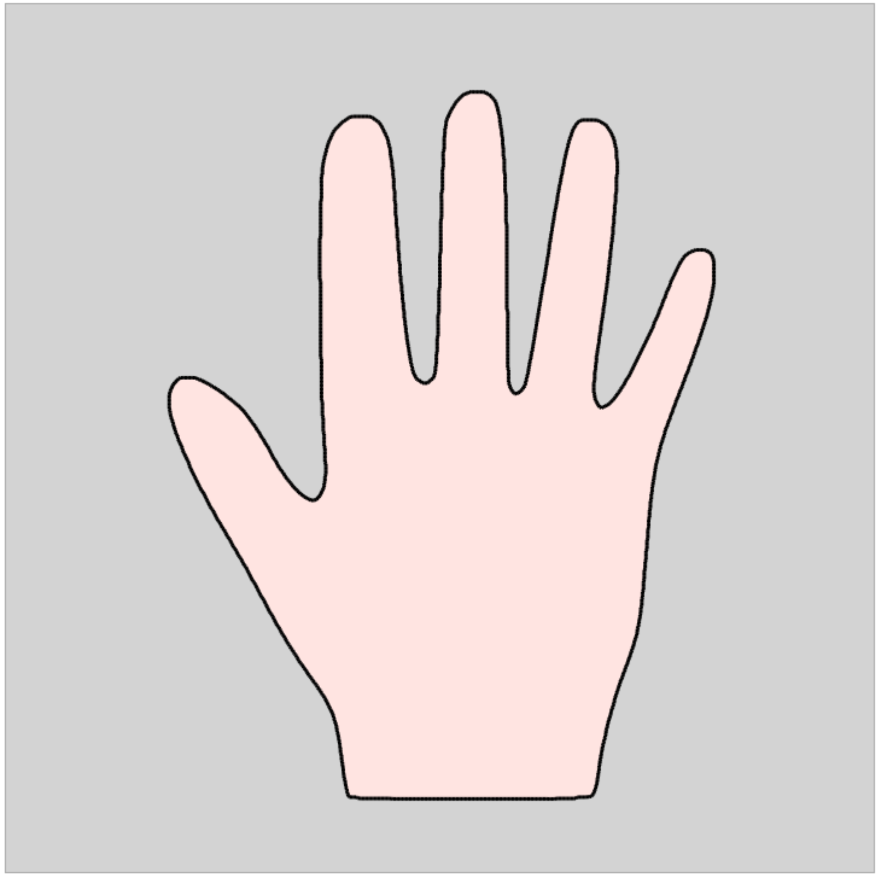
- How do we represent the mapping?
- How do we define the prior?
- What is the likelihood function?
- How can we solve the optimization problem?

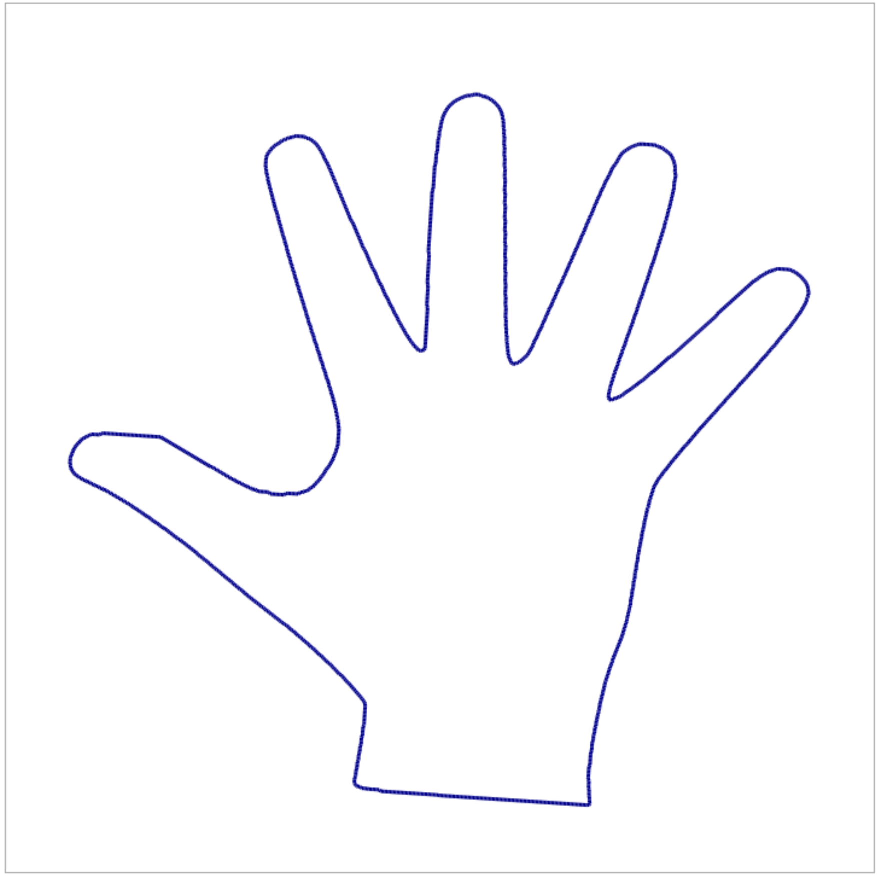
Representation of the mapping φ



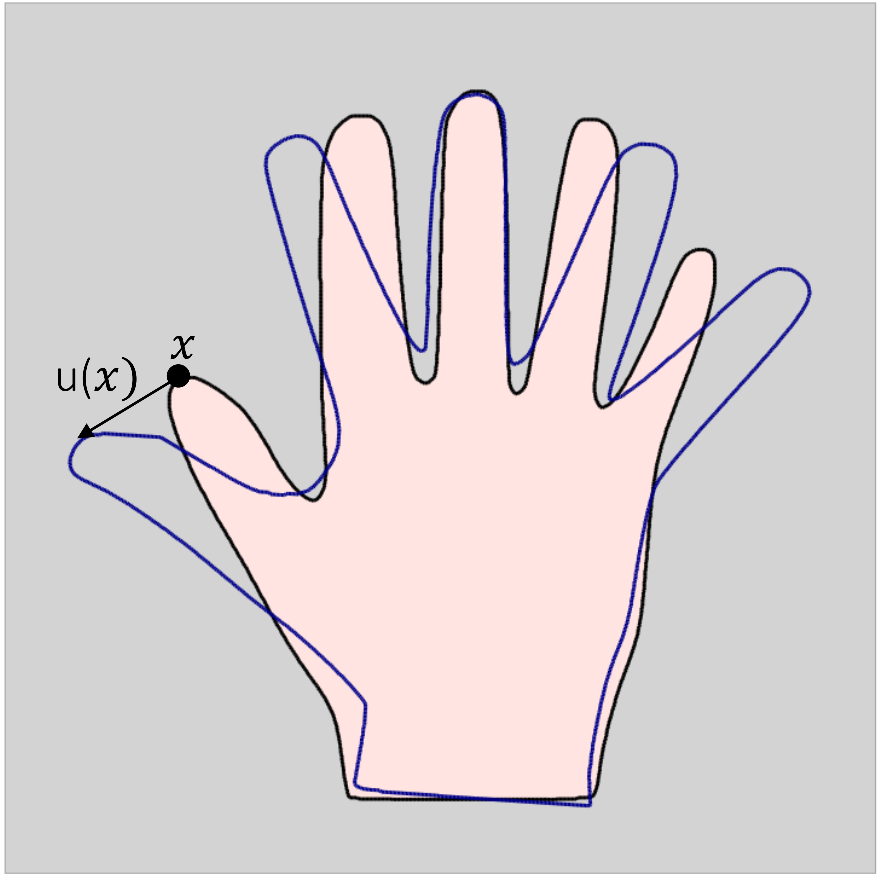


Representation of the mapping φ





Representation of the mapping φ

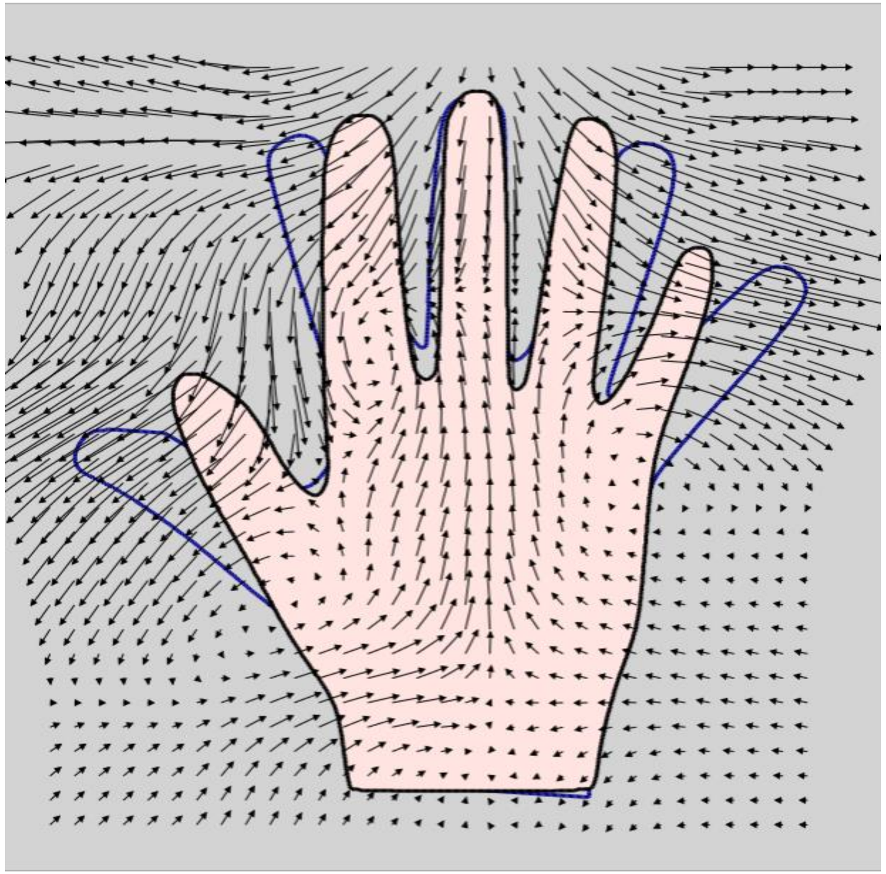


Assumption:
Images are rigidly aligned

- 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



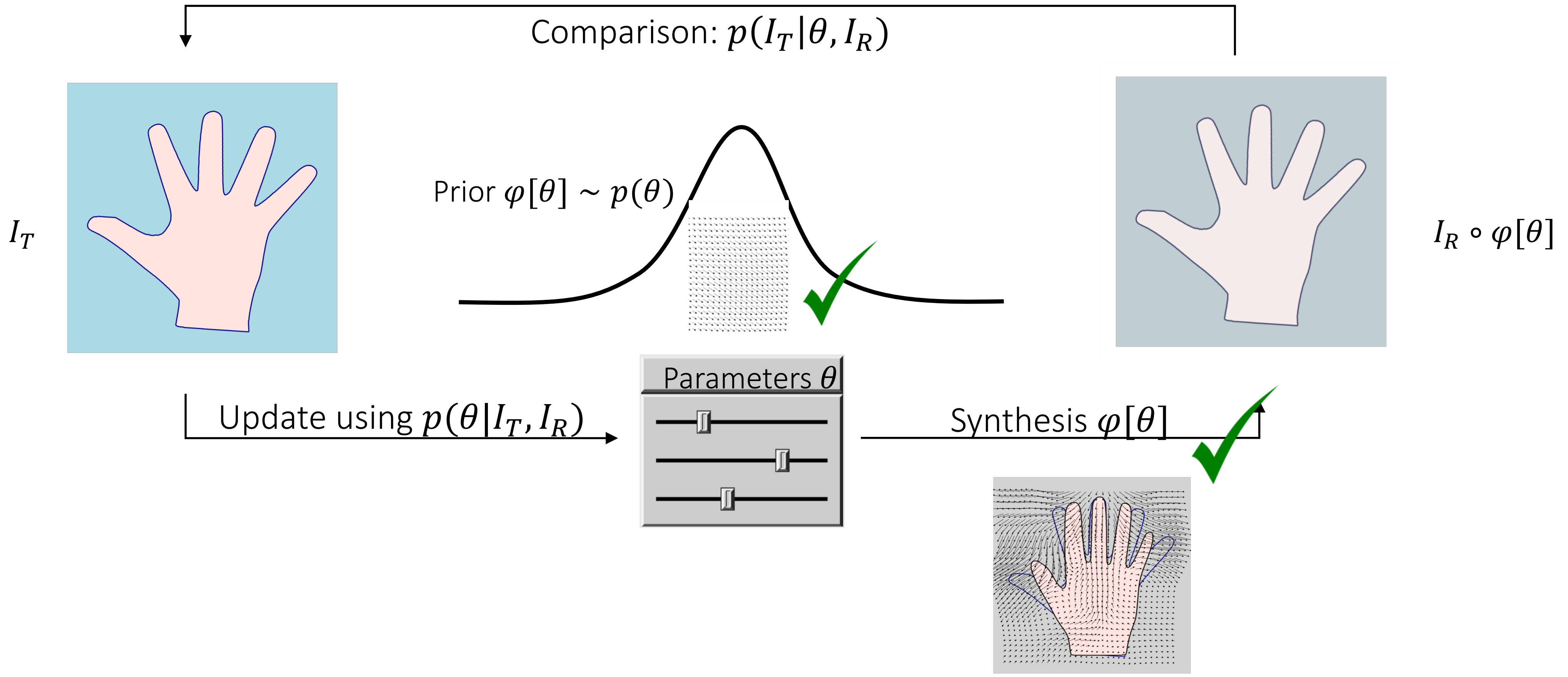
Assumption:
Images are rigidly aligned

- Mapping can be represented as a displacement vector field:

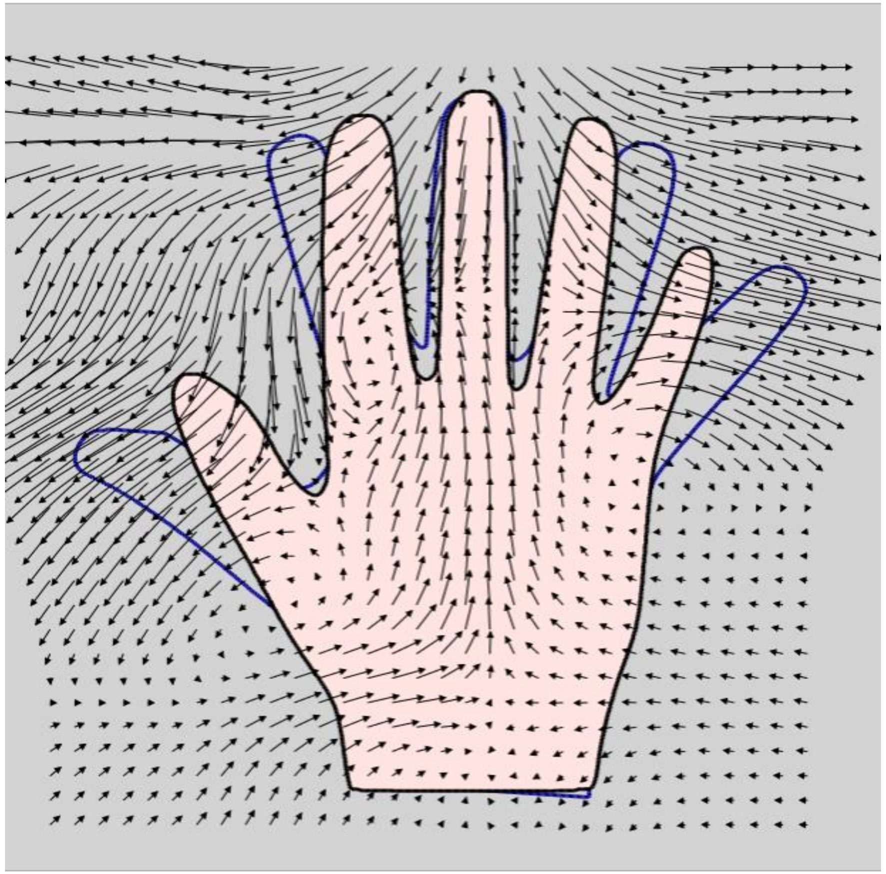
$$\varphi(x) = x + u(x)$$
$$u : \Omega \rightarrow \mathbb{R}^d$$

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

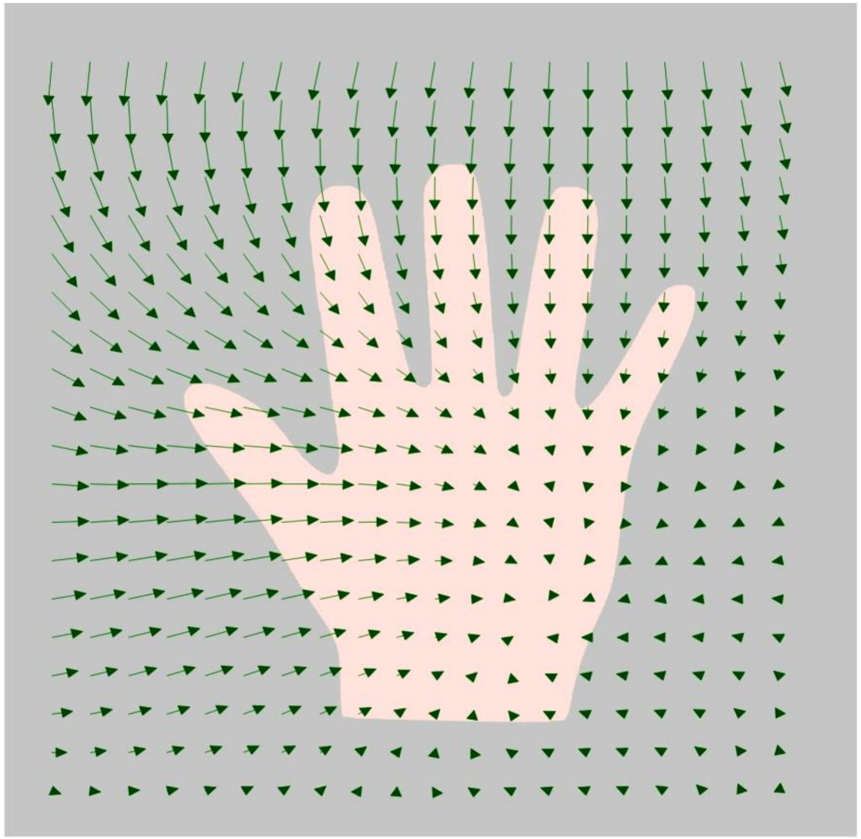
Zero mean:

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

Squared exponential covariance function (Gaussian kernel)

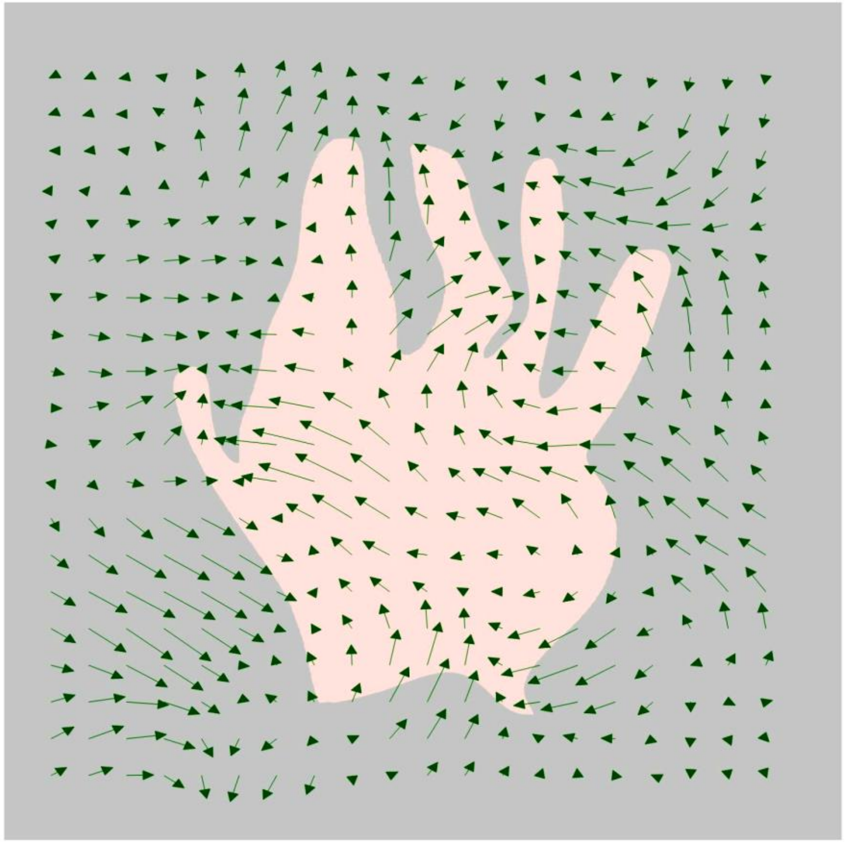
$$k(x, x') = \begin{pmatrix} \mathrm{s}_1 \exp\left(-\frac{\|x - x'\|^2}{\sigma_1^2}\right) & 0 \\ 0 & \mathrm{s}_2 \exp\left(-\frac{\|x - x'\|^2}{\sigma_2^2}\right) \end{pmatrix}$$

Example prior: Smooth 2D deformations



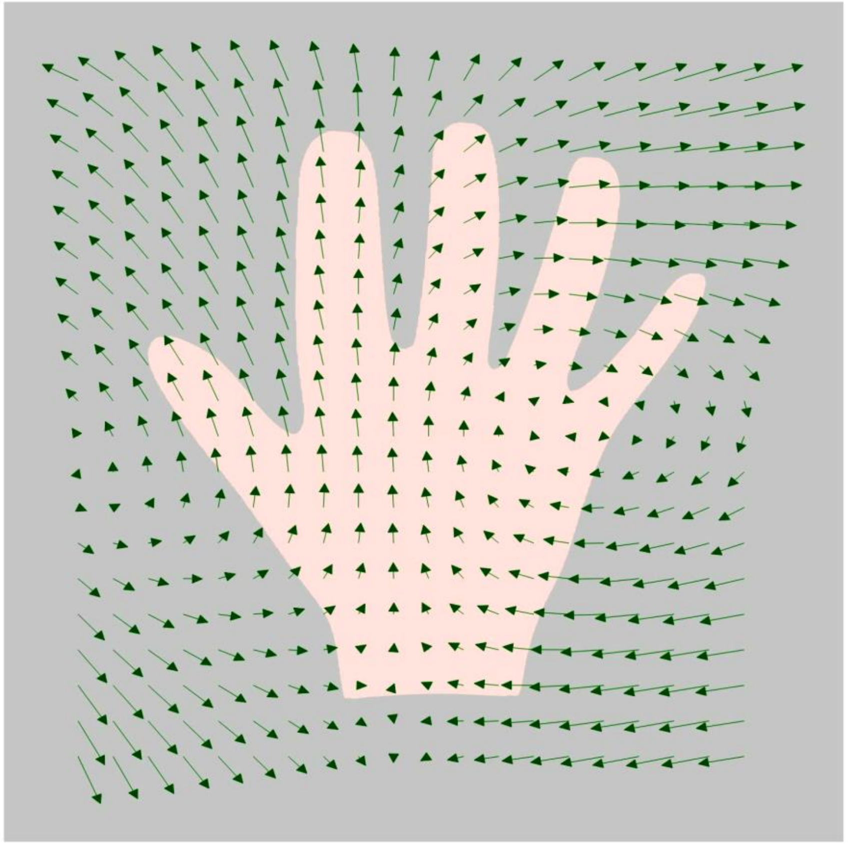
$s_1 = s_2$ small, $\sigma_1 = \sigma_2$ large

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

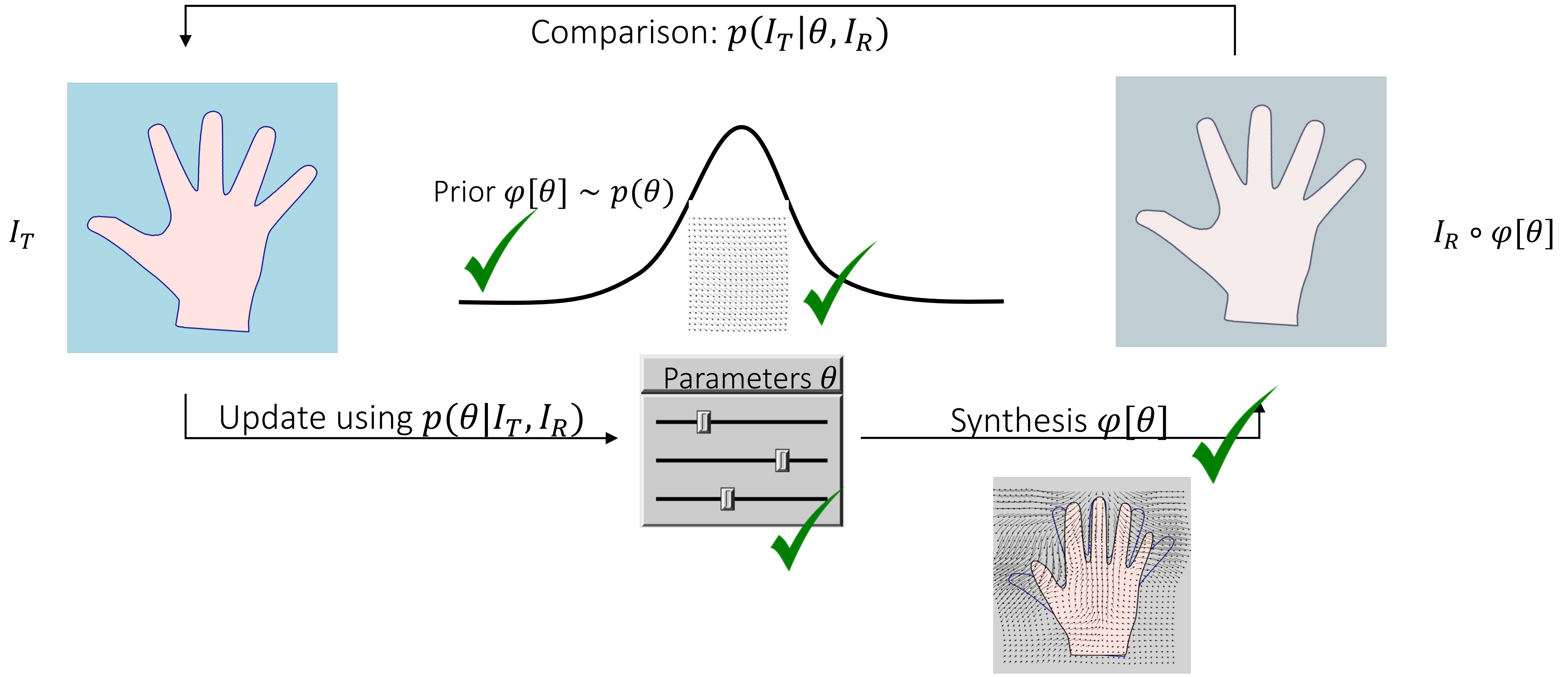
Represent $GP(\mu, k)$ using only the first r components of its KL-Expansion

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

- 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

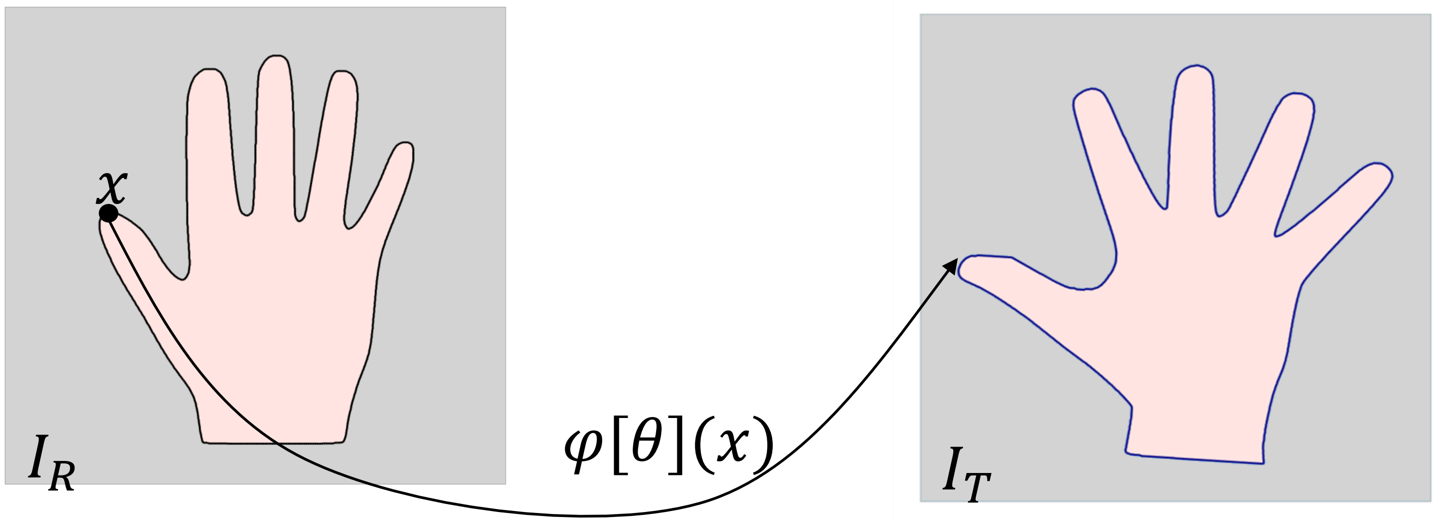


Likelihood function: Image registration

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)

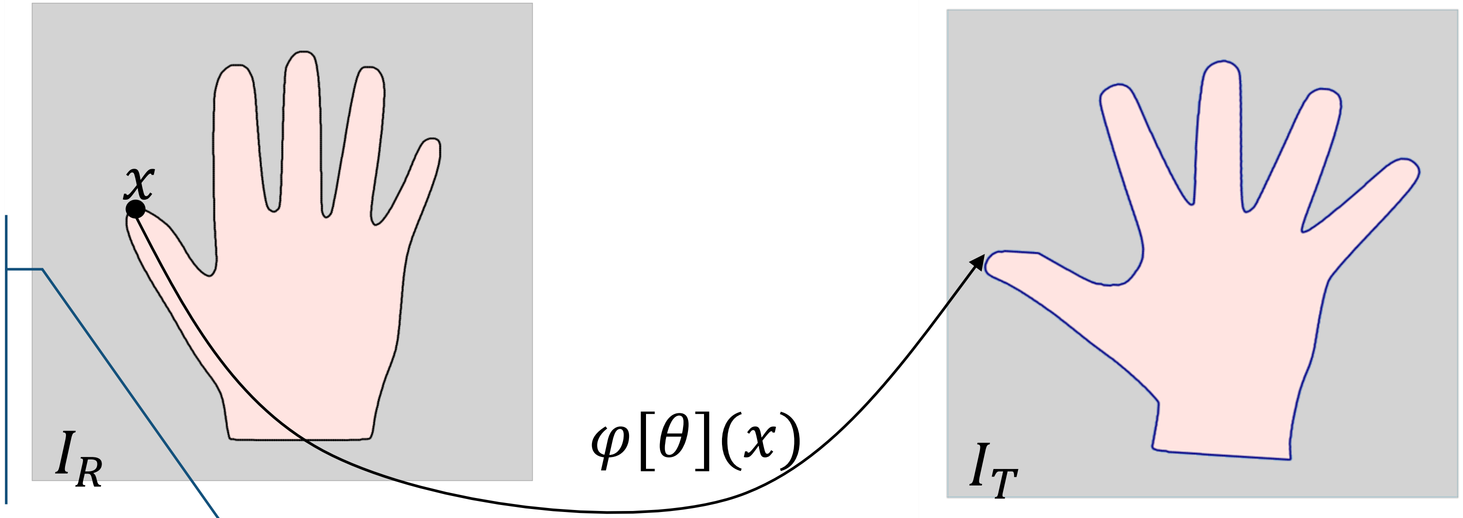
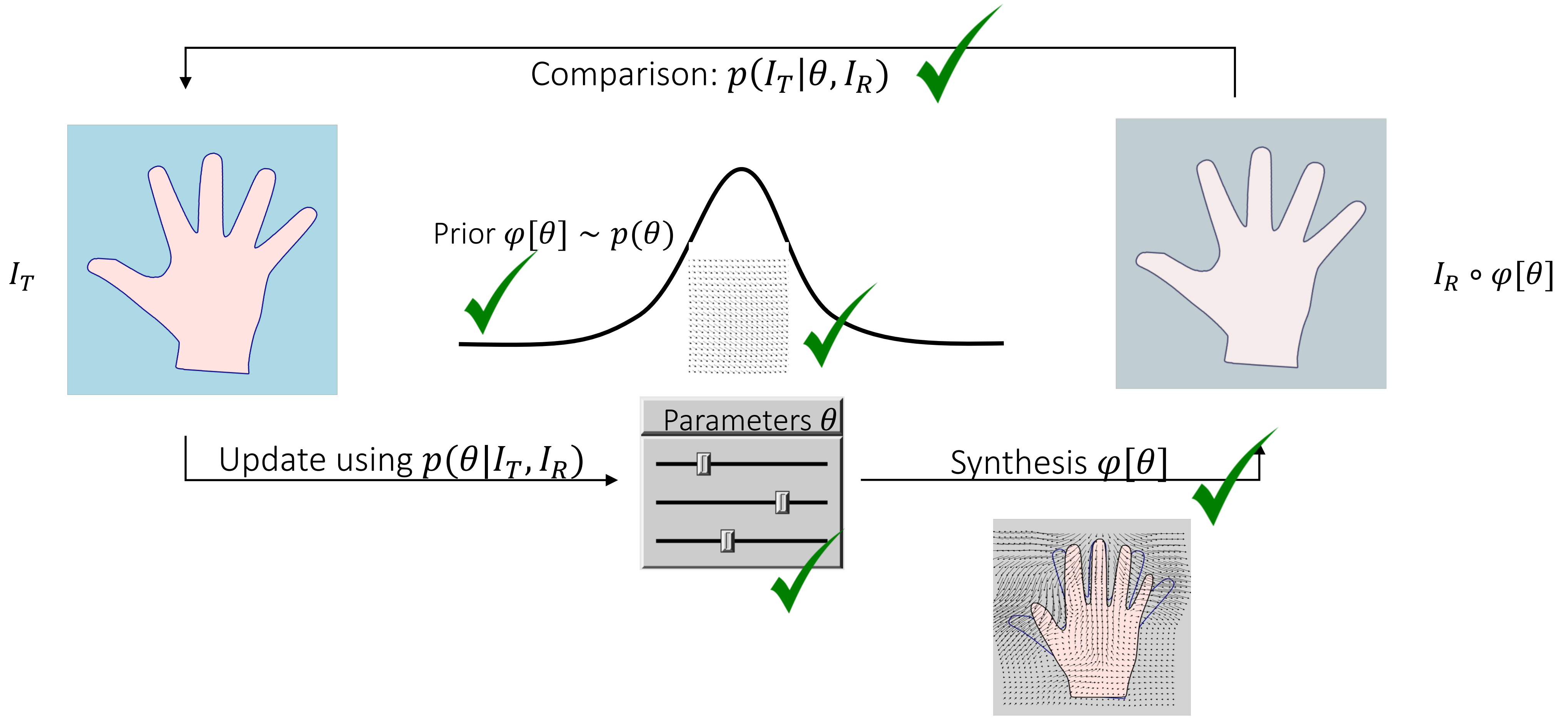


Image term outside mapping function. Makes problem really difficult

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

Registration as analysis by synthesis



Registration problem

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

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

- 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

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

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

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

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

Image metric

Regularizer

The registration problem

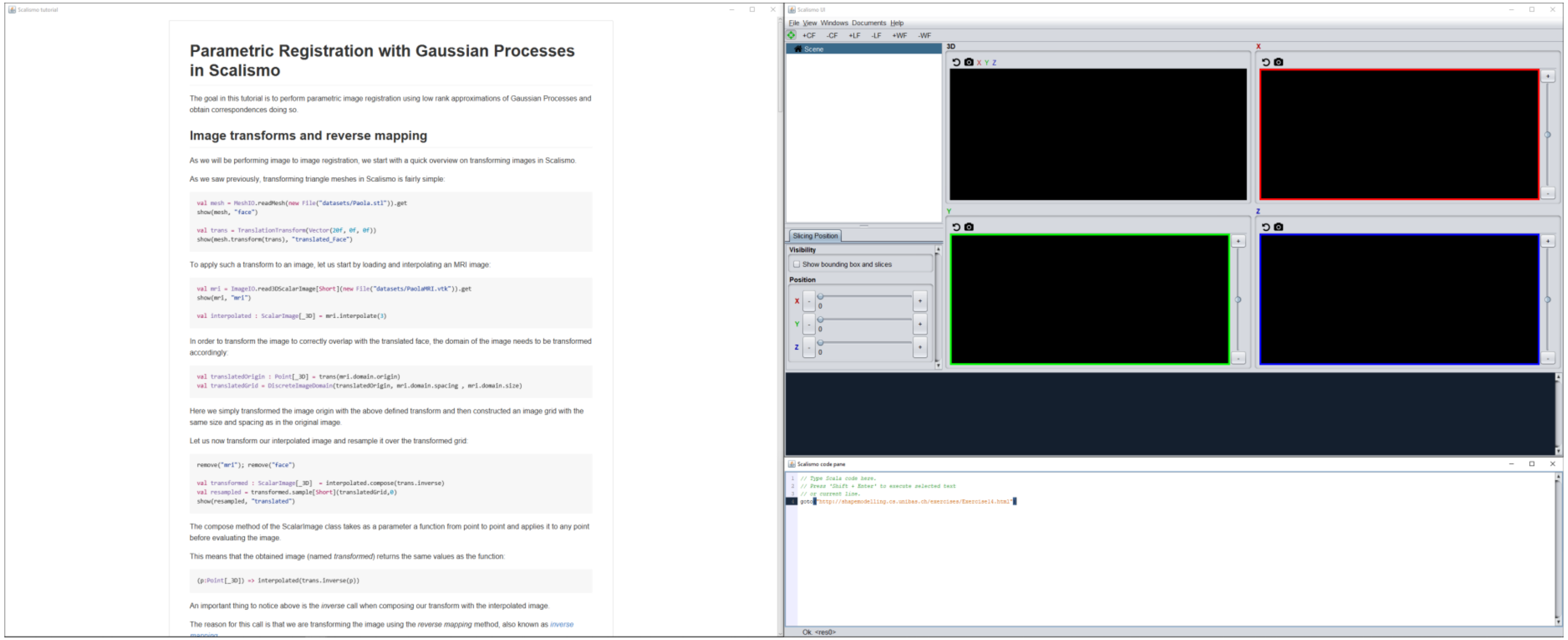
Probabilistic formulation

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

Variational formulation

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

Exercise: Registration in Scalismo



Type into the codepane:

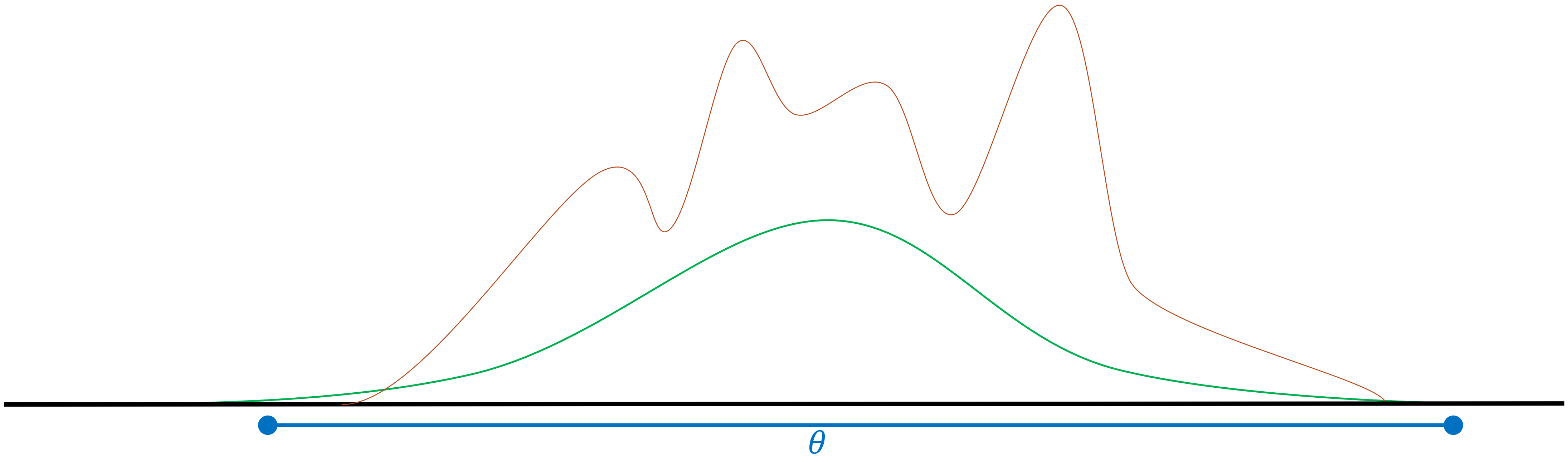
`goto(“http://shapemodelling.cs.unibas.ch/exercises/Exercise14.html”)`

Scalismo 0.16: Check examples in `https://github.com/unibas-gravis/pmm2018`

A selection of Gaussian process priors

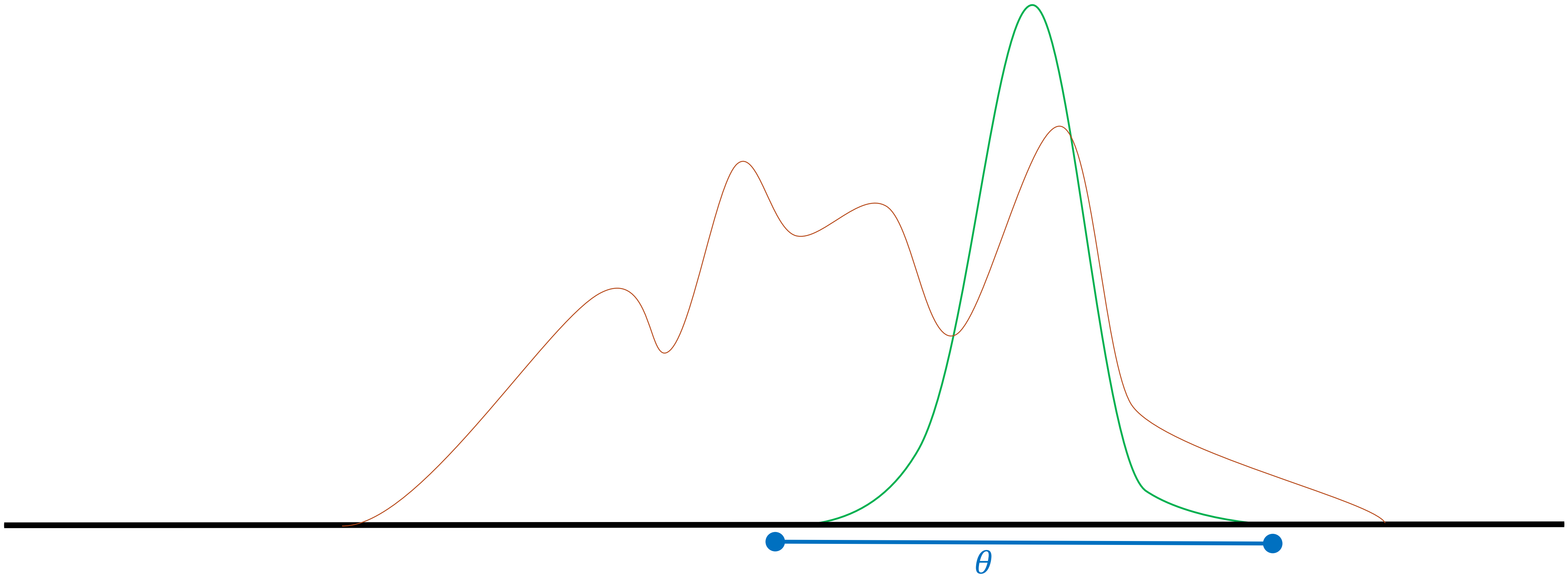
Why are priors interesting?

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



Why are priors interesting?

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



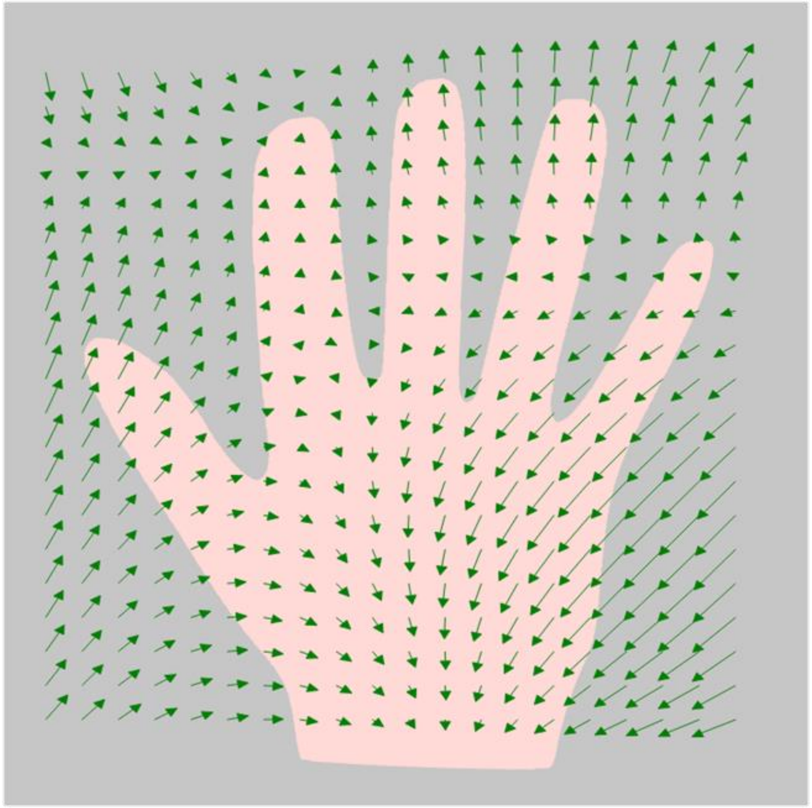
Models of smooth deformations

- Typical assumption:
  - Deformation field is smooth
- GP approach
  - Choose smooth kernel functions

$$k(x, x') = s\ \exp(-\frac{\|x - x'\|^2}{\sigma^2})$$

- Regularization operators
  - Penalize large derivatives

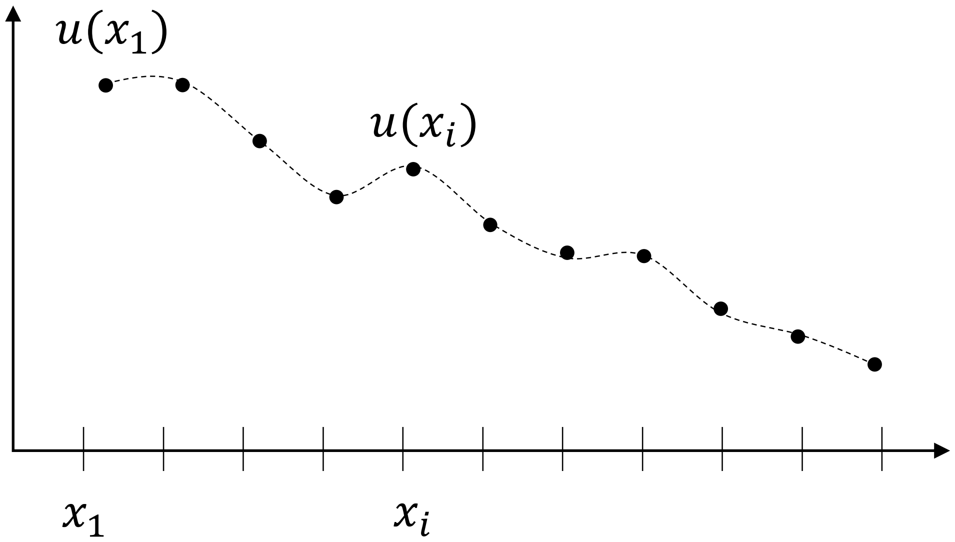
$$\mathcal{R}[u] = \|Ru\|^2 = \sum_{i=0}^{n} \alpha_i \|D_i u\|^2$$



Connection between regularizer and kernel

Discrete setting: Finite difference operators

Regularizer: $\mathcal{R}[\hat{u}] = \left\|\widehat{D}\hat{u}\right\|^2$



$$\widehat{D}\hat{u} = \begin{pmatrix} -2 & 1 & & & & \\ 1 & -2 & 1 & & 0 & \\ & 1 & -2 & 1 & & \\ & & 1 & -2 & 1 & \\ & 0 & & 1 & -2 & 1 \\ & & & & & \ddots \end{pmatrix} \begin{pmatrix} u(x_1) \\ u(x_2) \\ u(x_3) \\ u(x_4) \\ u(x_5) \\ \vdots \end{pmatrix}$$

Steinke, Florian, and Bernhard Schölkopf. "Kernels, regularization and differential equations." *Pattern Recognition* 41.11 (2008): 3271-3286.

Connection between regularizer and kernel

$$\mathrm{p}(\hat{\mathrm{u}}) = \exp(-\frac{1}{2}\mathcal{R}[\hat{u}]) = \exp(-\frac{1}{2}\left\|\widehat{D}\hat{u}\right\|^2)$$

$$\widehat{D}^T\widehat{D} = K^{-1}$$

$$= \exp(-\frac{1}{2}\left(\widehat{D}\hat{u}\right)^T(\widehat{D}\hat{u}) = \exp(-\frac{1}{2}\ \hat{u}^T\widehat{D}^T\widehat{D}\hat{u}\ )$$

- D specifies the “inverse” covariance matrix
- Can compute K from:

$$\widehat{D}^T\widehat{D}K = I$$

Green's functions and covariance functions

$$\mathcal{R}[u] = \|Ru\|^2 = \sum_{i=0}^{n} \alpha_i \|D_i u\|^2$$

Corresponding covariance function for GP is the Greens function G:

$$R^* R G(x, y) = \delta(x - y)$$

- We can define Gaussian processes, which mimic typical regularization operators.

T. Poggio and F. Girosi; Networks for Approximation and Learning, Proceedings of the IEEE, 1990

Example: Gaussian kernel

$$k(x, x') = \exp(-\frac{\|x - x'\|^2}{\sigma^2})$$

$$\mathcal{R}[u] = \|Ru\|^2 = \sum_{i=0}^{\infty} \frac{\sigma^{2i}}{i!\,2^i} \|D_i u\|^2$$

- Non-zero functions are penalized
  - pushes functions to zero away from data

Yuille, A. and Grzywacz M. A mathematical analysis of the motion coherence theory. International Journal of Computer vision

Example: Exponential kernel (1D case)

$$k(x, x') = \frac{1}{2\alpha} \exp(-\alpha \|x - x'\|)$$

$$\mathcal{R}[u] = \|Ru\|^2 = \alpha^2 u + \|D_1 u\|^2$$

Rasmussen, Carl Edward, and Christopher KI Williams. *Gaussian processes for machine learning*. Vol. 1. Cambridge: MIT press, 2006.

Matérn class of kernels

$$k(x,x') = s\,\frac{2^{1-\nu}}{\Gamma(\nu)}\left(2\sqrt{2\nu}\,\frac{\|x-x'\|}{\rho}\right)^{\nu} K_\nu(\sqrt{2\nu}\,\frac{\|x-x'\|}{\rho})$$

- Γ is the Γ function, k_ν the modified Bessel function and ν, ρ are parameters
- Process $u \sim GP(0,k)$ is $\nu - 1$ times m.s. differentiable

- Special cases:
  - $\nu = \frac{1}{2}$: $k(x,x') = s\exp(-\frac{\|x-x'\|}{\rho})$
  - $\nu = \frac{3}{2}$: $\mathrm{k}(\mathrm{x},\mathrm{x}') = s(1 + \frac{\sqrt{3}\|x-x'\|}{\rho})\exp(-\frac{\sqrt{3}\|x-x'\|}{\rho})$
  - $\nu \to \infty$ Gaussian kernel

Thin-plate splines

- Minimizes the bending energy of a metal sheet

$$R[u] = \|(\nabla^T \nabla u)\|^2$$

- Corresponding covariance function

$$k(x, x') = \frac{1}{12}(2\|x - x'\|^3 - 3R(\|x - x'\|^2 + R^3)$$

$$\text{where } R = \max_{x, x' \in \Omega} \|x - x'\|$$

Rohr, Karl, et al. "Landmark-based elastic registration using approximating thin-plate splines." *IEEE Transactions on medical imaging* 20.6 (2001): 526-534.
Williams, Oliver and Fitzgibbon Andrew, “Gaussian process implicit surfaces”

B-Splines

- We can build a covariance function from B-Spline basis functions β (s is a scaling constant)

$$k(x,y) = \sum_{k \in \mathbb{Z}^d} \beta(sx - k)\beta(sy - k)$$

- Corresponding deformation model often called “free form deformations”

- Rueckert, Daniel, et al. "Nonrigid registration using free-form deformations: application to breast MR images." *IEEE transactions on medical imaging* 18.8 (1999): 712-721.
- Klein, Stefan, et al. "Elastix: a toolbox for intensity-based medical image registration." *IEEE transactions on medical imaging* 29.1 (2010): 196-205.

Many standard models for registration can be formulated using Gaussian processes

- Yields probabilistic interpretation
- We can sample and visualize deformation fields
- Can use them as building blocks for more complicated priors

Constructing s.p.d. kernels

1. $k(x,x') = f(x)\, f(x')^T, f: X \to \mathbb{R}^d$
2. $k(x,x') = \alpha k_1(x,x'), \alpha \in \mathbb{R}_+$ (scaling)
3. $k(x,x') = B^T k_1(x,x') B,\ \ \mathrm{B} \in \mathbb{R}^{r \times d}$ (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

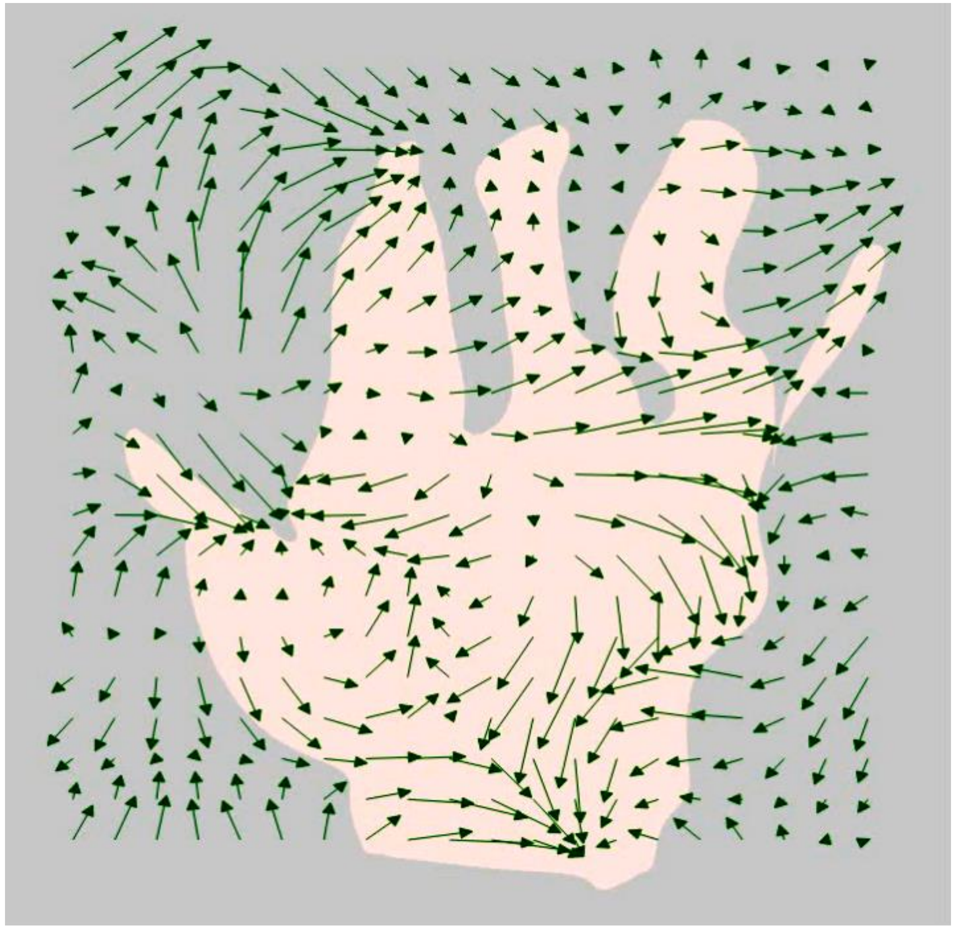
Add kernels that act on different scales:

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

- Wavelet like multiscale representation

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

Multi-scale kernel



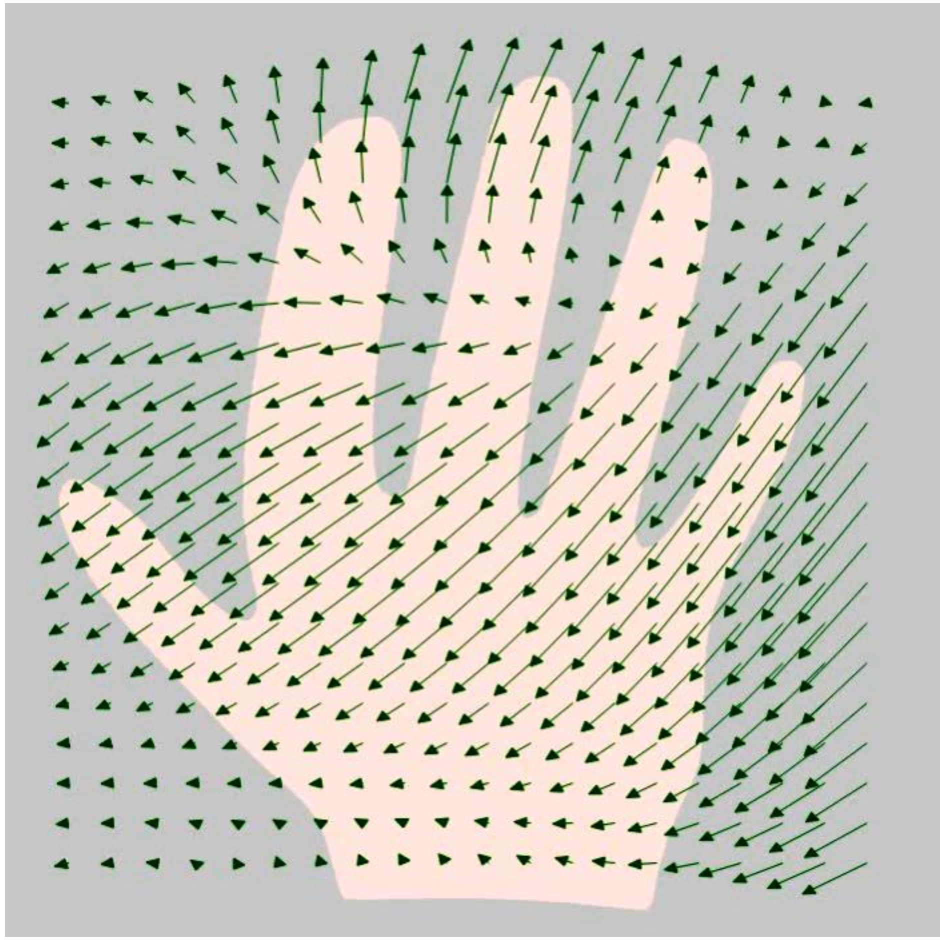
Anisotropic priors

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

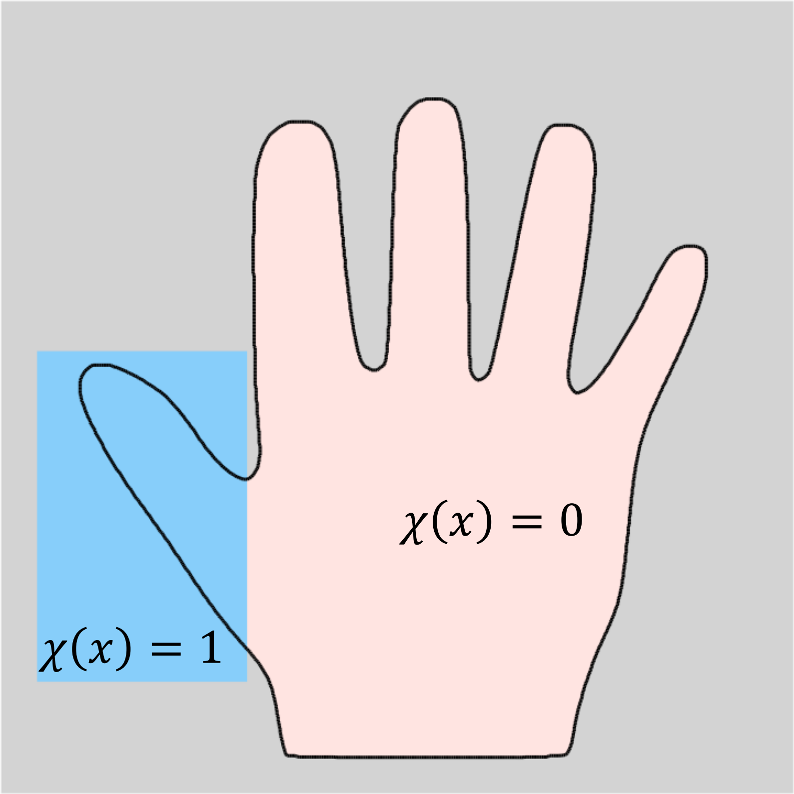


Spatially-varying priors

Use different models for different regions

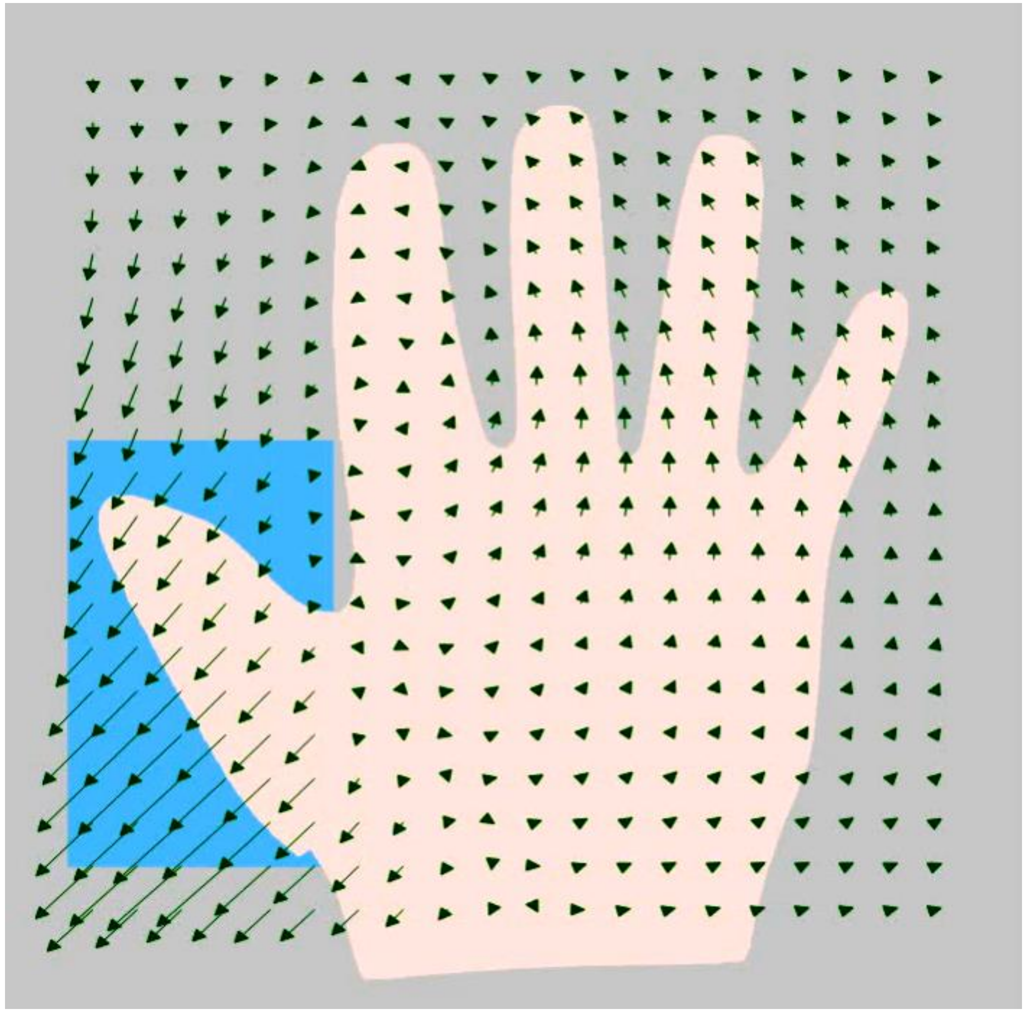
$$k(x,x') = \chi(x)\chi(x')k_1(x,x') \\ +\big(1-\chi(x)\big)(1-\chi(x'))\, k_2(x,x')$$

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



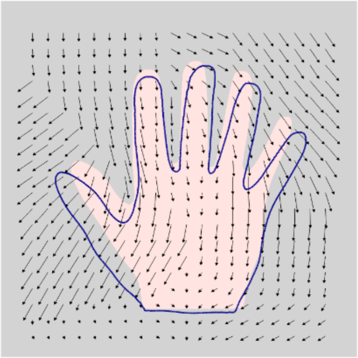
Freiman, Moti, Stephan D. Voss, and Simon K. Warfield. "Demons registration with local affine adaptive regularization: application to registration of abdominal structures." *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011.

Spatially-varying priors

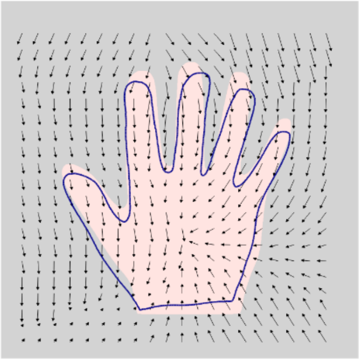


Statistical deformation models

Estimate mean and covariance function from data:

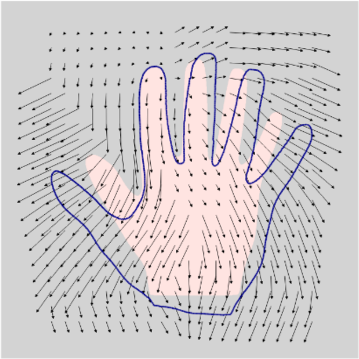


$u^1 : \Omega \rightarrow \mathbb{R}^2$



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

...

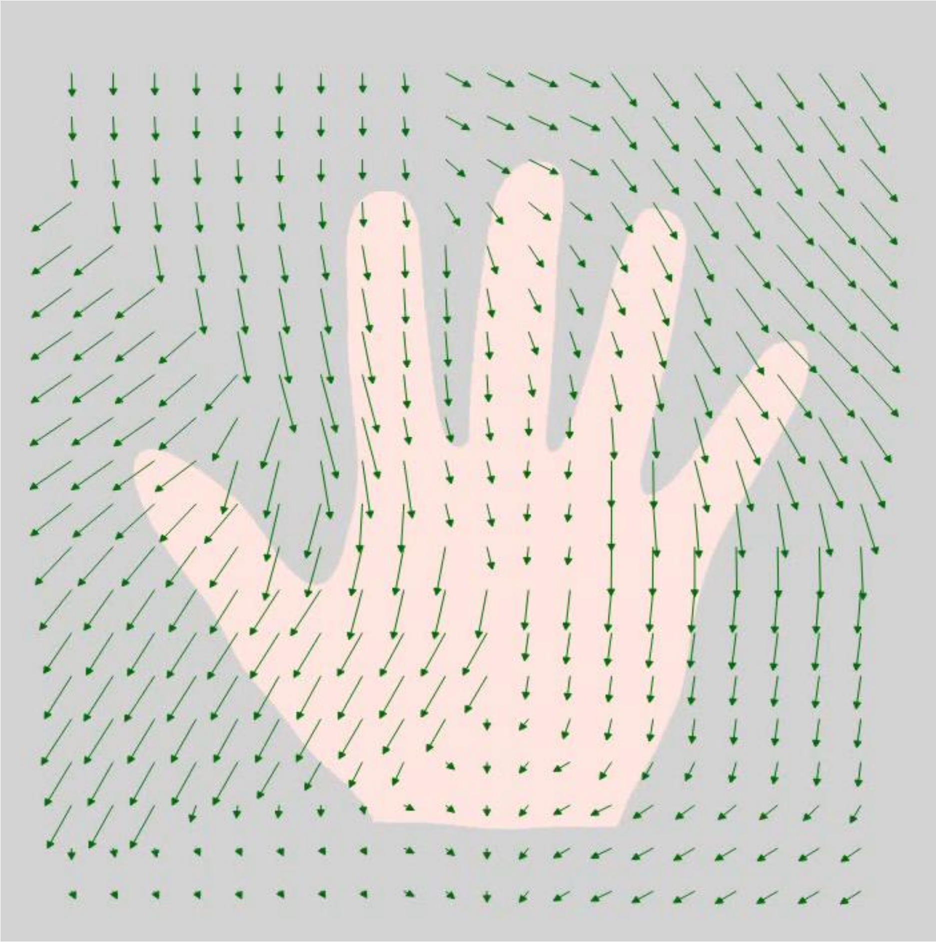


$u^n : \Omega \rightarrow \mathbb{R}^2$

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

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

Example 5: Statistical deformation models



A selection of likelihood functions

Landmark likelihood

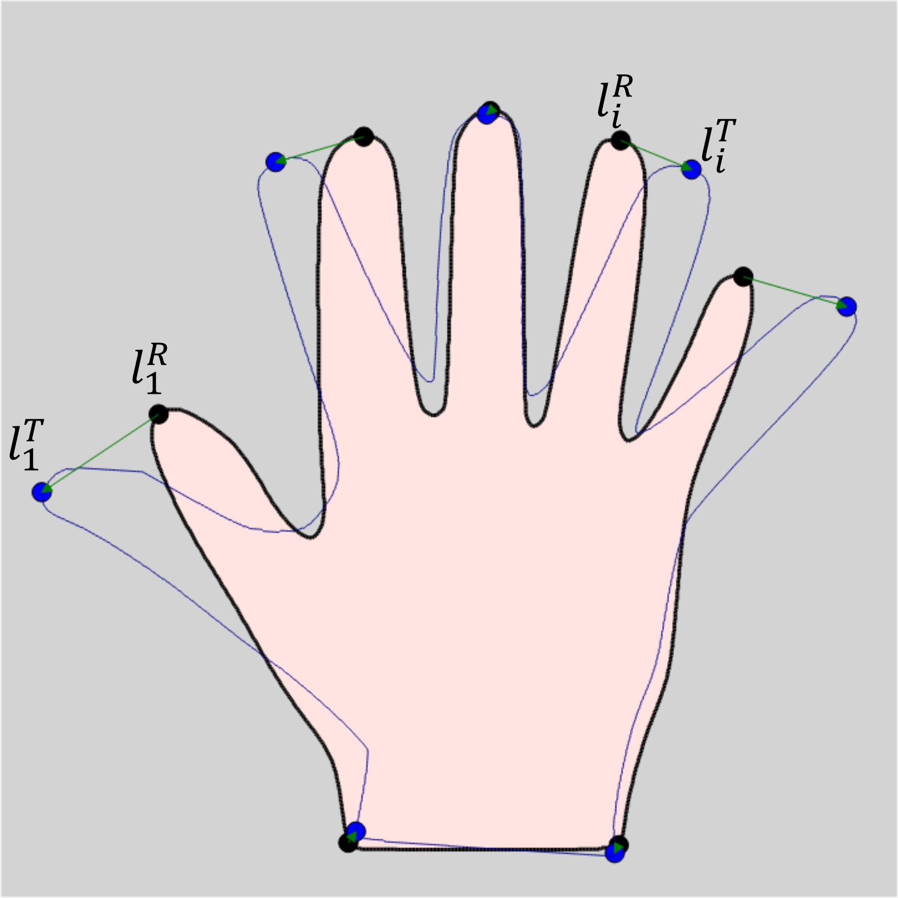
For one landmark pair (l_R, l_T):

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

For many landmarks

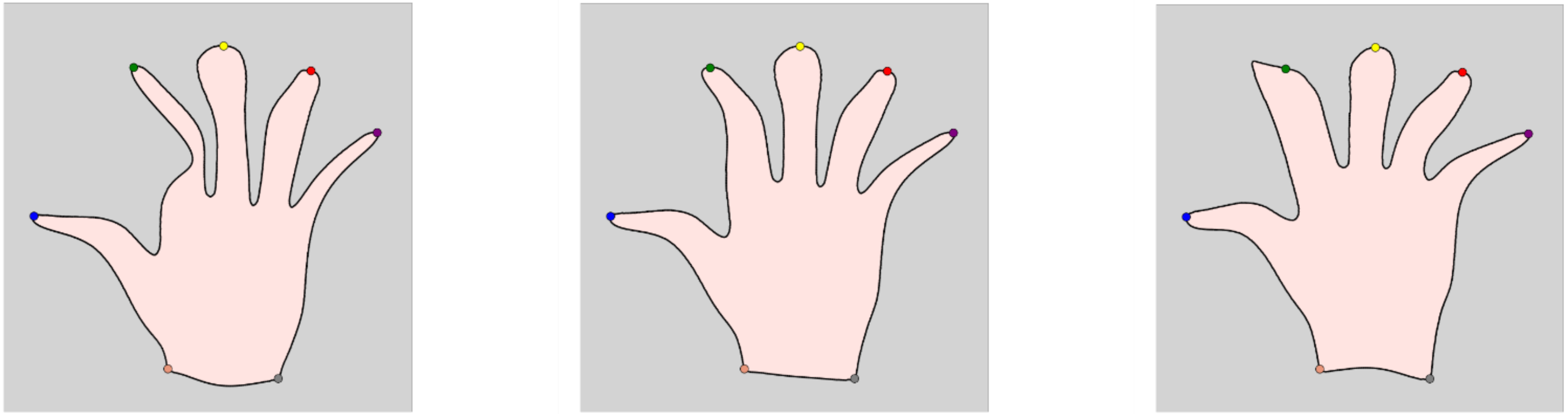
$$L = ((l_R^1, l_T^1), \dots, (l_R^n, l_T^n))$$

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



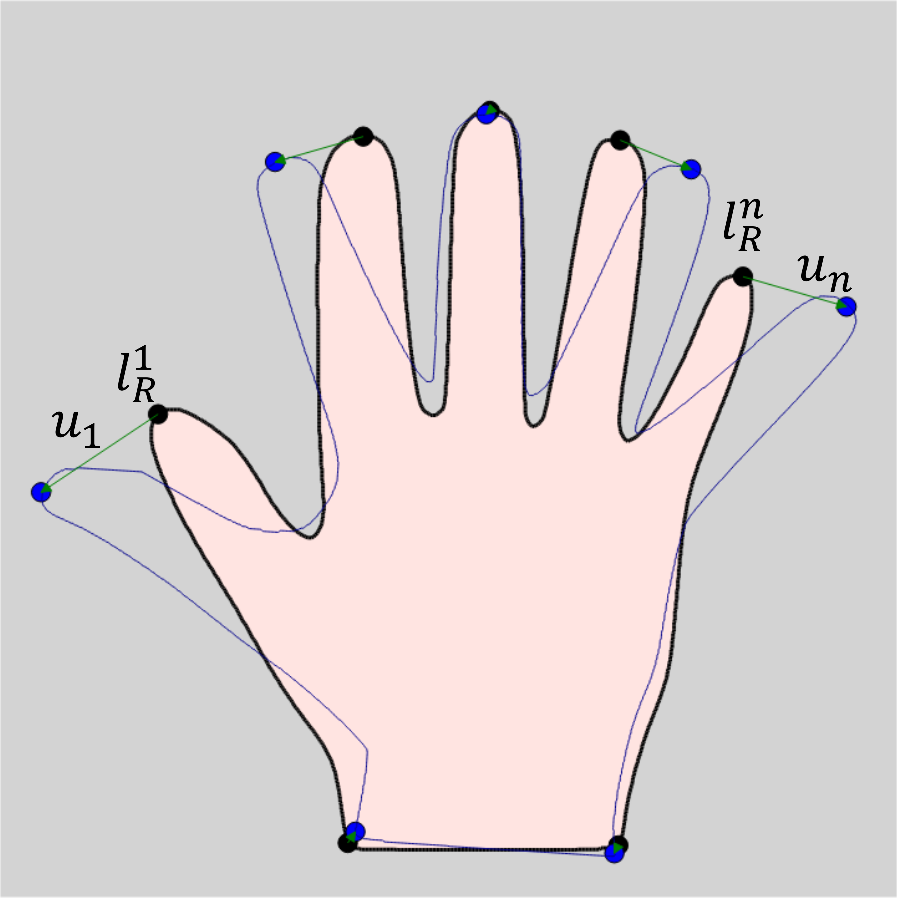
Landmark likelihood: Some remarks

- Classical problem in registration
- Needs either many landmark points or good structure of prior to achieve good results



Rohr, Karl, et al. "Landmark-based elastic registration using approximating thin-plate splines." *IEEE Transactions on medical imaging* 20.6 (2001): 526-534.

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:

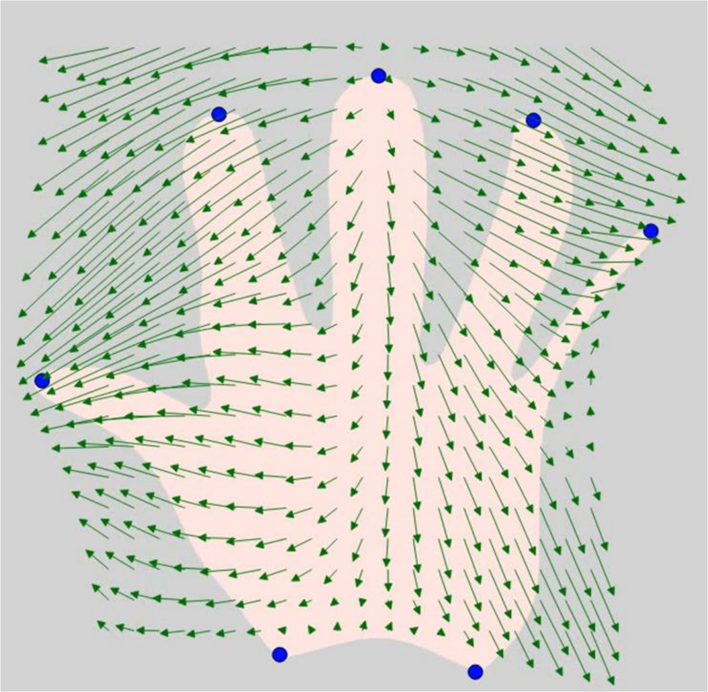
$$\tilde{u}_i = u(l_i) + \epsilon \text{ with } \epsilon \sim N(0, \sigma^2 I_{2\times 2}).$$

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

$$u \,|l_1^R, \dots, l_n^R, \tilde{u}_1, \dots, n$$

is a Gaussian process

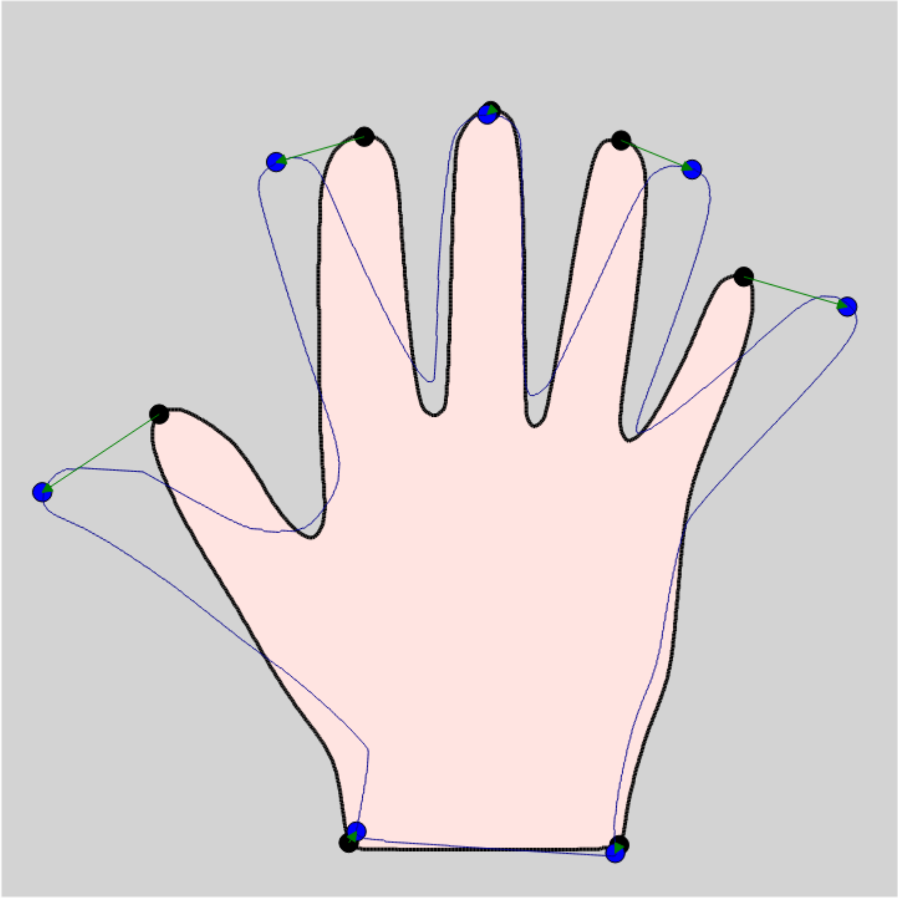
$$GP(\mu_p, k_p)$$

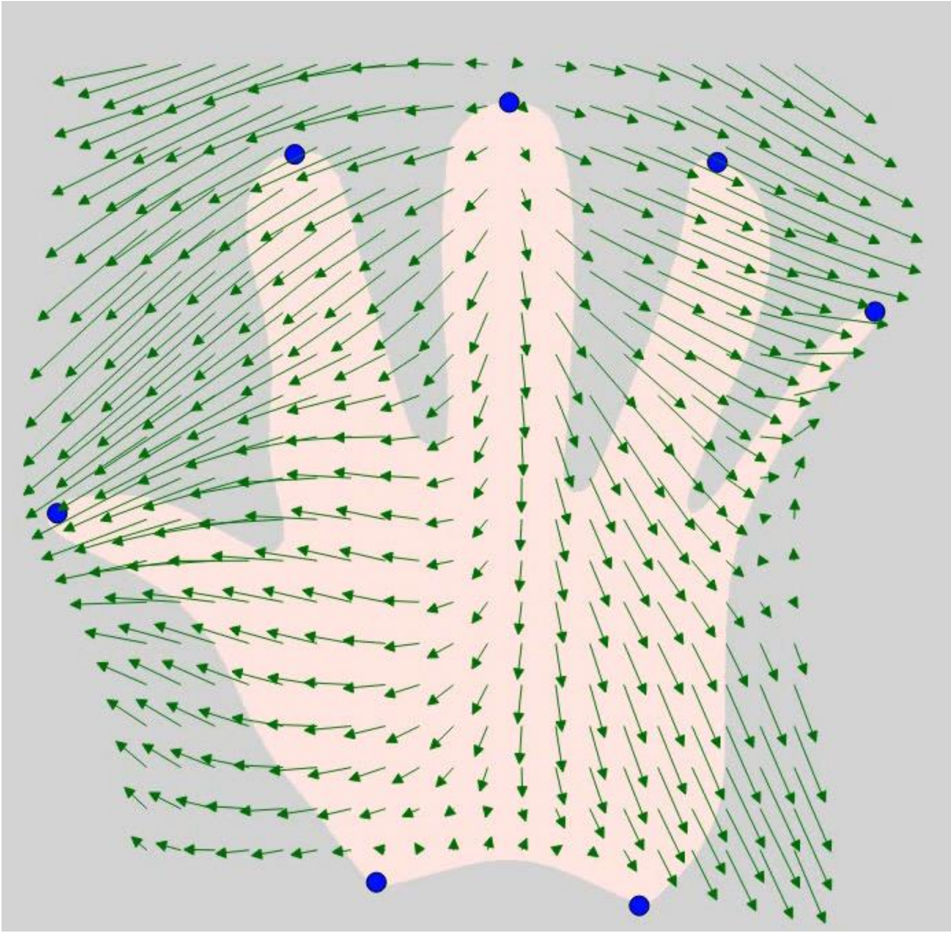
Its parameters are known analytically.

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

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

Landmark registration using GP Regression





Hybrid registration

- Combine landmark registration with intensity:
    1. Use Gaussian process regression to obtain posterior from $GP(\mu, k)$ from landmarks
    2. Use $GP(\mu_p, k_p)$ as new prior model for registration

- Simple solution to otherwise difficult problem

Wörz, Stefan, and Karl Rohr. "Hybrid spline-based elastic image registration using analytic solutions of the navier equation." *Bildverarbeitung für die Medizin 2007*. Springer Berlin Heidelberg, 2007. 151-155.

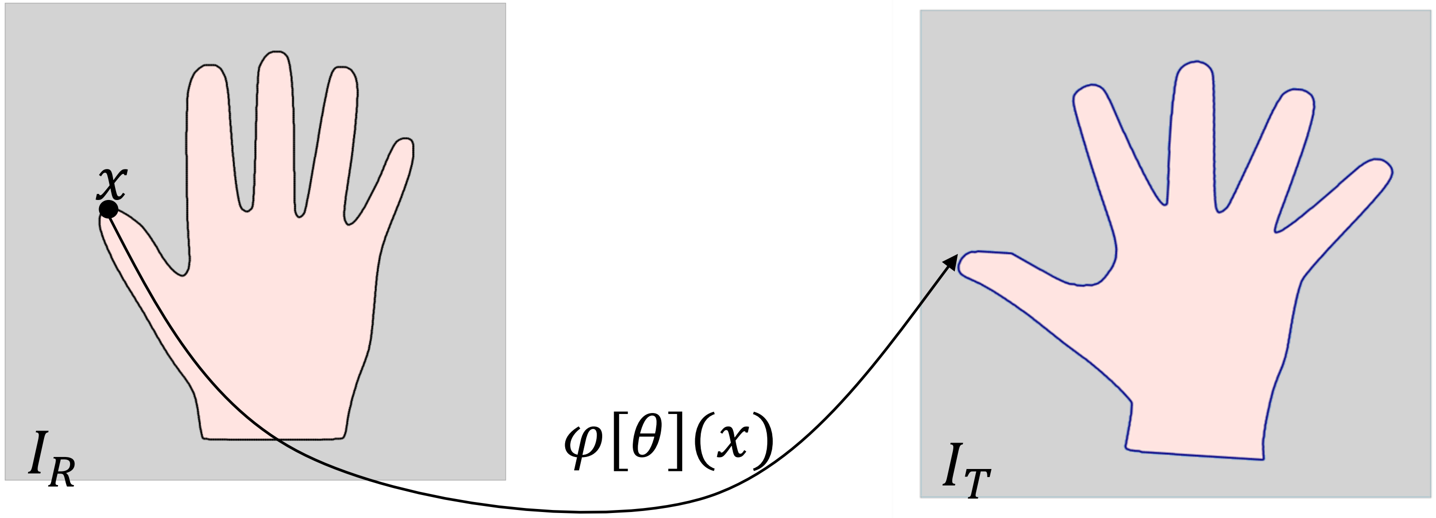
Lu, Huanxiang, Philippe C. Cattin, and Mauricio Reyes. "A hybrid multimodal non-rigid registration of MR images based on diffeomorphic demons." *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010.

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)



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

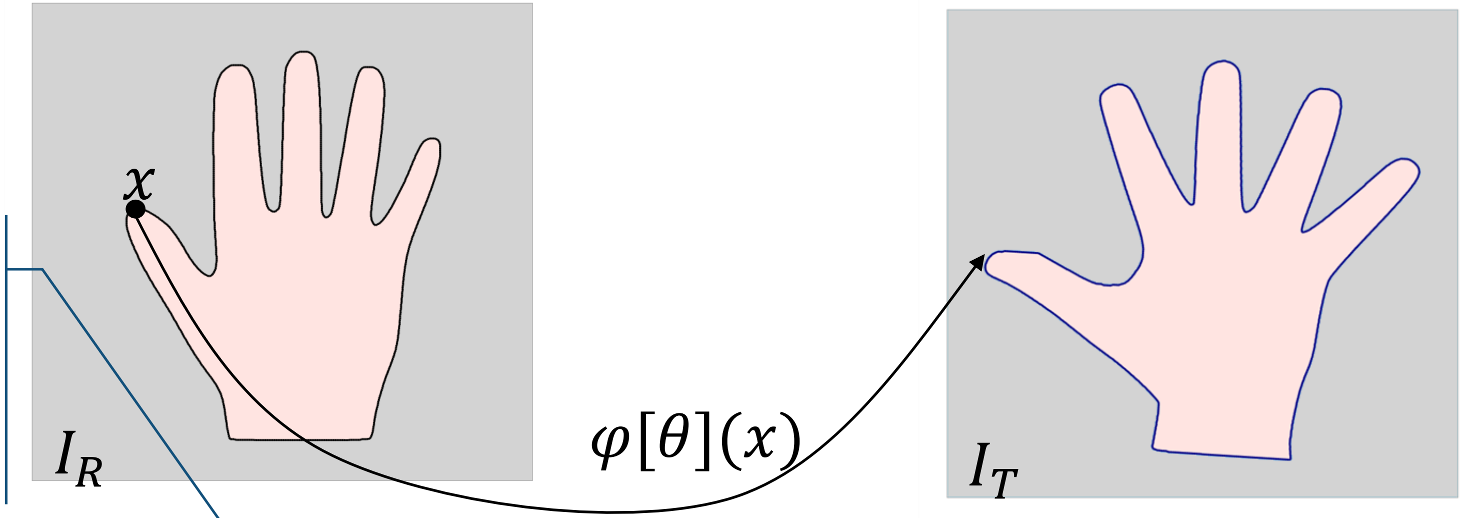
Likelihood function: Image registration

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



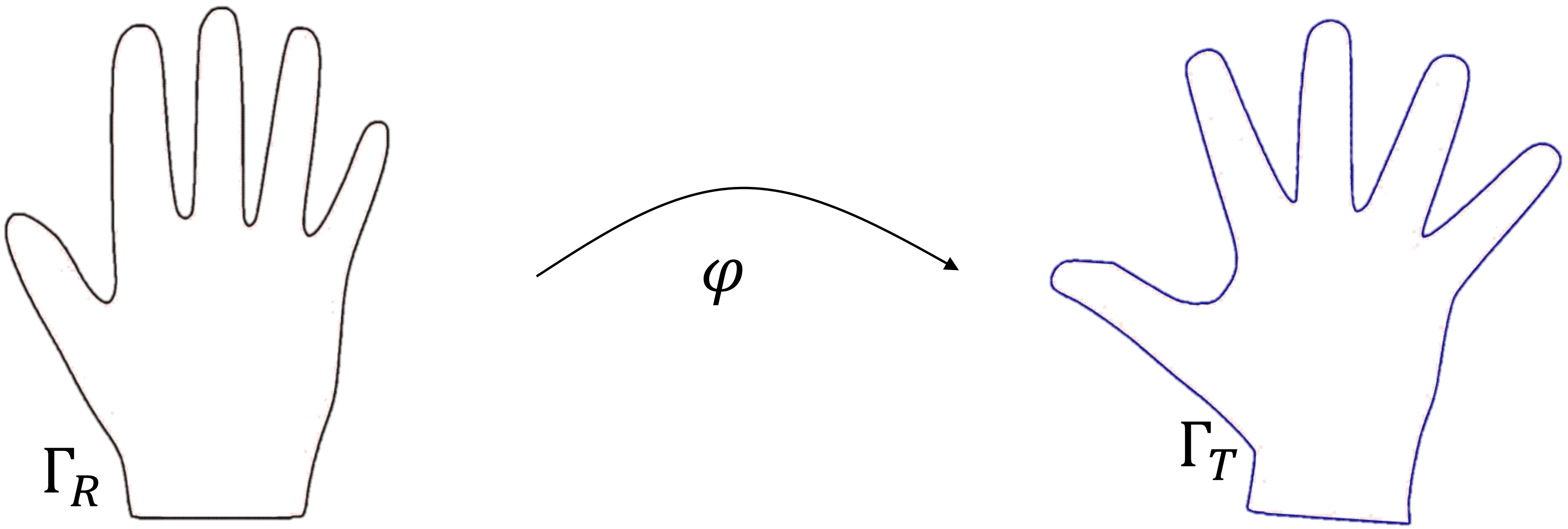
$$p(I_T|I_R,\theta) = \prod_{x\in\Omega} p(I_T(\varphi[\theta](x))|I_R,\theta,x) = \prod_{x\in\Omega} \frac{1}{Z}\exp\left(-\frac{\left(I_T\big(\varphi(x)\big) - I_R(x)\right)^2}{\sigma^2}\right)$$

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?



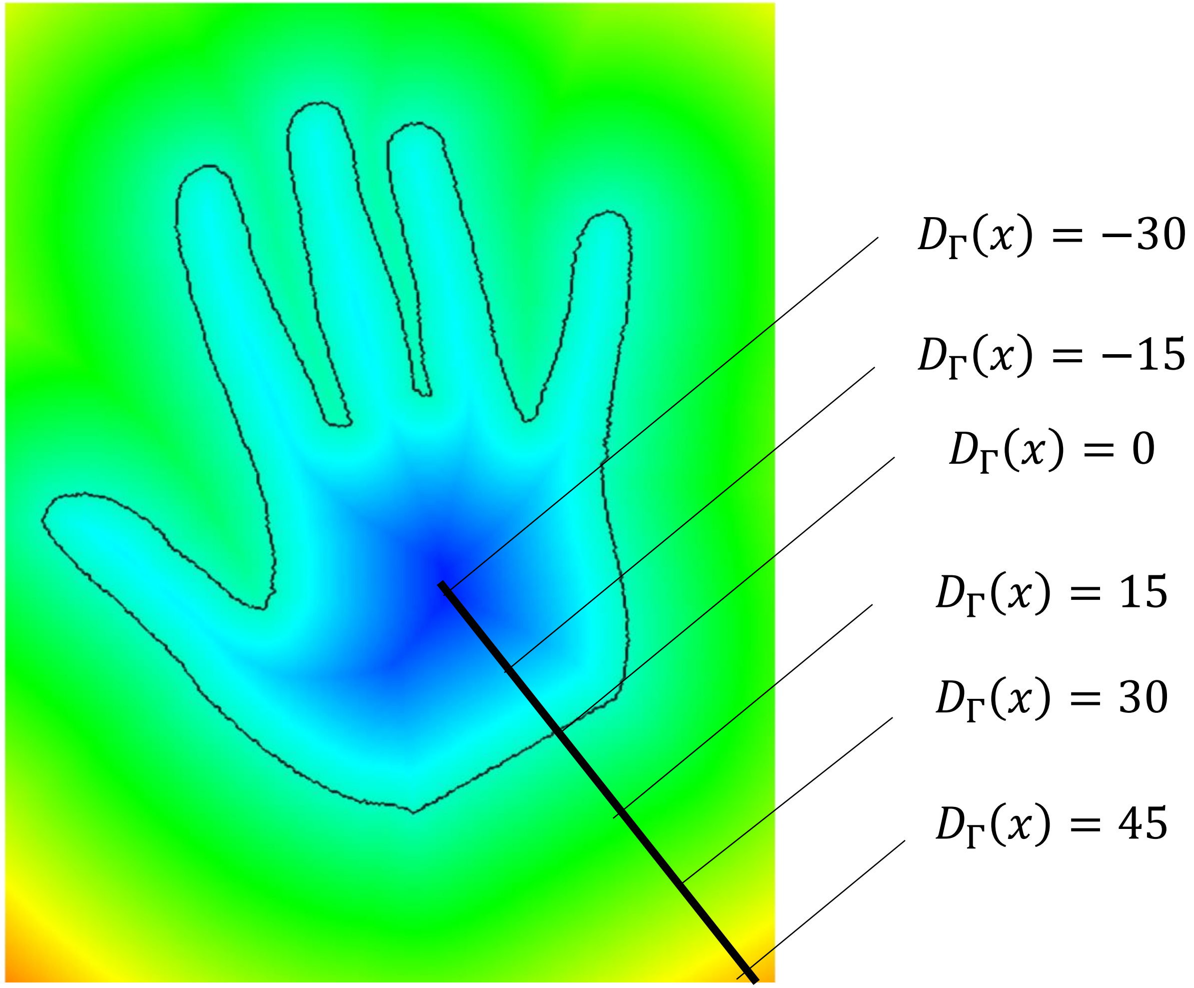
Reference (surface): Γ_R

Target (surface): Γ_T

A trick: Implicit definition of a surface

- Any surface Γ can be represented as the zero level set of a level set function Φ
$$\Gamma = \{x \mid \Phi(x) = 0\}$$

- Popular choice is the signed distance function defined as
$$D_\Gamma(x) = \|CP_\Gamma(x) - x\|$$
with
$$CP_\Gamma(x) = \arg\min_{x' \in \Gamma} \|x - x'\|$$

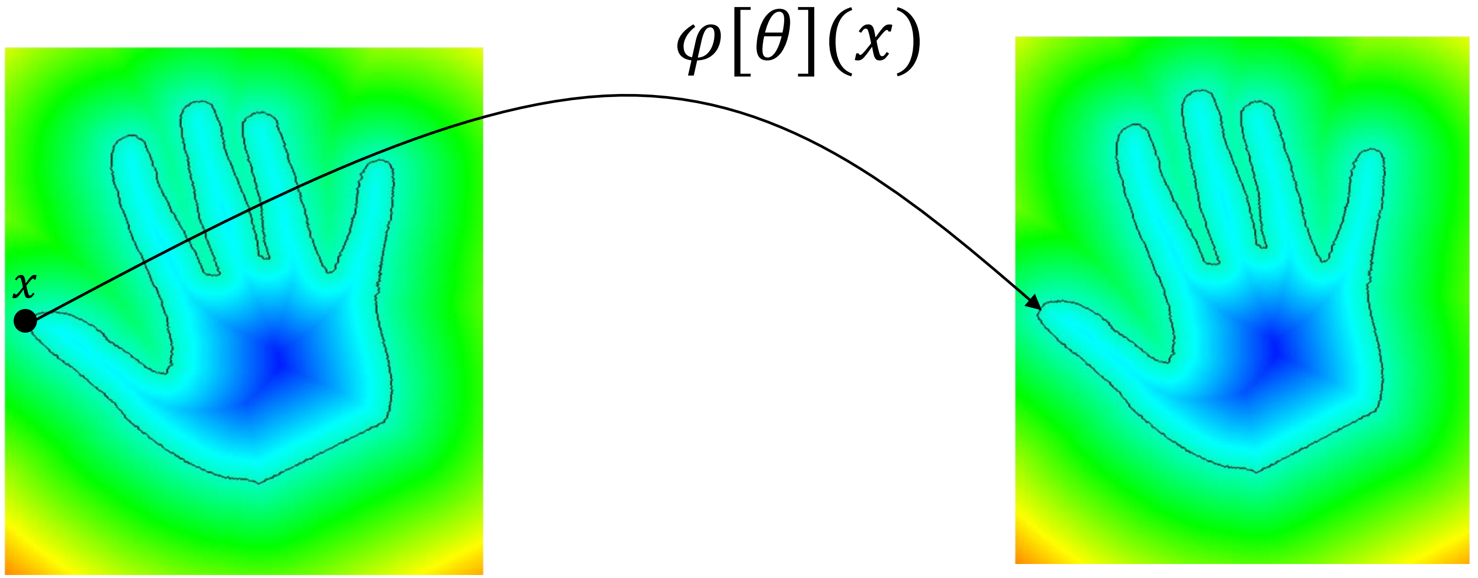


Likelihood function: Surface registration

- We define the distance functions and use image to image likelihoods

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

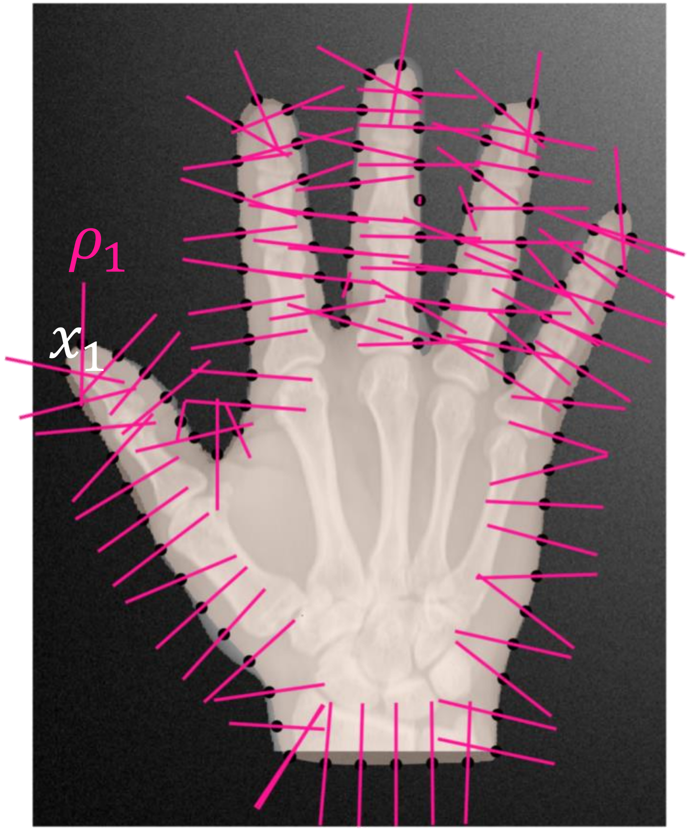
- Most likely solution will map points on the zero level-sets to each other
  - Noise parameter σ^2 has geometric interpretation (variance of distance between the mapped points)



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

Likelihood function: Active shape models

Shape is well matched if environment around profile points is likeli 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

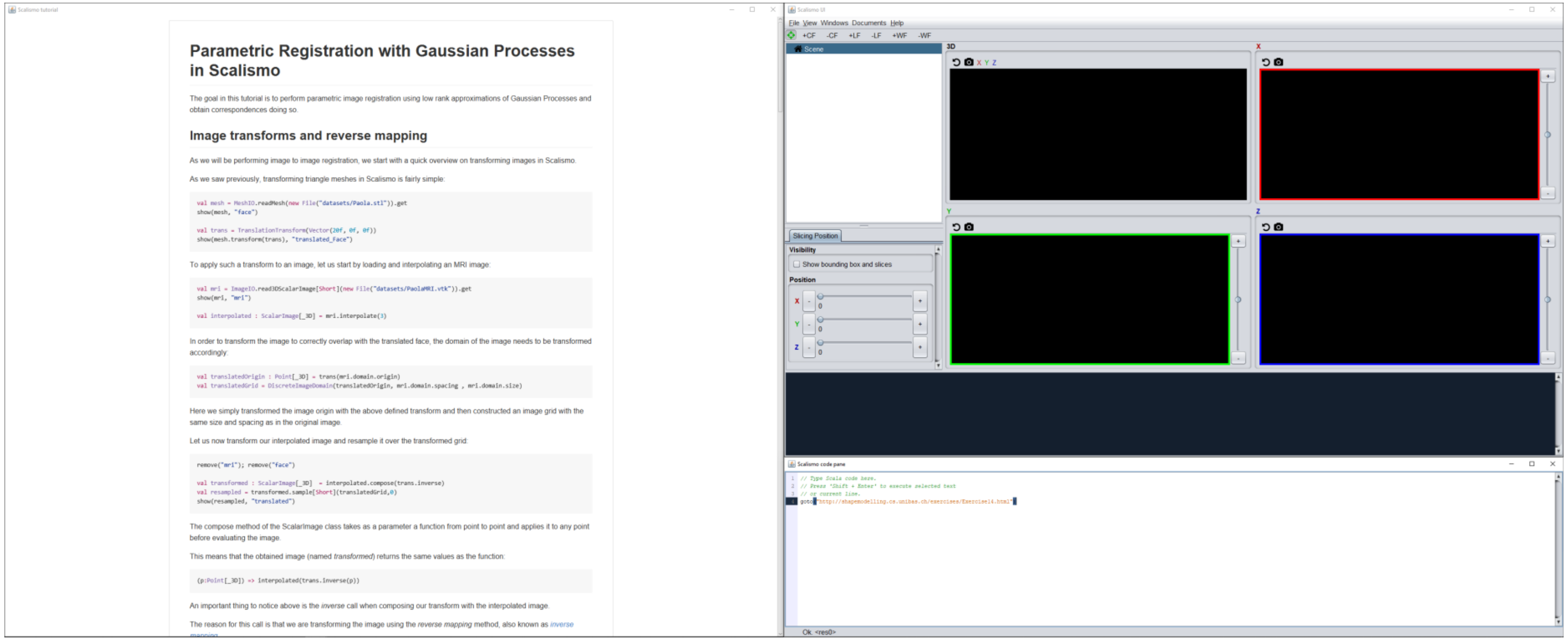
- Single profile point x_i:

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

- Image likelihood:

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

Exercise: Active Shape Models in Scalismo

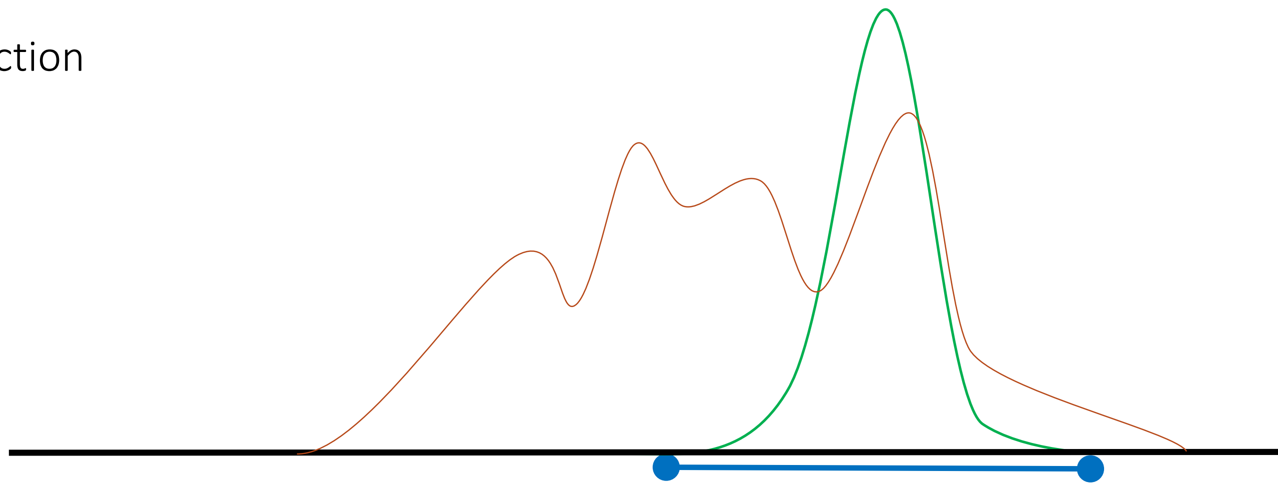


Type into the codepane:
`goto(“http://shapemodelling.cs.unibas.ch/exercises/Exercise16.html”)`

Scalismo 0.16: Check examples in `https://github.com/unibas-gravis/pmm2018`

Summary: Modelling

- GPs provide probabilistic interpretation to classic registration models
  - Can visualize assumptions
  - New ways to combine priors to individual applications.
- Modelling of prior is separate from likelihood
  - Any GP model can be combined with any likelihood function
- Flexible framework to tailor model and algorithm to needs of applications
  - No increase in complexity
- Modelling and model fitting are separated
  - Gradient based methods are important, but not the only method

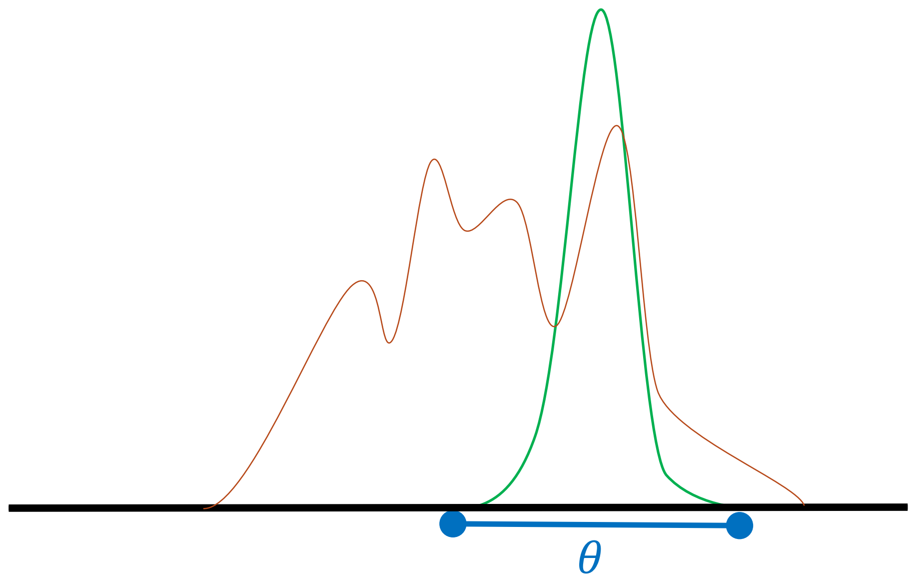


Optimization

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



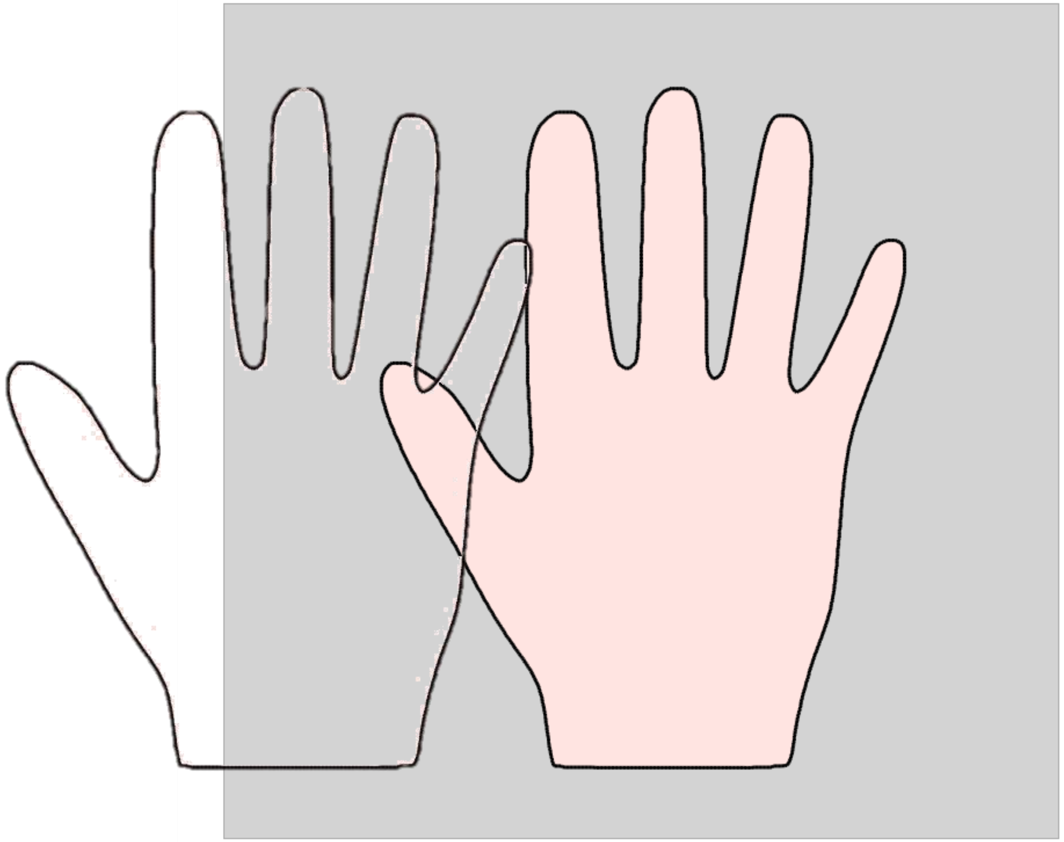
Local minima

- Rigid Transformation
  - Minima due to structure of object

  Possible approach: Multi-resolution

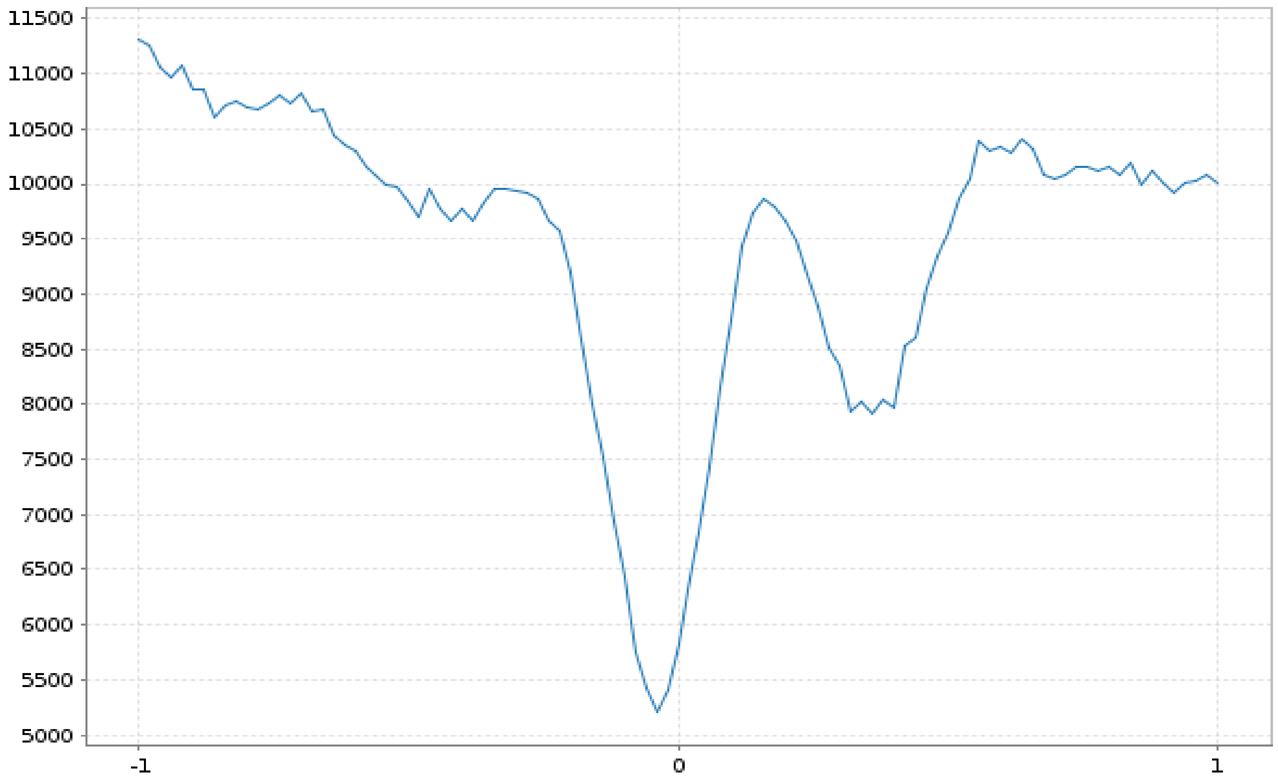
- Non-rigid Transformation
  - Minima appear/dissappear when shape changes

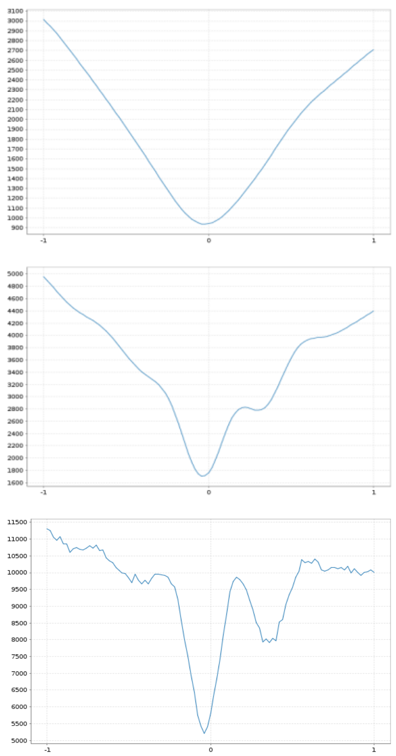
  Possible approach: Multi-scale models, regularization



Multi resolution

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





Implementation

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

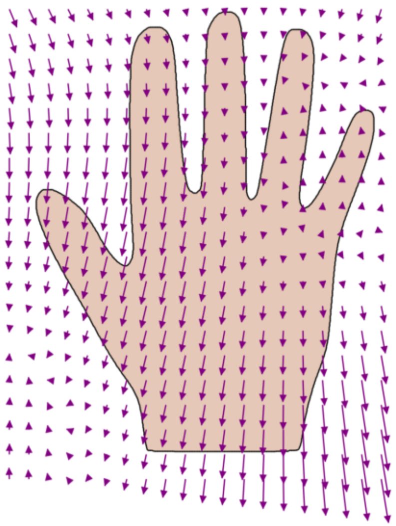
Initial registration — Final registration

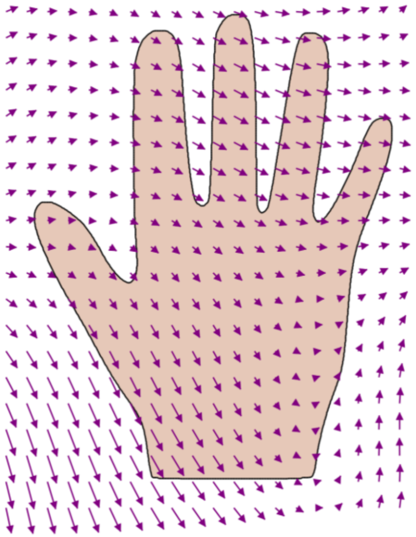
Almost no local minima
No-details

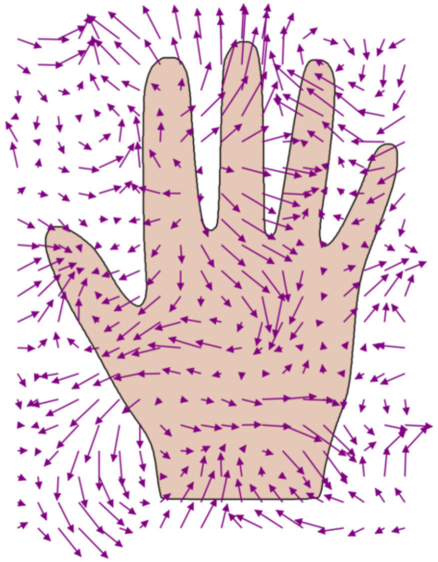
Many local minima
All-details

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

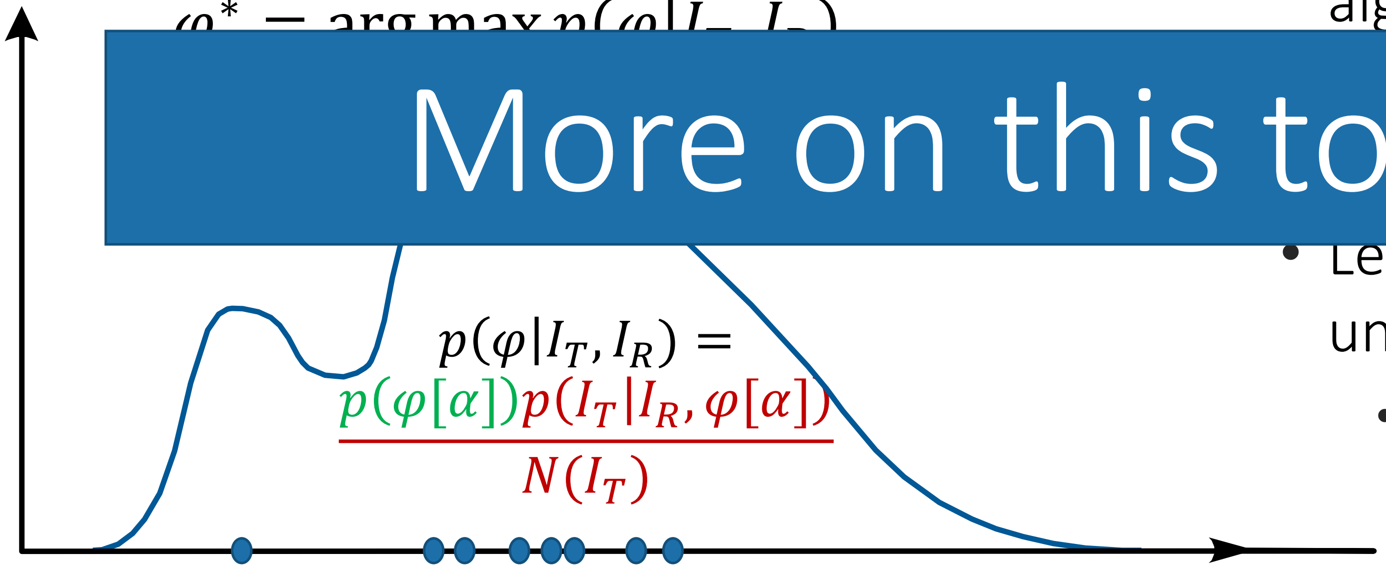
Strategies:

- 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

MAP Solution

$$\varphi^* = \arg\max p(\varphi|I_T, I_R)$$



$$p(\varphi|I_T, I_R) = \frac{p(\varphi[\alpha])p(I_T|I_R, \varphi[\alpha])}{N(I_T)}$$

- Can obtain full posterior distribution using the Metropolis Hastings algorithm
- Leads to principled way to integrate unreliable bottom up methods
  - Automatically detected landmarks

More on this tomorrow!