Statistical Warps

A Least Committed Model

Mads Nielsen[1] Peter Johansen[2] Andrew Jackson[3] Benny Lautrup[3]

[1]ITU [2]DIKU [3] NBI

Copenhagen, Denmark

Outline

- Motivation

Outline

- Motivation
- Bayesian Warps

Outline

- Motivation
- Bayesian Warps
- – Thin Plate Splines

Outline

- Motivation
- Bayesian Warps
- – Thin Plate Splines
- Brownian Euclidean Model

Outline

- Motivation
- Bayesian Warps
- – Thin Plate Splines
- Brownian Euclidean Model
- – Geometry Constrained Diffusion

Outline

- Motivation
- Bayesian Warps
- – Thin Plate Splines
- Brownian Euclidean Model
- – Geometry Constrained Diffusion
- Brownian Warps

Outline

- Motivation
- Bayesian Warps
- – Thin Plate Splines
- Brownian Euclidean Model
- – Geometry Constrained Diffusion
- Brownian Warps
- – Least Committed Distribution

Outline

- Motivation
- Bayesian Warps
- – Thin Plate Splines
- Brownian Euclidean Model
- – Geometry Constrained Diffusion
- Brownian Warps
- – Least Committed Distribution
- – – Invariances

Outline

- Motivation
- Bayesian Warps
- – Thin Plate Splines
- Brownian Euclidean Model
- – Geometry Constrained Diffusion
- Brownian Warps
- – Least Committed Distribution
- – – Invariances
- – – Implementation

Outline

- Motivation
- Bayesian Warps
- – Thin Plate Splines
- Brownian Euclidean Model
- – Geometry Constrained Diffusion
- Brownian Warps
- – Least Committed Distribution
- – – Invariances
- – – Implementation
- – – Examples

Outline

- Motivation
- Bayesian Warps
- – Thin Plate Splines
- Brownian Euclidean Model
- – Geometry Constrained Diffusion
- Brownian Warps
- – Least Committed Distribution
- – – Invariances
- – – Implementation
- – – Examples
- Wrap up

Perceptual Signals

Theory of shape (Grenander 1976)

- Noise and Blur $I \mapsto I * h + \eta$

Solutions:

- Noise and Blur: Tikhonov regularisation

Perceptual Signals

Theory of shape (Grenander 1976)

- Noise and Blur $I \mapsto I * h + \eta$
- Superposition $I_1, I_2 \mapsto I_1 + I_2$

Solutions:

- Noise and Blur: Tikhonov regularisation
- Superposition: Unmixing, ICA

Perceptual Signals

Theory of shape (Grenander 1976)

- Noise and Blur $I \mapsto I * h + \eta$
- Superposition $I_1, I_2 \mapsto I_1 + I_2$
- Interruptions $I_1(x), I_2(x) \mapsto I_1(x)$ or $I_2(x)$

Solutions:

- Noise and Blur: Tikhonov regularisation
- Superposition: Unmixing, ICA
- Interruptions: Mumford-Shah ....

Perceptual Signals

Theory of shape (Grenander 1976)

- Noise and Blur $I \mapsto I * h + \eta$
- Superposition $I_1, I_2 \mapsto I_1 + I_2$
- Interruptions $I_1(x), I_2(x) \mapsto I_1(x)$ or $I_2(x)$
- Domain warping $I(x) \mapsto I(w(x))$

Solutions:

- Noise and Blur: Tikhonov regularisation
- Superposition: Unmixing, ICA
- Interruptions: Mumford-Shah ....
- Domain warping: ?????? Mumford 1996: The best approach is unclear

Perceptual Signals

Theory of shape (Grenander 1976)

- Noise and Blur $I \mapsto I * h + \eta$
- Superposition $I_1, I_2 \mapsto I_1 + I_2$
- Interruptions $I_1(x), I_2(x) \mapsto I_1(x)$ or $I_2(x)$
- Domain warping $I(x) \mapsto I(w(x))$

Solutions:

- Noise and Blur: Tikhonov regularisation
- Superposition: Unmixing, ICA
- Interruptions: Mumford-Shah ....
- Domain warping: ?????? Mumford 1996: The best approach is unclear

Perceptual Signals

Theory of shape (Grenander 1976)

- Noise and Blur $I \mapsto I * h + \eta$
- Superposition $I_1, I_2 \mapsto I_1 + I_2$
- Interruptions $I_1(x), I_2(x) \mapsto I_1(x)$ or $I_2(x)$
- Domain warping $I(x) \mapsto I(w(x))$

Solutions:

- Noise and Blur: Tikhonov regularisation
- Superposition: Unmixing, ICA
- Interruptions: Mumford-Shah ....
- Domain warping: ?????? Mumford 1996: The best approach is unclear

Perceptual Signals

Theory of shape (Grenander 1976)

- Noise and Blur $I \mapsto I * h + \eta$
- Superposition $I_1, I_2 \mapsto I_1 + I_2$
- Interruptions $I_1(x), I_2(x) \mapsto I_1(x)$ or $I_2(x)$
- Domain warping $I(x) \mapsto I(w(x))$

Solutions:

- Noise and Blur: Tikhonov regularisation
- Superposition: Unmixing, ICA
- Interruptions: Mumford-Shah ....
- Domain warping: ?????? Mumford 1996: The best approach is unclear

Perceptual Signals

Theory of shape (Grenander 1976)

- Noise and Blur $I \mapsto I * h + \eta$
- Superposition $I_1, I_2 \mapsto I_1 + I_2$
- Interruptions $I_1(x), I_2(x) \mapsto I_1(x)$ or $I_2(x)$
- Domain warping $I(x) \mapsto I(w(x))$

Solutions:

- Noise and Blur: Tikhonov regularisation
- Superposition: Unmixing, ICA
- Interruptions: Mumford-Shah ....
- Domain warping: ?????? Mumford 1996: The best approach is unclear

Warps

A warp is a geometrical transformation:

$$W : \mathbb{R}^D \mapsto \mathbb{R}^D$$

If the warp W is continuous and invertible it's a *Homeomorphism*

If the warp W is also C^∞ it's a *Diffeomorphism*

Our choice: a diffeomorphism of Jacoby determinant

$$\det(\partial x_i W^j) > 0$$

Bayesian Warps

Given some (stochastic) constraints C on the warp W we seek

$$p(W|C) = \frac{1}{Z} p(C|W) p(W)$$

$p(C|W)$ is a matching term:
How well does the warped image match the goal?

$p(W)$ is the prior on warps:
What is the likelihood of a given warp?

Motivation

Earlier work on $p(W)$ does not seem adequate for all situations:

- Not derived from first principles

Motivation

Earlier work on $p(W)$ does not seem adequate for all situations:

- Not derived from first principles
- Not invariant with respect to Euclidean coordinate system

Motivation

Earlier work on $p(W)$ does not seem adequate for all situations:

- Not derived from first principles
- Not invariant with respect to Euclidean coordinate system
- $p(W) \neq p(W^{-1})$

Motivation

Earlier work on $p(W)$ does not seem adequate for all situations:

- Not derived from first principles
- Not invariant with respect to Euclidean coordinate system
- $p(W) \neq p(W^{-1})$
- $p(W = W_2 \circ W_1) \neq \int p(W_2) dW_1$

Motivation

Earlier work on $p(W)$ does not seem adequate for all situations:

- Not derived from first principles
- Not invariant with respect to Euclidean coordinate system
- $p(W) \neq p(W^{-1})$
- $p(W = W_2 \circ W_1) \neq \int p(W_2) dW_1$
- W not invertible $\not\Rightarrow$ $p(W) = 0$

Euler-Lagrange formulation

We define the energy functional

$$E(W|C) = -\log p(W|C) = E(C|W) + E(W)$$

For Markov models of p(W), we find

$$E(W) = \int_\Omega F(W(x))dx$$

and the gradient descend algorithm

$$\partial_t W = -\frac{\delta E(W)}{\delta W} + \lambda \text{ constraints}$$

Glasbey and Mardia

[Journal of Royal Stat. Soc. B (2001), vol 63, part 3]
Define a *base distortion criterion*

$$\begin{aligned} F_B(W) &\geq 0 \\ F_B(I) &= 0 \end{aligned}$$

and the *null set distortion criterion*

$$F(W) = \min_{g \in C} F_B(W \circ g)$$

where C is a group of invariance.

1st order model

Glasbey and Mardia use

$$F_B = \sum_i \sum_j \left(\frac{\partial W^i}{\partial x_j} \right)^2$$

This may be viewed as a Brownian motion model in Euclidean coordinates, and

$$\partial_t W^i = \triangle W^i$$

This is a simple diffusion as in [Andresen and Nielsen, 1999]

1st order model, problems

The basic problems with this model are

- Derived from first principles
- Invariant with respect to Euclidean coordinate system,
- $p(W) \neq p(W^{-1})$
- $p(W = W_2 \circ W_1) \neq \int p(W_2) dW_1$
- W not invertible $\not\Rightarrow \ p(W) = 0$

2nd Order Model

Glasbey and Mardia use

$$F_B = \sum_i \sum_j \sum_k \left(\frac{\partial^2 W^i}{\partial x_j \partial x_k} \right)^2$$

This is the bending energy of a thin plate [Bookstein 91].

Advantage:
Very fast implementation
Affine invariance

2nd order model, problems

The basic problems with this model are

- Derived from physical analogy
- Invariant with respect to Euclidean coordinate system,
- $p(W) \neq p(W^{-1})$
- $p(W = W_2 \circ W_1) \neq \int p(W_2) dW_1$
- W not invertible $\not\Rightarrow \; p(W) = 0$

Constrained Diffusion

Glasbey and Mardia's 1st order model leads to

$$\partial_t W^i = \triangle W^i$$

Suppose constraint are given as

$$W(S_1 : \mathbb{R}^2 \mapsto \mathbb{R}^3) \quad = \quad S_2 : \mathbb{R}^2 \mapsto \mathbb{R}^3$$

Then we may construct

$$\partial_t W^i = \triangle W^i - (\triangle W^i \perp S_2) \quad \text{if} \quad x \in S_1$$

Growing Mandible

- The mandibular bone is identified as an iso-intensity surface in 3D CT-scans.

Growing Mandible

- The mandibular bone is identified as an iso-intensity surface in 3D CT-scans.
- 5 CT scans are given pr. patient: 2, 9, 22, 48, 84 months

Growing Mandible

- The mandibular bone is identified as an iso-intensity surface in 3D CT-scans.
- 5 CT scans are given pr. patient: 2, 9, 22, 48, 84 months
- Initial registration is performed as similarity transform to minimize distance.

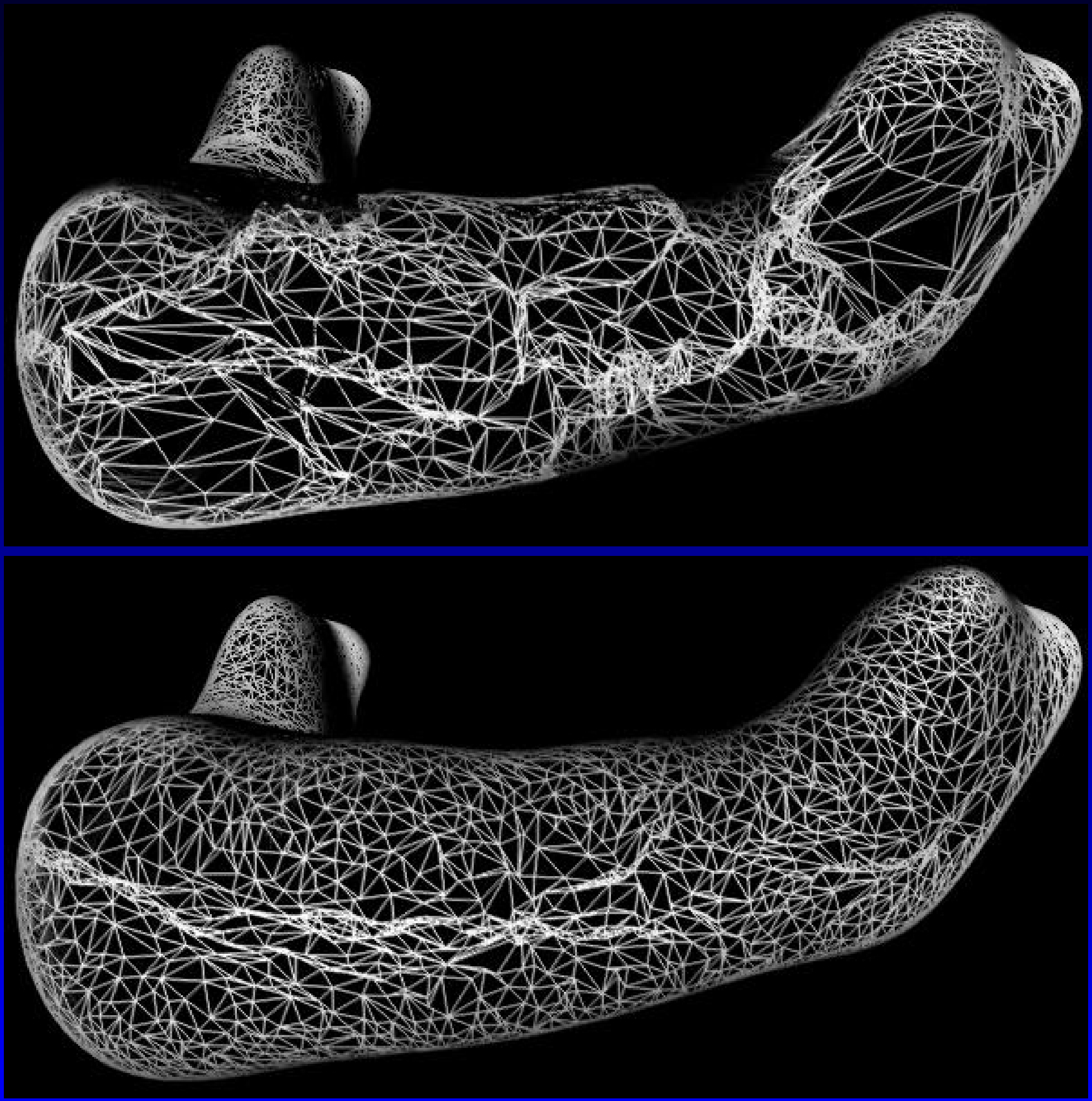
Growing Mandible

- The mandibular bone is identified as an iso-intensity surface in 3D CT-scans.
- 5 CT scans are given pr. patient: 2, 9, 22, 48, 84 months
- Initial registration is performed as similarity transform to minimize distance.
- Closest point projection to make $S_1 \mapsto S_2$.

Growing Mandible

- The mandibular bone is identified as an iso-intensity surface in 3D CT-scans.
- 5 CT scans are given pr. patient: 2, 9, 22, 48, 84 months
- Initial registration is performed as similarity transform to minimize distance.
- Closest point projection to make $S_1 \mapsto S_2$.
- Geometry constrained diffusion to make W as smooth as possible.

Growing Mandible



Infinitessimal I.I.D. Warps

Just like the brownian motion model arrises as the sum of infinitessimal statistically independent motion:

$$X = \lim_{N \to \infty} \sum_{i=0}^{N} \frac{X_i}{\sqrt{N}} \quad \Rightarrow \quad p(X) = \mathrm{Gauss}(X)$$

We construct a warp as

$$W_B = \lim_{N \to \infty} \prod_{i=0}^{N} \circ W_i$$

where W_i are infinitessimal independent warps.

1st order structure

Let

$$J_W = \partial_{x_i} W^j$$

be the Jacobian of W.

Then

$$J_{W_B} = \lim_{N \to \infty} \prod_{i=0}^{N} J_{W_i}$$

We may model

$$J_{W_i} = I + \sigma \frac{1}{\sqrt{N}} H_i$$

Limiting Distribution

The limiting distribution of

$$J_{W_B} = \lim_{N\to\infty} \prod_{i=0}^{N} I + \sigma \frac{1}{\sqrt{N}} H_i$$

when H_i are independent and $W : \mathbb{R}^2 \mapsto \mathbb{R}^2$, is given as

$$p(J_{W_B}) = \text{Gauss}(S) \sum_{n=0}^{\infty} g_n(F) \cos(n\theta)$$

[Jackson, Lautrup, Johansen, Nielsen 2001]

Parameters

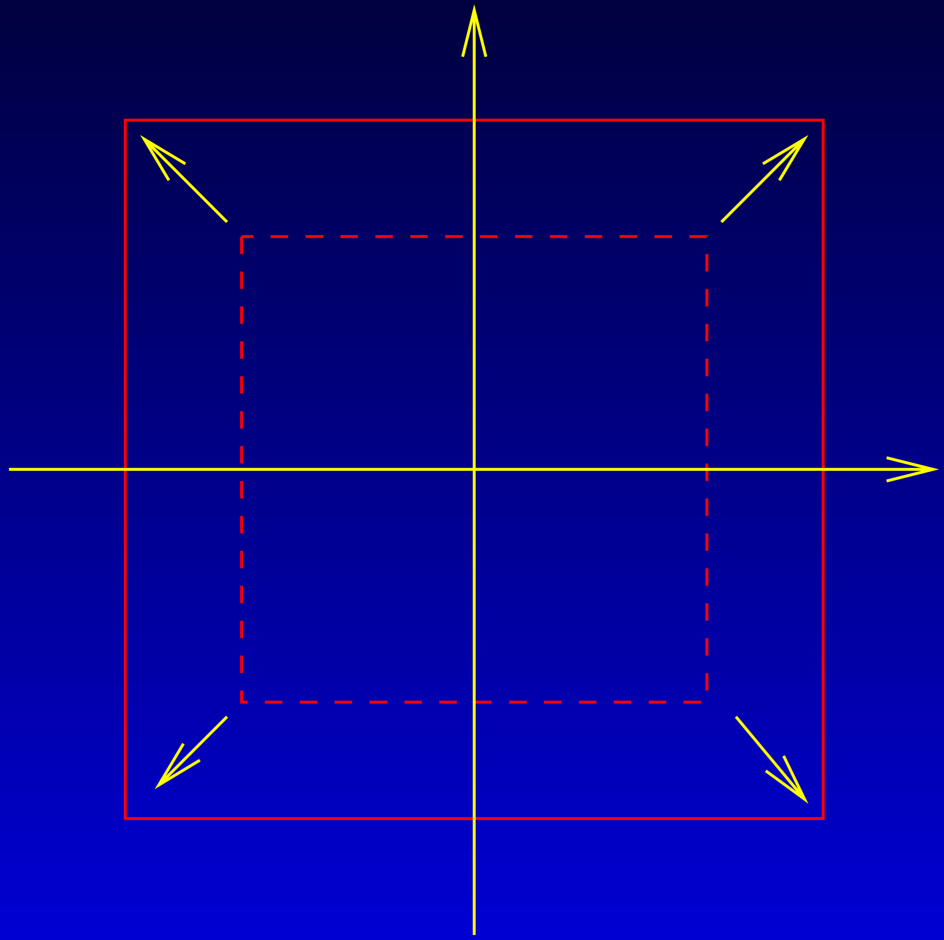
$$p(J_{W_B}) = \text{Gauss}(S) \sum_{n=0}^{\infty} g_n(F) \cos(n\theta)$$

where

$$\begin{aligned}
\text{Scaling} \quad S &= \log(\det(J_{W_B})) \\
\text{Skewness} \quad F &= \frac{1}{2\det(J_{W_B})} \| J_{W_B} \|_2^2 \\
\text{Rotation} \quad \theta &= \arctan\left(\frac{j_{12} - j_{21}}{j_{11} + j_{22}}\right)
\end{aligned}$$

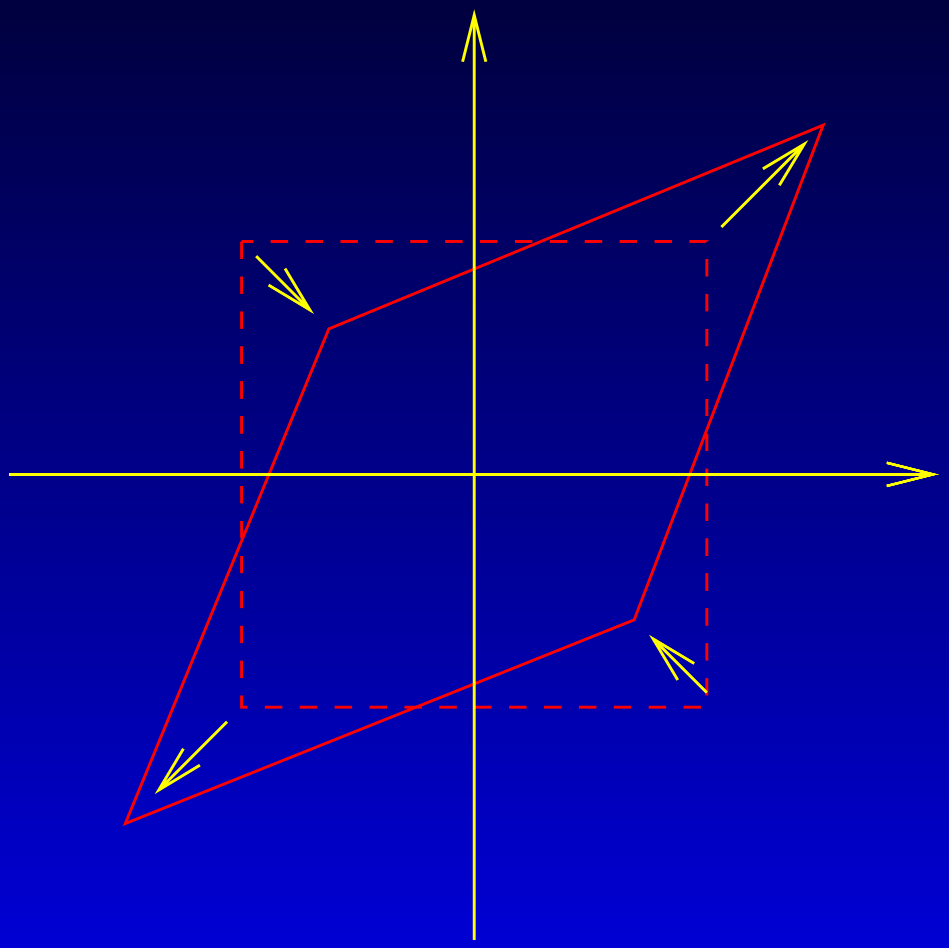
Scaling

$S \approx 0.8, \;\; F = 1, \;\; \theta = 0$



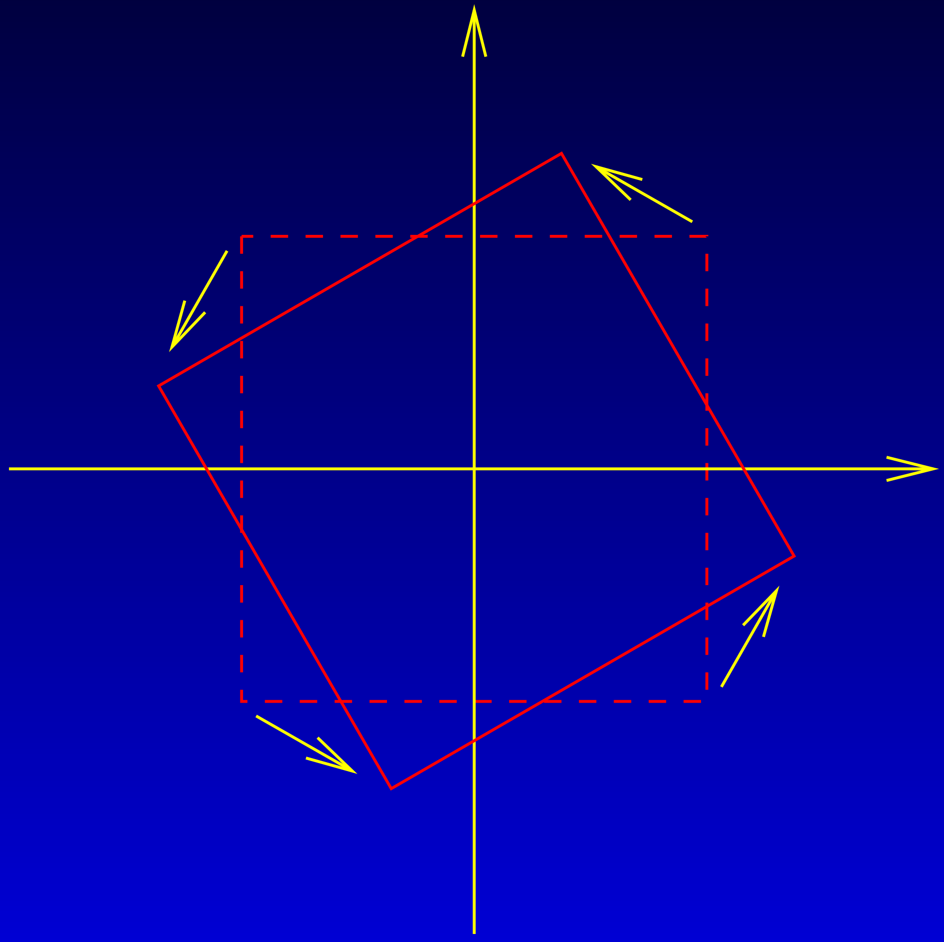
Skewness

$S = 0,\ F \approx 2,\ \theta = 0$



Rotation

$$S = 0,\ F = 1,\ \theta \approx 0.5$$



Inversion symmetry

Define $W'_B = W_B^{-1}$, then $J'_{W_B} = J^{-1}_{W_B}$

and

$$\begin{aligned} S' &= -S \\ F' &= F \\ \theta' &= -\theta \end{aligned}$$

Since $p(J_{W_B})$ is even in S and θ:

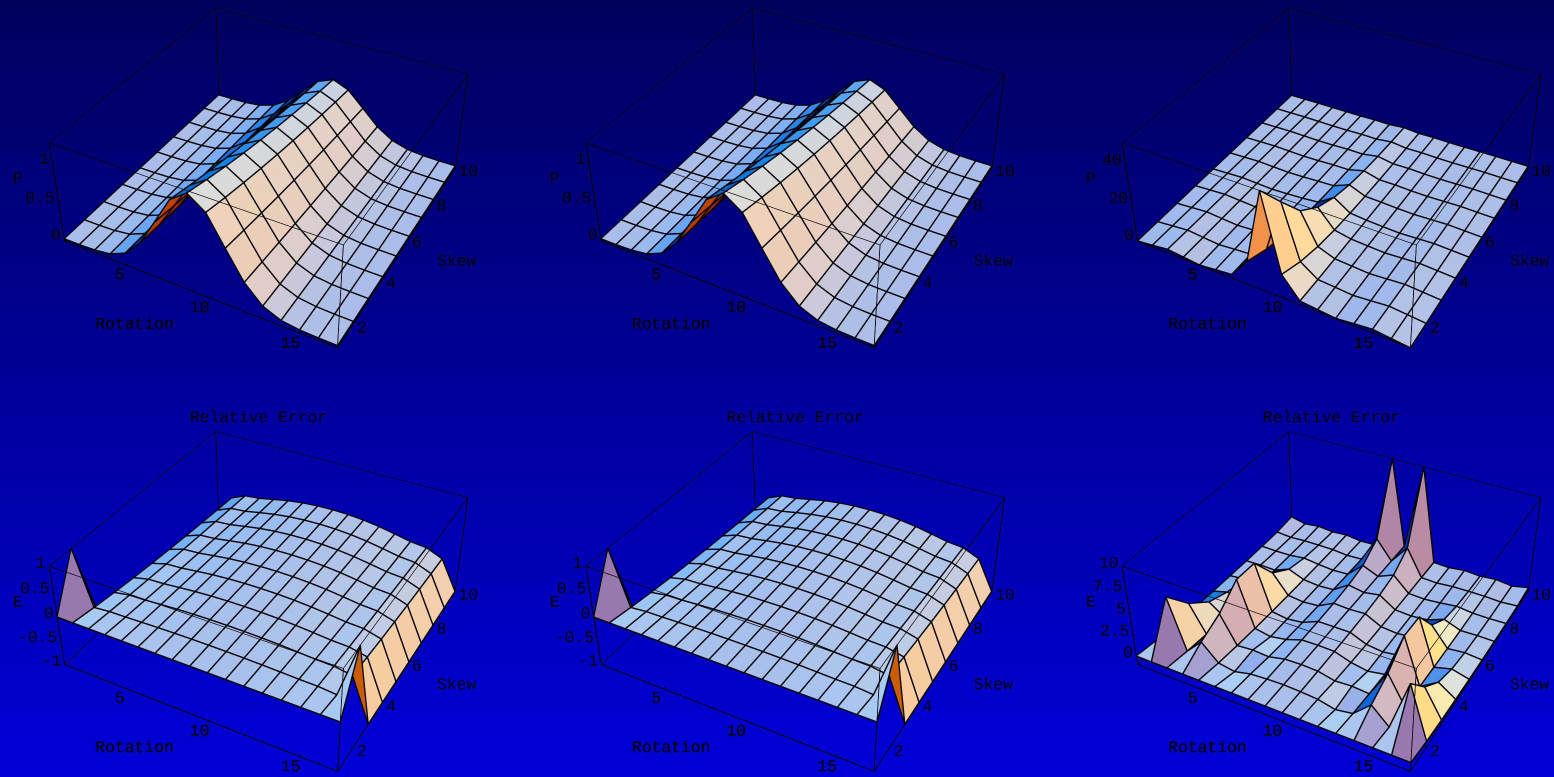
$$p(J_{W_B}) = p(J^{-1}_{W_B})$$

$p(J_{W_B})$

For $\sigma > 0.4$ we approximate:

$p(J_{W_B}) = \text{Gauss}(S) \times \text{Gauss}(\theta) \times \text{Exp}(-F^{\alpha})$

$\sigma = 1$ $\sigma = 0.6$ $\sigma = 0.3$



Warp Distribution

Assumption:
All local linear transformations are statistically independent.

$$\begin{aligned} p(W) &= \prod_{\Omega} p(J(W)) \\ &\Rightarrow \\ E(W) &\approx \int_{\Omega} S^2 + 2\theta^2 + 2\sigma^{1.33} F^{0.67} d\tilde{x} \end{aligned}$$

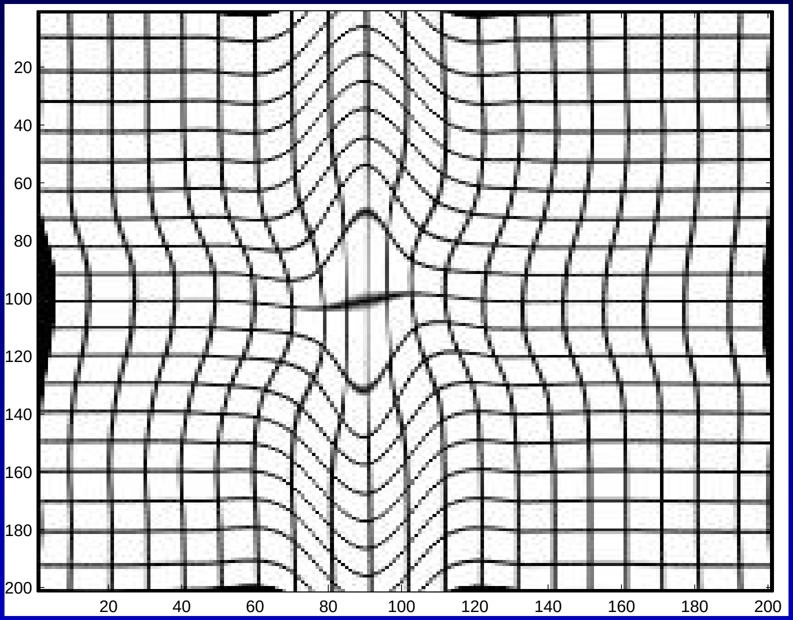
where

$$d\tilde{x} = \sqrt{\det(J)}dx$$

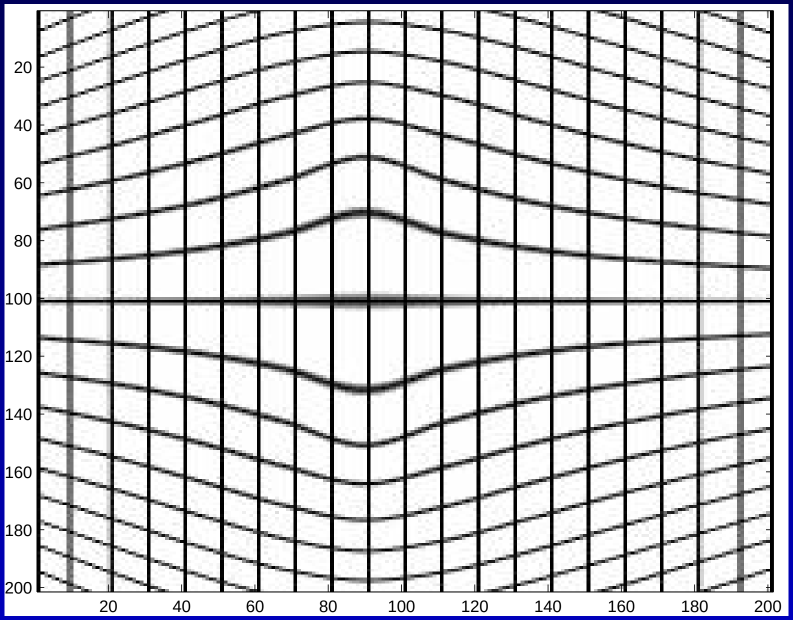
Scaling term

The scaling term S^2 aims at keeping the local area constant:

S^2



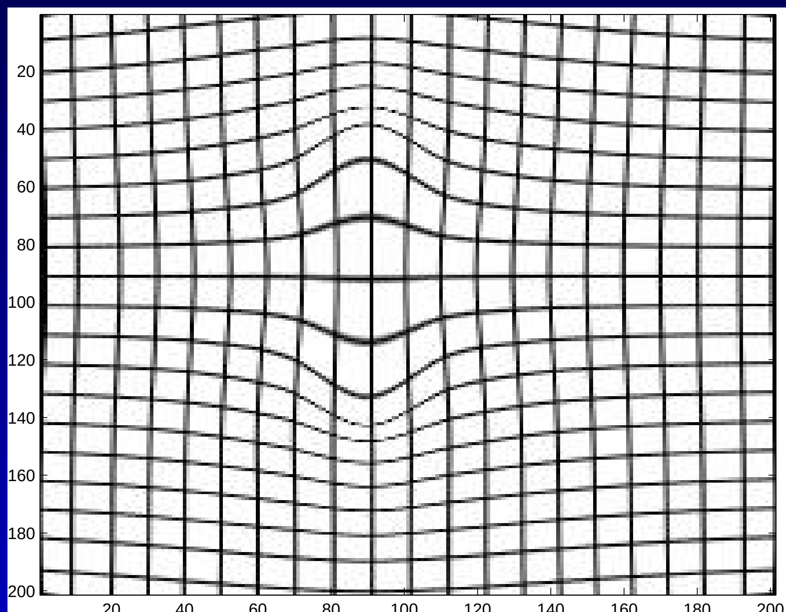
Thin-Plate



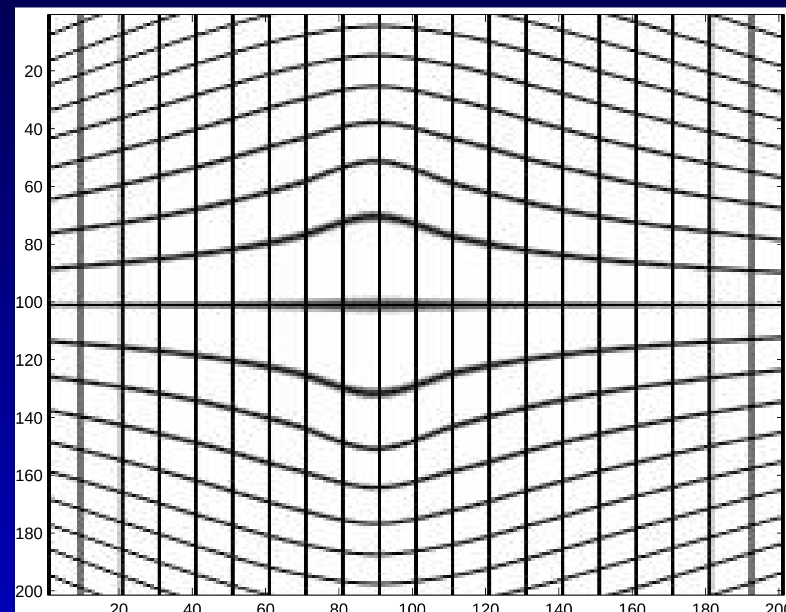
Skewness term

The skewness term F aims at keeping the local skew low:

F



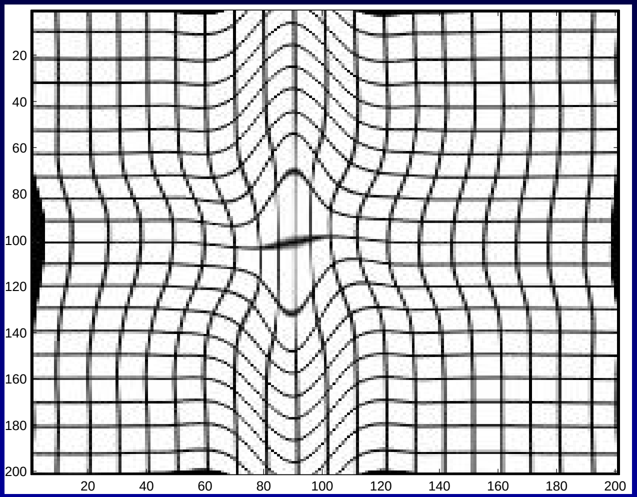
Thin-Plate

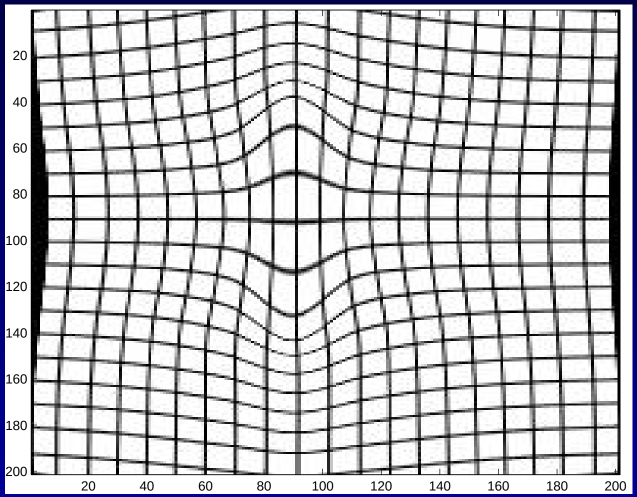


$n = 0$ approximation

$$p(J_{W_B}) = \text{Gauss}(S)\, g_0(F)$$

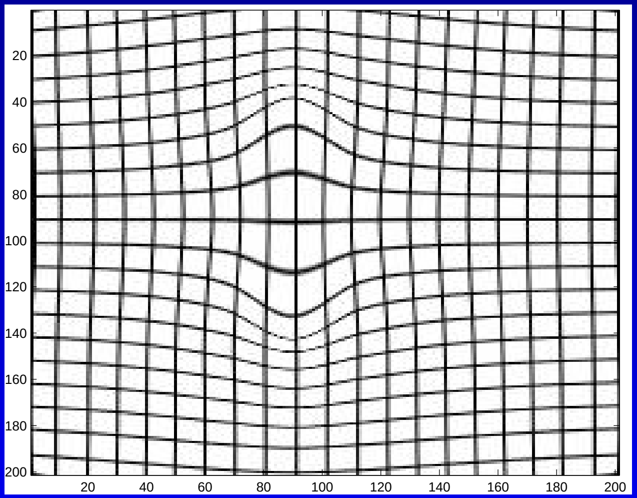
S^2

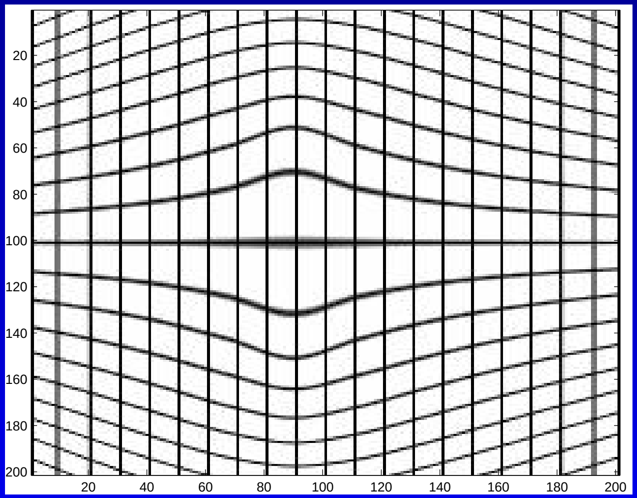




S^2, F

F

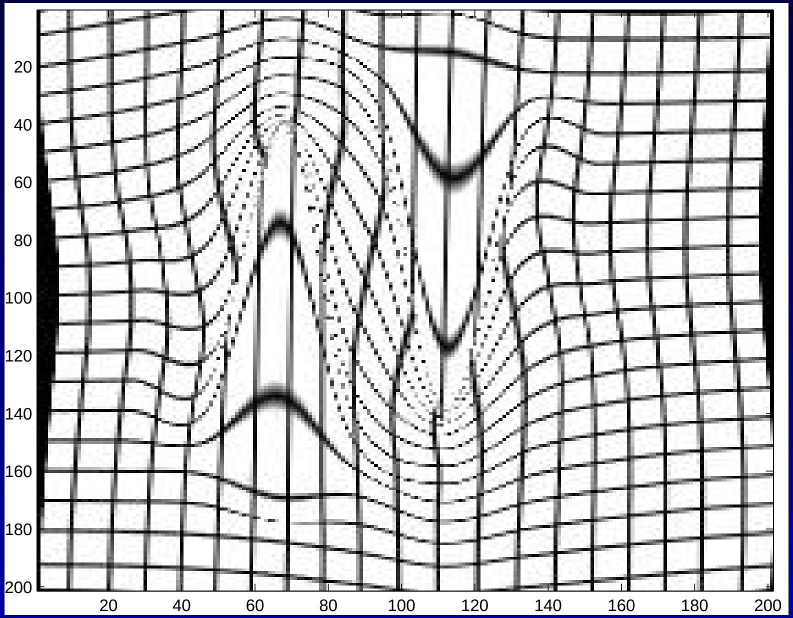




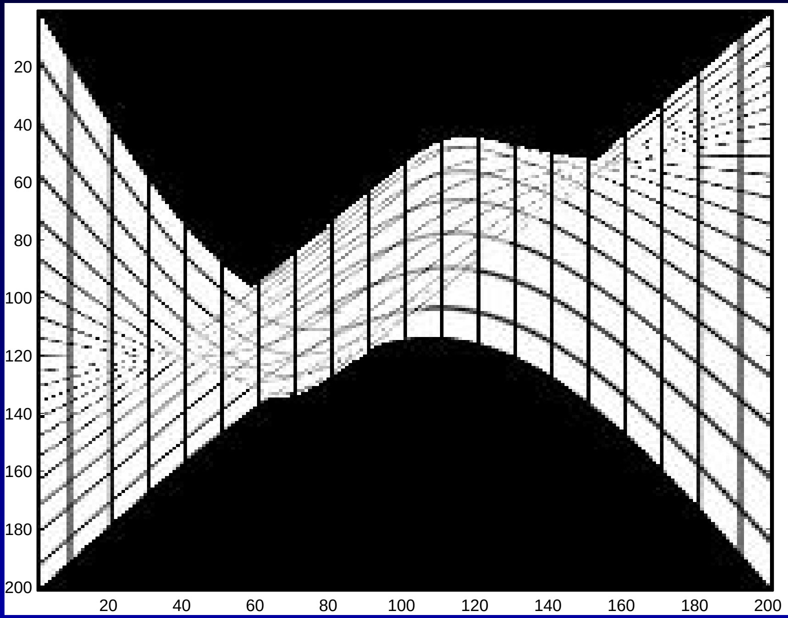
Thin-Plat

Large Deformation

S^2, F



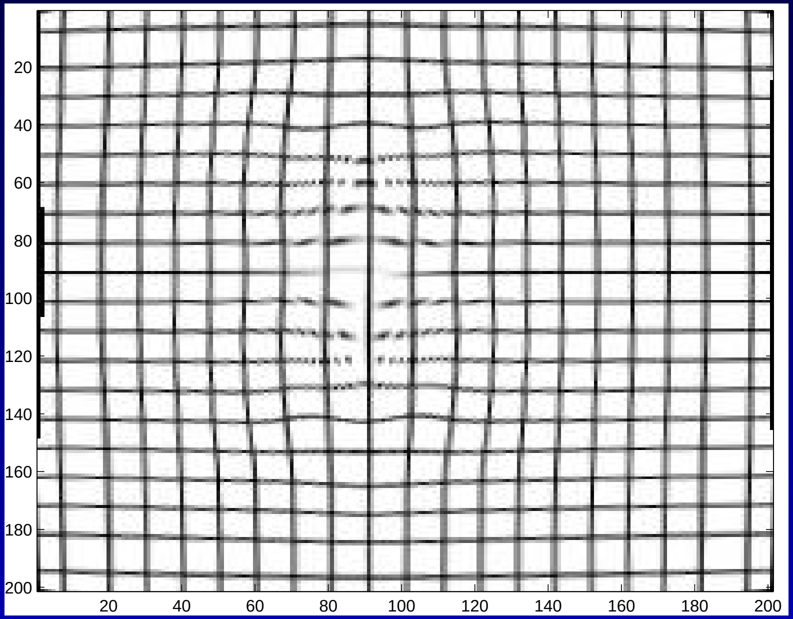
Thin-Plate

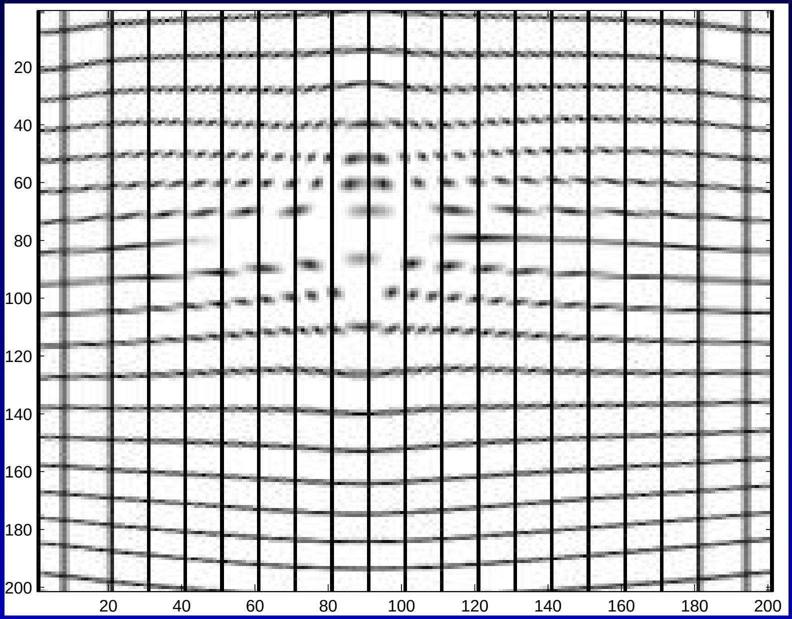


Inversion symmetry

Implementation not perfect,.....yet

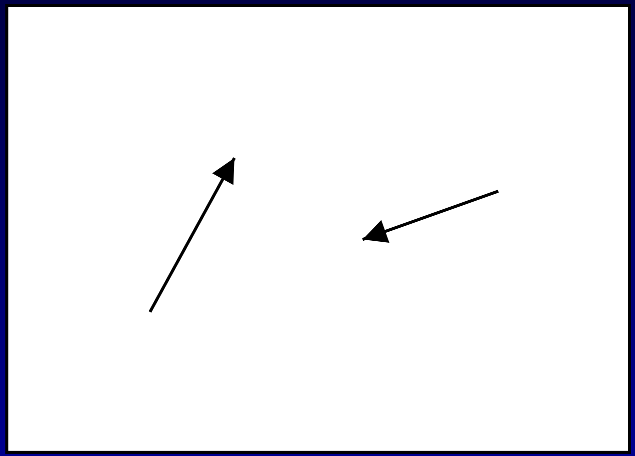
S^2, F | Thin-Plate

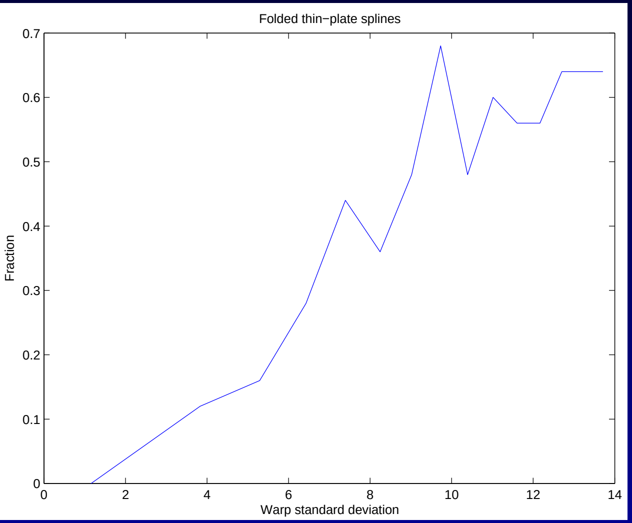


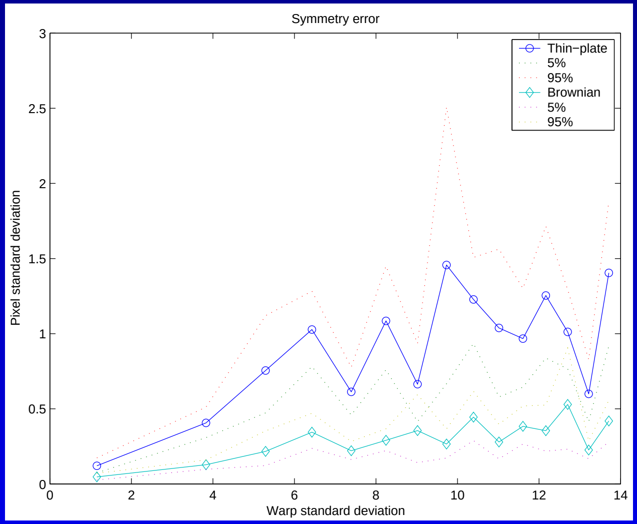


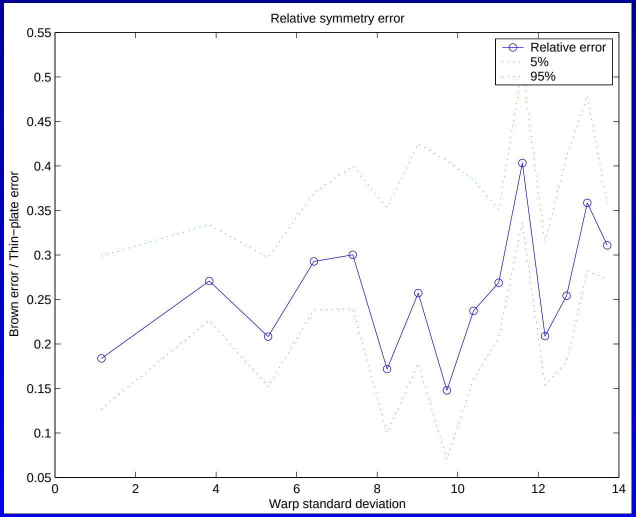
Inversion statistics

25 trials on 50×50 grid for each complexity



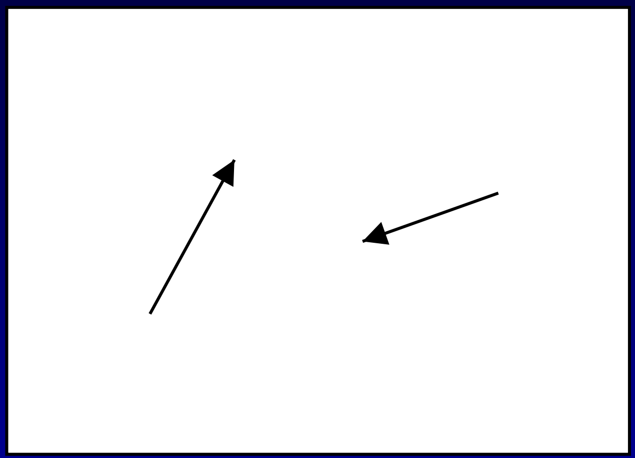


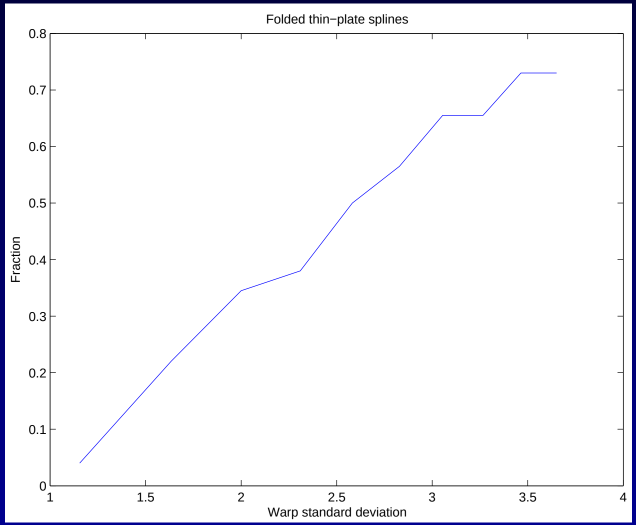


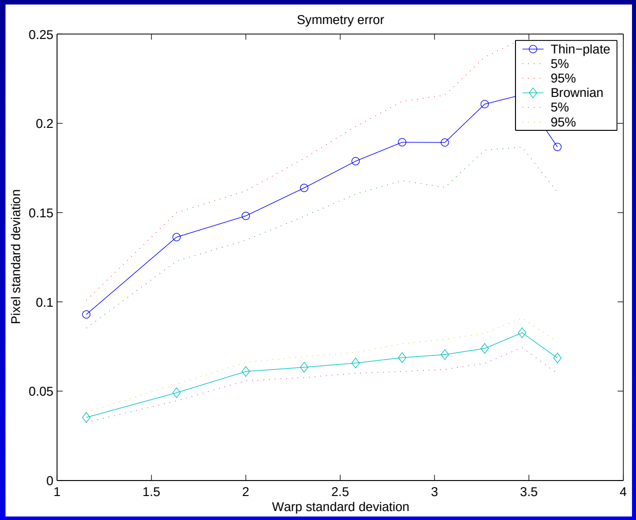


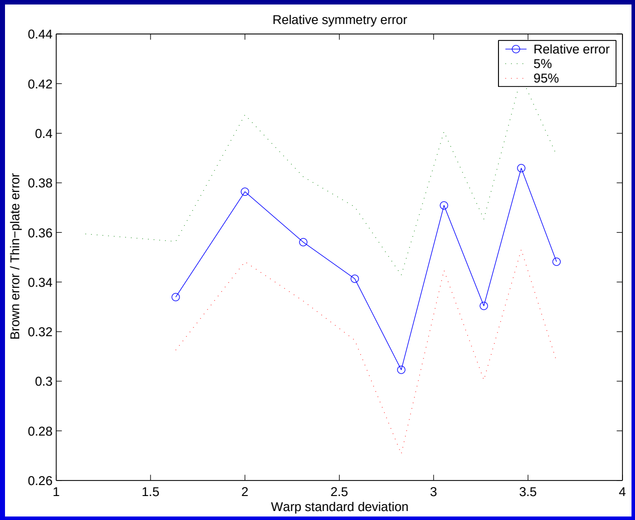
Inversion statistics

200 trials on 10×10 grid for each complexity









Wrap up

- Definition of Least Committed Warps
- Inversion symmetry
- Distribution approximated
- Maximum Likelihood interpolation
- Needs: Better implementation, expansion to nD
- Application: registration, flow, shape complexity
- Physics: chaotic flows
-
-