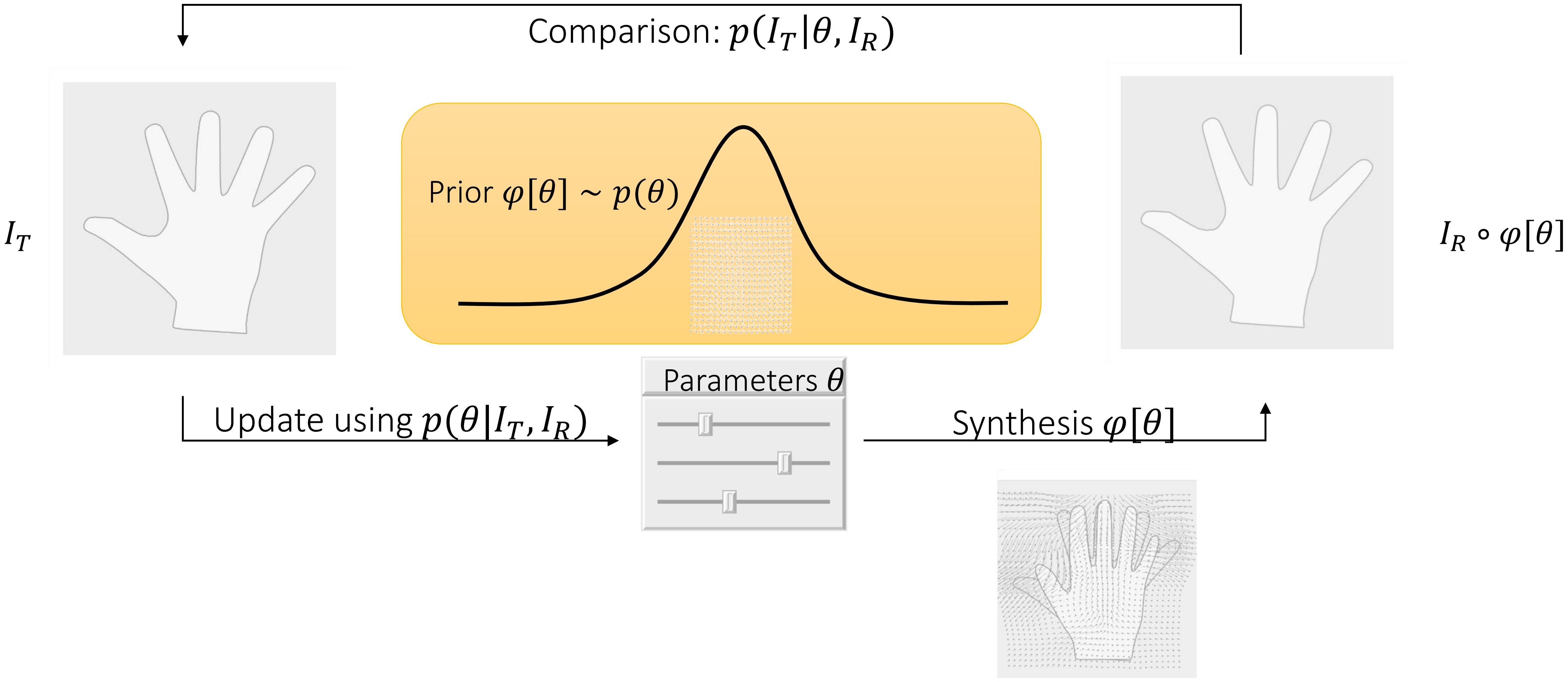
Registration – Deformation models

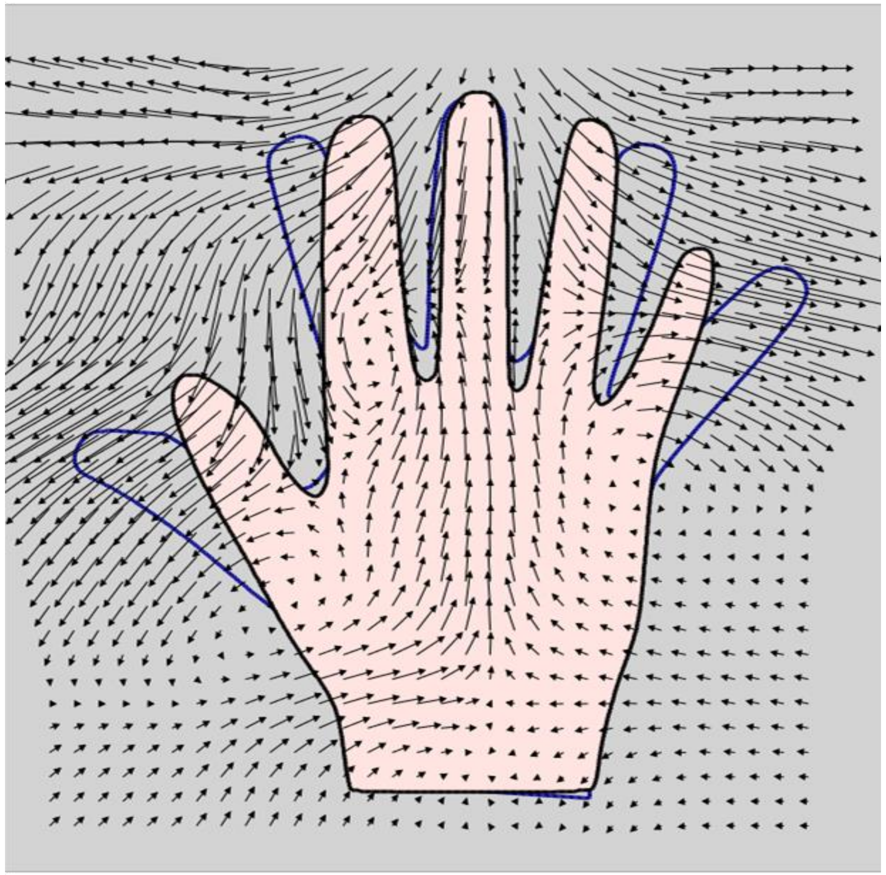
Marcel Lüthi

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

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

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

Characteristics of deformation fields

Example prior: Smooth 2D deformations

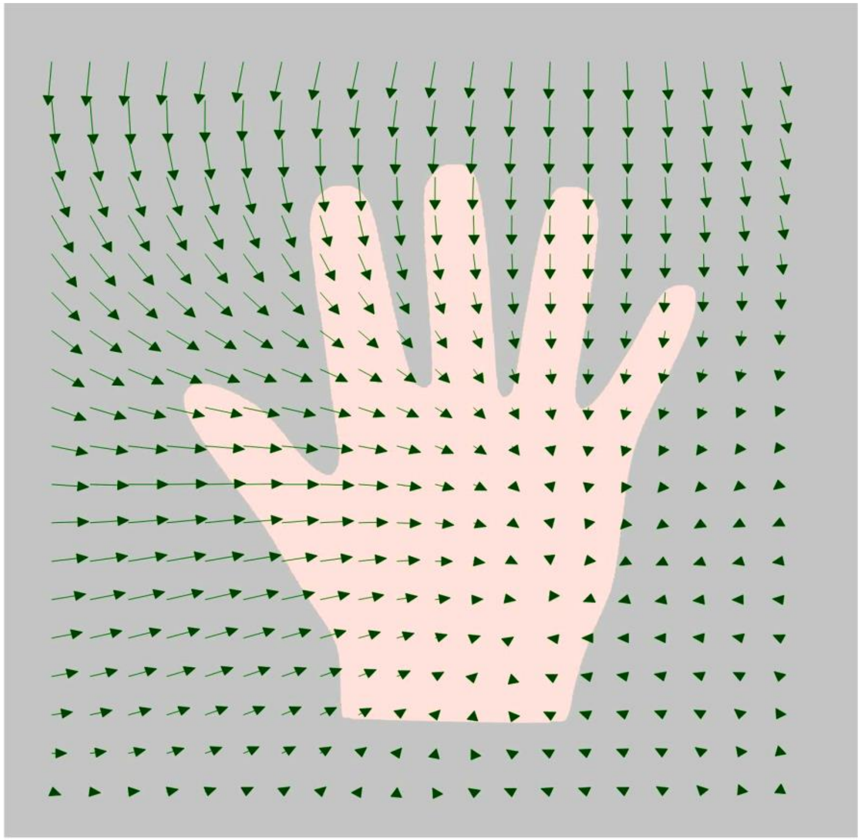
Zero mean:

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

Squared exponential covariance function (Gaussian kernel)

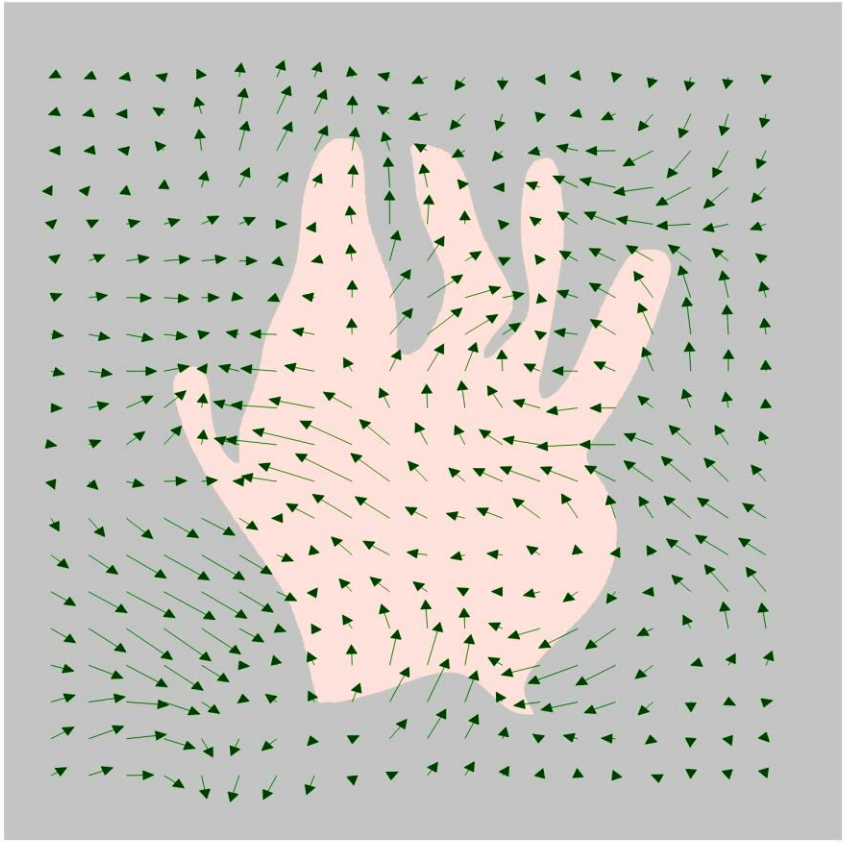
$$k(x, x') = \begin{pmatrix} s_1 \exp\left(-\frac{\|x - x'\|^2}{\sigma_1^2}\right) & 0 \\ 0 & s_2 \exp\left(-\frac{\|x - x'\|^2}{\sigma_2^2}\right) \end{pmatrix}$$

Example prior: Smooth 2D deformations



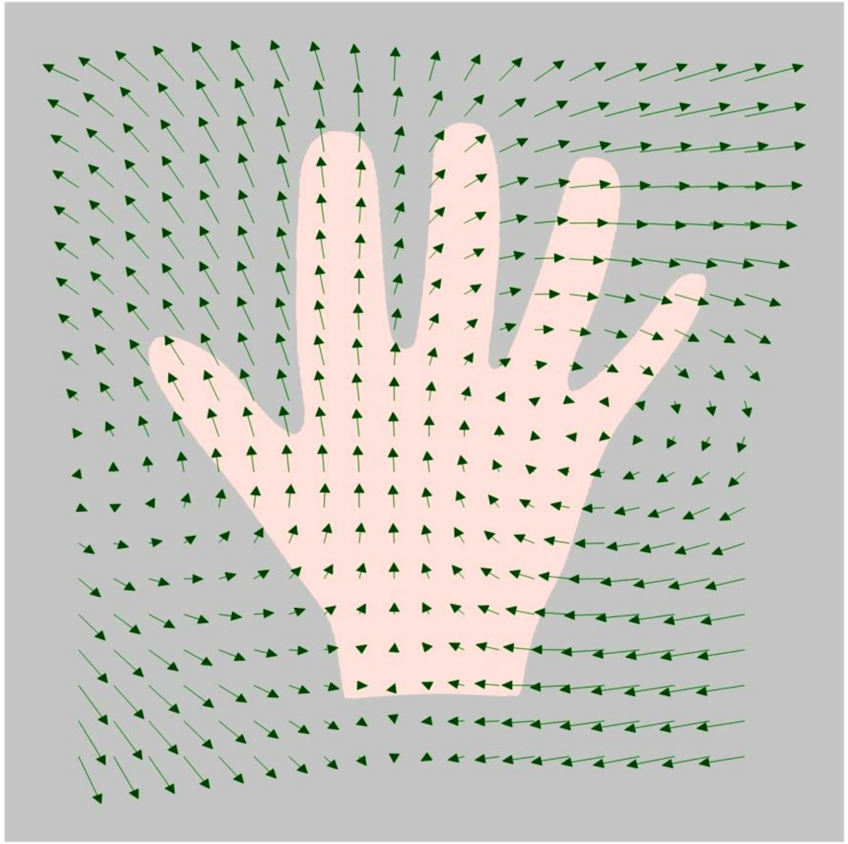
$s_1 = s_2$ small, $\sigma_1 = \sigma_2$ large

Example prior: Smooth 2D deformations



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

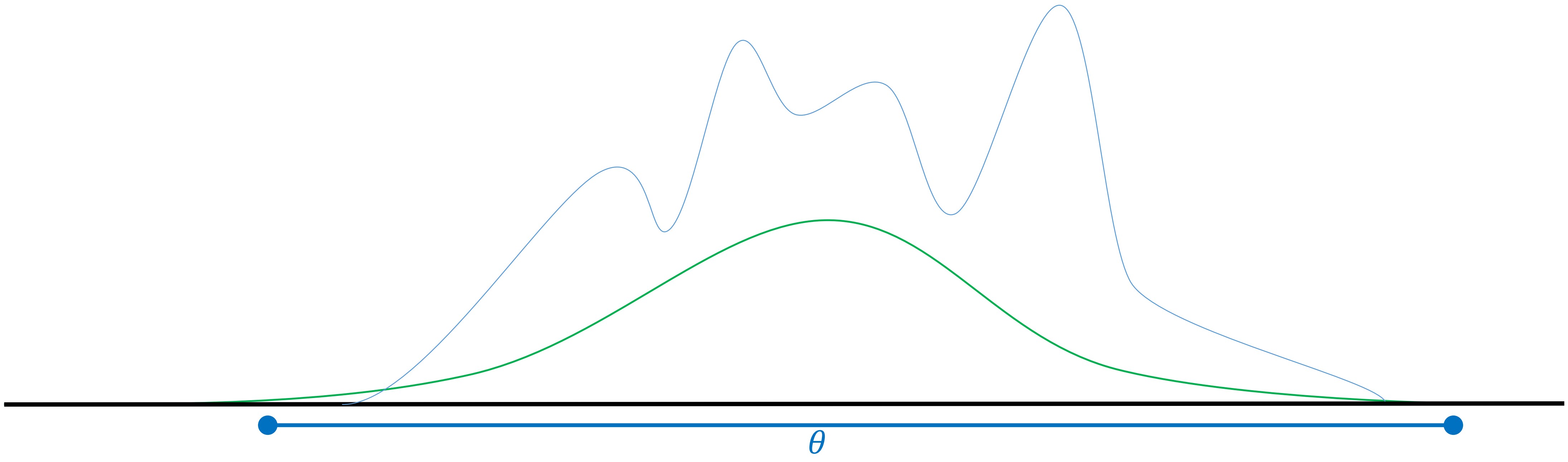
Example prior: Smooth 2D deformations



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

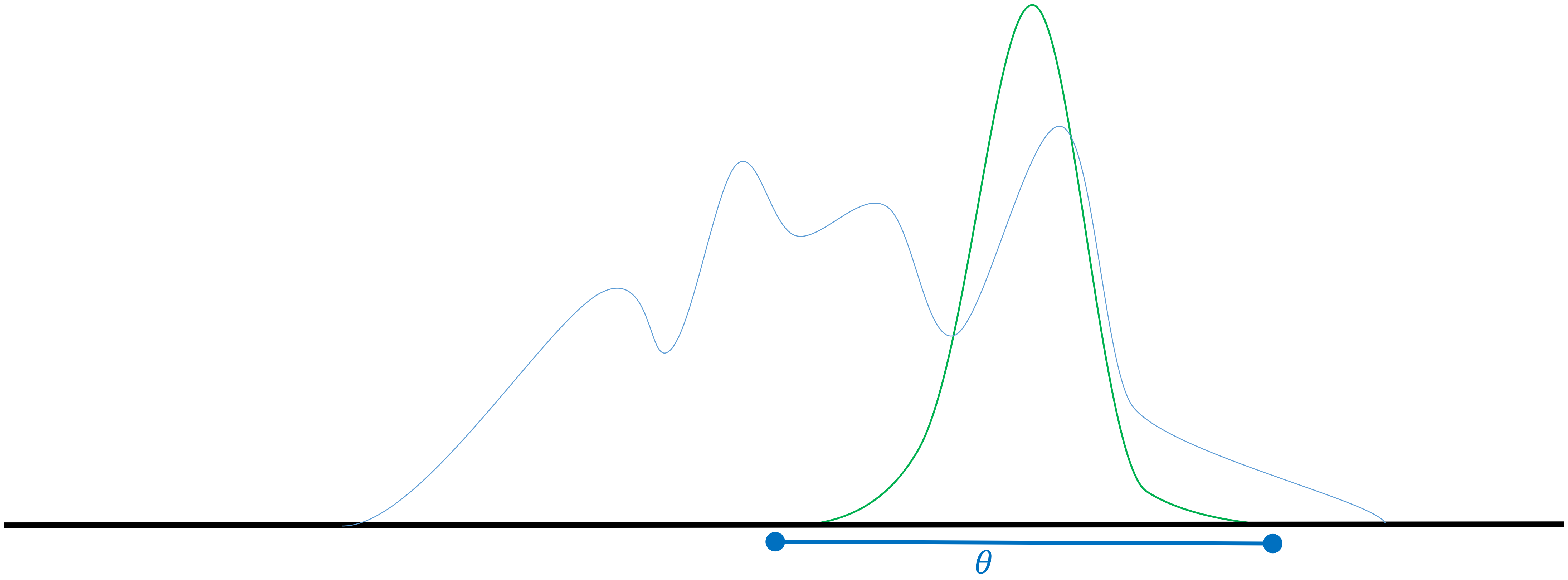
Why are priors interesting?

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



Why are priors interesting?

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



Intermezzo – The space of samples

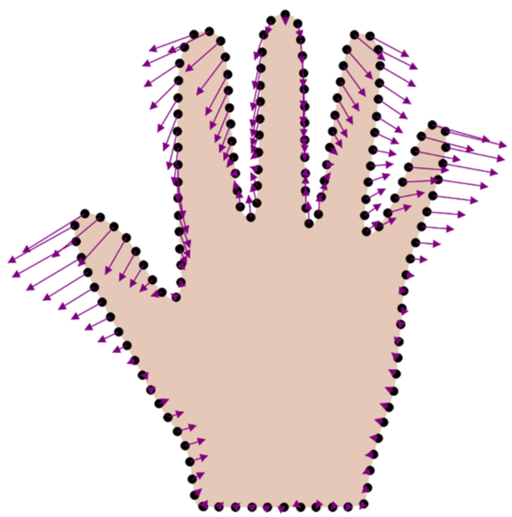
Gaussian processes - Deeper Insights

Scalar-valued Gaussian processes

Vector-valued (this course)

- Samples u are deformation fields:

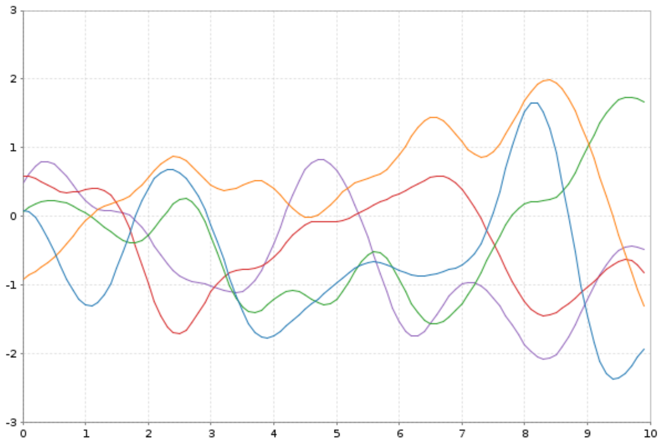
$$u: \mathbb{R}^n \rightarrow \mathbb{R}^d$$



Scalar-valued (more common)

- Samples f are real-valued functions

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$



The space of samples

$u \sim \mu + \sum_i \alpha_i \sqrt{\lambda_i}\, \phi_i = \sum_i \beta_i k(x_i, \cdot)$ for some β

Argument:

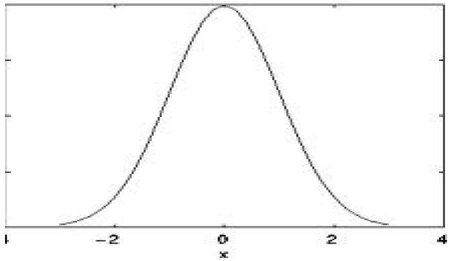
- Covariance function k is symmetric and positive definite
- For any finite sample it holds that:
  => the covariance matrix is symmetric
  => rowspace = columnspace = eigenspace

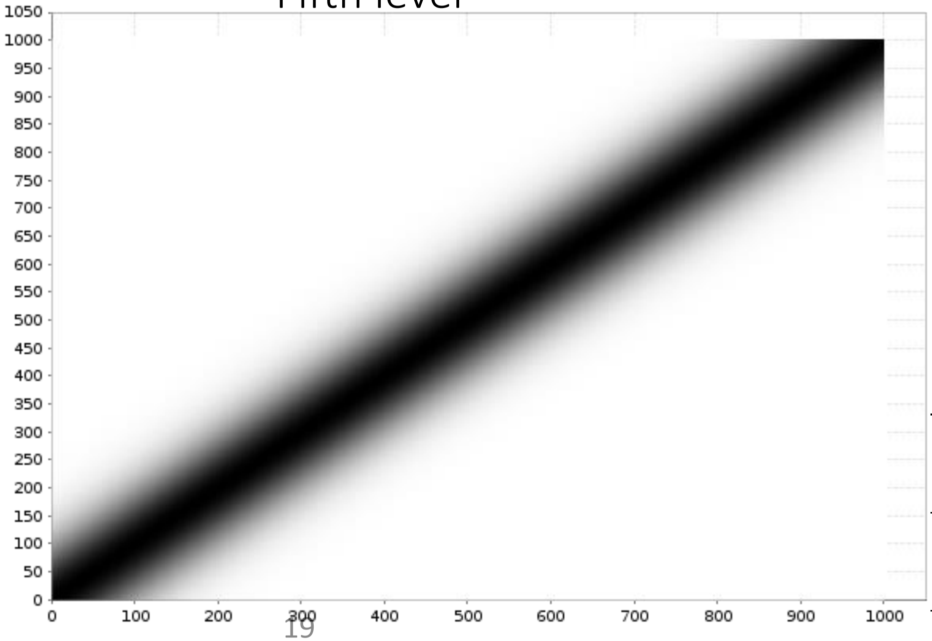
Samples are linear combinations of the "rows" of k

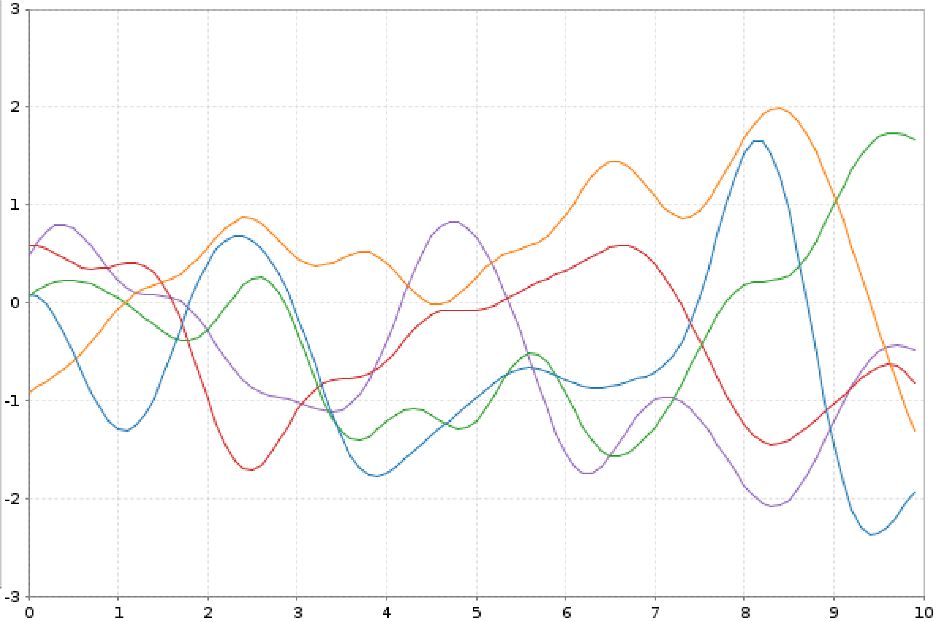
Example: Gaussian kernel

- Click to edit Master text styles
  - Second level
    - Third level
      - Fourth level
        - Fifth level

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

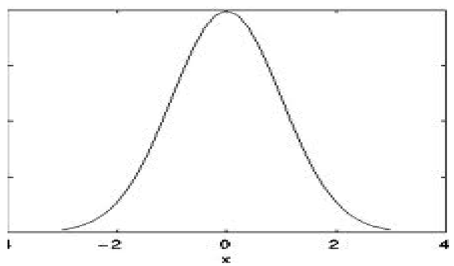




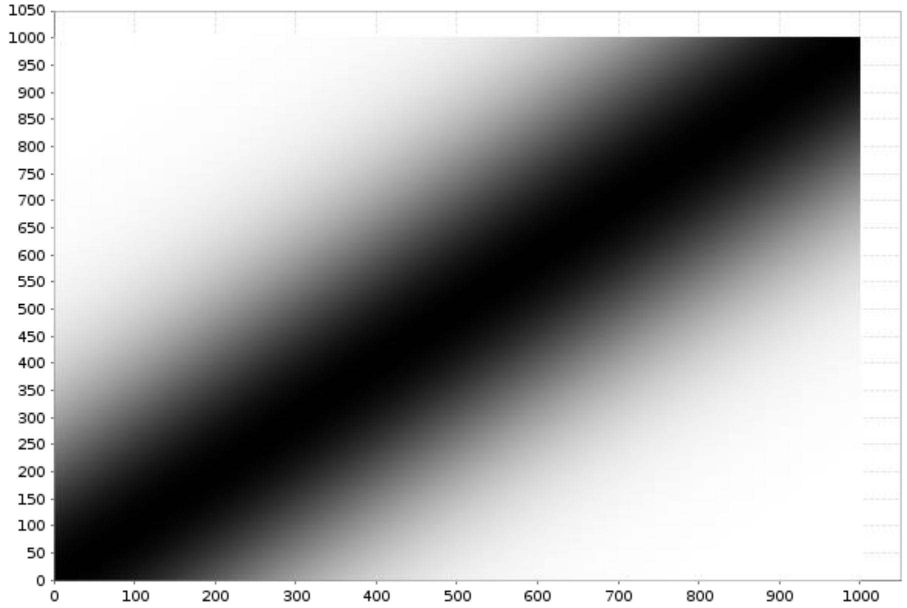


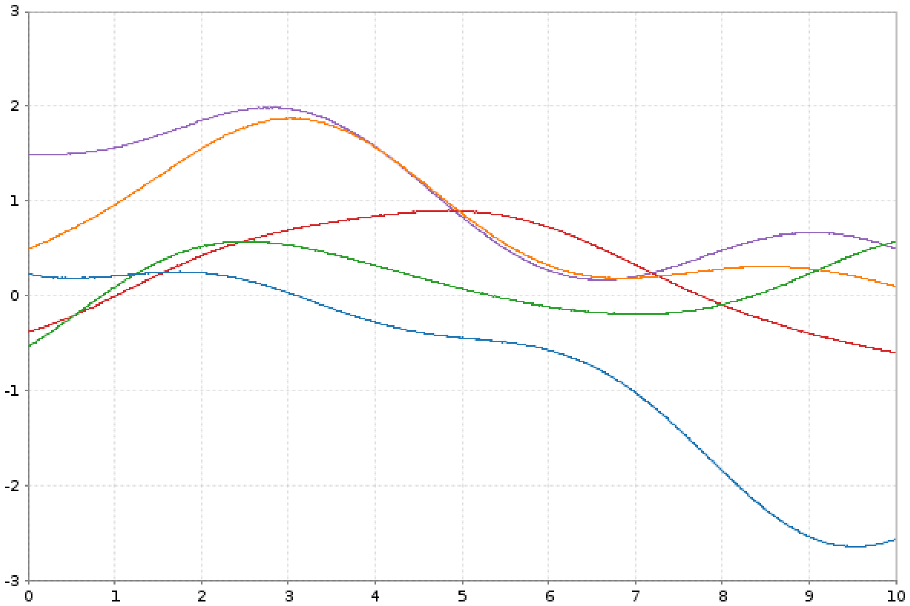
Example: Gaussian kernel

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



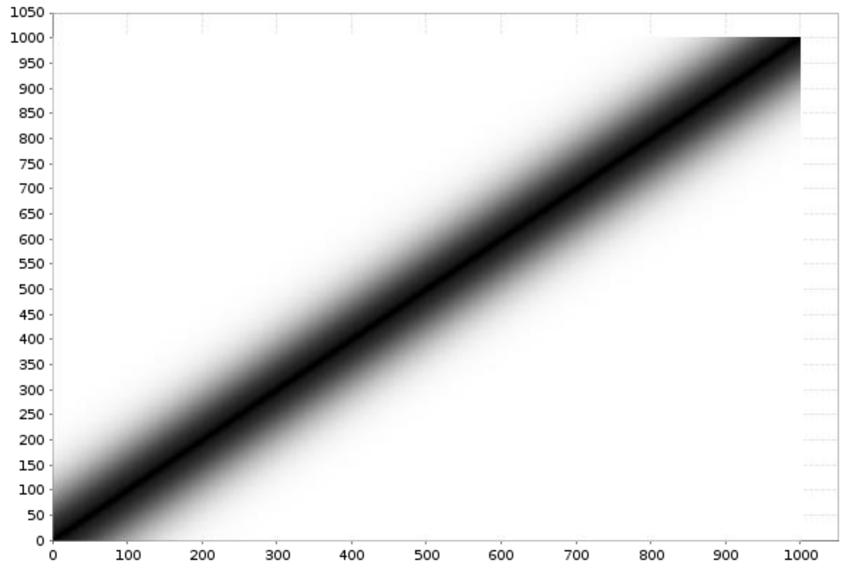
$$\sigma = 3$$

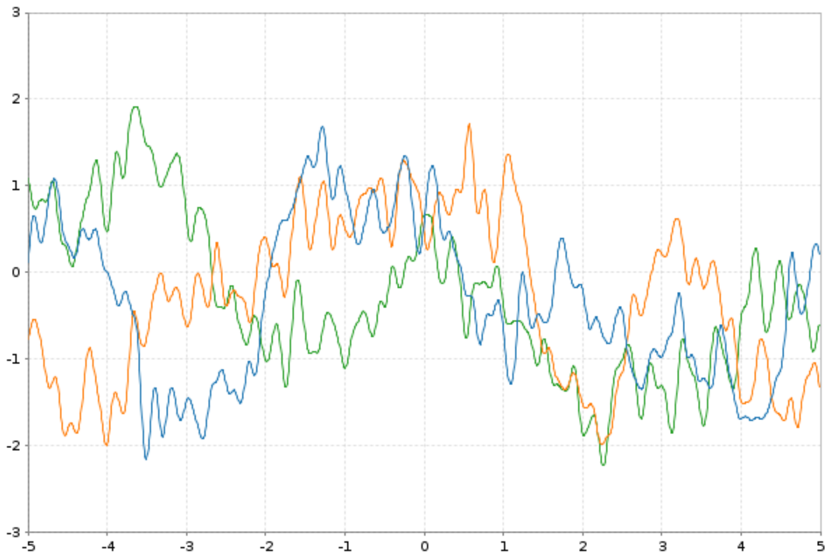




Multi-scale signals

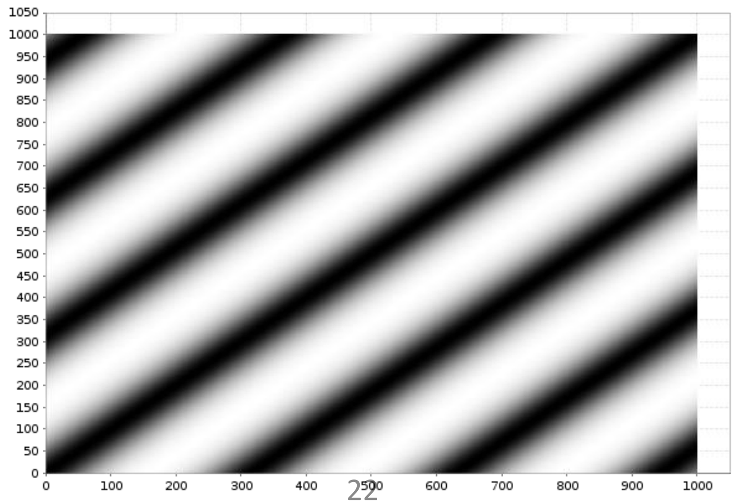
- $\mathrm{k(x, x')} = \exp\left(-\left\|x - \frac{x'}{1}\right\|^2\right) + 0.1 \exp\left(-\left\|x - \frac{x'}{0.1}\right\|^2\right)$

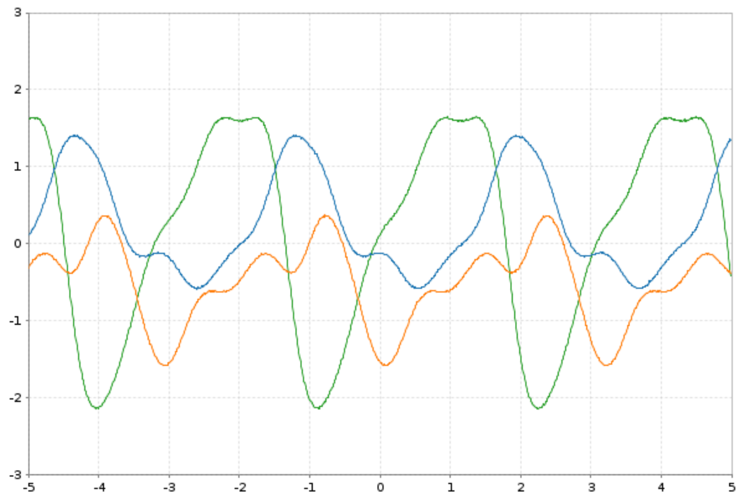




Periodic kernels

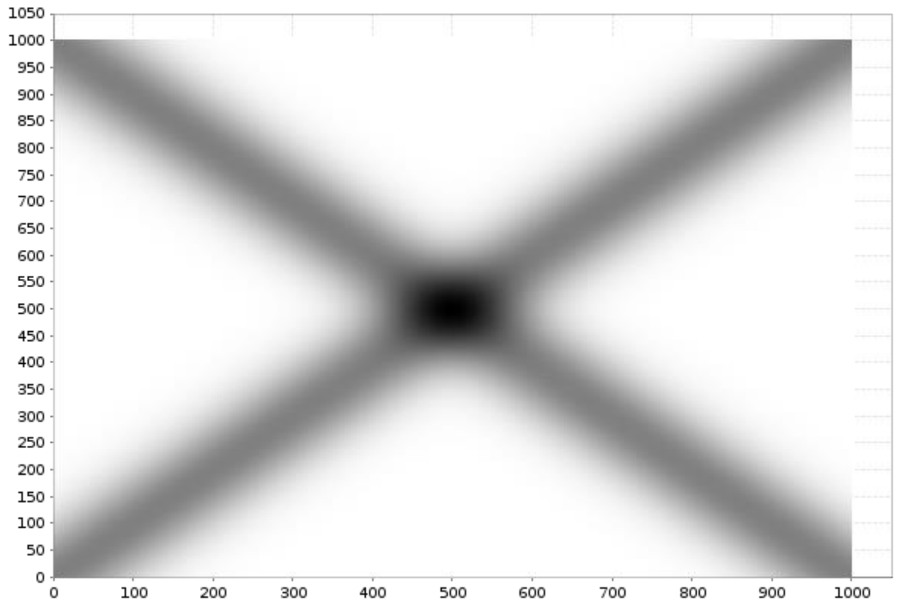
- Define $u(x) = \begin{pmatrix} \cos(x) \\ \sin(x) \end{pmatrix}$
- $k(x, x') = \exp(-\|(u(x) - u(x')\|^2 = \exp(-4\sin^2\left(\frac{\|x - x'\|}{\sigma^2}\right))$

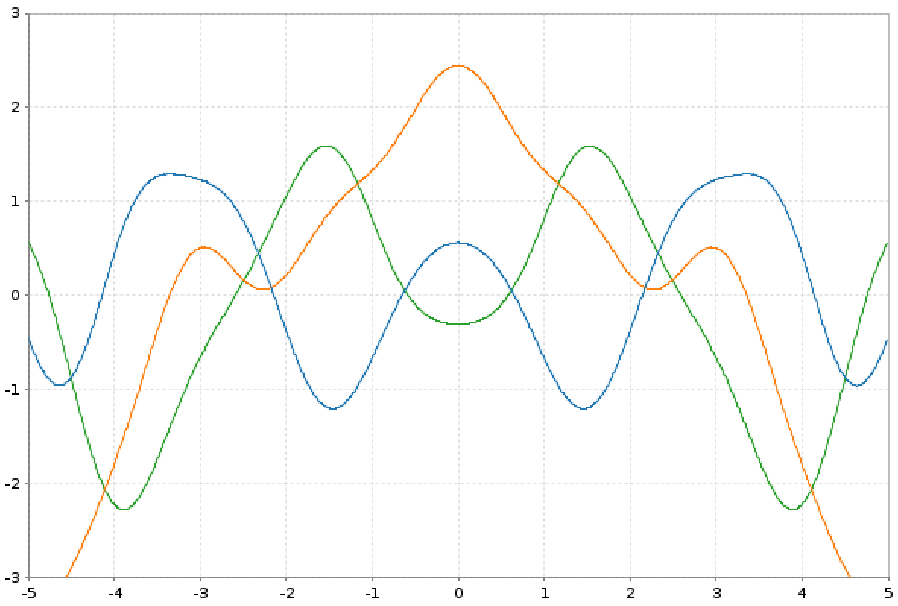




Symmetric kernels

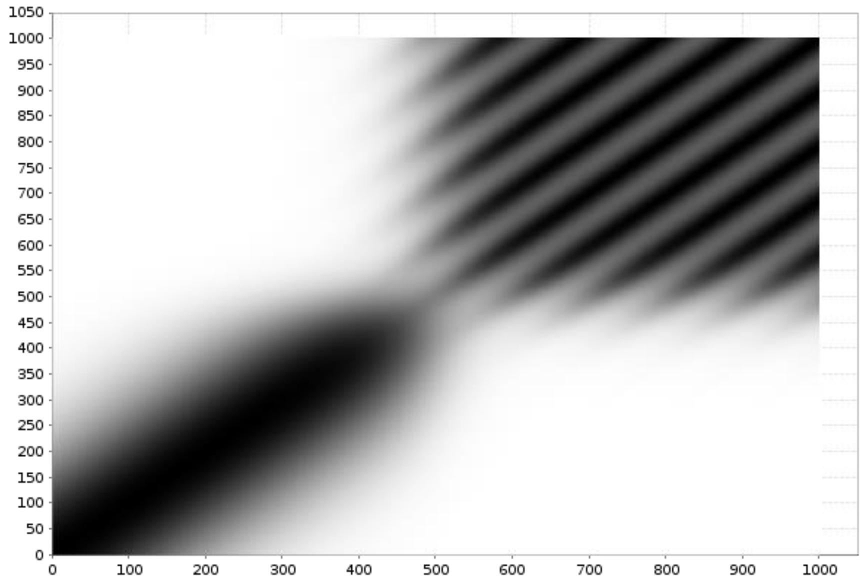
- *Enforce that f(x) = f(-x)*
- $k(x, x') = k(-x, x') + k(x, x')$

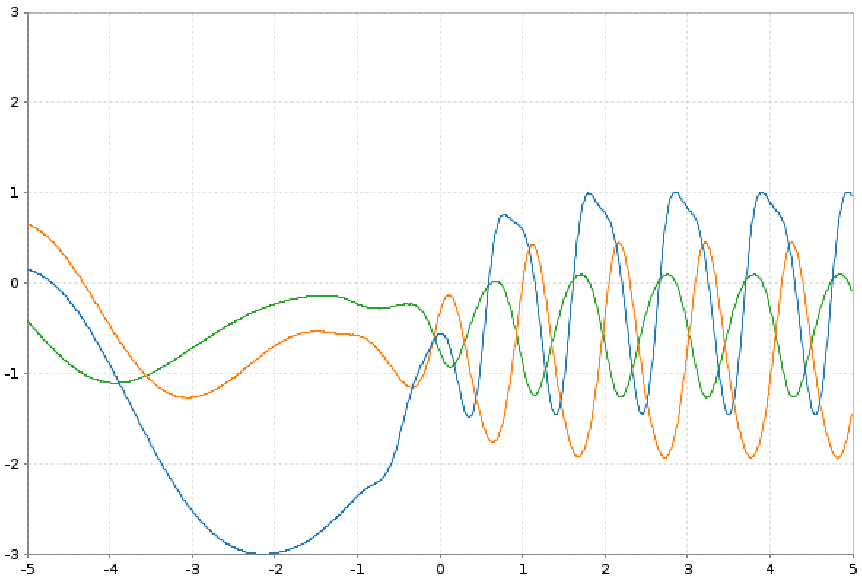




Changepoint kernels

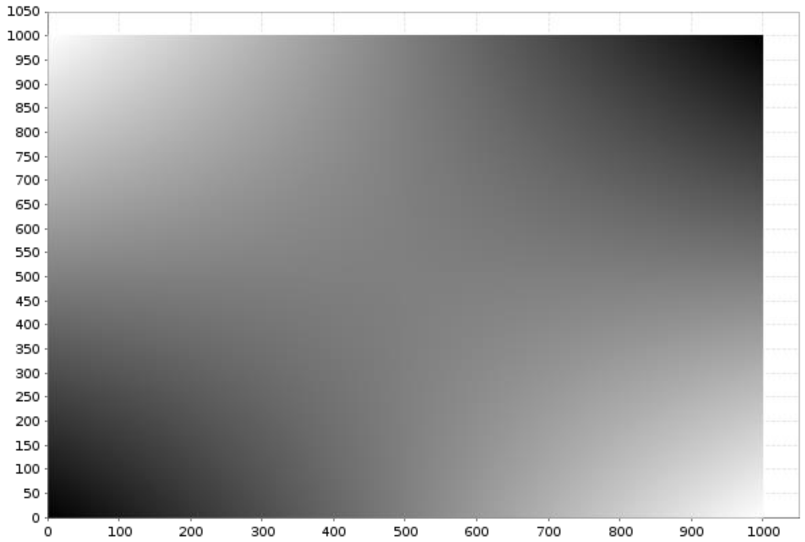
- $k(x, x') = s(x)k_1(x, x')s(x') + (1 - s(x))k_2(x, x')(1 - s(x'))$
- $\mathrm{s}(x) = \frac{1}{1+\exp(-x)}$

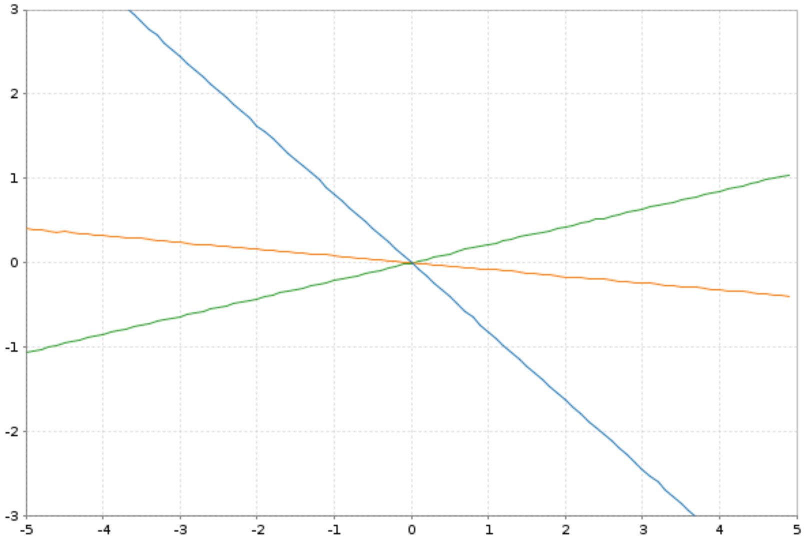




Combining existing functions

$$k(x, x') = f(x)f(x')$$

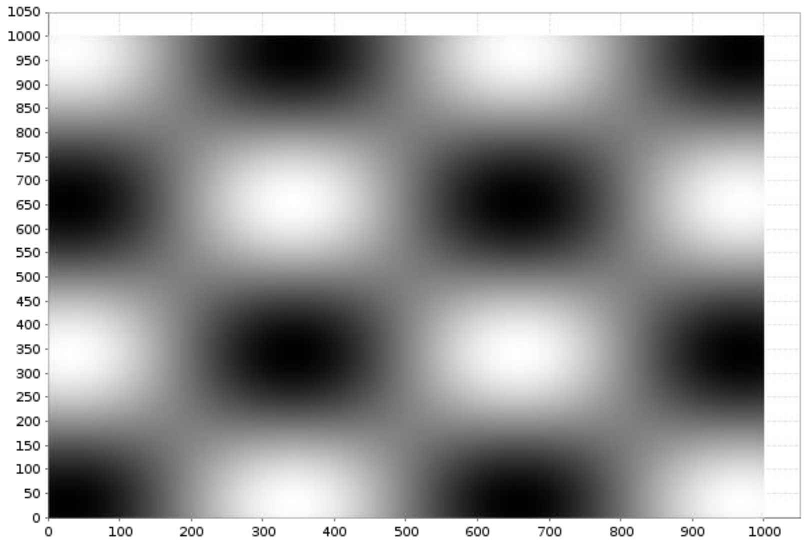


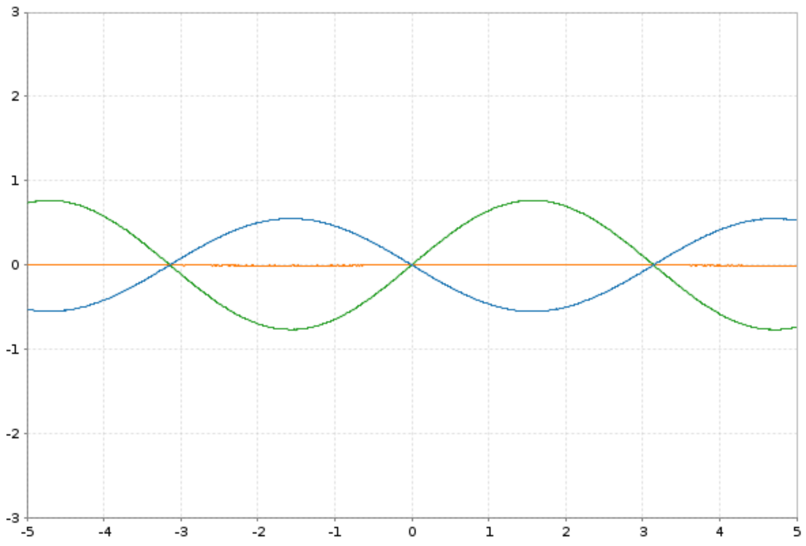


$$f(x) = x$$

Combining existing functions

$$k(x, x') = f(x)f(x')$$

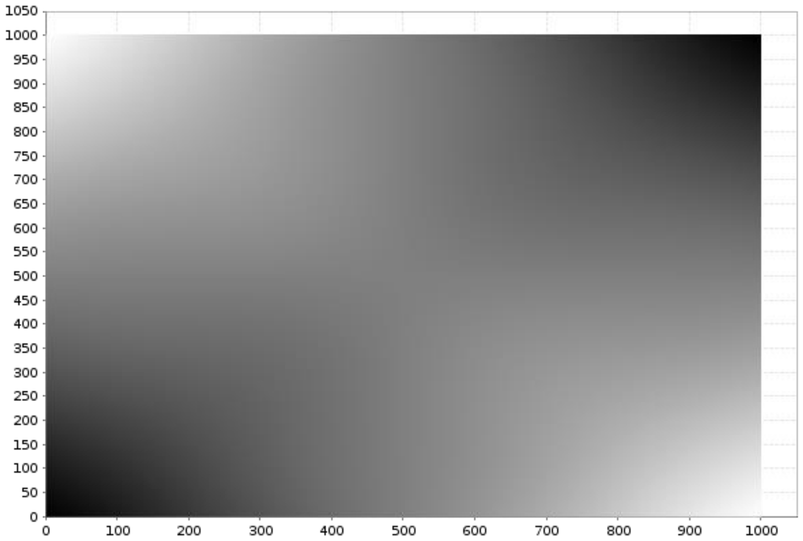


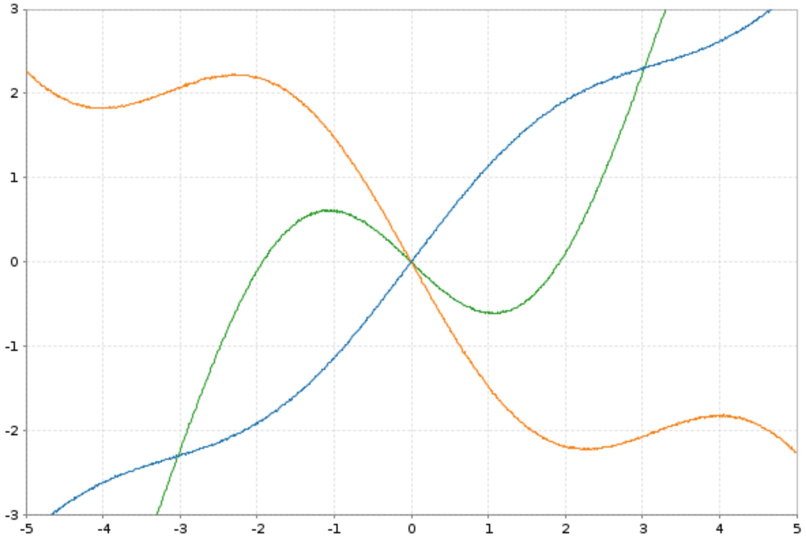


$$\mathrm{f(x) = sin(x)}$$

Combining existing functions

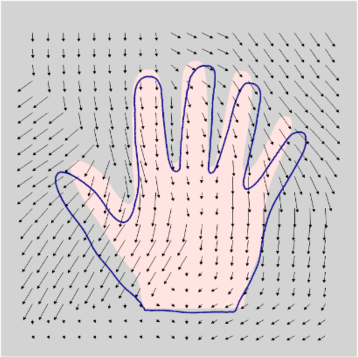
$$k(x,x') = \sum_i f_i(x) f_i(x')$$



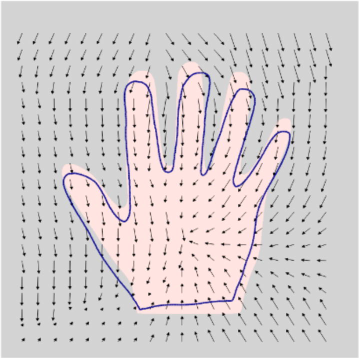


$$\{f_1(x) = x, f_2(x) = \sin(x)\}$$

Statistical models

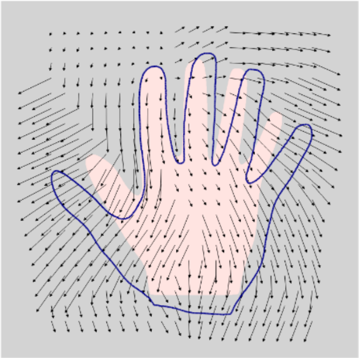


$u^1 : \Omega \to \mathbb{R}^2$



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

…



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

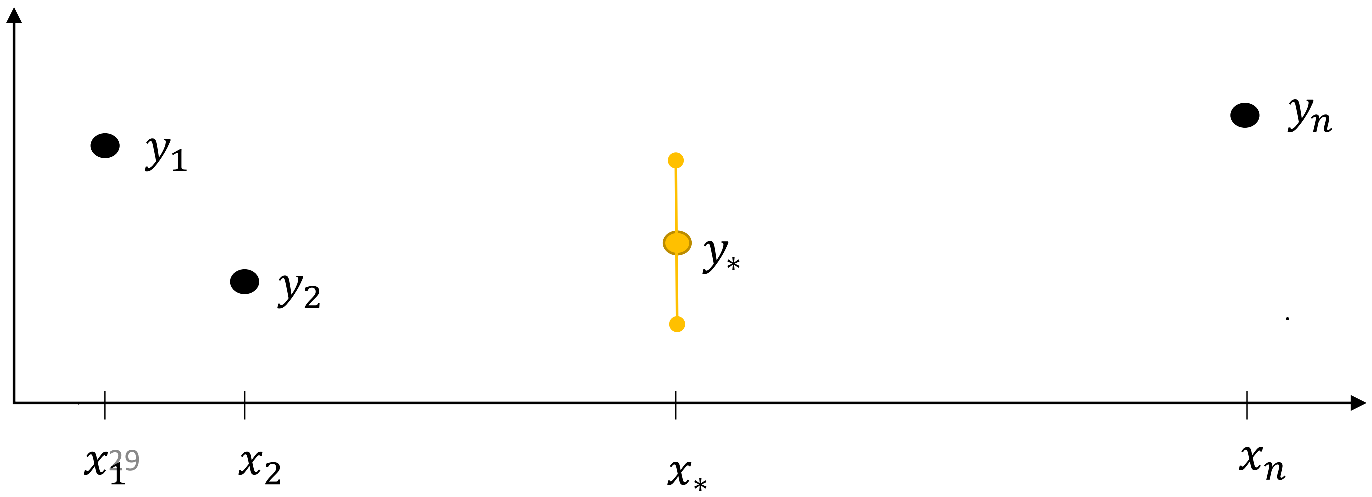
$$\mu(x) = \overline{u}(x) = \frac{1}{n}\sum_{i-1}^{n} u^i\,(x)$$

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

Statistical shape models are linear combinations of example deformations $u^1, \ldots u^n$.

Gaussian process regression

- Given: observations $\{(x_1, y_1), \dots, (x_n, y_n)\}$
- Model: $y_i = f(x_i) + \epsilon,\ f \sim GP(\mu, k)$
- Goal: compute p$(y_* | x_*, x_1, \dots, x_n, y_1, \dots, y_n)$



Gaussian process regression

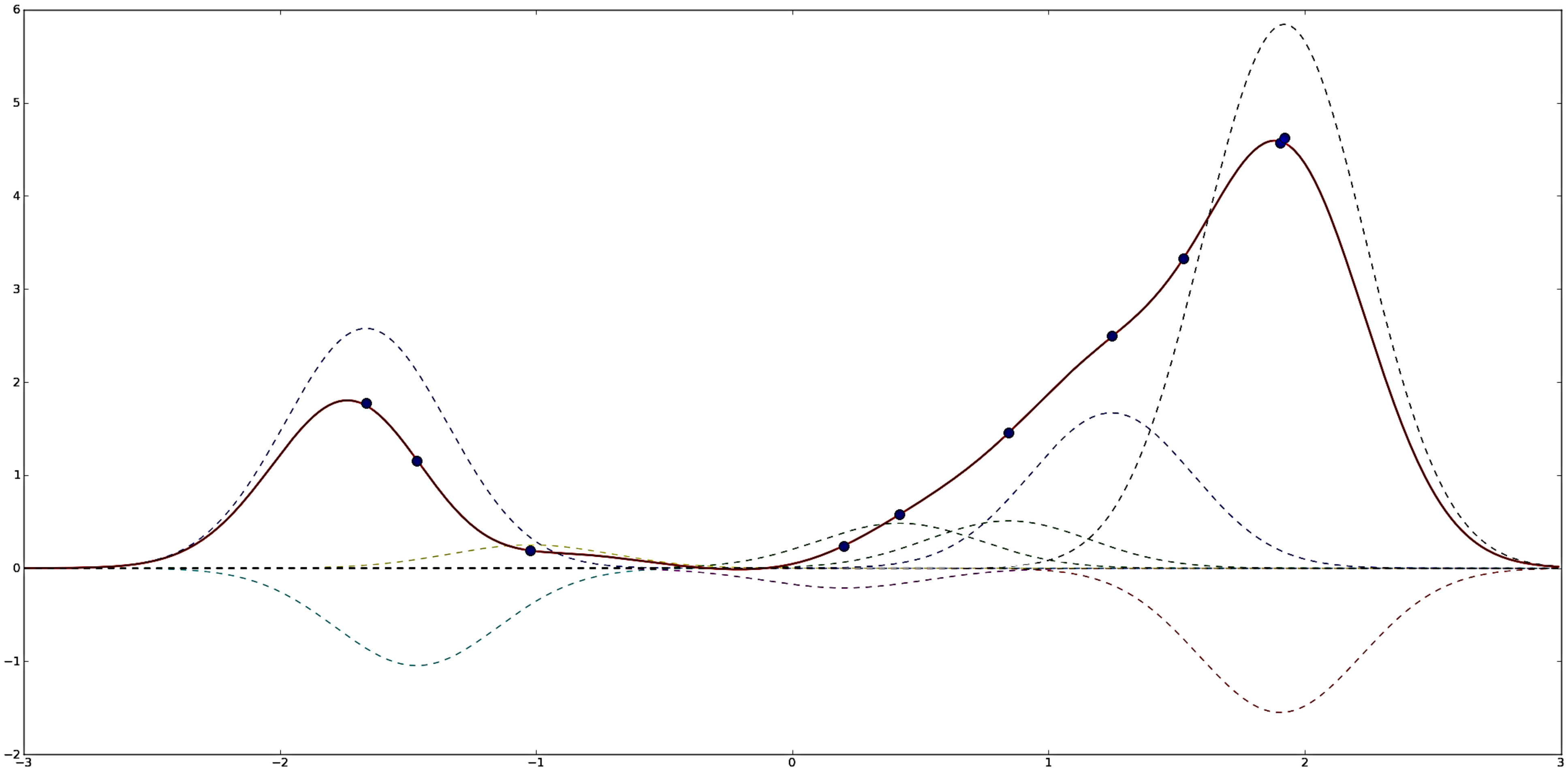
- Solution given by posterior process $GP(\mu_p, k_p)$ with

$$\mu_p(x_*) = K(x_*, X)[K(X, X) + \sigma^2 I]^{-1} y$$

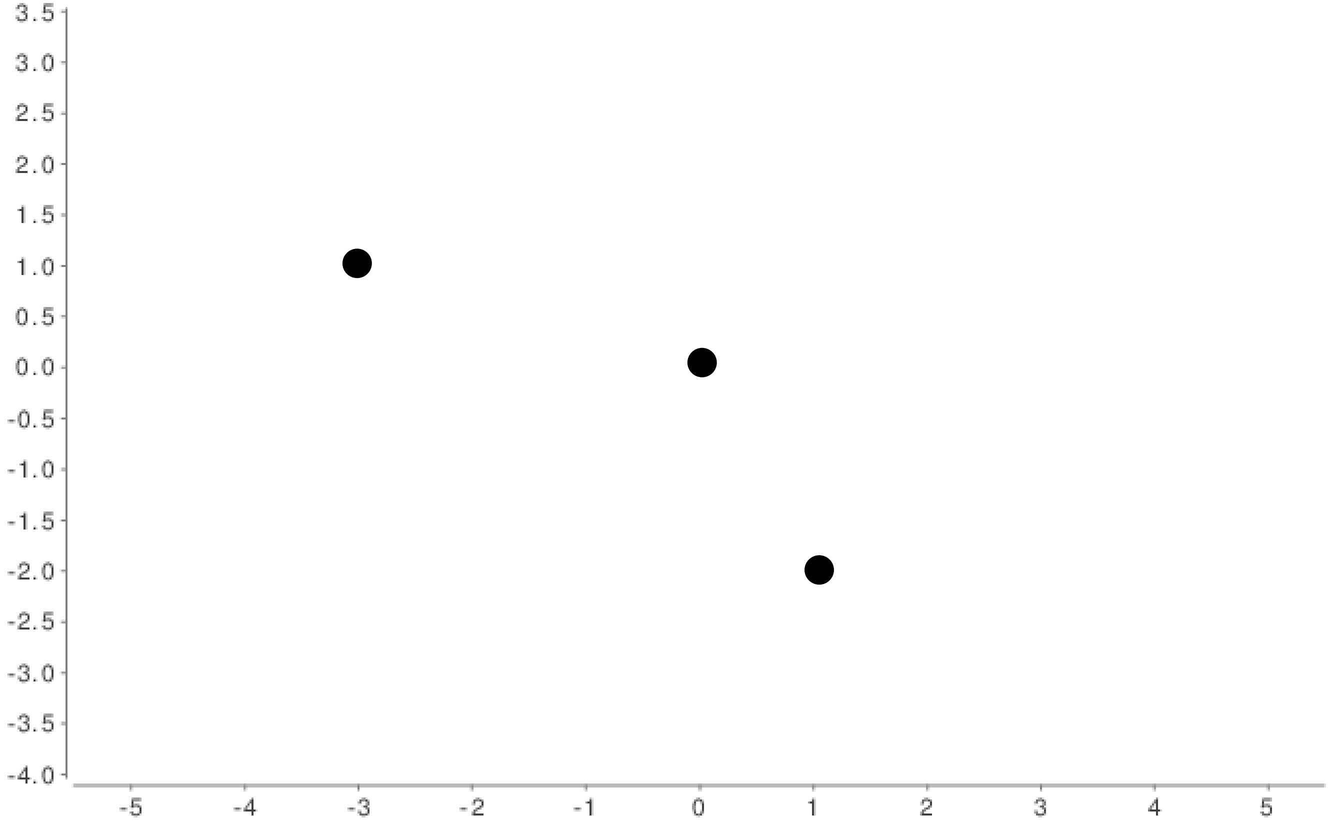
$$k_p(x_*, x_*{}') = k(x_*, x_*{}') - K(x_*, X)[K(X, X) + \sigma^2 I]^{-1} K(X, x_*')$$

- The covariance is independent of the value at the training points
  - Structure of posterior GP determined solely by kernel.
- The most likely solution is a linear combination of kernels evaluated at the training points
  - This is known as the Representer Theorem in machine learning.
  - Structure of solution determined solely by kernel.

Illustration: Representer theorem

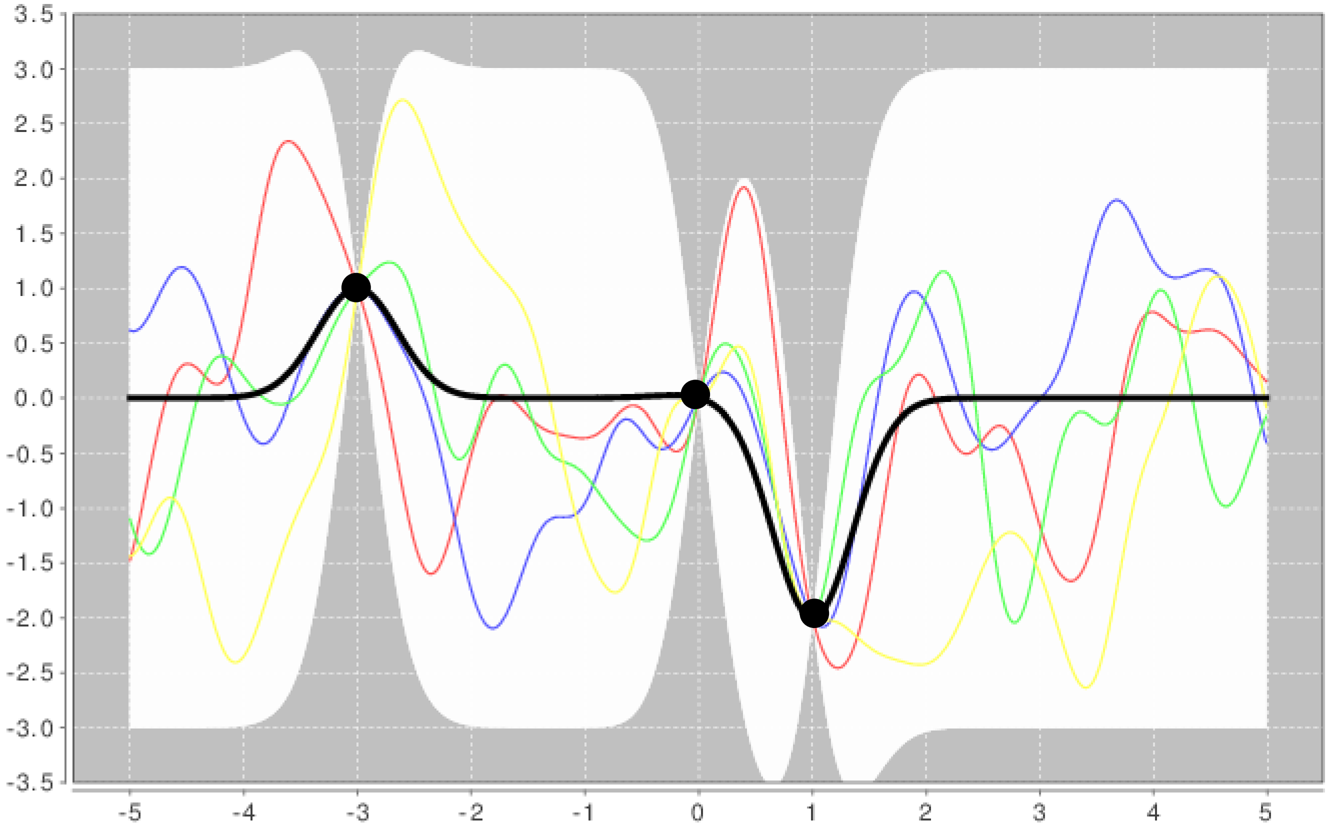


Examples



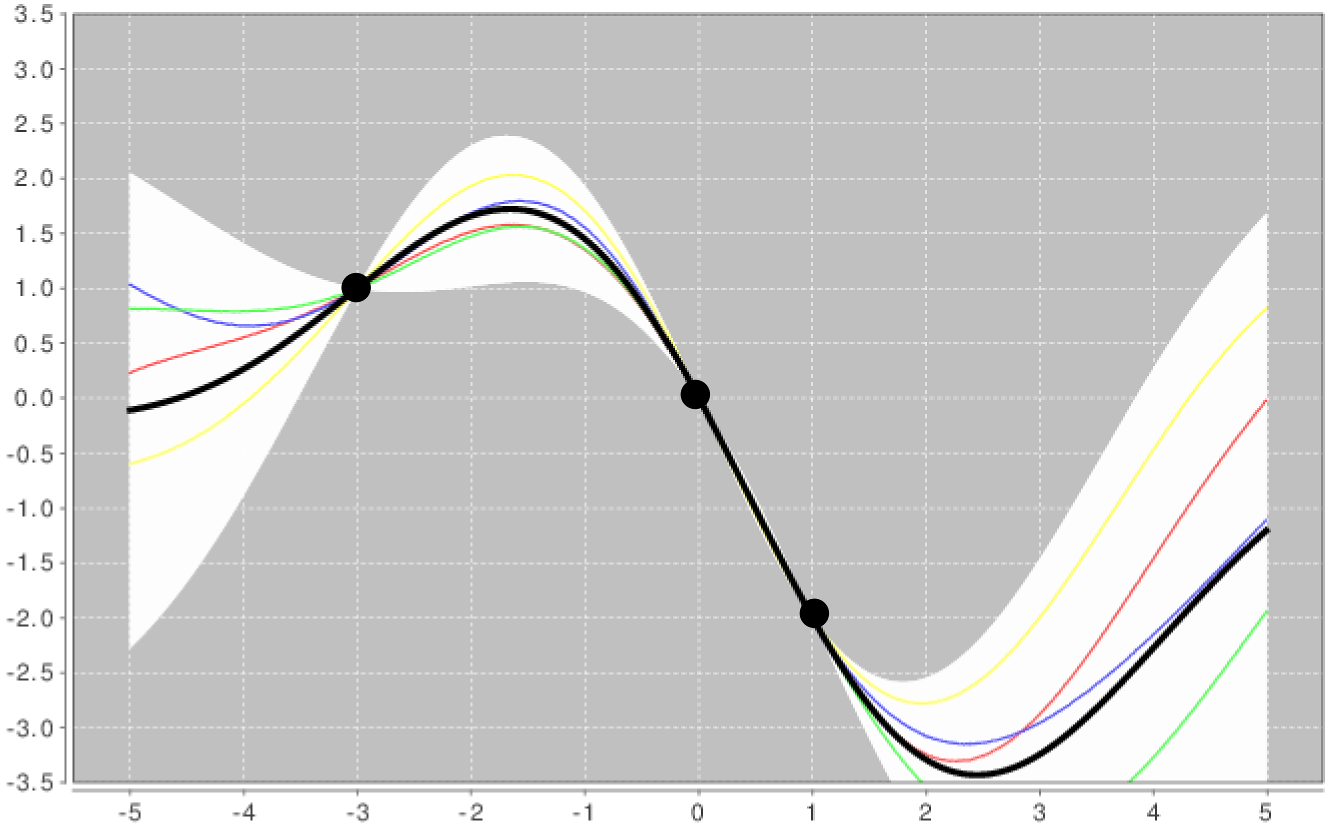
Examples

- Gaussian kernel ($\sigma = 1$)



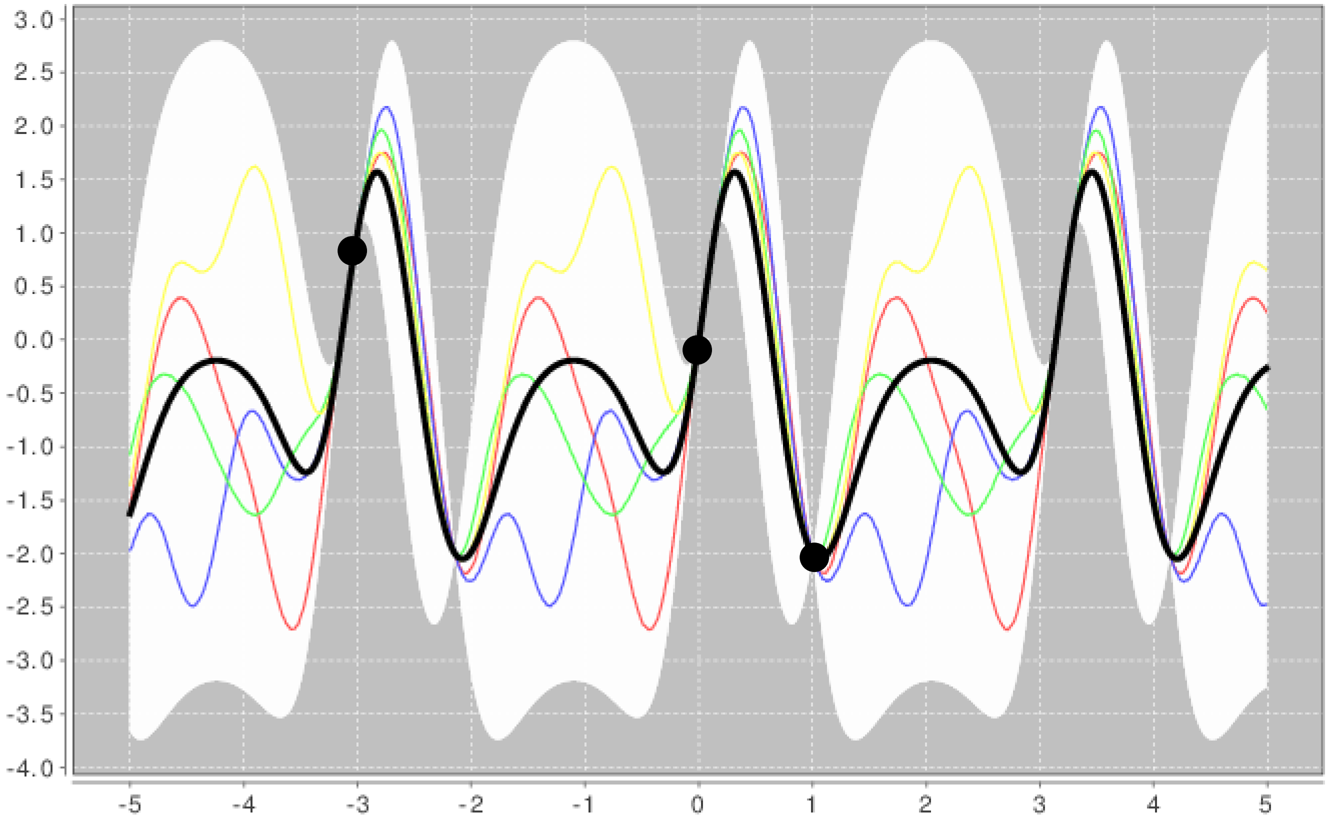
Examples

- Gaussian kernel ($\sigma = 5$)



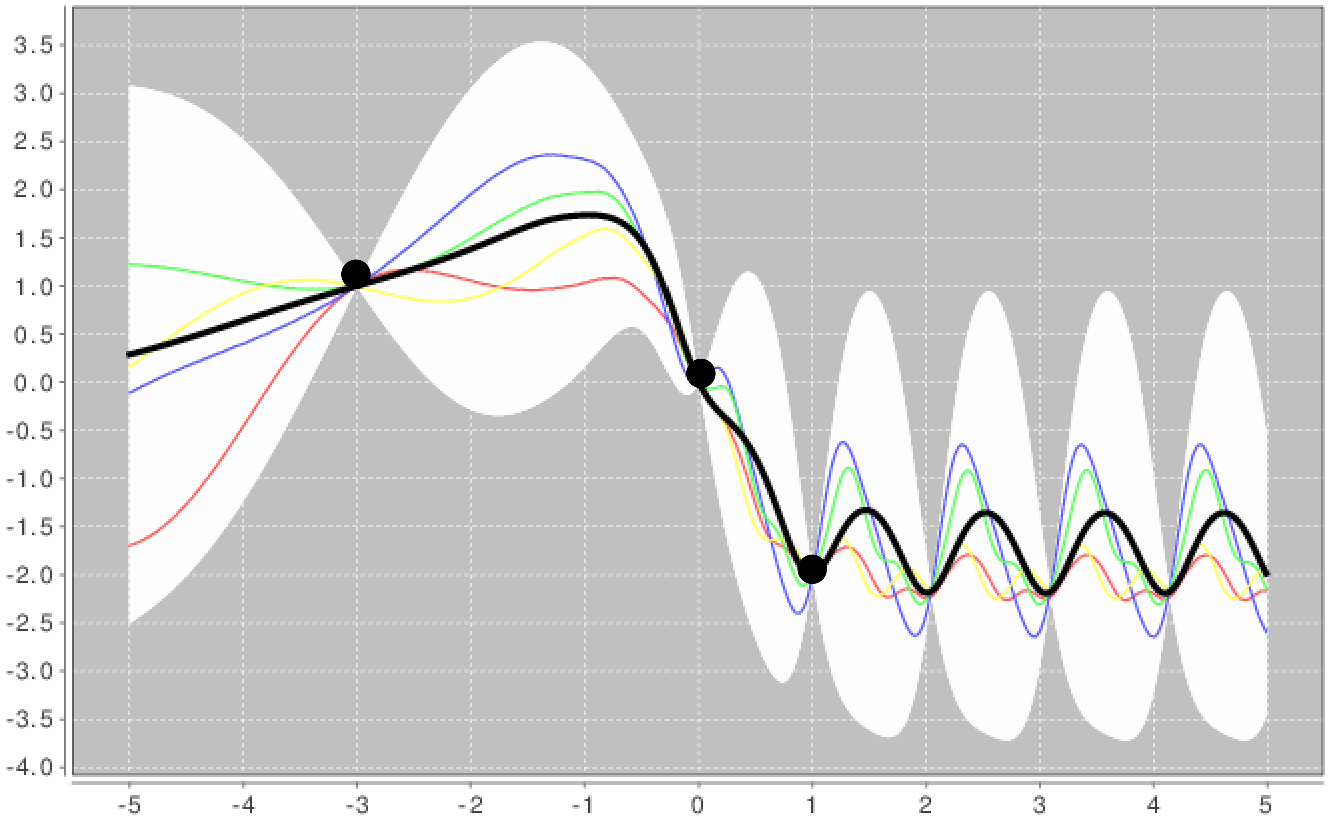
Examples

- Periodic kernel



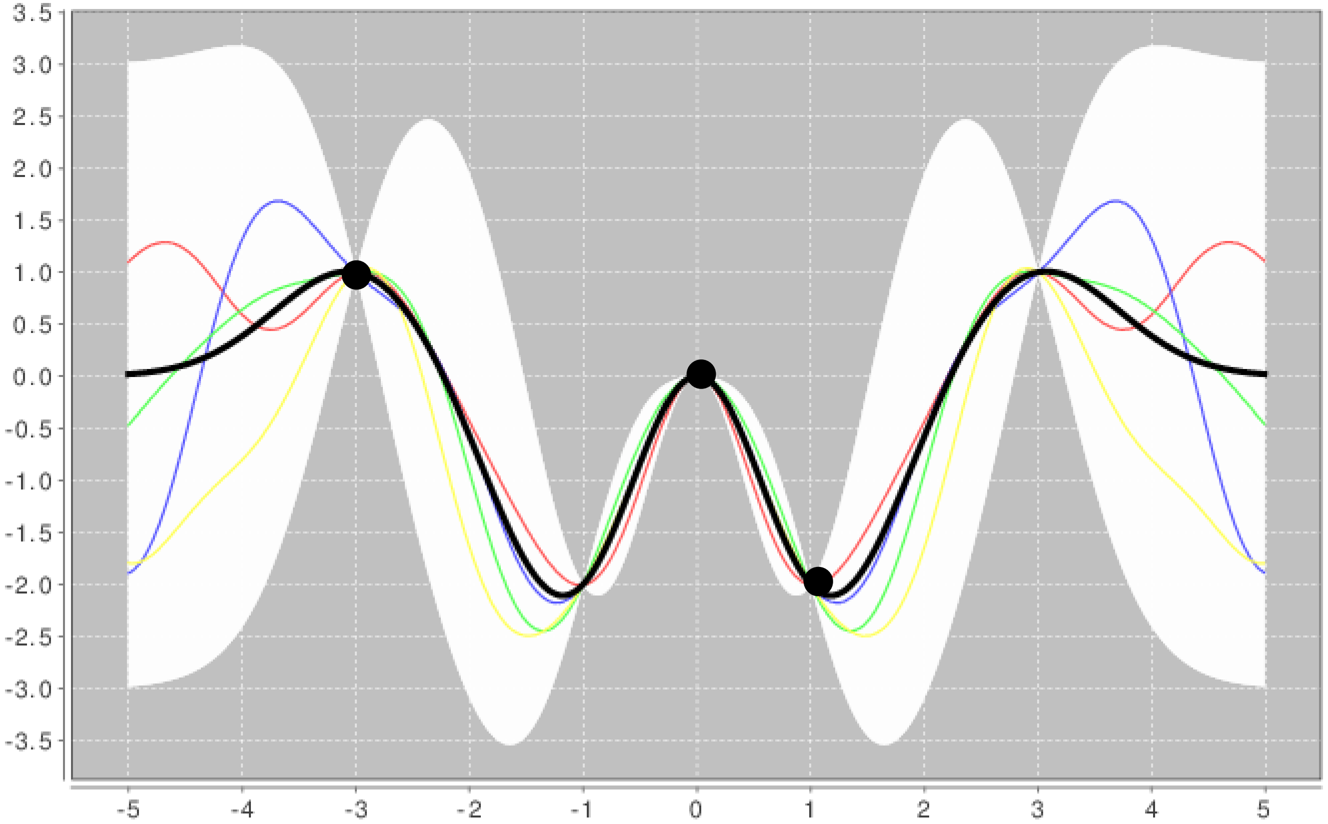
Examples

- Changepoint kernel



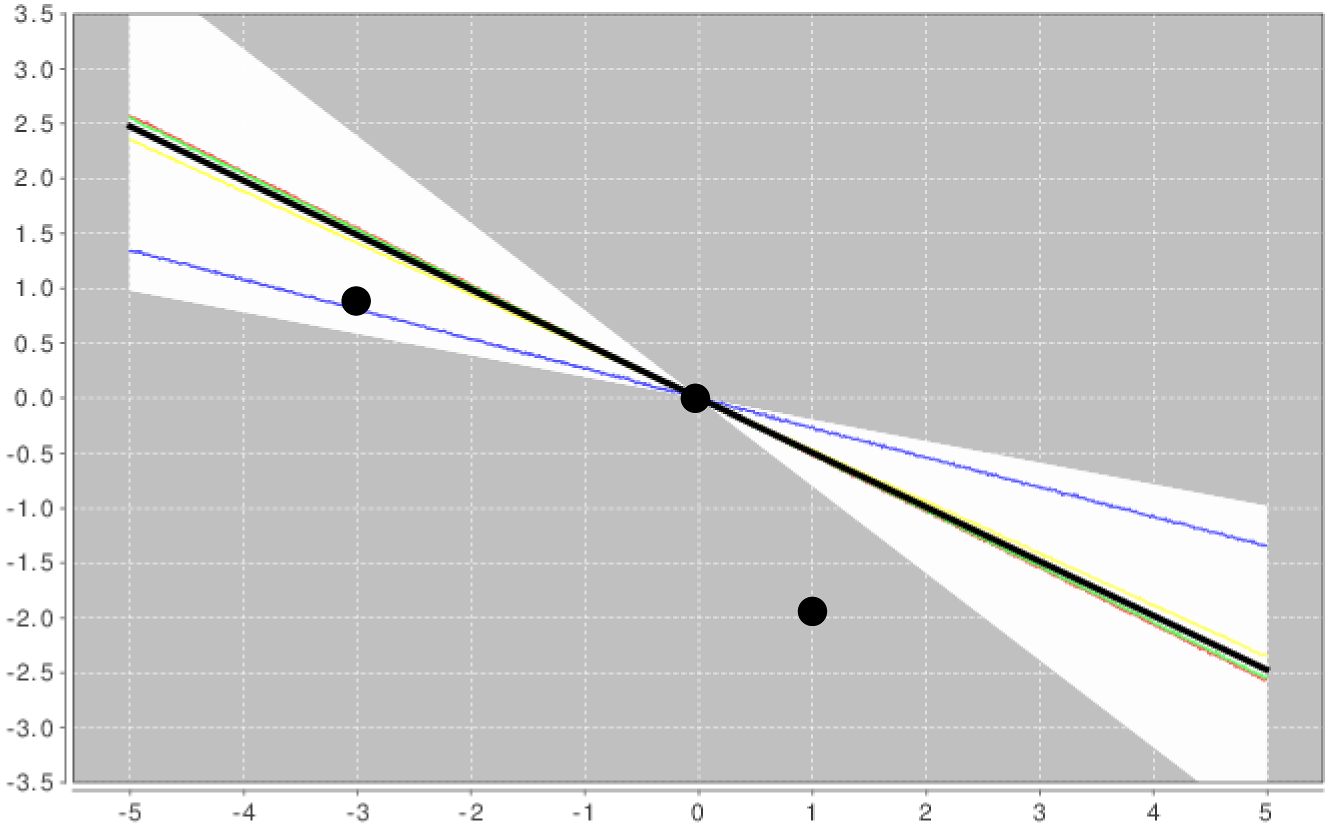
Examples

- Symmetric kernel



Examples

- Linear kernel



Deformation models for registration

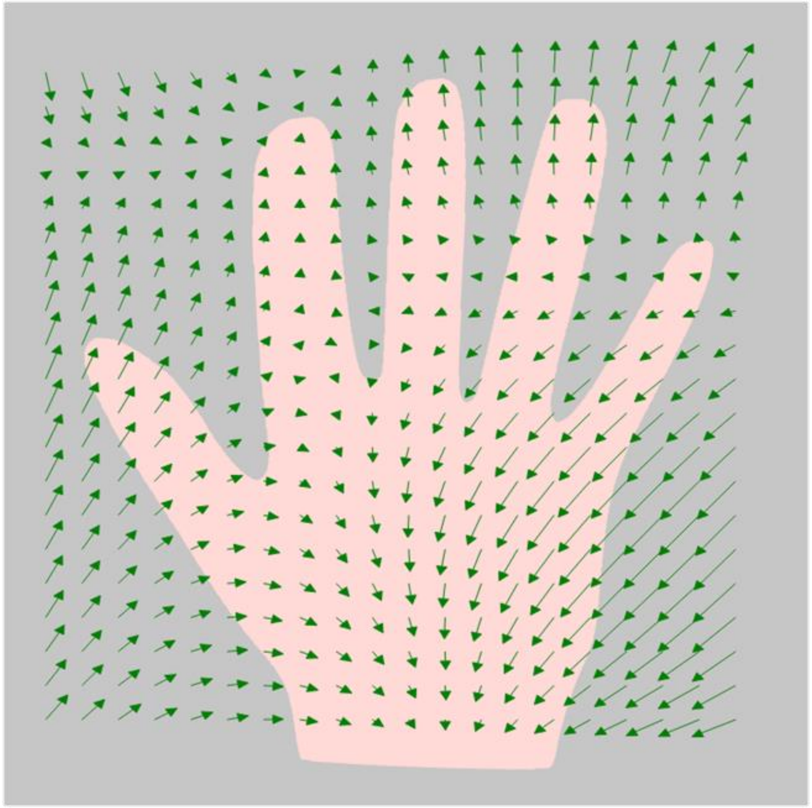
Basic assumption: Deformation fields are smooth

- 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

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



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

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

- Corresponding covariance function

$$k(x, x') = \frac{1}{12}\left(2\|x - x'\|^3 - 3R(\|x - x'\|^2 + R^3\right)$$
$$\text{where } R = \max_{x, x' \in \Omega} \|x - x'\|$$

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

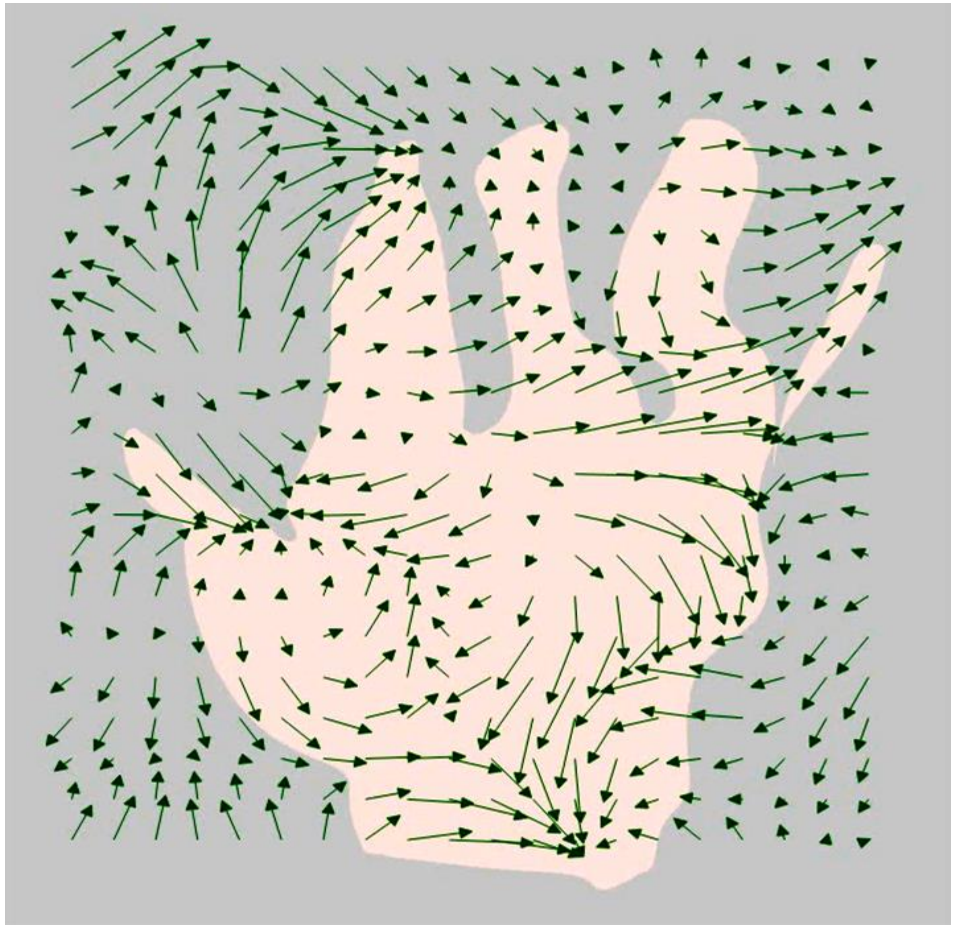
Add kernels that act on different scales:

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

- Wavelet like multiscale representation

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

Multi-scale kernel



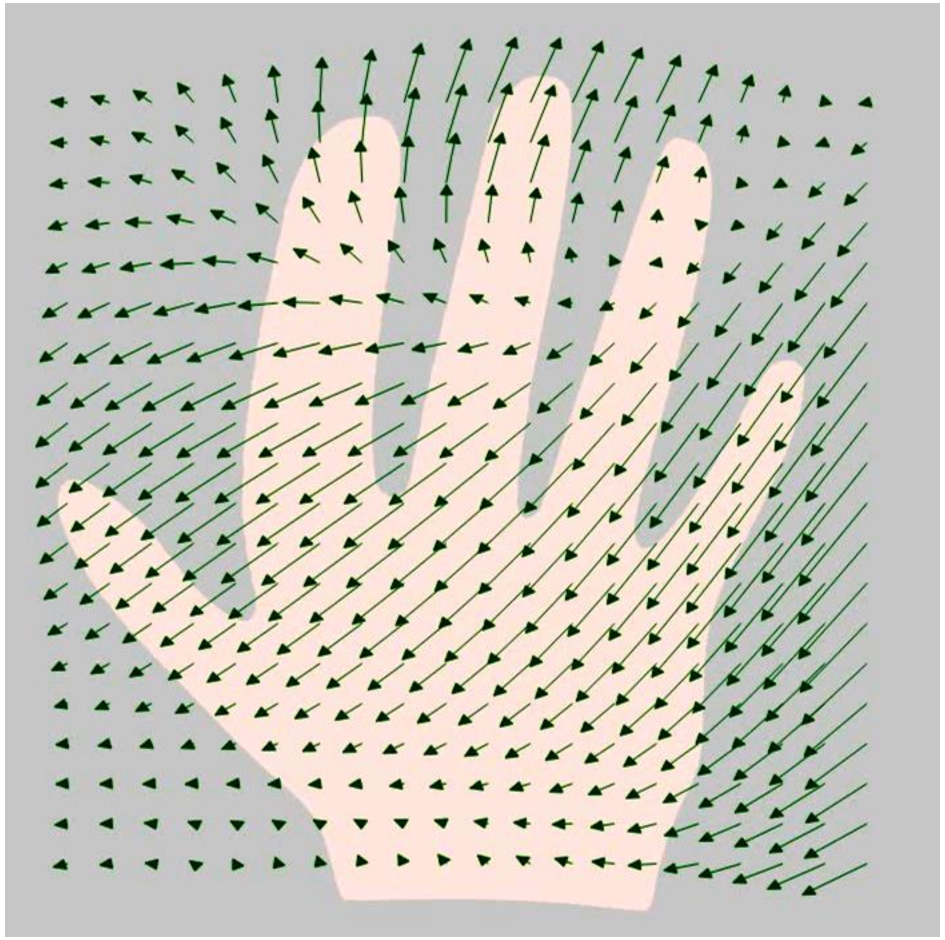
Anisotropic priors

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

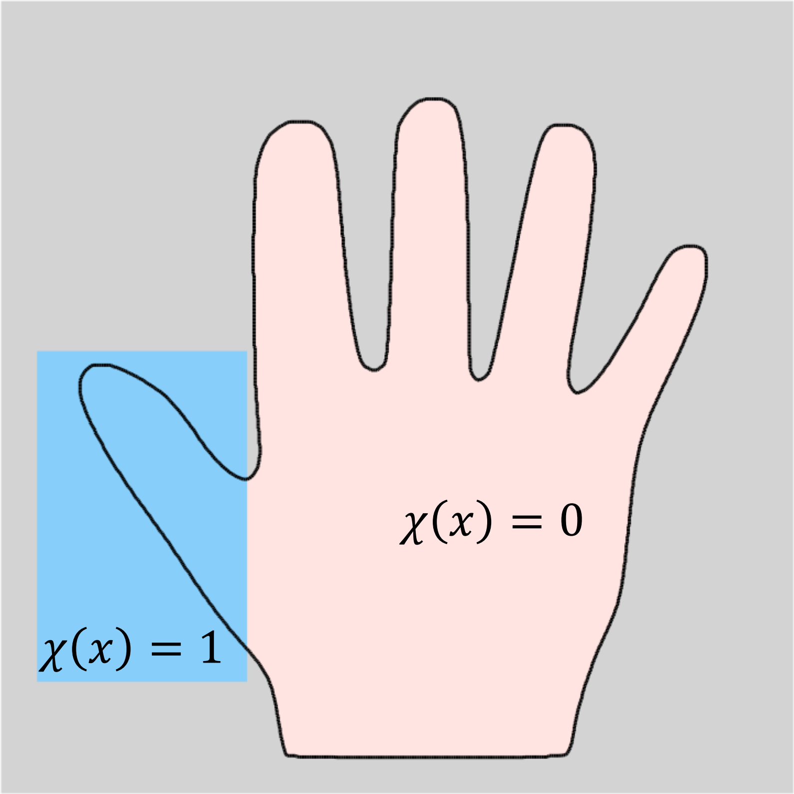


Spatially-varying priors

Use different models for different regions

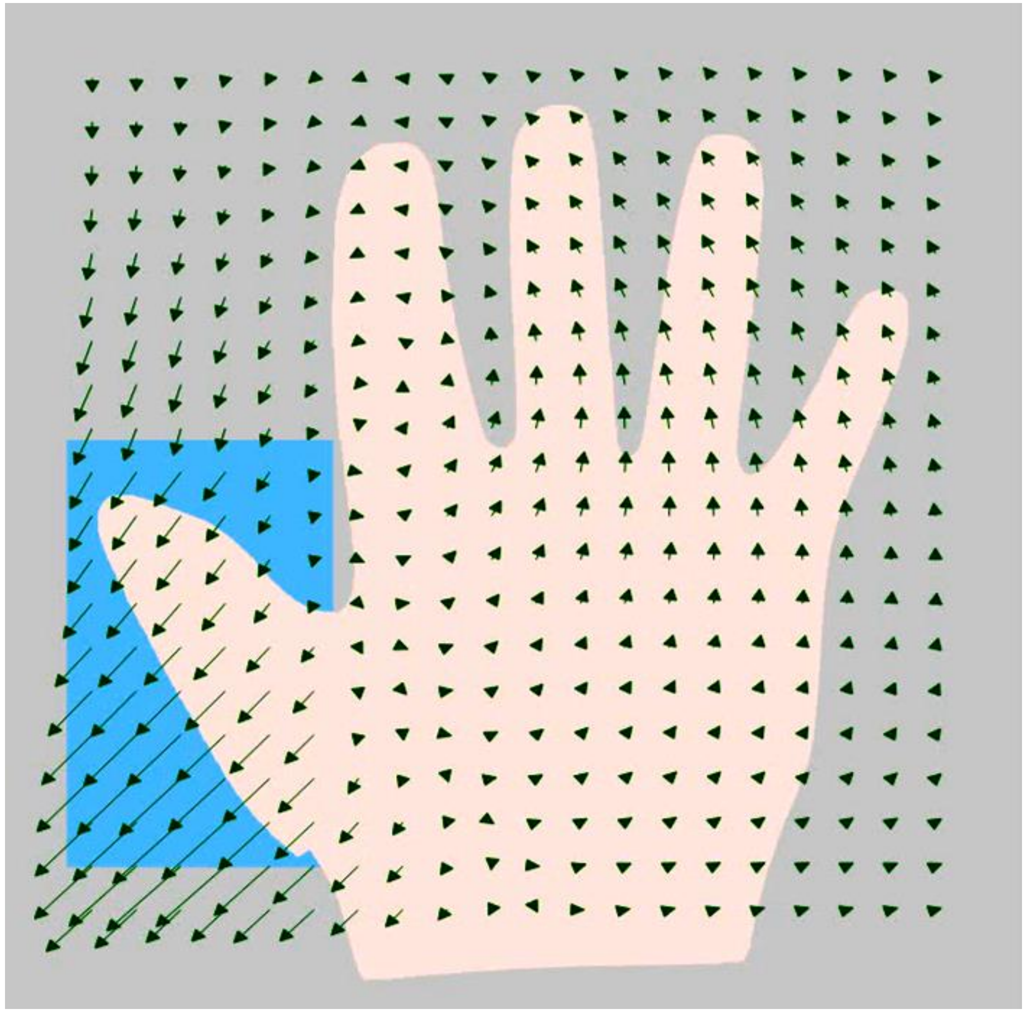
$$k(x,x') = \chi(x)\chi(x')k_1(x,x') + (1-\chi(x))(1-\chi(x'))\,k_2(x,x')$$

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



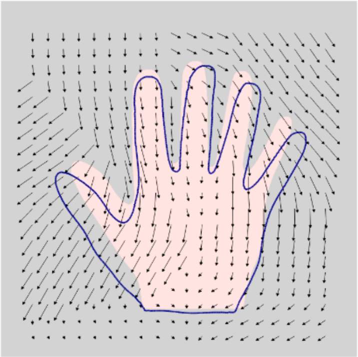
Freiman, Moti, Stephan D. Voss, and Simon K. Warfield. "Demons registration with local affine adaptive regularization: application to registration of abdominal structures." *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011.

Spatially-varying priors

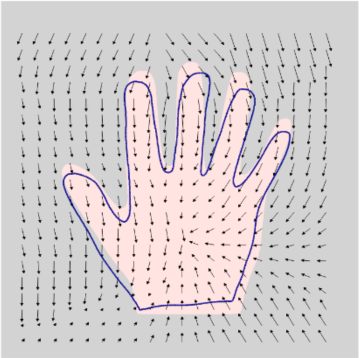


Statistical deformation models

Estimate mean and covariance function from data:

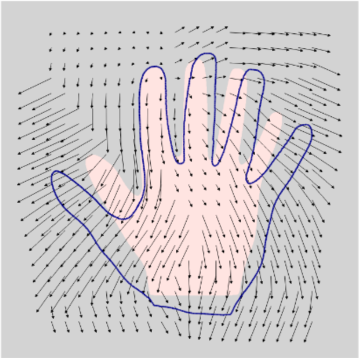


$u^1 : \Omega \rightarrow \mathbb{R}^2$



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

…

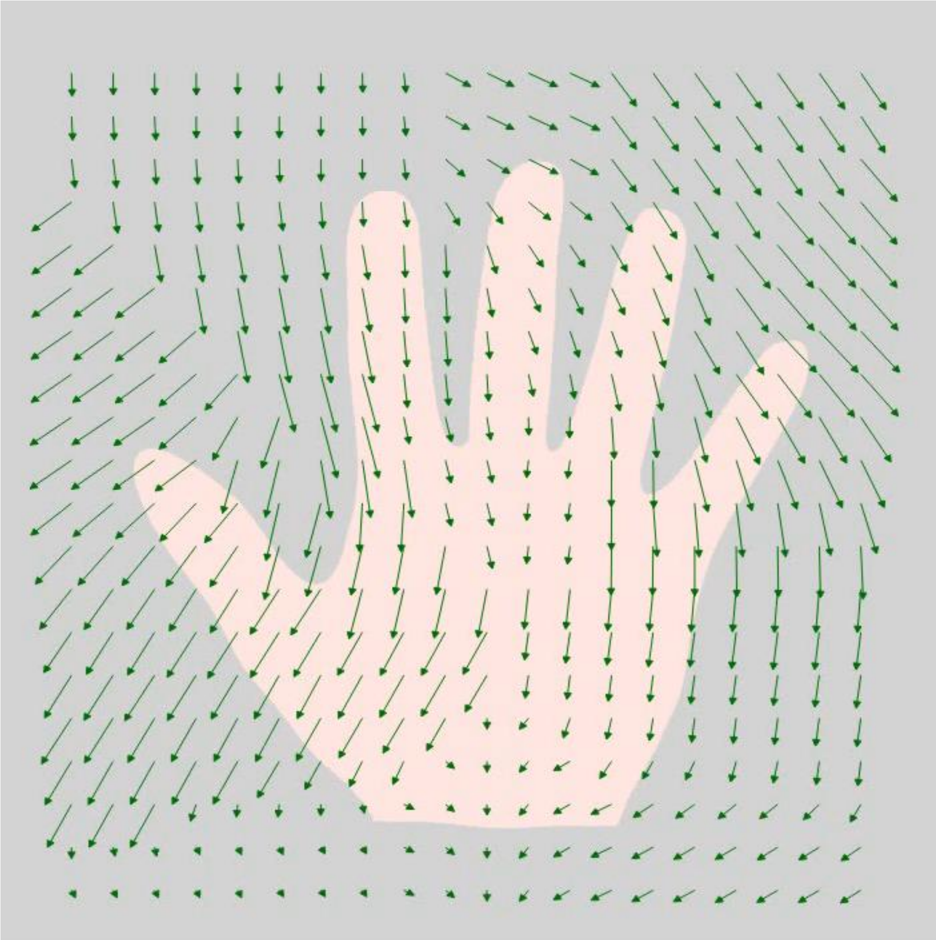


$u^n : \Omega \rightarrow \mathbb{R}^2$

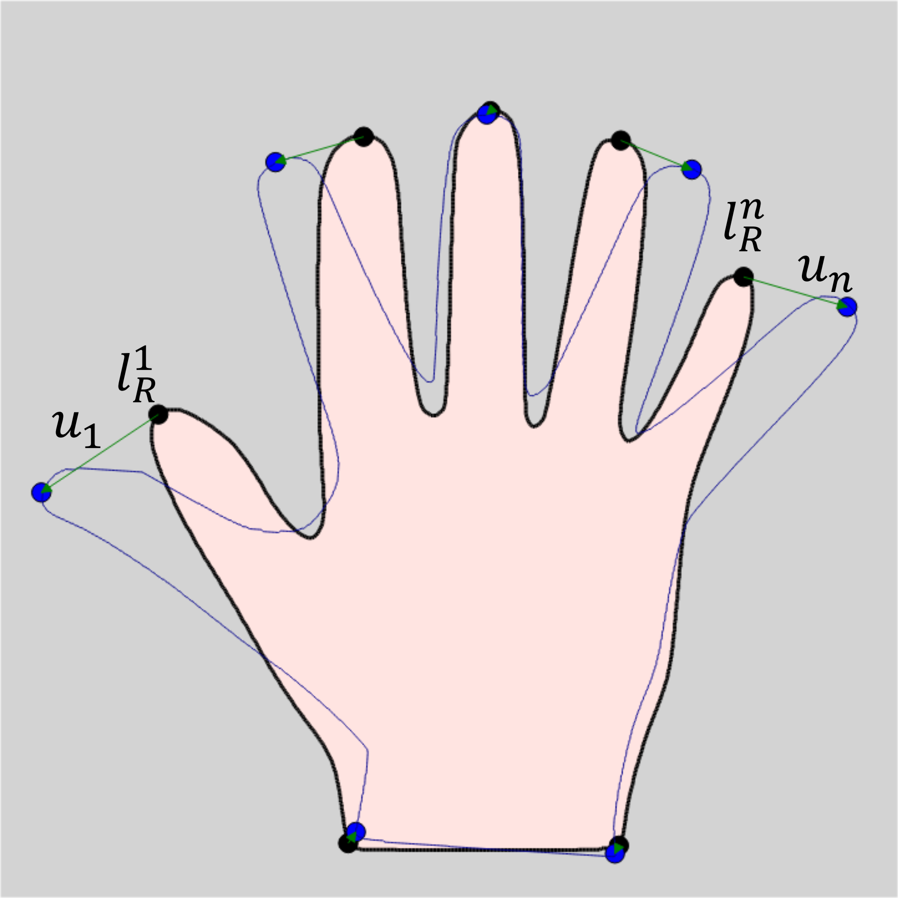
$$\mu(x) = \overline{u}(x) = \frac{1}{n} \sum_{i-1}^{n} u^i(x)$$

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

Example 5: Statistical deformation models



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:

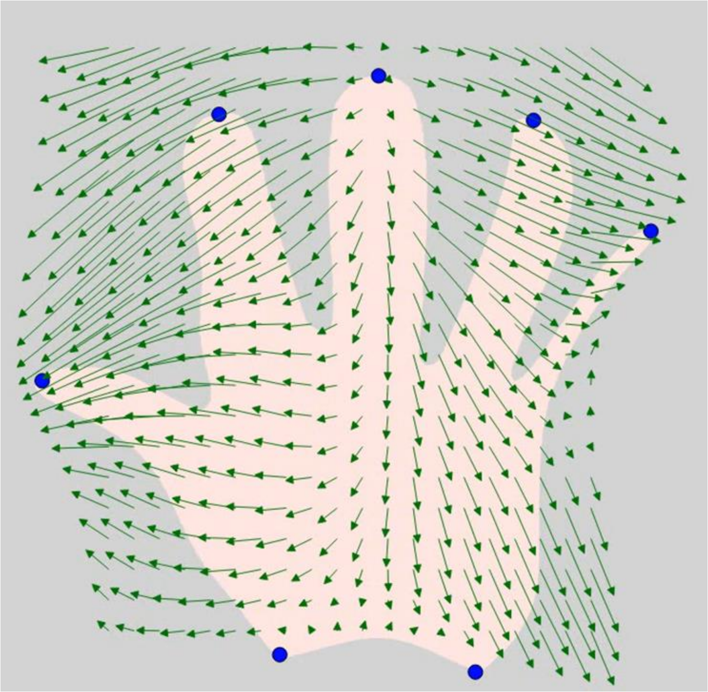
$$\tilde{u}_i = u(l_i) + \epsilon \text{ with } \epsilon \sim N(0, \sigma^2 I_{2\times 2}).$$

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

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

is a Gaussian process

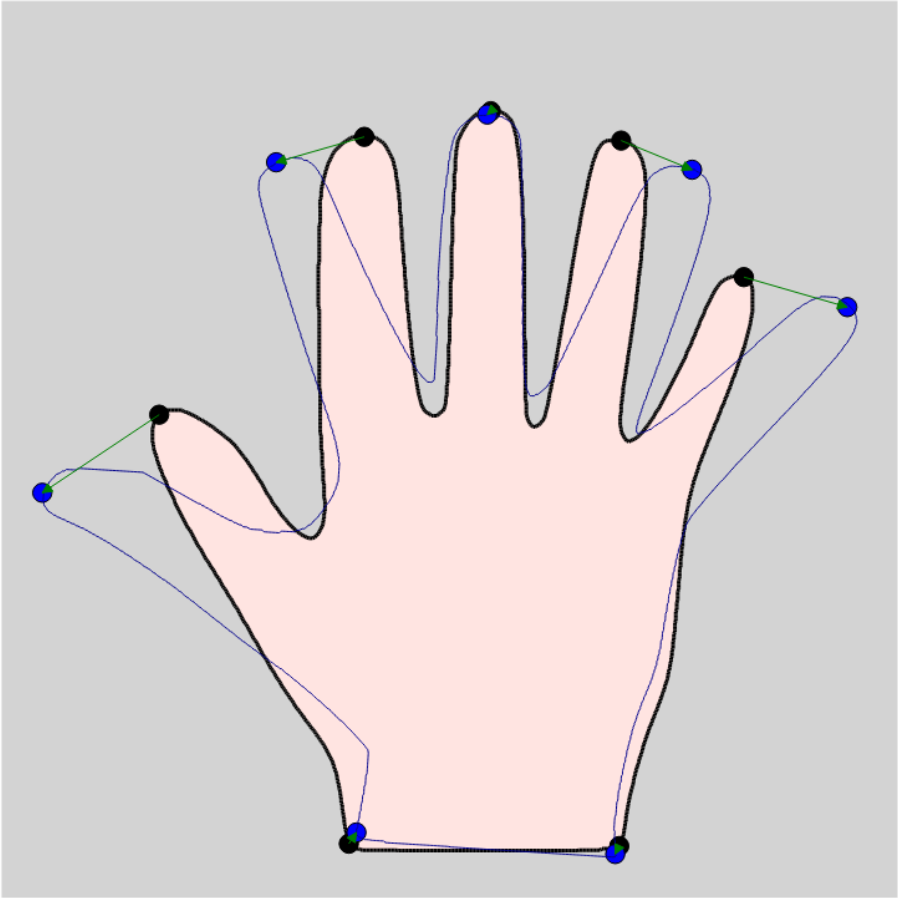
$$GP(\mu_p, k_p)$$

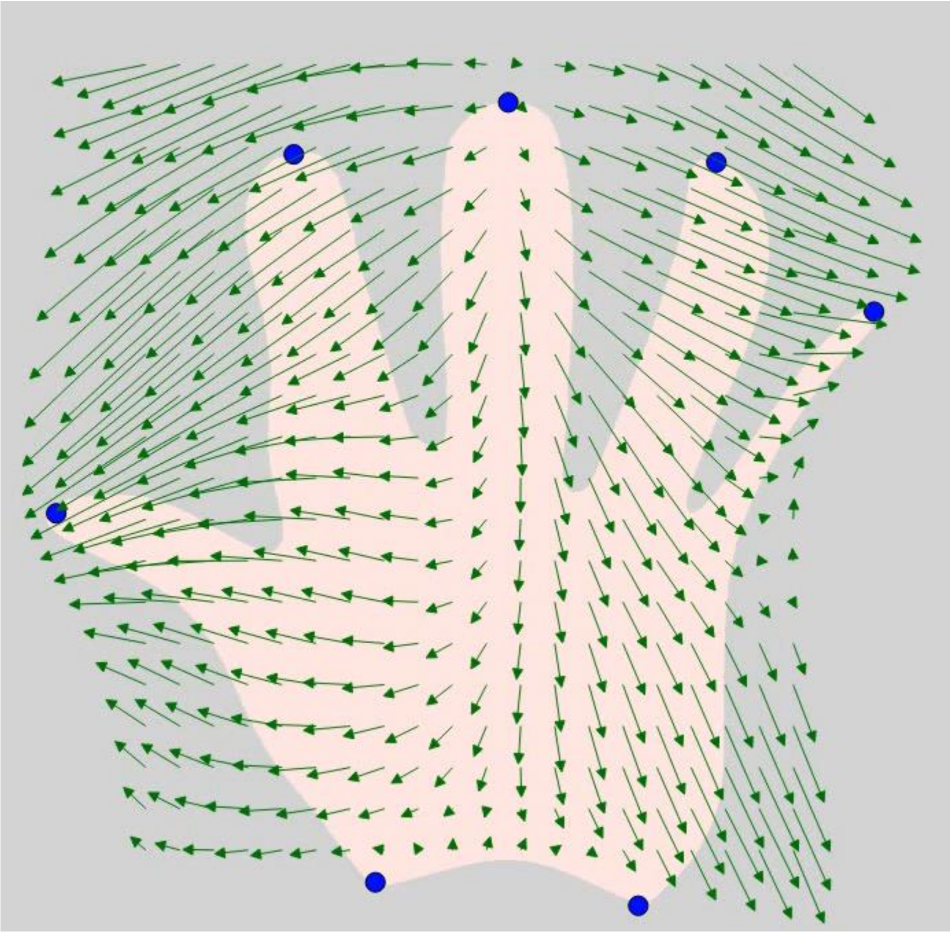
Its parameters are known analytically.

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

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

Landmark registration using GP Regression





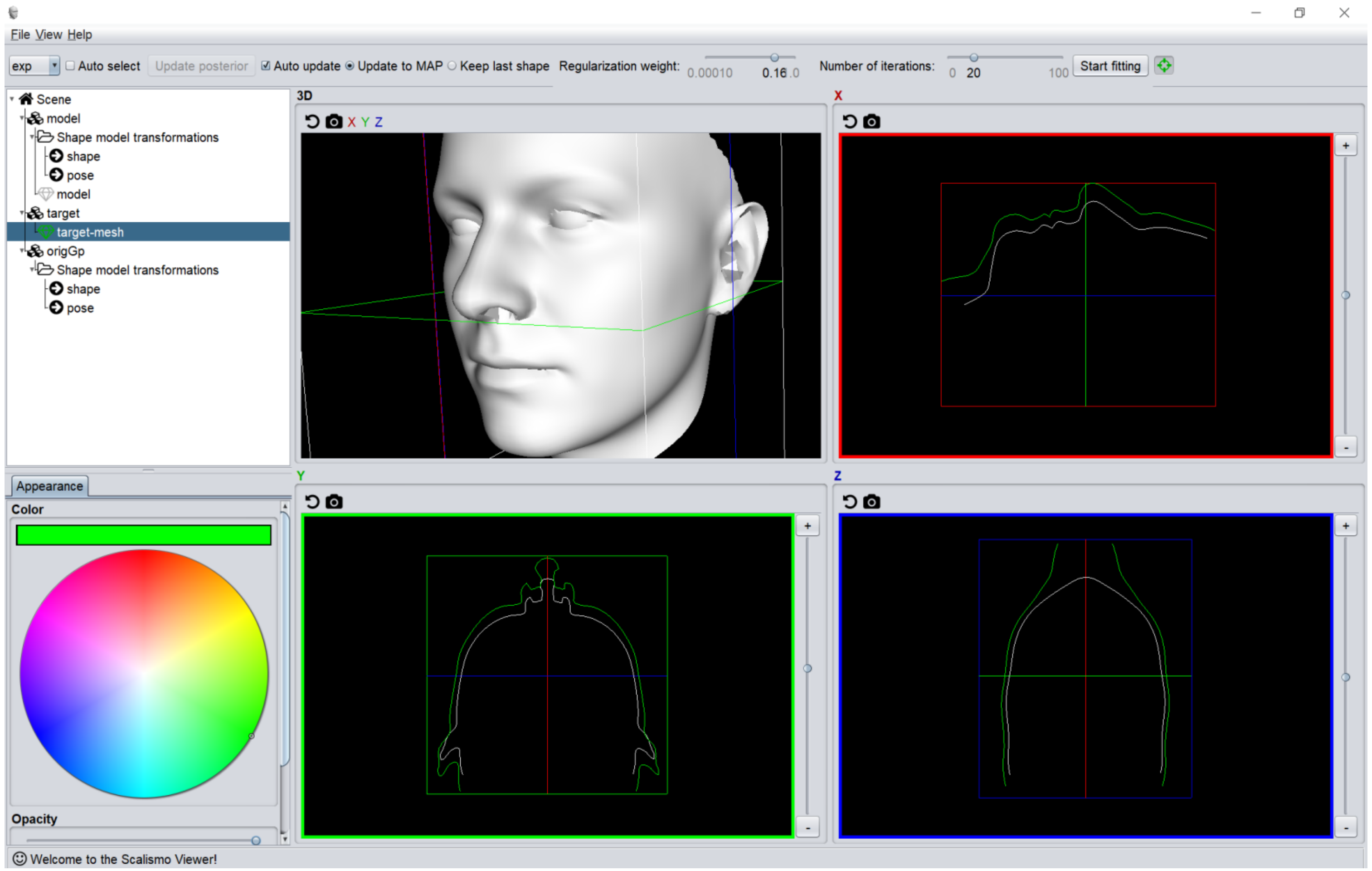
Hybrid registration

- We can now 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

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.

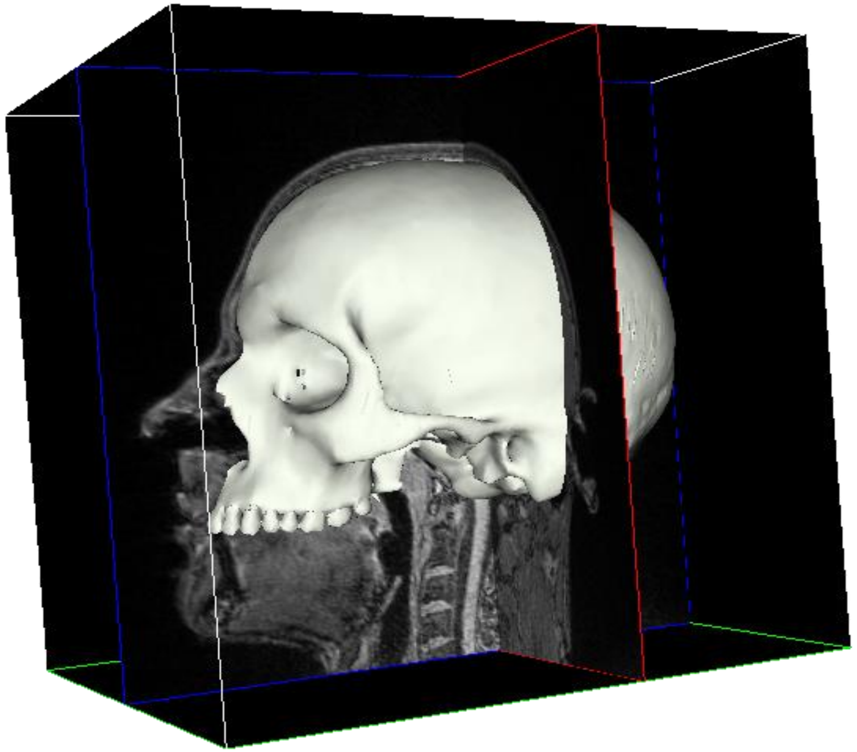
Lu, Huanxiang, Philippe C. Cattin, and Mauricio Reyes. "A hybrid multimodal non-rigid registration of MR images based on diffeomorphic demons." *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010.

Demo: Priors and interactive registration



Skull Segmentation in MRI

Lab-meeting

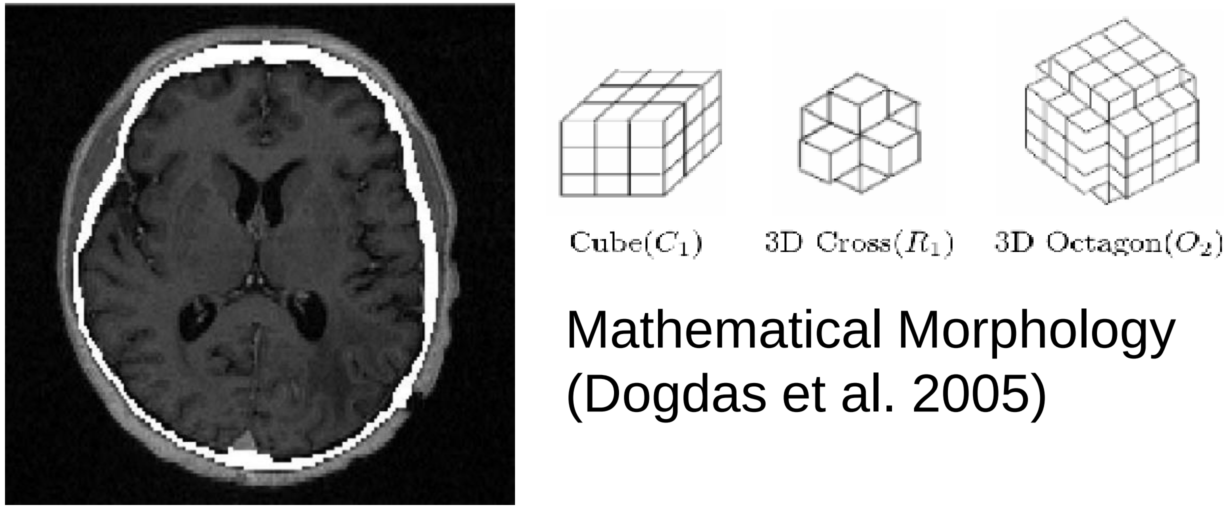


Slides by Patrick Kahr

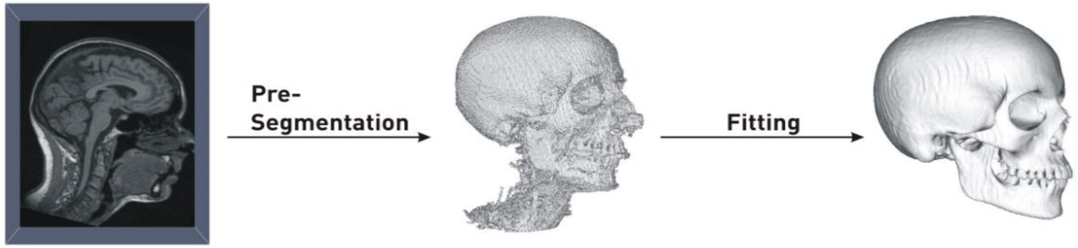
Skull Segmentation in MRI

Problem: Bone and air have similar intensities in MRI

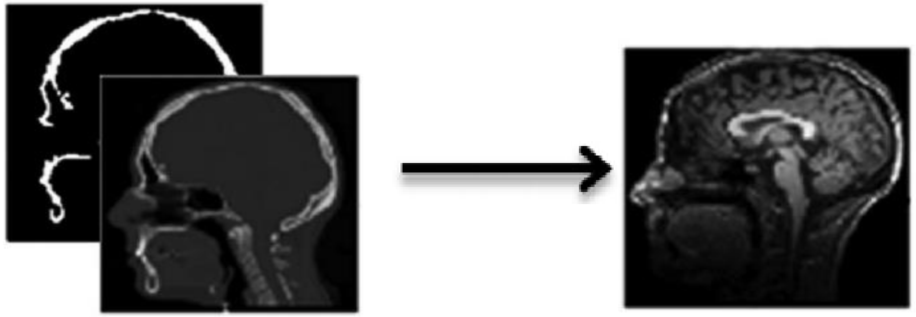
→ unlike CT, no threshold segmentation possible



Mathematical Morphology (Dogdas et al. 2005)

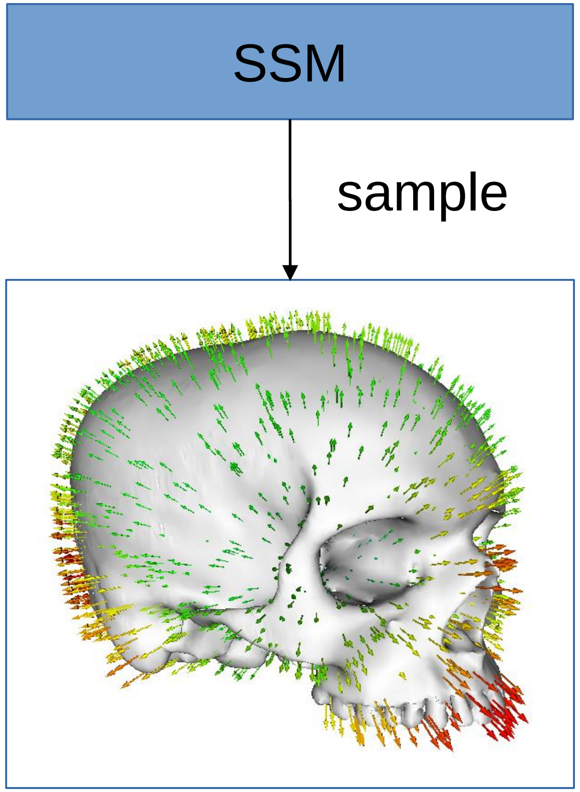


- Multi-level model fitting (Lerch, Lüthi 2008)



- Multi-Atlas matching (Torrado-Carvajal et al. 2015)

Modeling deformations with SSMs



deformation field defined on reference mesh

Model deformations of reference shape using an SSM

$$GP(\mu_{ssm}, k_{ssm})$$

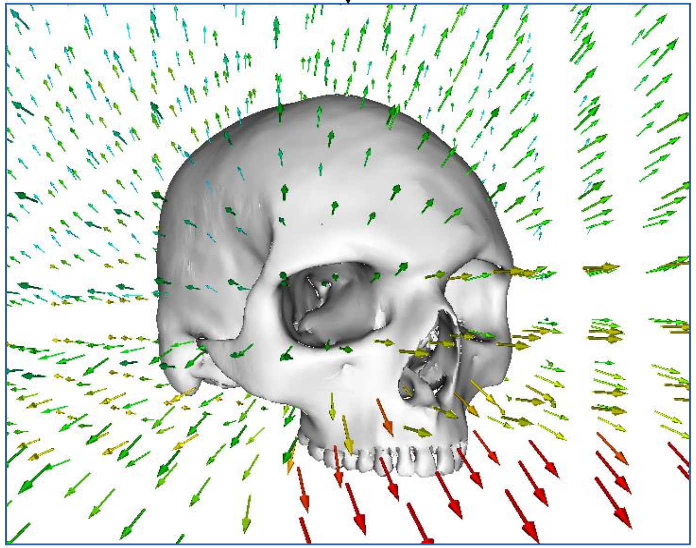
Deformations are only defined on surface of reference shape:

→ deformation field needs to be interpolated for the rest of the image.

Modeling deformations with SSMs

SSM + deformation model

sample



deformation field defined on complete image domain

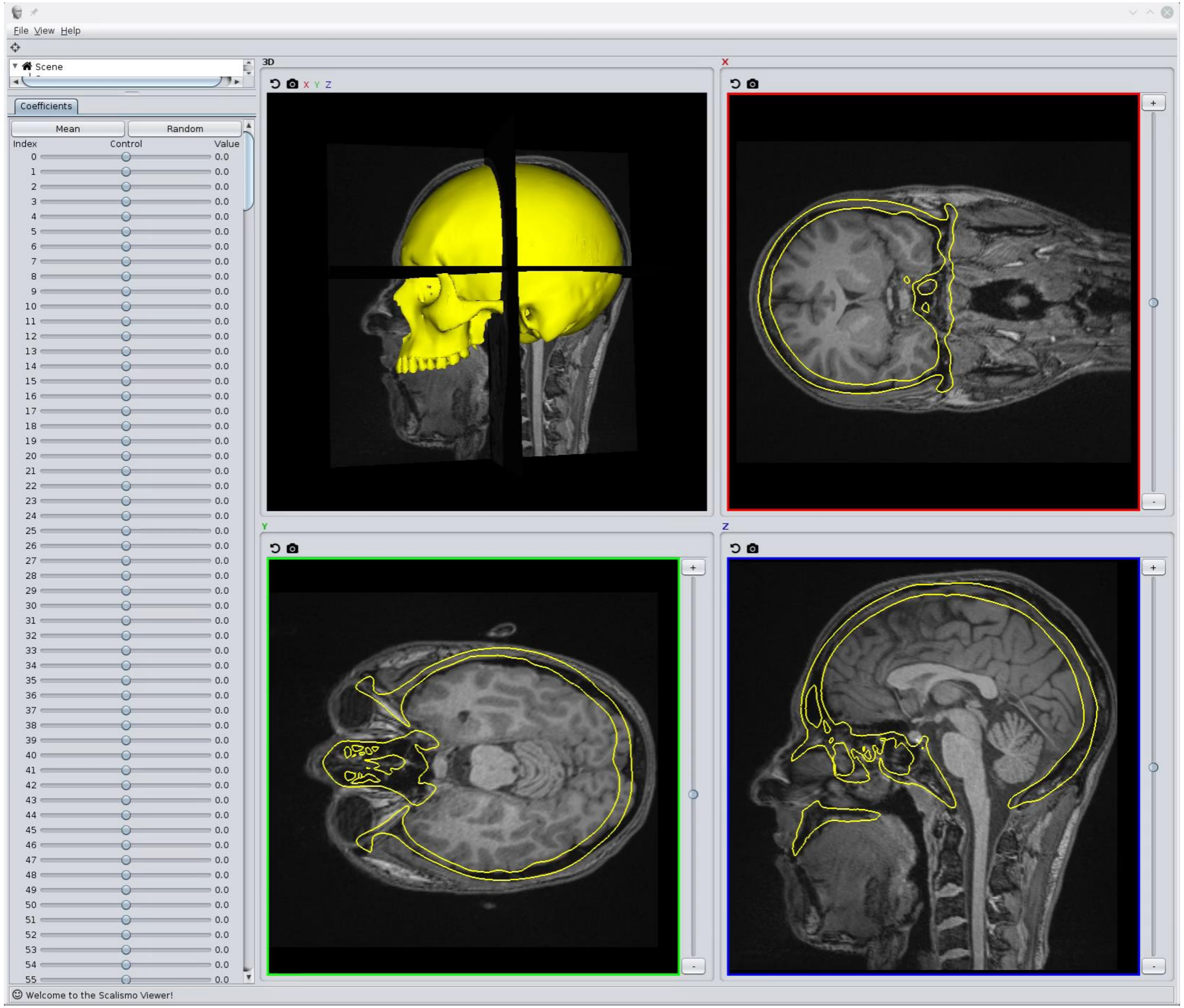
Hybrid kernel: mix SSM with smooth Gaussian kernel

$$(1 - w(x))k_{gaussian}(x, x')(1 - w(x')) + w(x)k_{SSM}(x, x')w(x')$$

where

x', y': closest points to x,y on the surface,

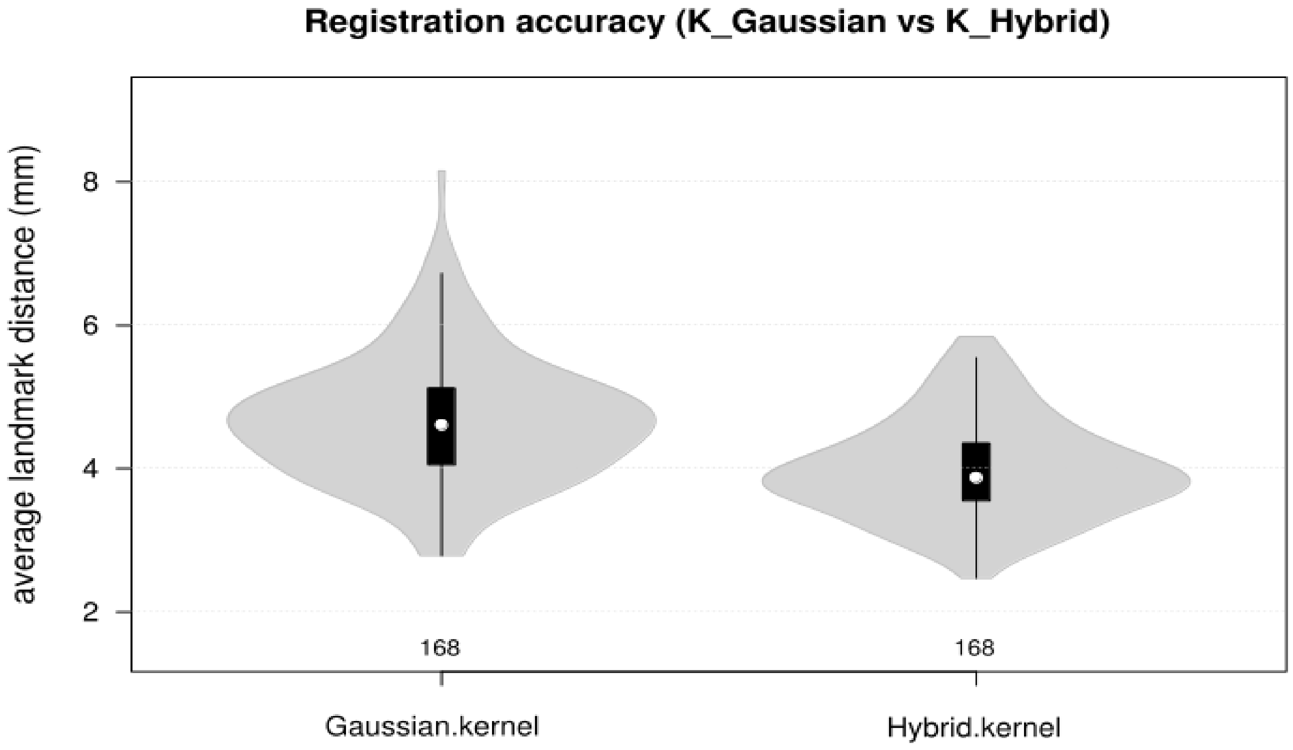
w = 1 if x, y on the surface,
w→0 for x, y far away from surface.

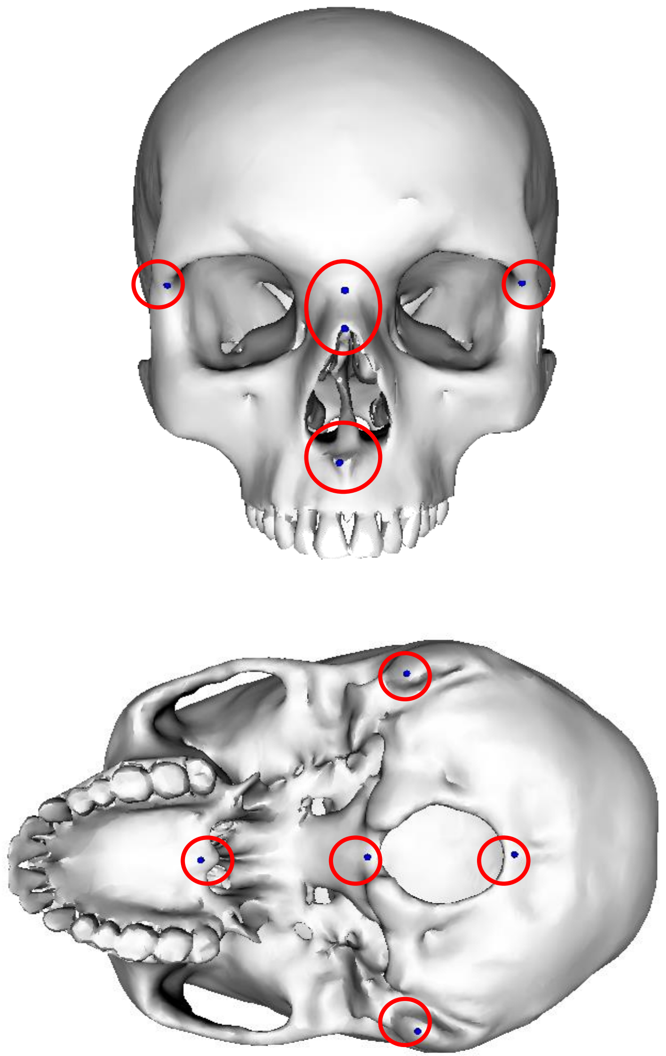

File View Help
Scene
Coefficients
Mean
Random
Index
Control
Value
3D
X
Y
Z
Welcome to the Scalismo Viewer!

Registration accuracy: KGaussian vs KHybrid

Register 14 CT templates to 12 MR targets (168 registrations).

Measure average distance to a set of 10 anatomical landmarks.

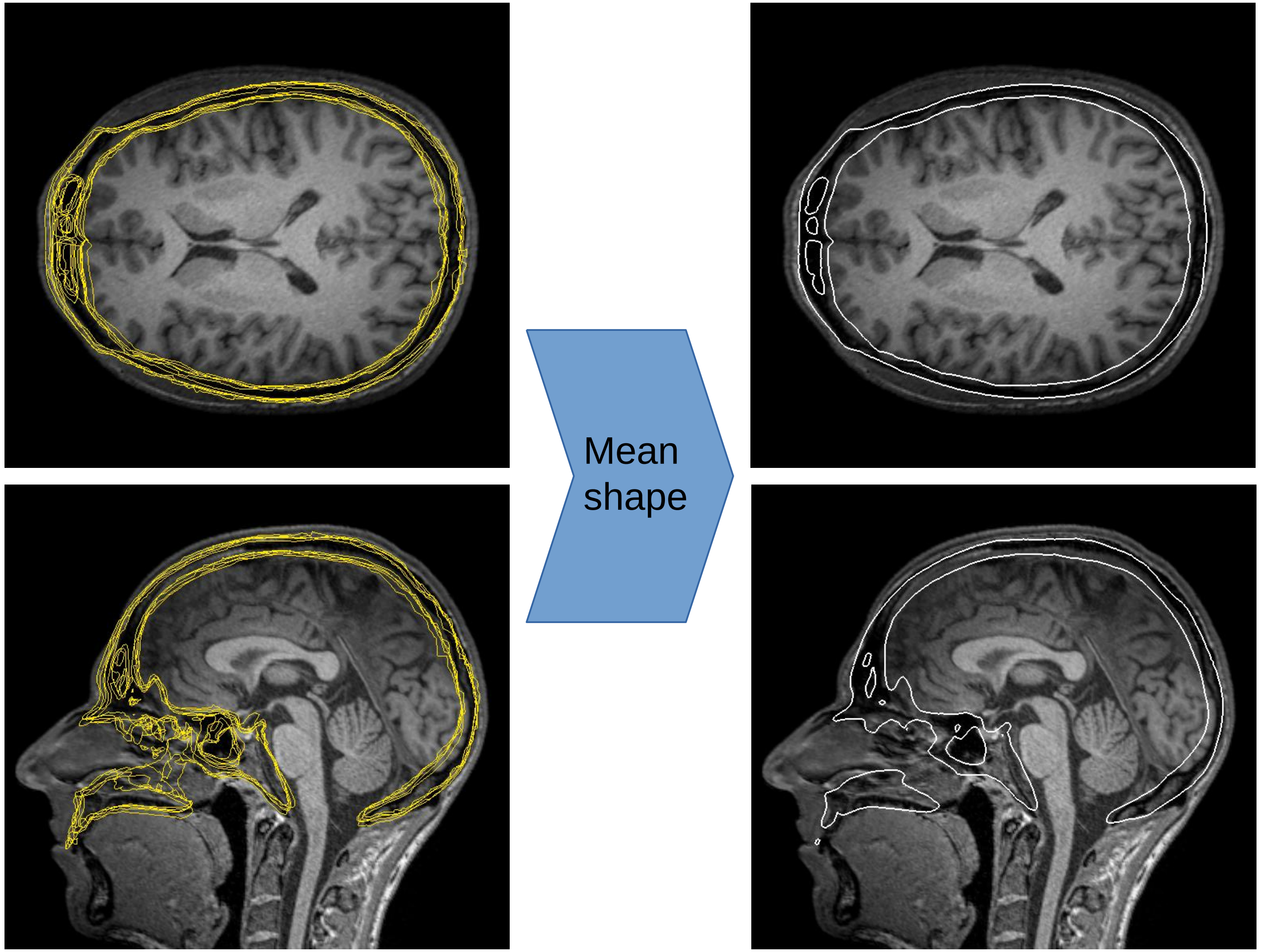




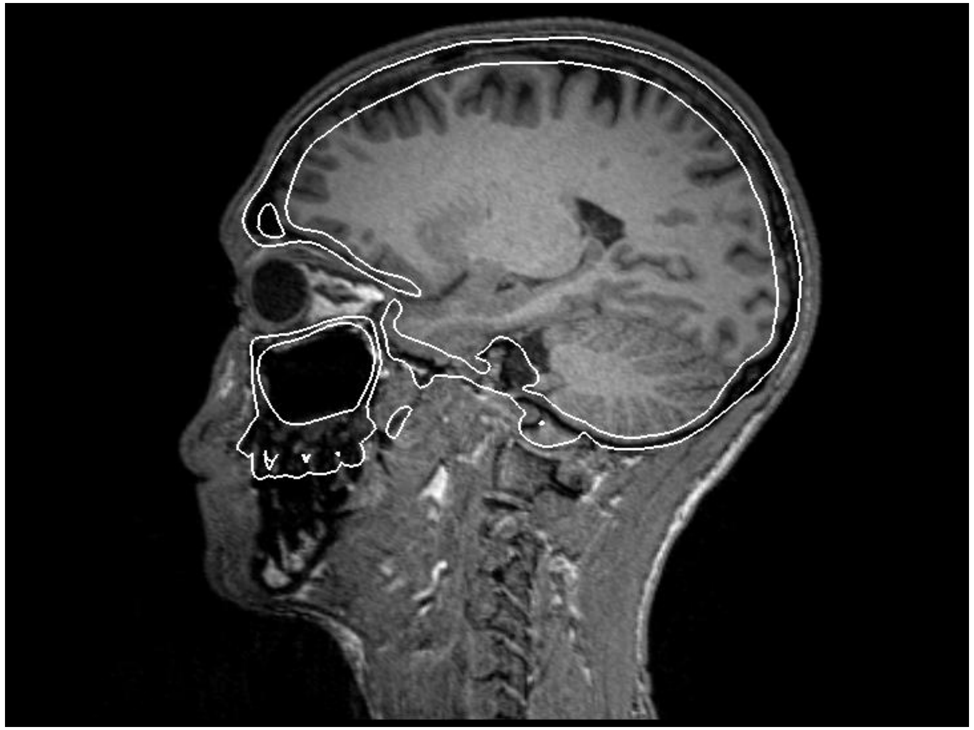
Segmentation accuracy

Multi-atlas matching:

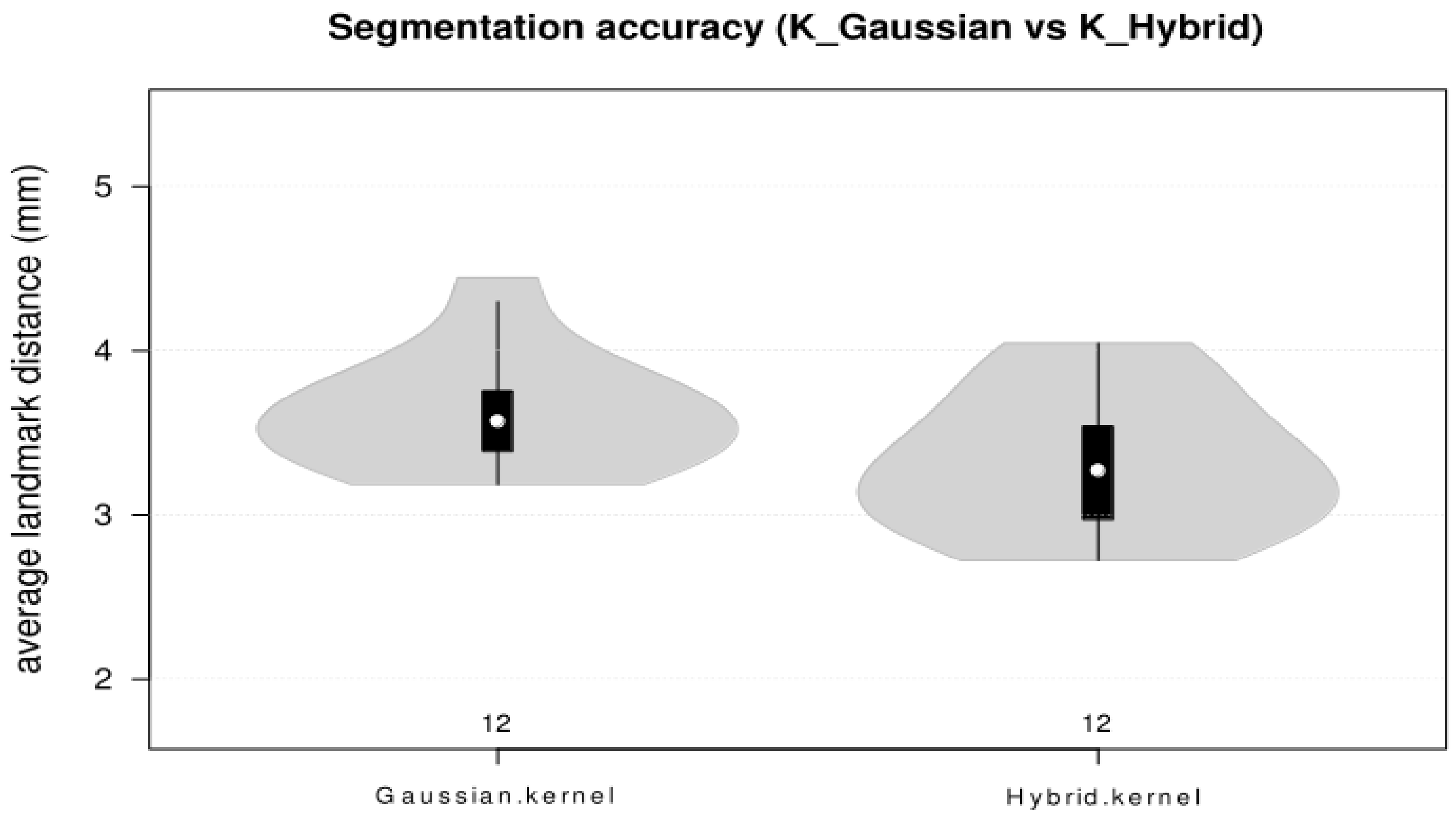
Target segmentation is mean shape obtained from 14 CT registrations.



Segmentation example

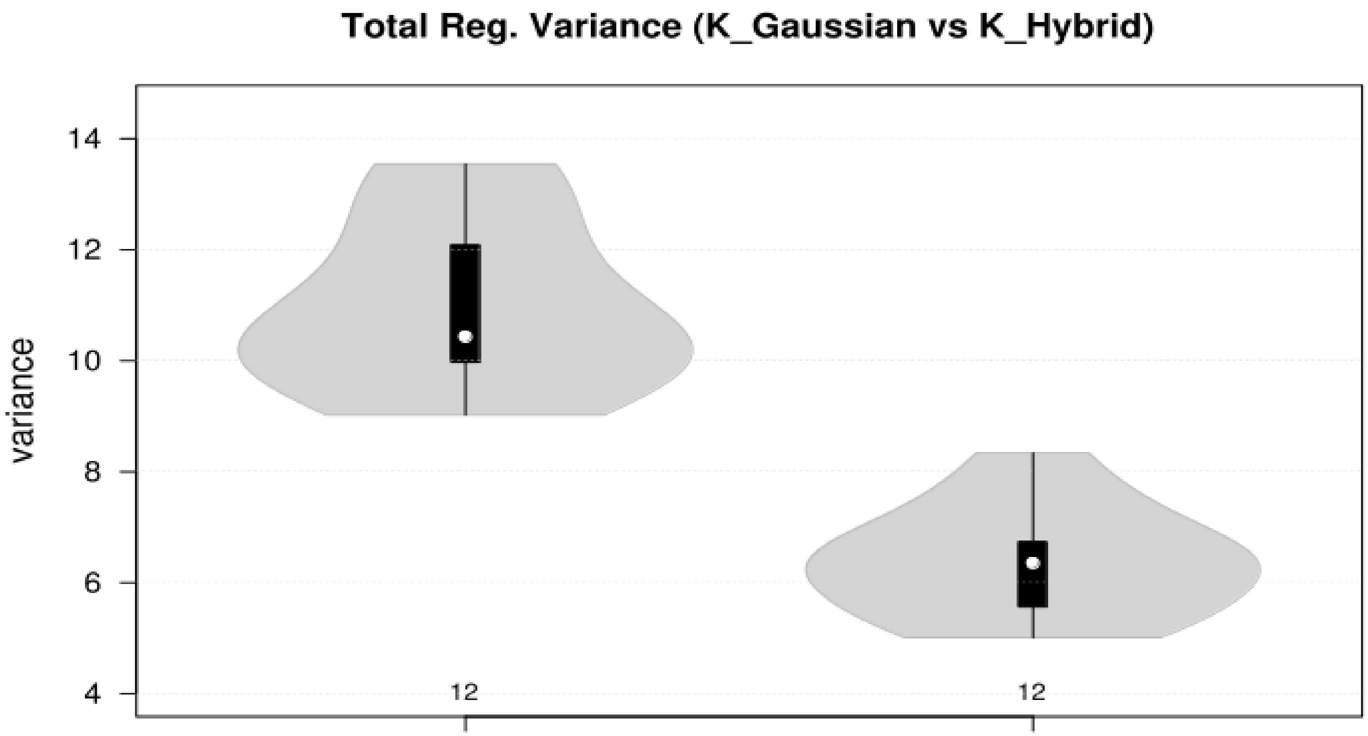


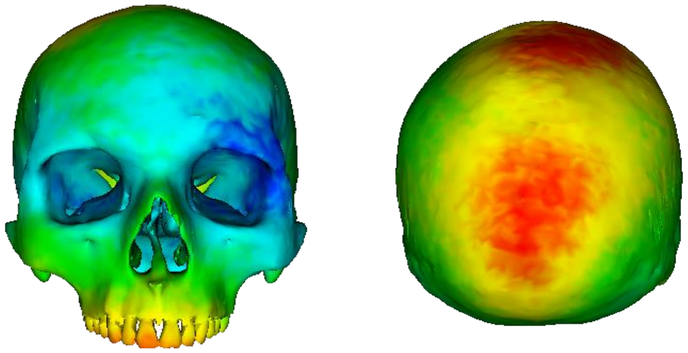
Segmentation accuracy



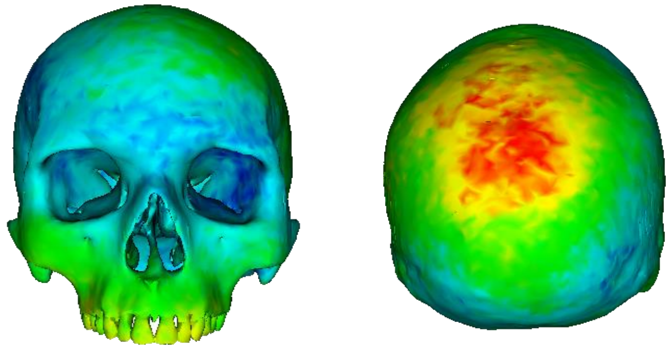
Variance

Lower variance between registrations for each of the 12 targets with K_{hybrid}.





$K_{Gaussian}$



K_{Hybrid}

Summary

- GPs provide probabilistic interpretation to classic registration models
  - But, can visualize assumptions
  - New ways to combine priors to individual applications.
- Modelling and model fitting are separated
  - Change in prior does not lead to change in algorithm
  - No increase in complexity
  - Can tailor model to application without increase in complexity
- Can use SSMs of individual organs to guide image registration