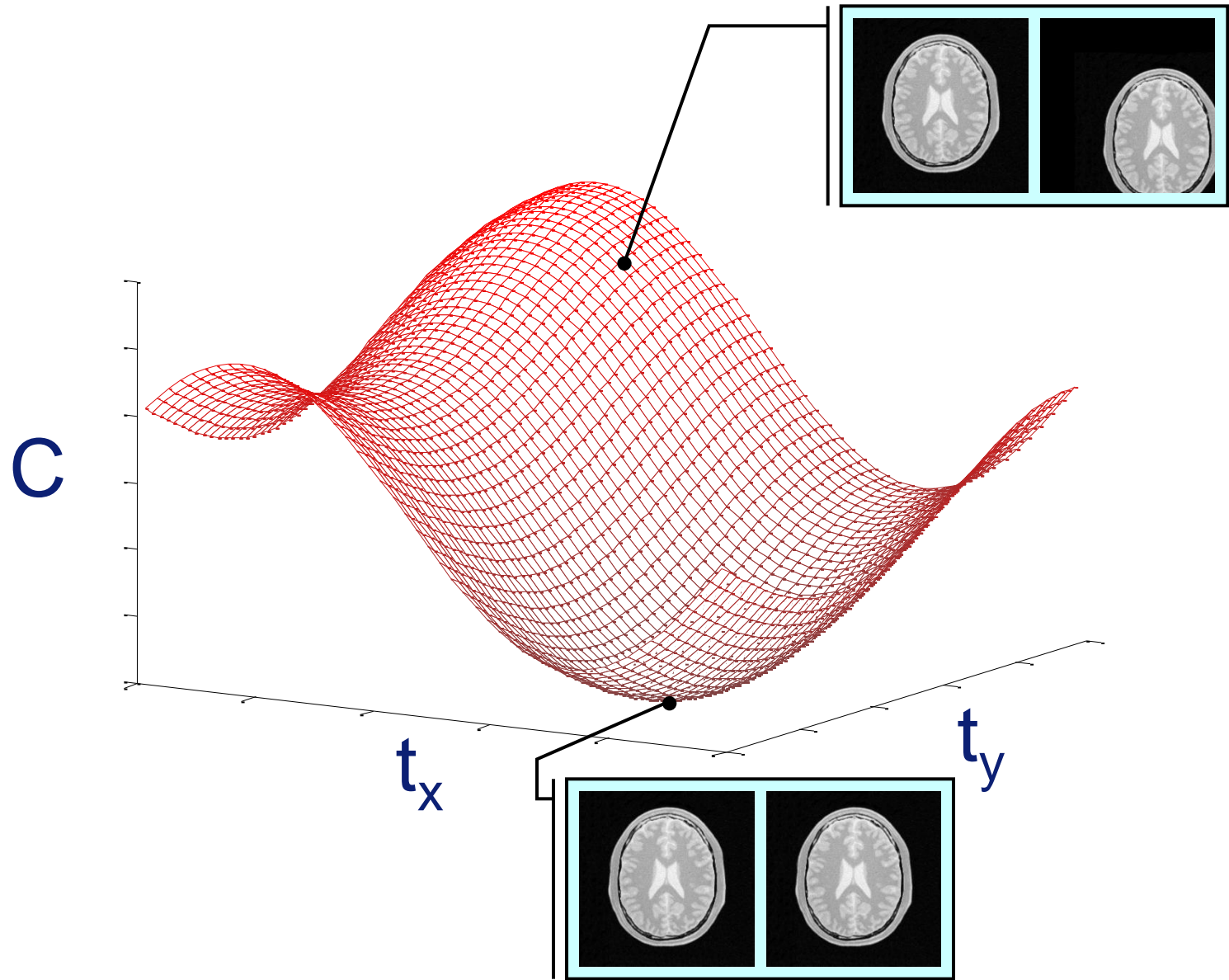


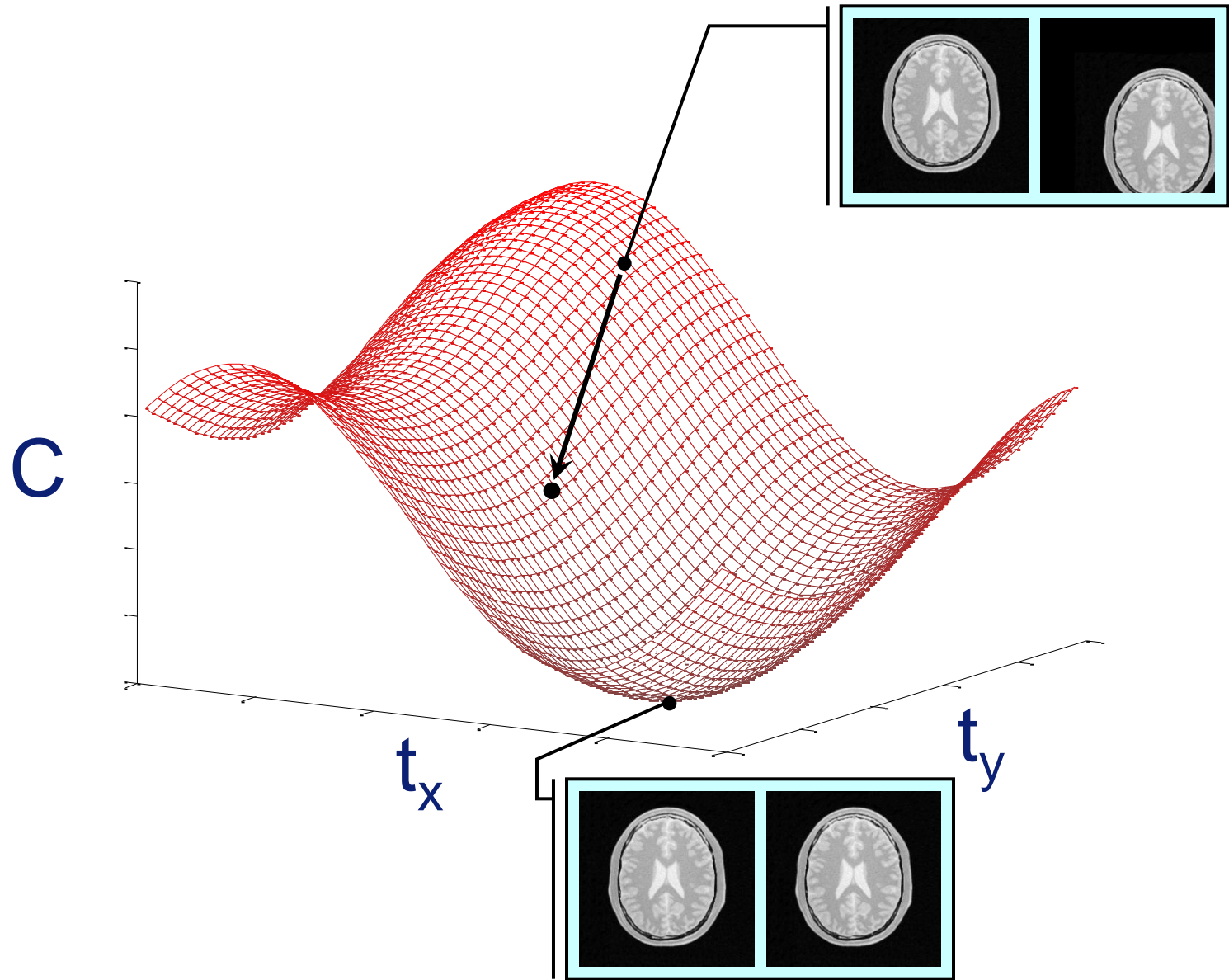
# Registration by continuous optimisation

**Stefan Klein**  
**Erasmus MC, the Netherlands**  
**Biomedical Imaging Group Rotterdam (BIGR)**

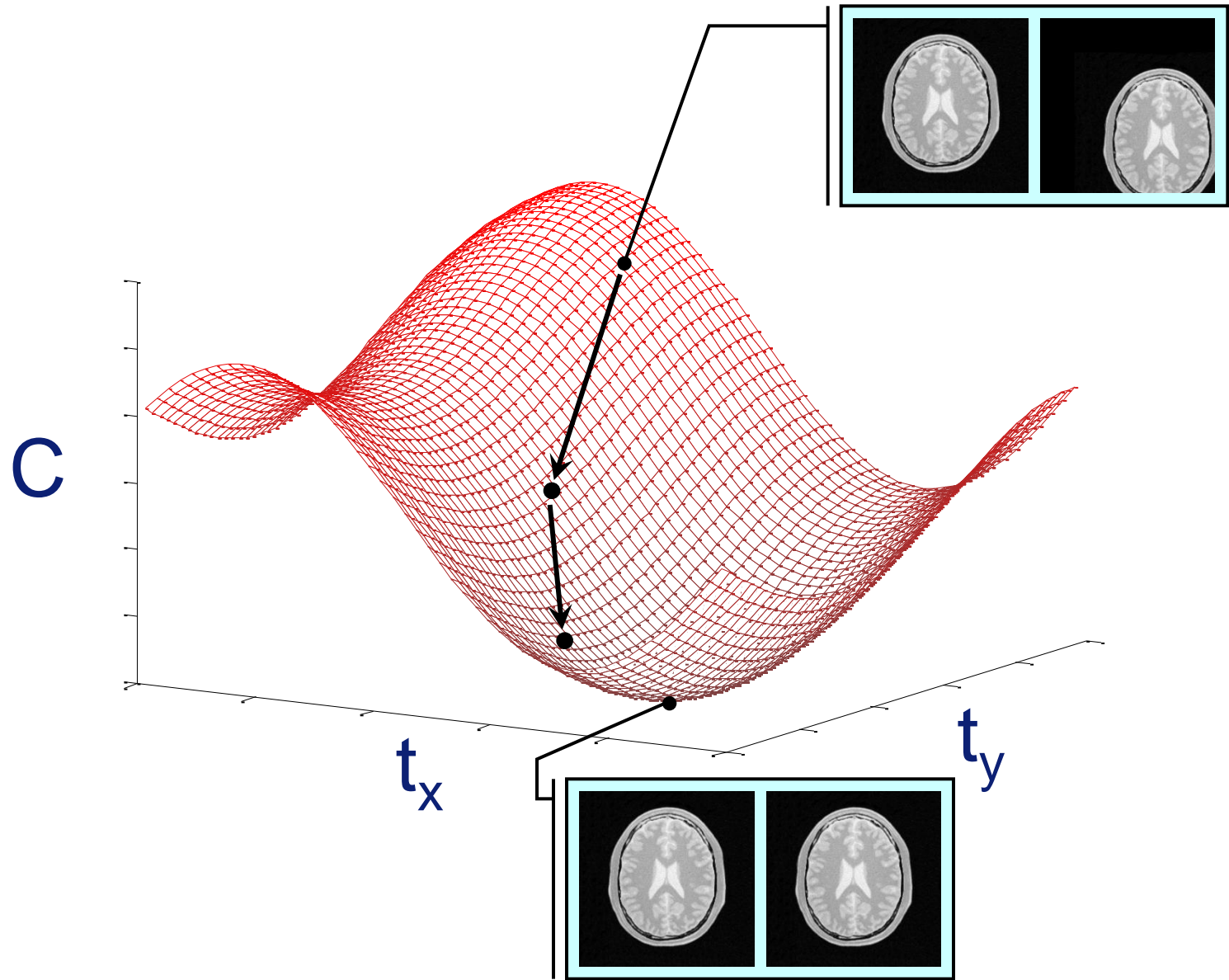
# Registration = optimisation



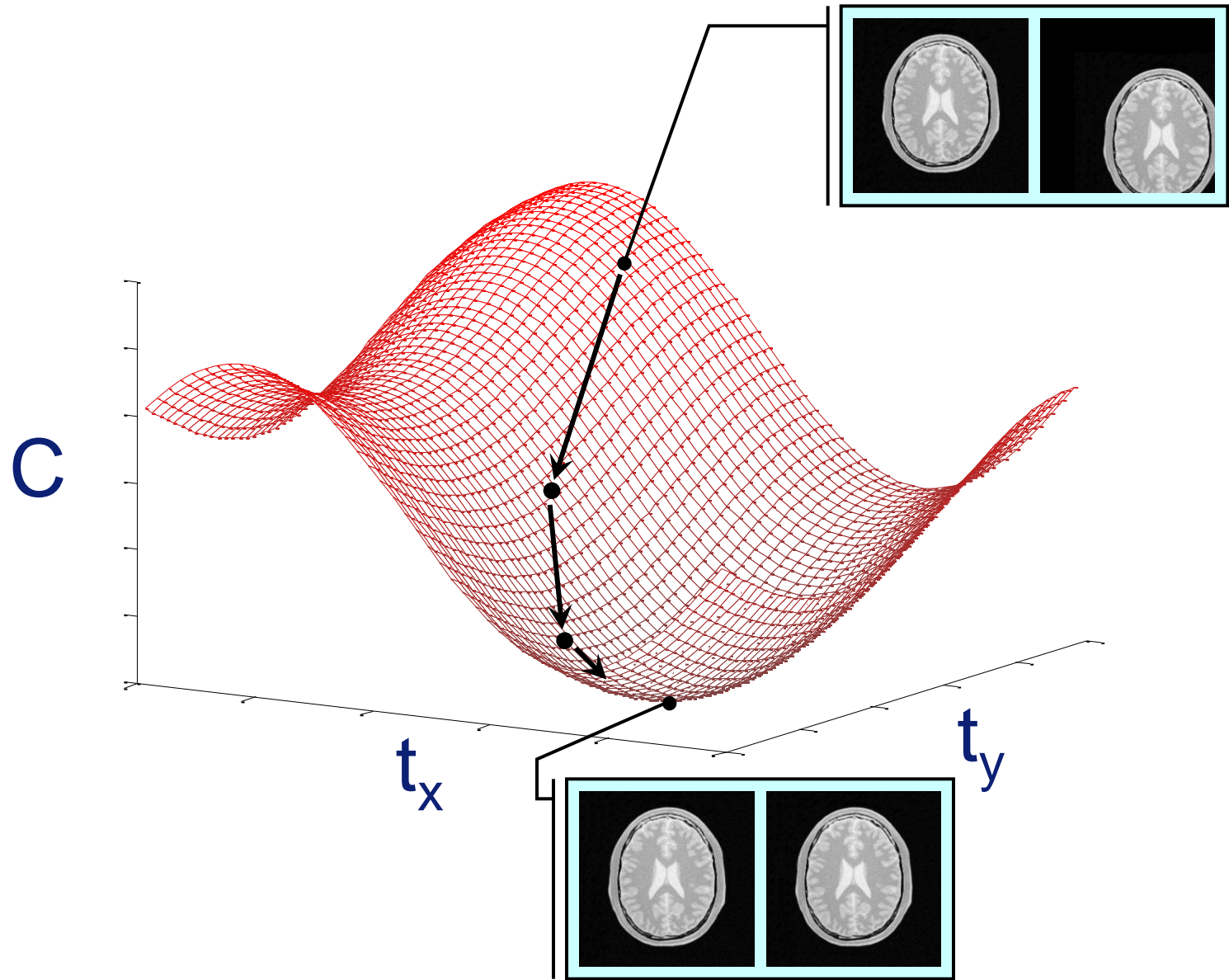
# Registration = optimisation



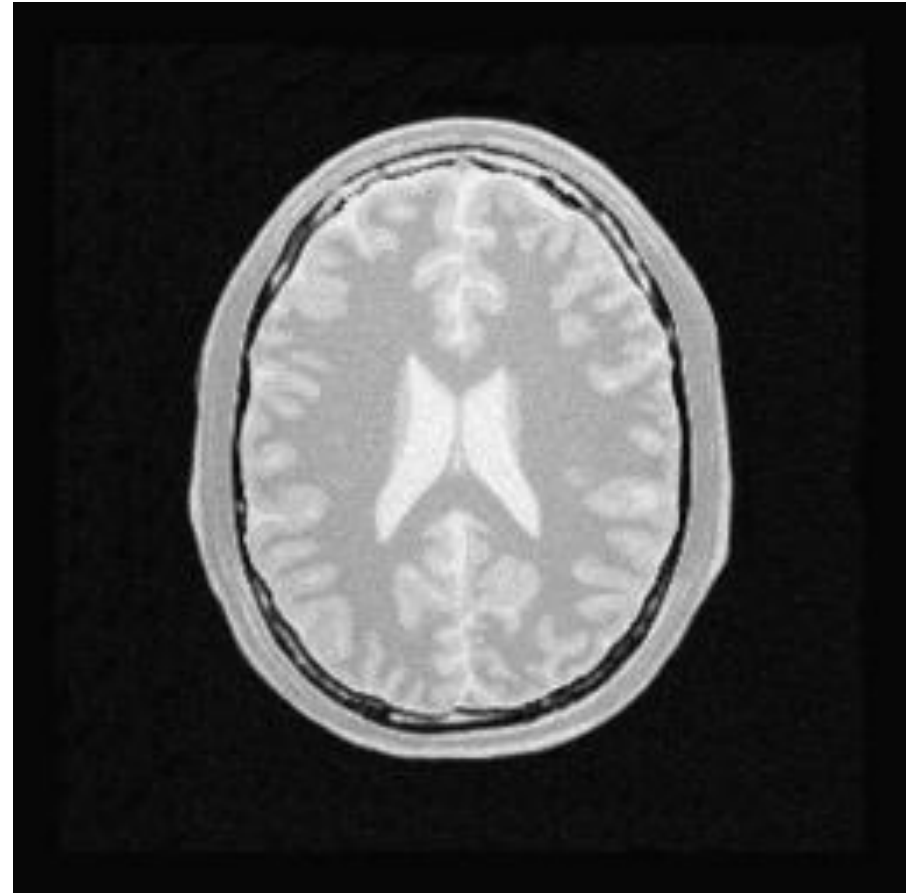
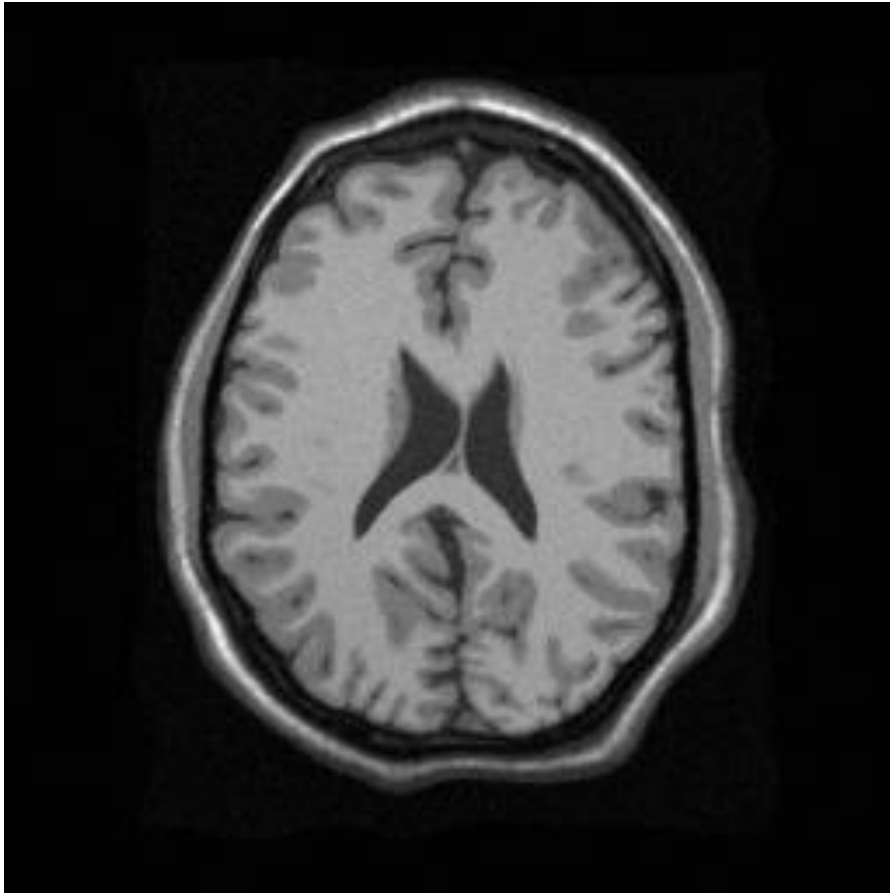
# Registration = optimisation



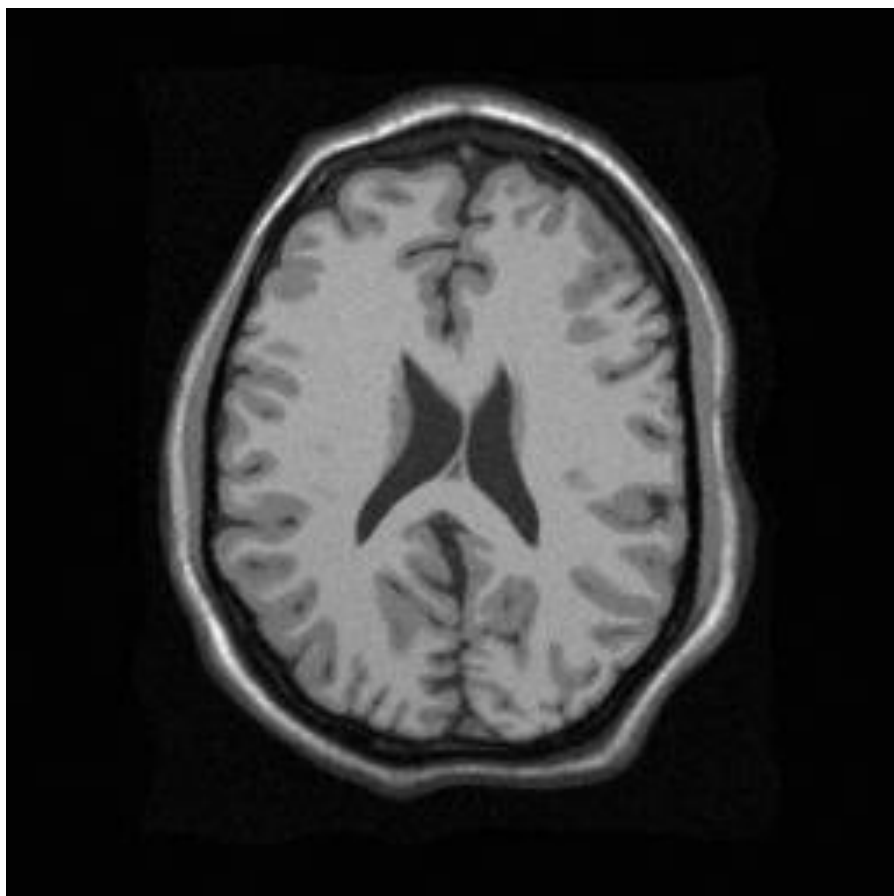
# Registration = optimisation



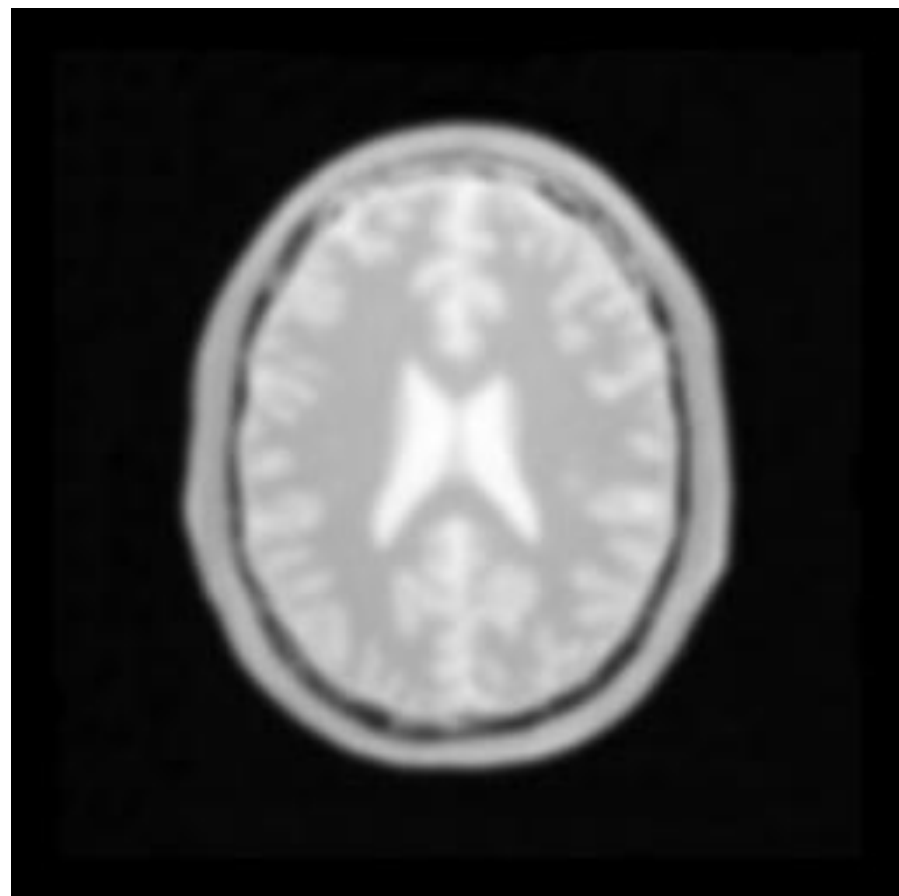
# Example



# Example

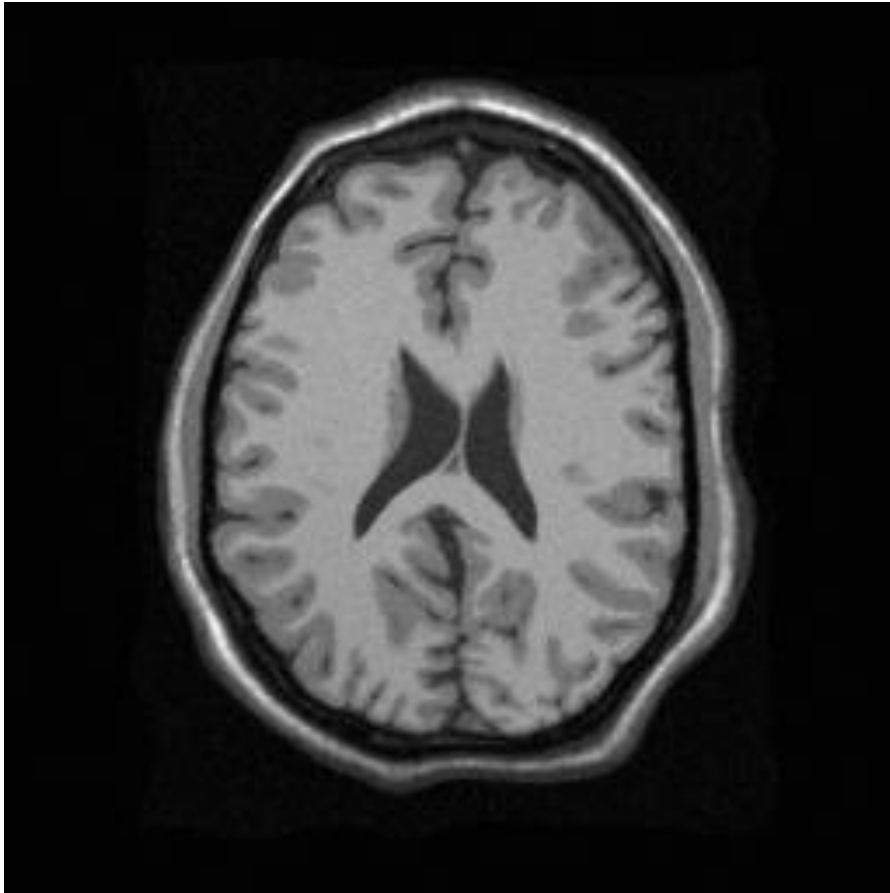


fixed image

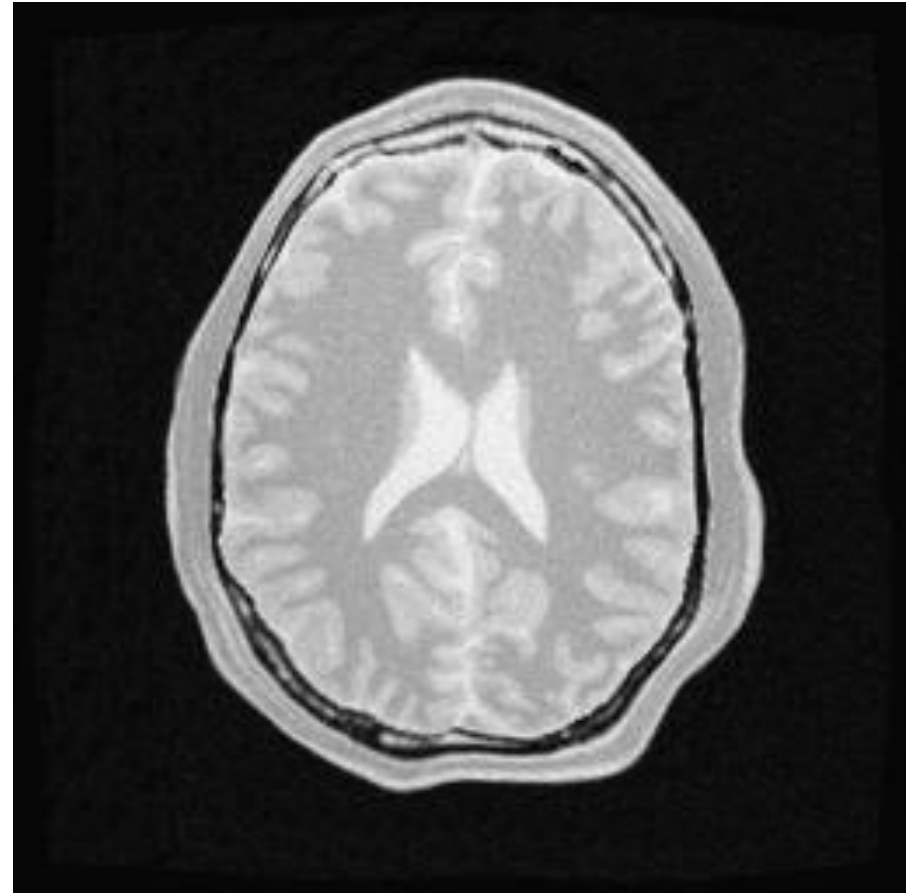


moving image

# Example



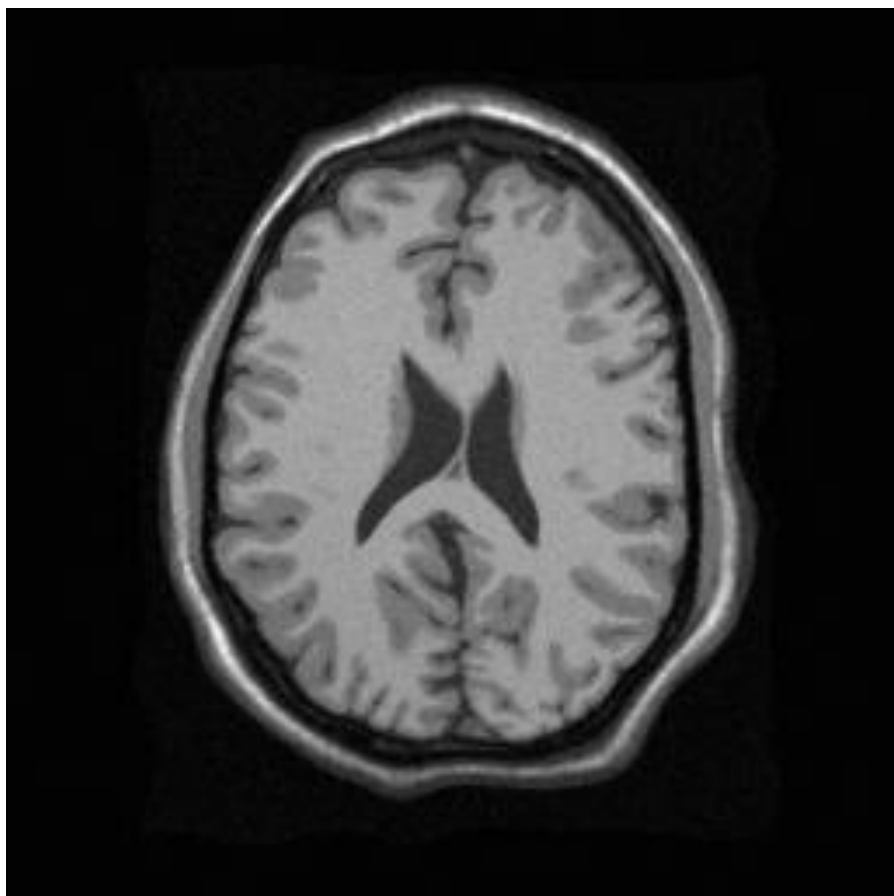
fixed image



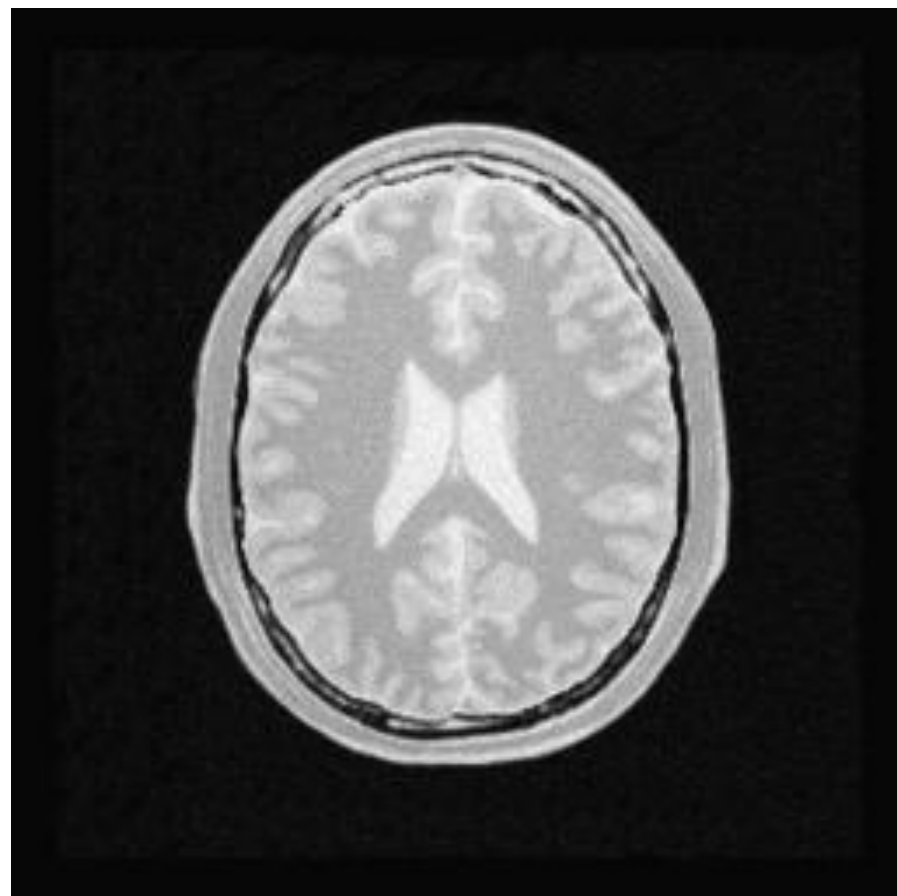
moving image



# Example



fixed image



moving image

# Math

- $F(\mathbf{x})$  = fixed image,  $M(\mathbf{x})$  = moving image  
 $\mathbf{x}$  = voxel coordinate
- Transformation function:  $T(\mathbf{x}; \mathbf{p})$   
 $\mathbf{p}$  = vector of transformation parameters
- Cost function:  $C(\mathbf{p})$   
measures similarity of fixed image  $F(\mathbf{x})$  and  
deformed moving image  $M(T(\mathbf{x}; \mathbf{p}))$
- Find  $\mathbf{p}$  that minimises  $C$

# Iterative optimisation

$$\mathbf{p}_{k+1} = \mathbf{p}_k + a_k \cdot \mathbf{d}_k$$

$\mathbf{d}_k$  = search direction

$a_k$  = step size

gradient descent:  $\mathbf{d}_k = -\frac{\partial \mathcal{C}}{\partial \mathbf{p}}(\mathbf{p}_k) \equiv -\mathbf{g}_k$

# Gradient descent

$$\mathbf{p}_{k+1} = \mathbf{p}_k - a_k \cdot \mathbf{g}_k$$

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ \vdots \end{bmatrix}_{k+1} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ \vdots \end{bmatrix}_k - a_k \cdot \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ \vdots \end{bmatrix}_k$$

# Gradient descent

$$\mathbf{p}_{k+1} = \mathbf{p}_k - a_k \cdot \mathbf{g}_k$$

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ \vdots \end{bmatrix}_{k+1} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ \vdots \end{bmatrix}_k - a_k \cdot \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \vdots \\ \vdots \end{bmatrix}_k$$

$g_1 = \left[ \frac{\partial C}{\partial p_1} \right]_k$

# Cost function derivative

Example for mean of squared differences:

$$C(\mathbf{p}) = \frac{1}{N} \sum_{\mathbf{x}} (\mathbf{F}(\mathbf{x}) - \mathbf{M}(\mathbf{T}(\mathbf{x}; \mathbf{p})))^2$$

$$\frac{\partial C}{\partial \mathbf{p}} = -\frac{2}{N} \sum_{\mathbf{x}} (\mathbf{F}(\mathbf{x}) - \mathbf{M}(\mathbf{T}(\mathbf{x}; \mathbf{p}))) \frac{\partial \mathbf{M}}{\partial \mathbf{p}}$$

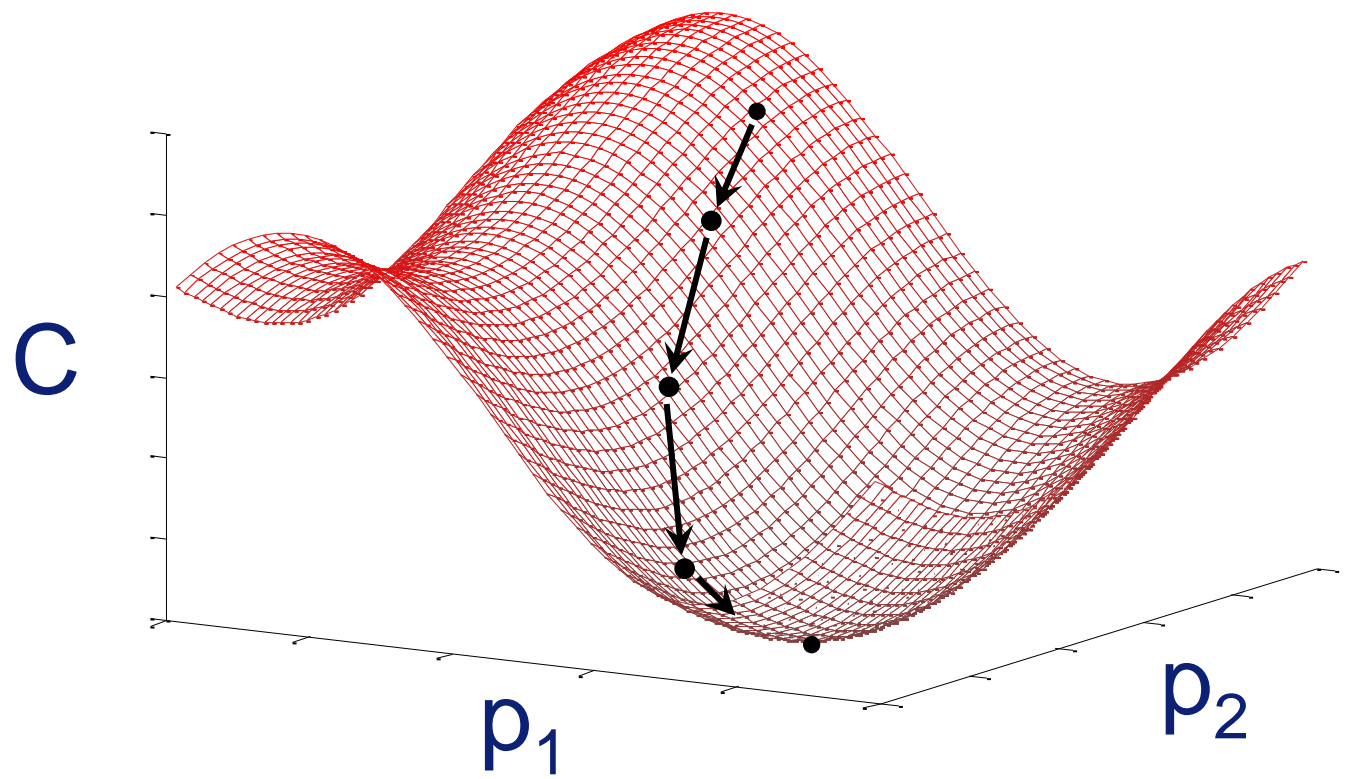
$$= -\frac{2}{N} \sum_{\mathbf{x}} (\mathbf{F}(\mathbf{x}) - \mathbf{M}(\mathbf{T}(\mathbf{x}; \mathbf{p}))) \left( \frac{\partial \mathbf{T}}{\partial \mathbf{p}} \right)^t \frac{\partial \mathbf{M}}{\partial \mathbf{x}}$$

# Choice of $d_k$

$$\mathbf{p}_{k+1} = \mathbf{p}_k + a_k \cdot \mathbf{d}_k$$

# Choice of $d_k$

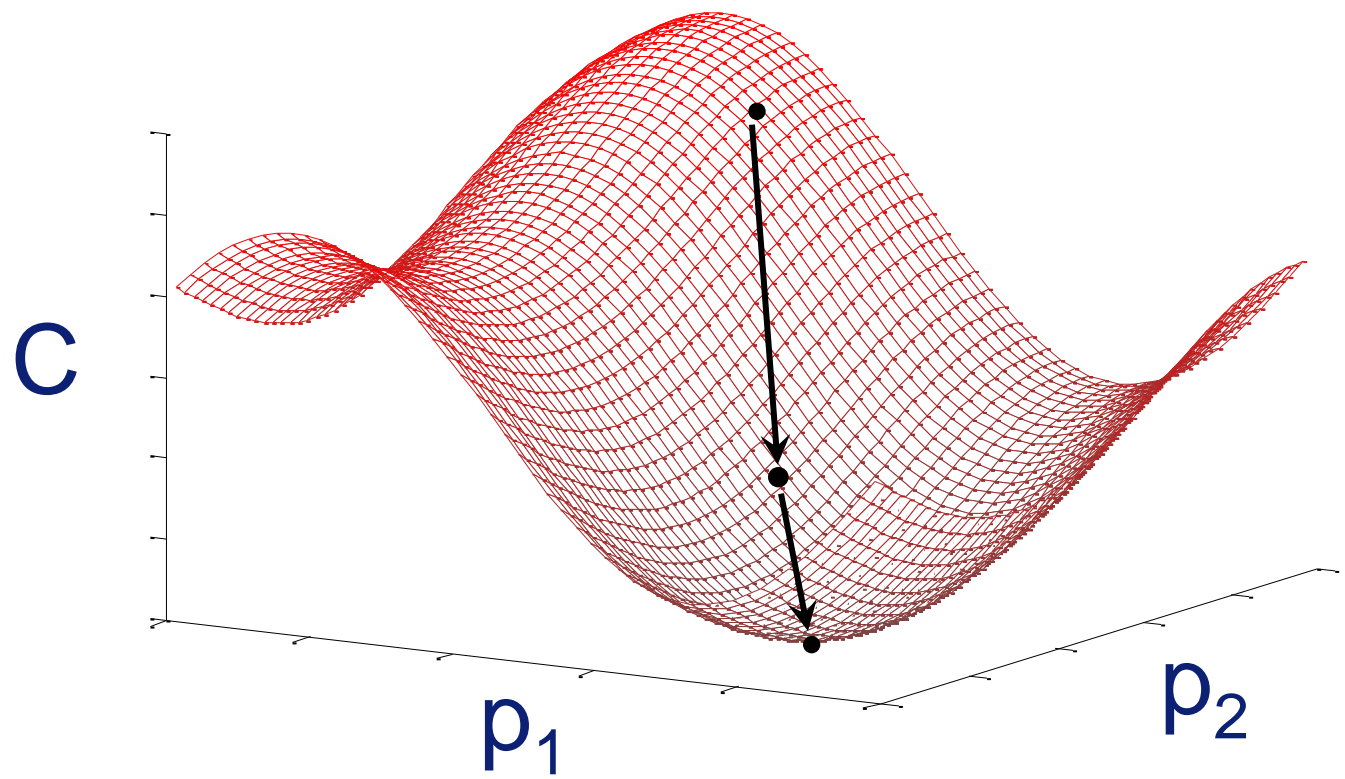
*gradient descent*





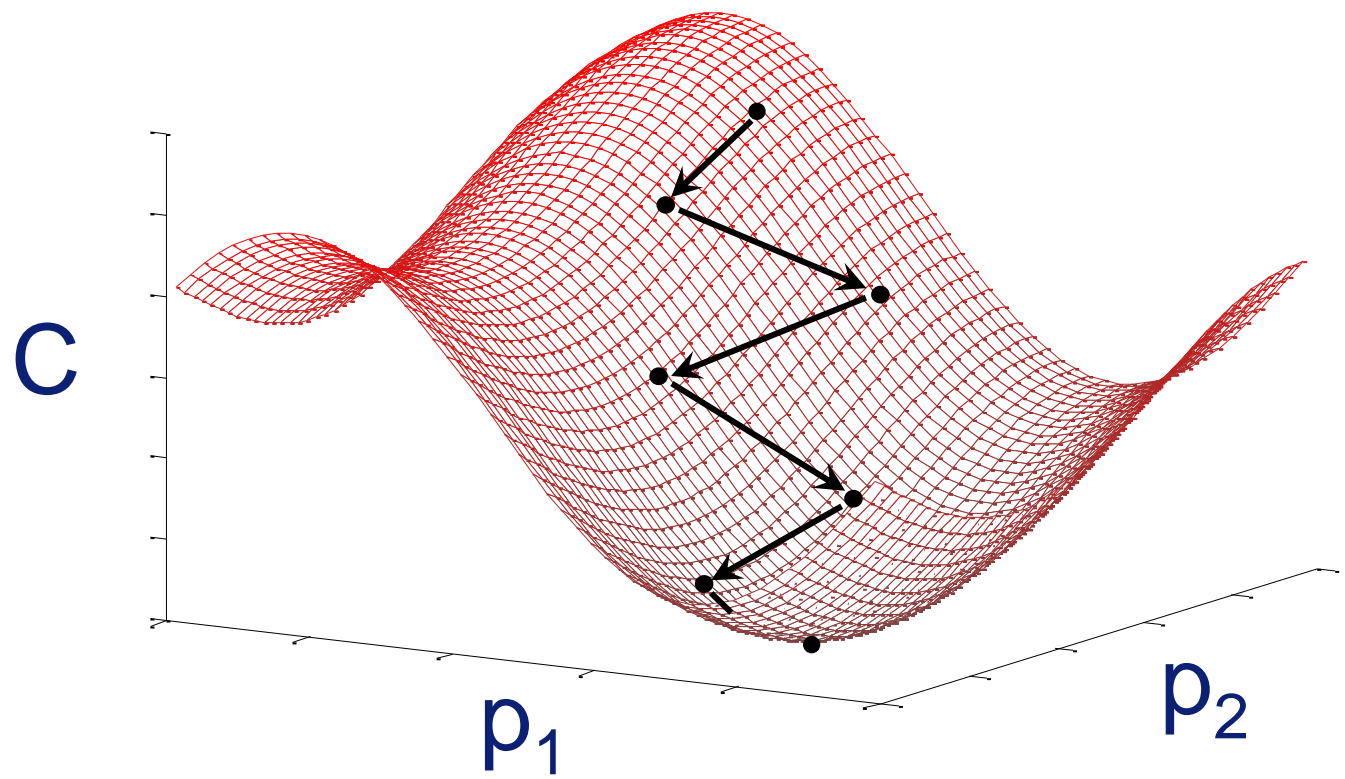
# Choice of $d_k$

*smarter steps*



# Choice of $d_k$

*cheaper steps*



# Choice of $\mathbf{d}_k$

$$\mathbf{p}_{k+1} = \mathbf{p}_k + a_k \cdot \mathbf{d}_k$$

gradient descent:  $\mathbf{d}_k = -\mathbf{g}_k$

Newton:  $\mathbf{d}_k = -[\mathbf{H}_k]^{-1} \mathbf{g}_k$

quasi-Newton:  $\mathbf{d}_k = -\mathbf{B}_k \mathbf{g}_k$

conjugate gradient:  $\mathbf{d}_k = -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}$

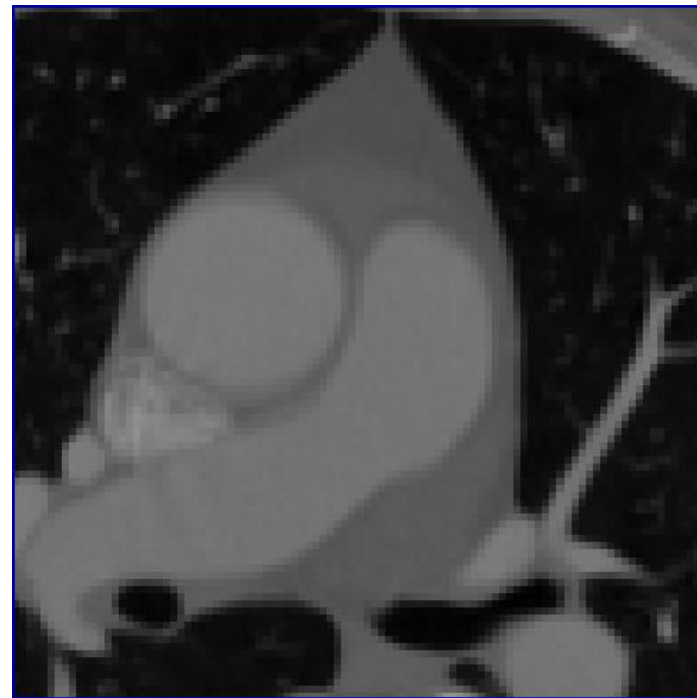
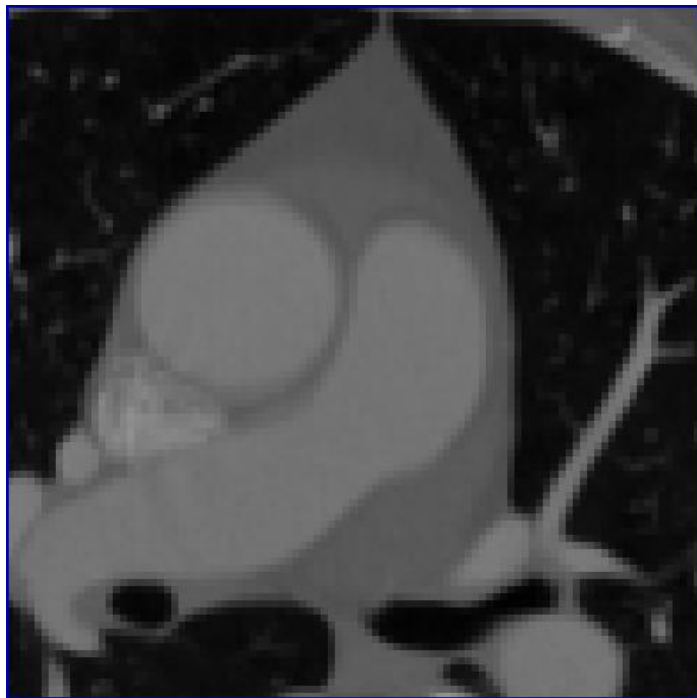
stochastic gradient:  $\mathbf{d}_k \approx -\mathbf{g}_k$

} smarter steps

→ cheaper steps

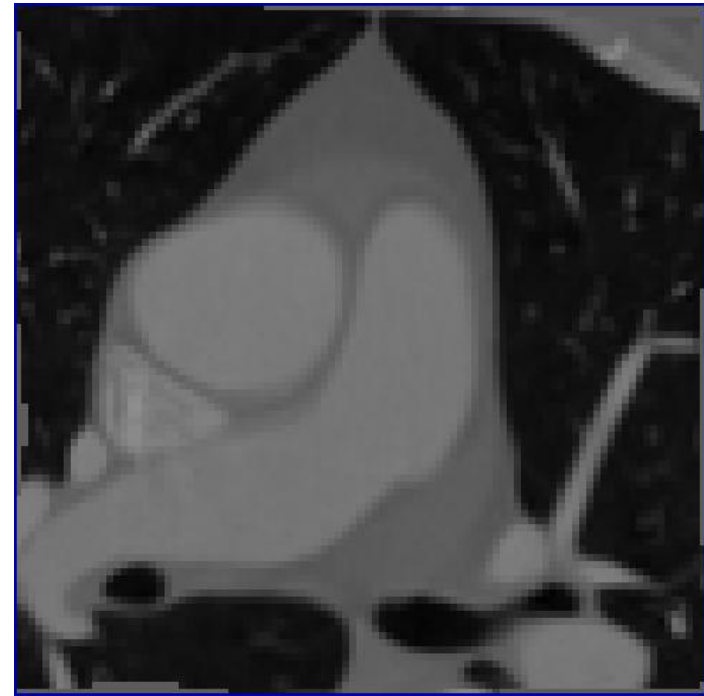
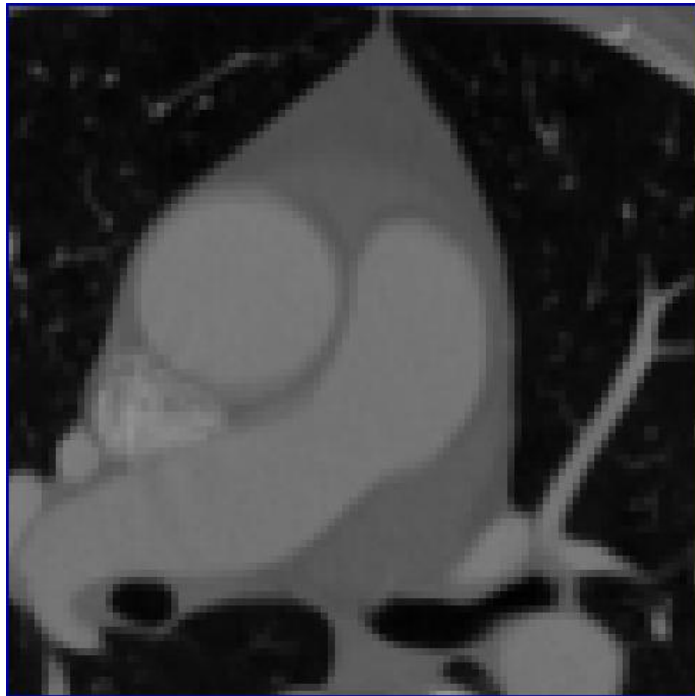
# Experimental comparison

*Cardiac CT, 97x97x97 voxels, artificially deformed*



# Experimental comparison

*Cardiac CT, 97x97x97 voxels, artificially deformed*



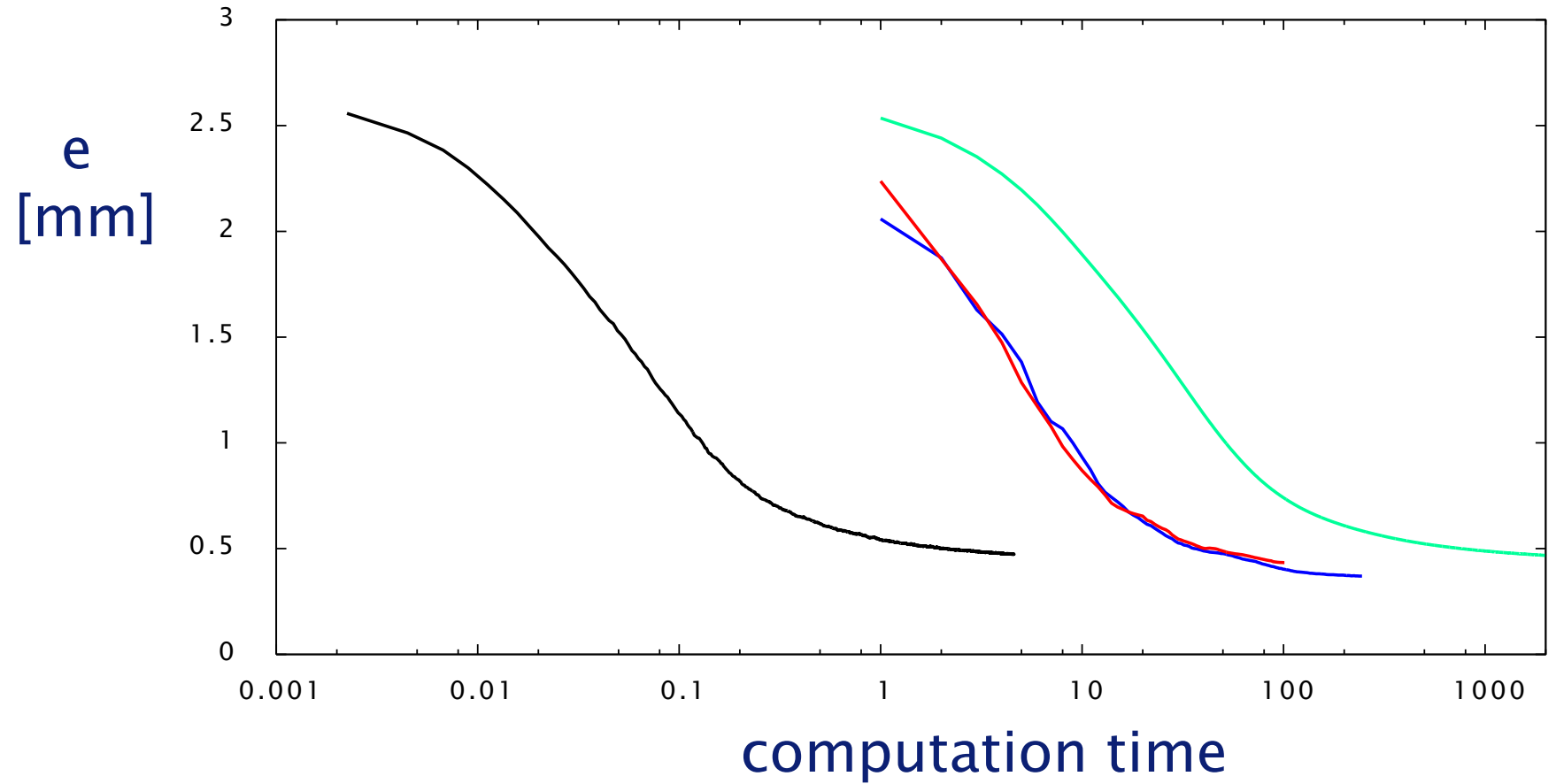
# Experimental comparison

*Error measure:*

$$e = \frac{1}{N} \sum_{\mathbf{x}} \left\| \mathbf{T}(\mathbf{x}) - \hat{\mathbf{T}}(\mathbf{x}) \right\|$$

# Experimental comparison

- gradient descent
- quasi-Newton
- conjugate gradient
- stochastic gradient



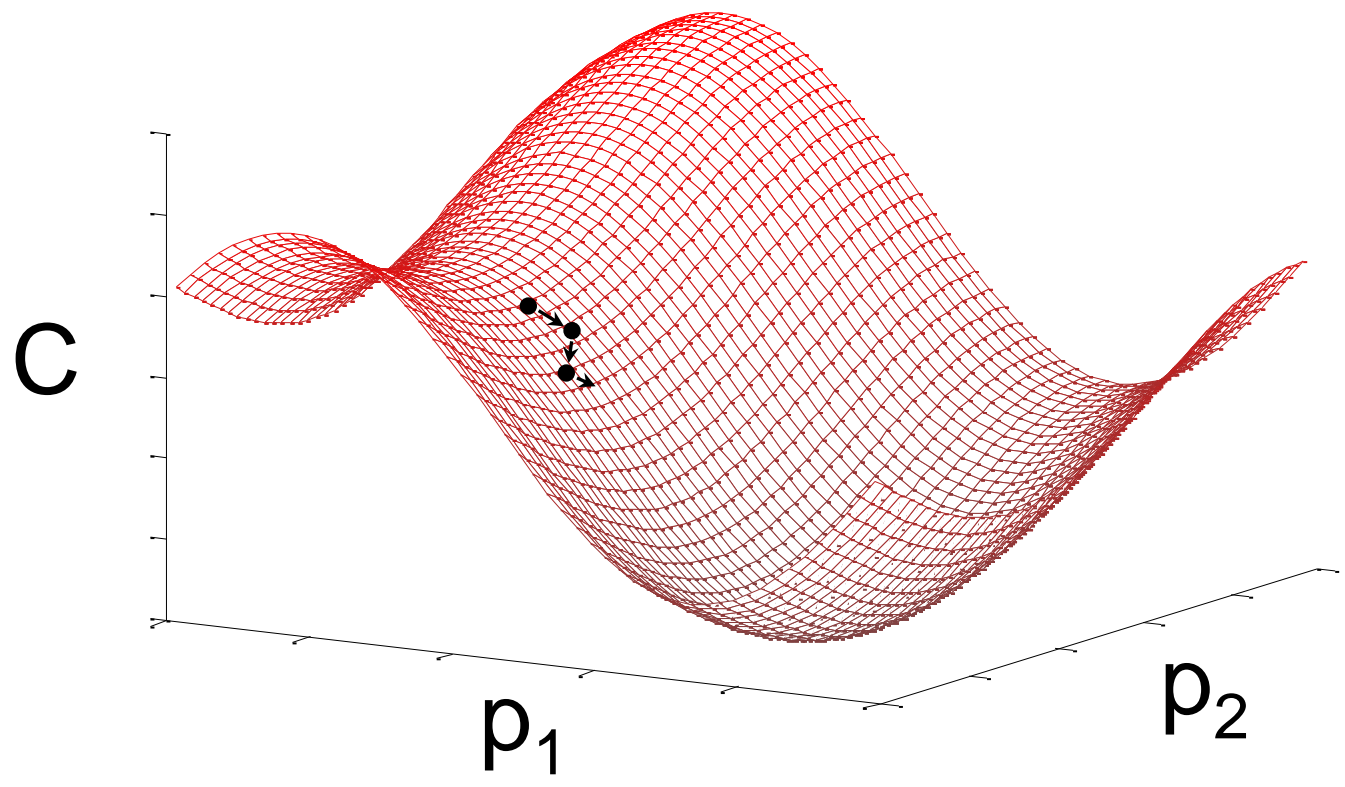
# Choice of $a_k$

$$\mathbf{p}_{k+1} = \mathbf{p}_k + a_k \cdot \mathbf{d}_k$$



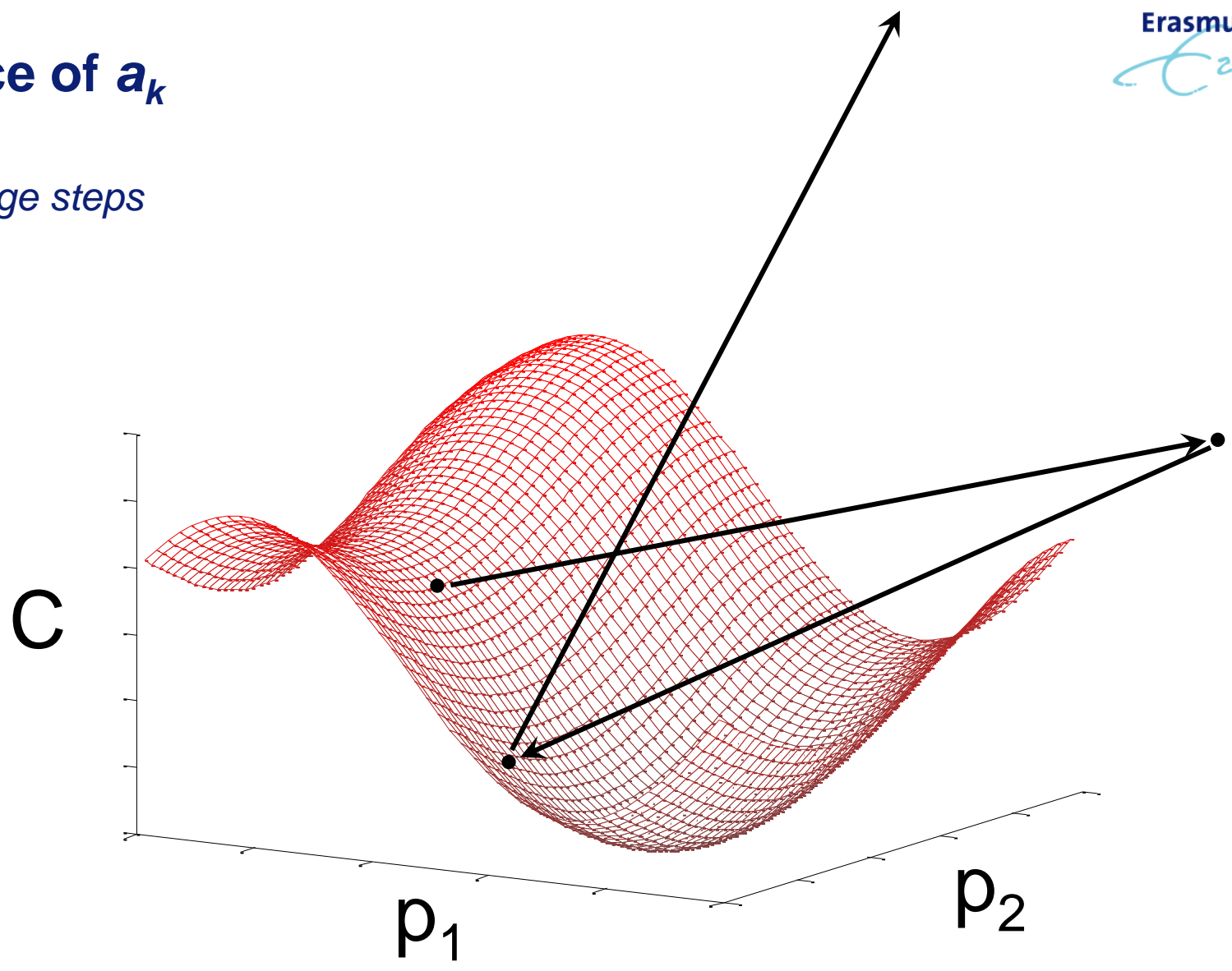
# Choice of $a_k$

*Too small steps*



# Choice of $a_k$

*Too large steps*



# Choice of $a_k$

$$\mathbf{p}_{k+1} = \mathbf{p}_k + a_k \cdot \mathbf{d}_k$$

constant:  $a_k = a$

slowly decaying:  $a_k = f(k) = a / (A + k)^\alpha$

exact line search:  $a_k = \operatorname{argmin}_a C(\mathbf{p}_k + a \mathbf{d}_k)$

inexact line search:  $a_k \approx \operatorname{argmin}_a C(\mathbf{p}_k + a \mathbf{d}_k)$  *[Wolfe conditions]*

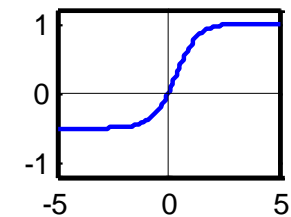
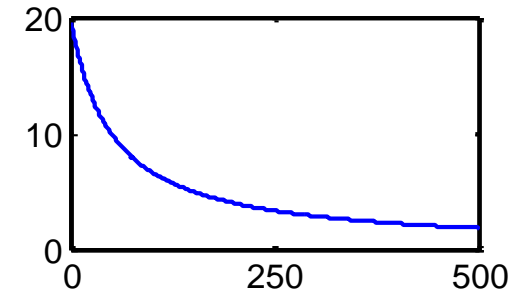
adaptive:  $a_k = F(\text{progress in previous iterations})$

# Stochastic gradient descent with adaptive strategy for $a_k$

$$\mathbf{p}_{k+1} = \mathbf{p}_k - f(t_k) \mathbf{g}_k$$

$$f(t_k) = a / (A + t_k)^\alpha$$

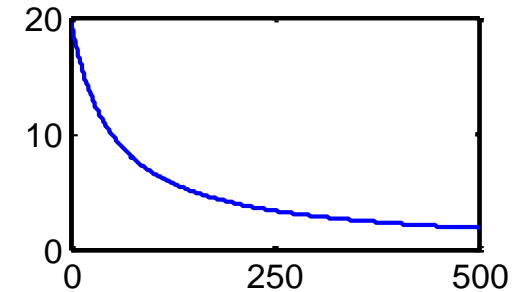
$$t_{k+1} = t_k + \text{sigmoid}(-\mathbf{g}_k^T \mathbf{g}_{k-1})$$



# Stochastic gradient descent with adaptive strategy for $a_k$

$$\mathbf{p}_{k+1} = \mathbf{p}_k - f(t_k) \mathbf{g}_k$$

$$f(t_k) = a / (A + t_k)^\alpha$$

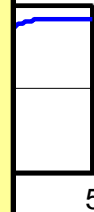


t

Choose  $a$  such that:

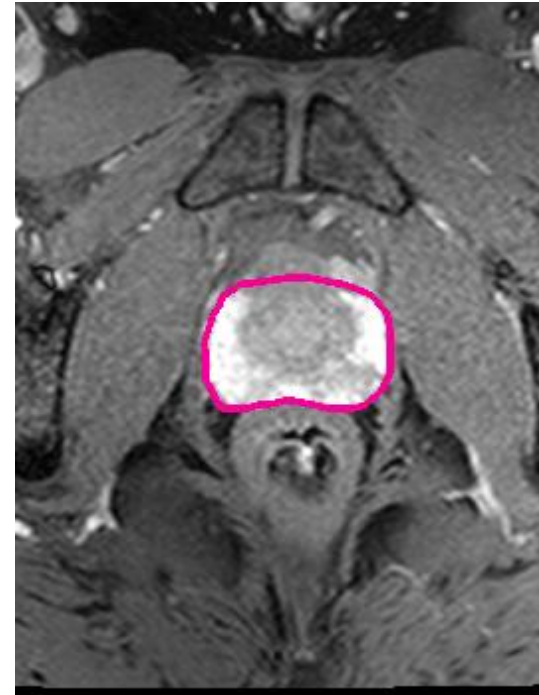
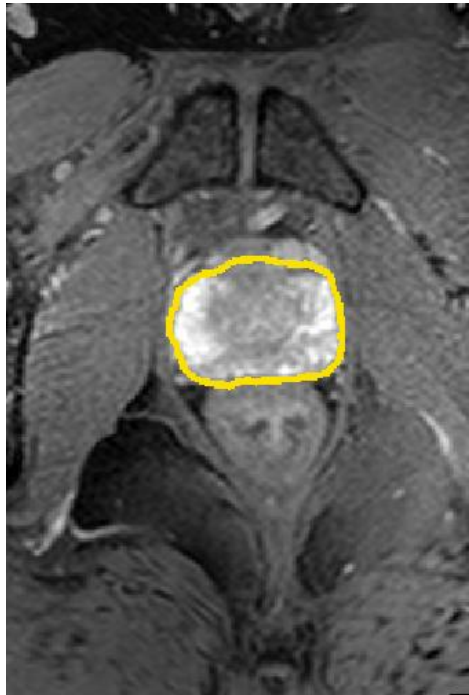
max. voxel displacement per iteration  $< \delta$  [mm]

(with 95% probability)



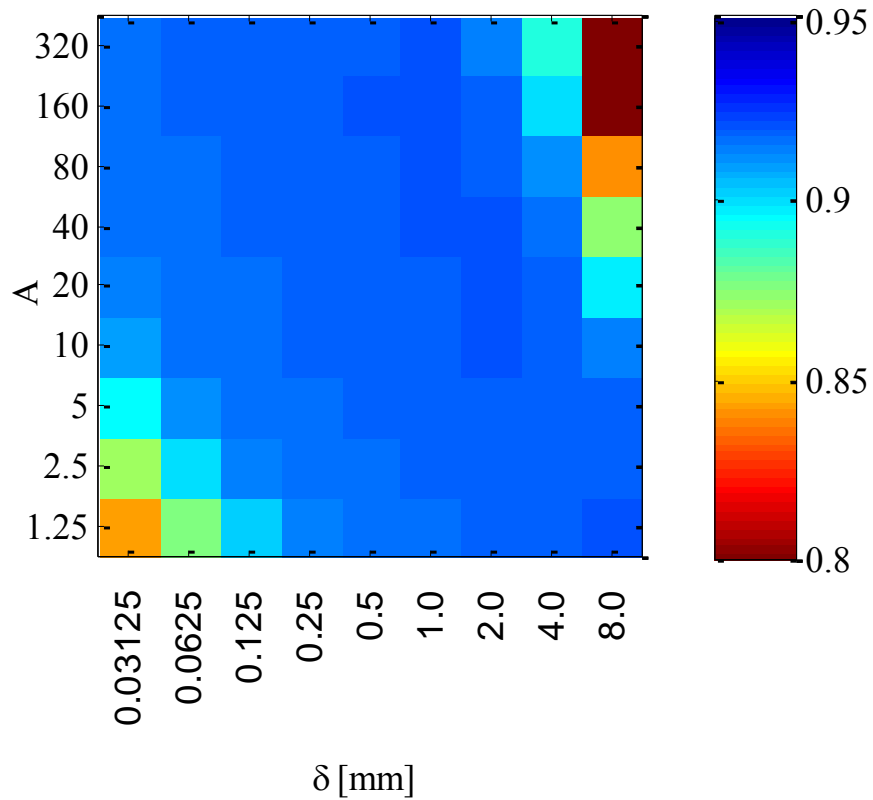
## Experimental comparison

- 6 prostate MR image pairs: nonrigid registration
- evaluation measure:  
    overlap of manual segmentations after registration

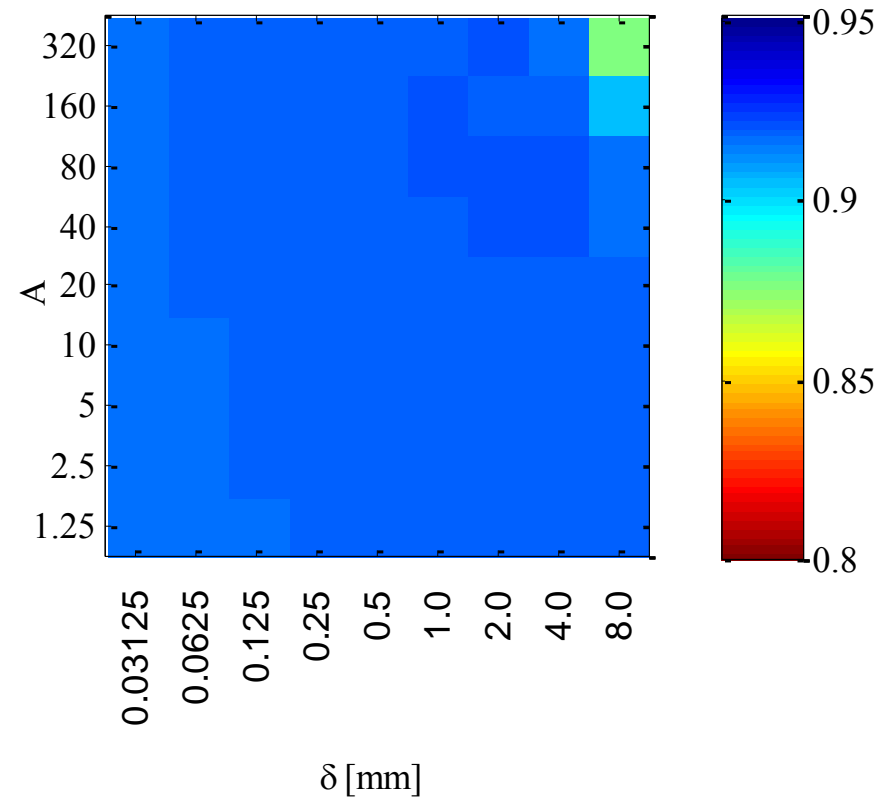


# Experimental comparison

non-adaptive



adaptive



# Experimental comparison

Experiments with:

- brain, lung, prostate
- CT, MRI
- sum of squared differences, mutual information, normalized mutual information
- rigid, nonrigid

→  $A = 20$ ,  $\delta = \text{voxelsize}$

→ good results in all experiments!



# Local similarity measures

- MI = mutual information
  - assumes grey-value distribution does not vary over image domain
- LMI = localised mutual information

$$= \frac{1}{N} \sum_x MI(\Omega_x)$$

(aka: regional MI, conditional MI, spatial information encoded MI)

# Local similarity measures

- MI = mutual information
  - assumes grey-value distribution does not vary over image domain
- LMI = localised mutual information

$$= \frac{1}{N} \sum_x MI(\Omega_x)$$

(aka: regional MI, conditional MI, spatial information encoded MI)

→ *can be efficiently implemented with stochastic gradient descent!*

# Summary

- Parametric formulation can be solved by continuous optimisation
- Derivative-based methods: require  $\partial C / \partial \mathbf{p}$
- Extensive literature
  
- Basic method: gradient descent
- Popular choice:
  - quasi-Newton or conjugate gradient icm inexact line search
- “Recommended”:
  - stochastic gradient descent with adaptive step sizes

# Literature

- Nocedal & Wright: Numerical Optimization
- IEEE Trans. Image Processing 2007 - Klein, Staring, Pluim  
*Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines*
- Int. J. Computer Vision 2009 - Klein, Pluim, Staring, Viergever  
*Adaptive stochastic gradient descent optimisation for image registration*
- IEEE Trans. Image Processing 2000 - Thevenaz, Unser  
*Optimization of mutual information for multiresolution image registration*



- Rigid and nonrigid registration
- Various cost functions, transformation models, multiresolution strategies etc.
- Many optimisation algorithms implemented
- Free: <http://elastix.isi.uu.nl>
- Based on Insight ToolKit (ITK): <http://www.itk.org>
- IEEE Trans. Medical Imaging 2010 - Klein, Staring, Murphy, Viergever, Pluim – elastix: a toolbox for intensity based medical image registration