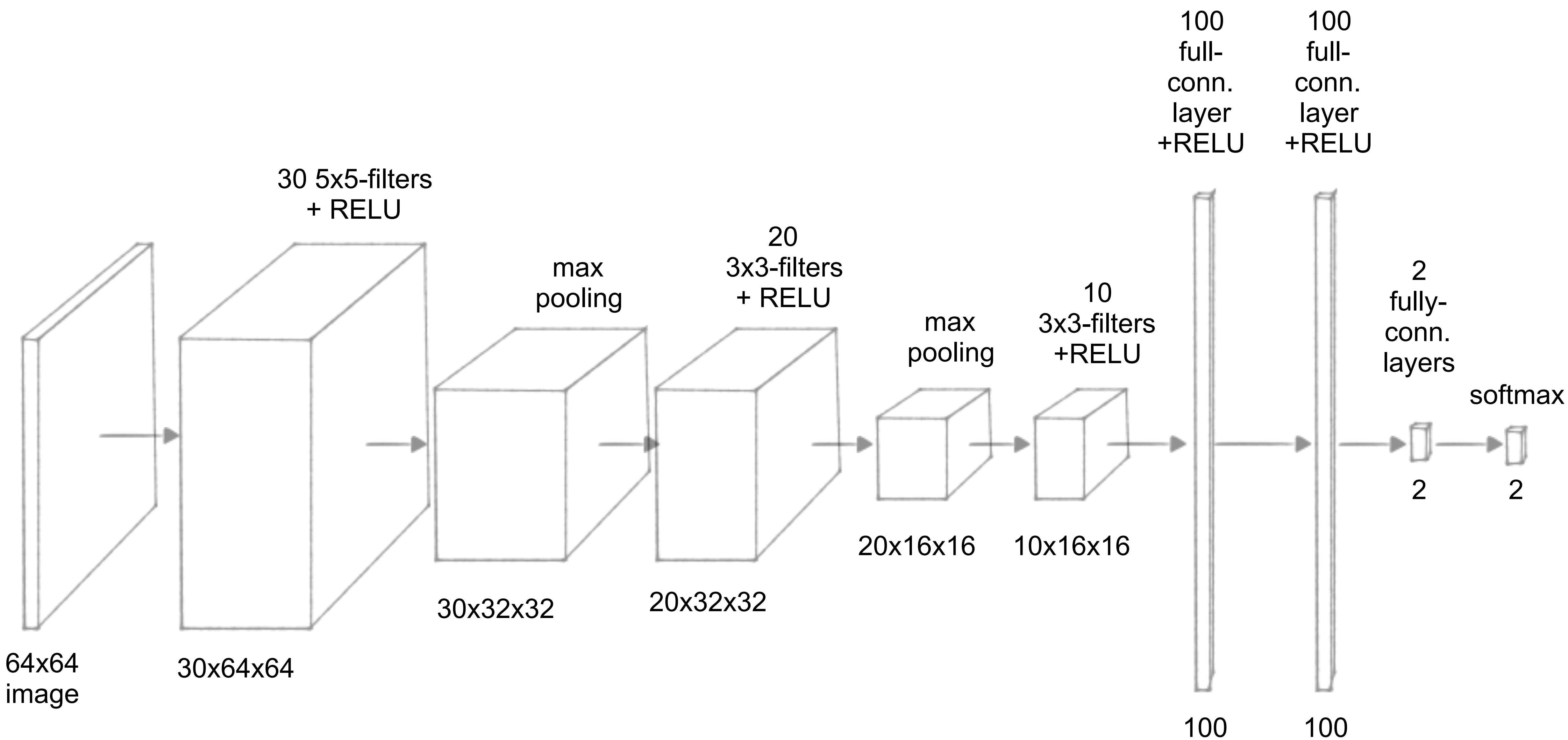
SSY097 - Image Analysis

Lecture 7 - Robust Model Fitting

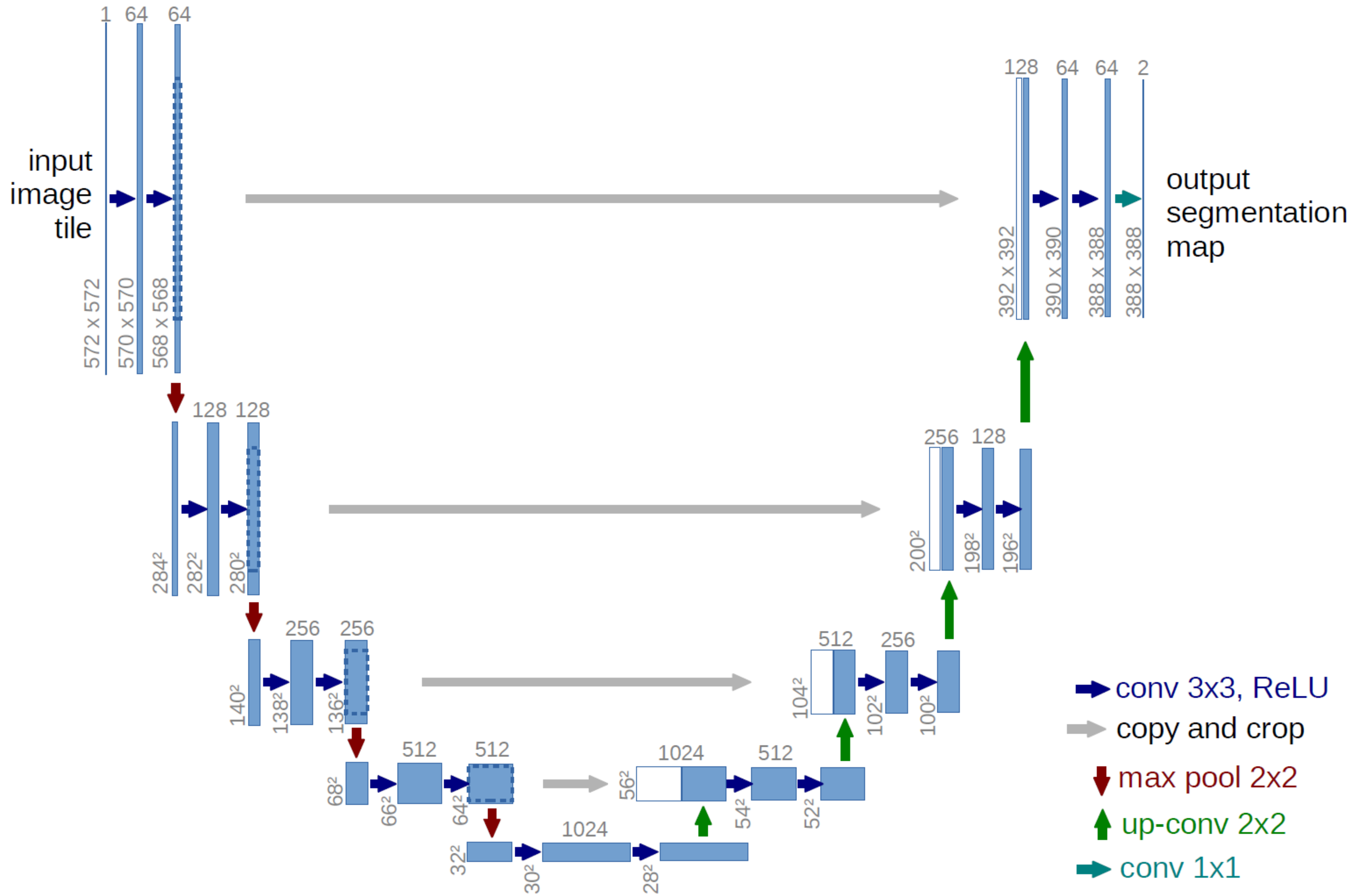
Torsten Sattler
(slides adapted from Olof Enqvist)

Last Lecture



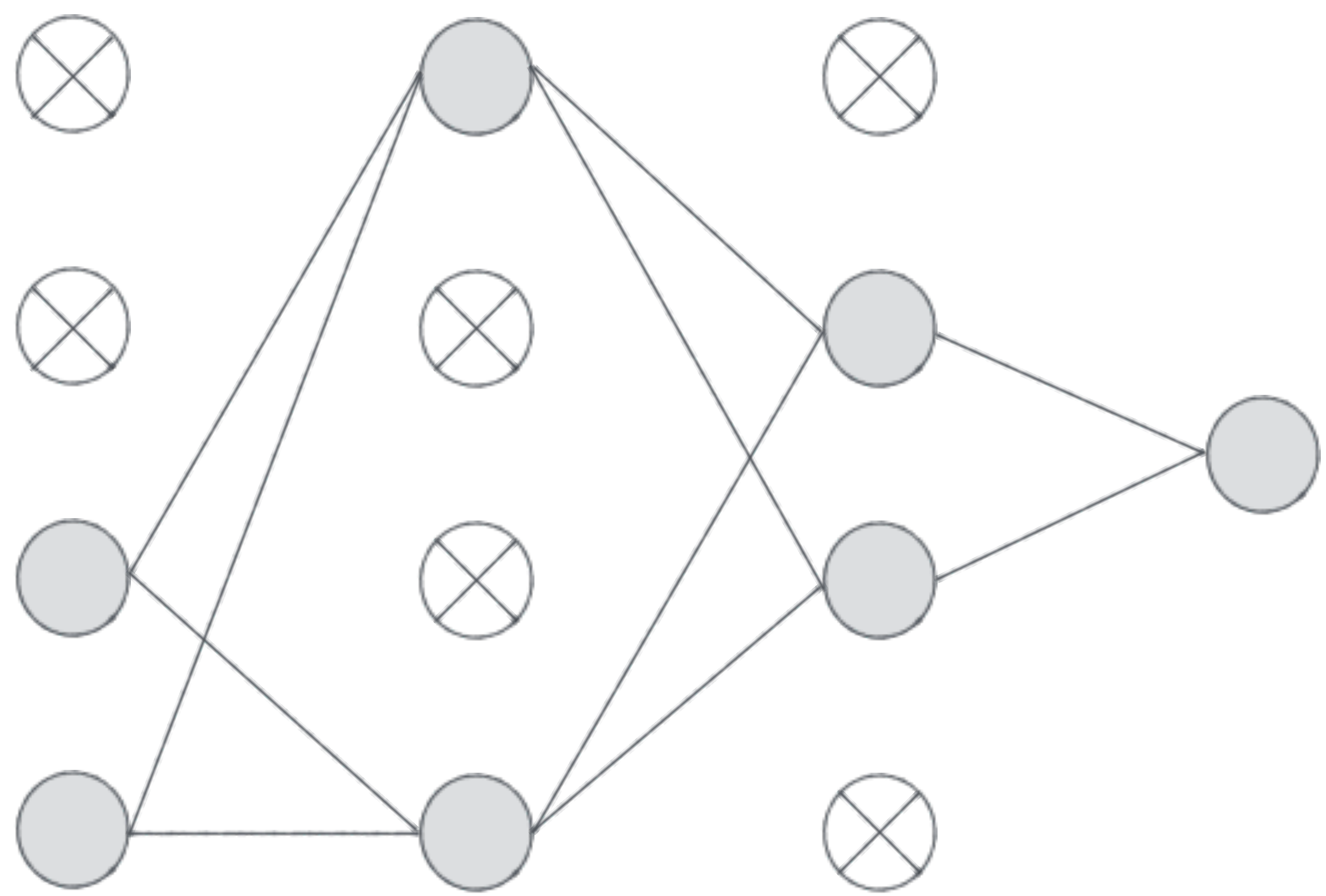
Convolutional Neural Networks

Last Lecture



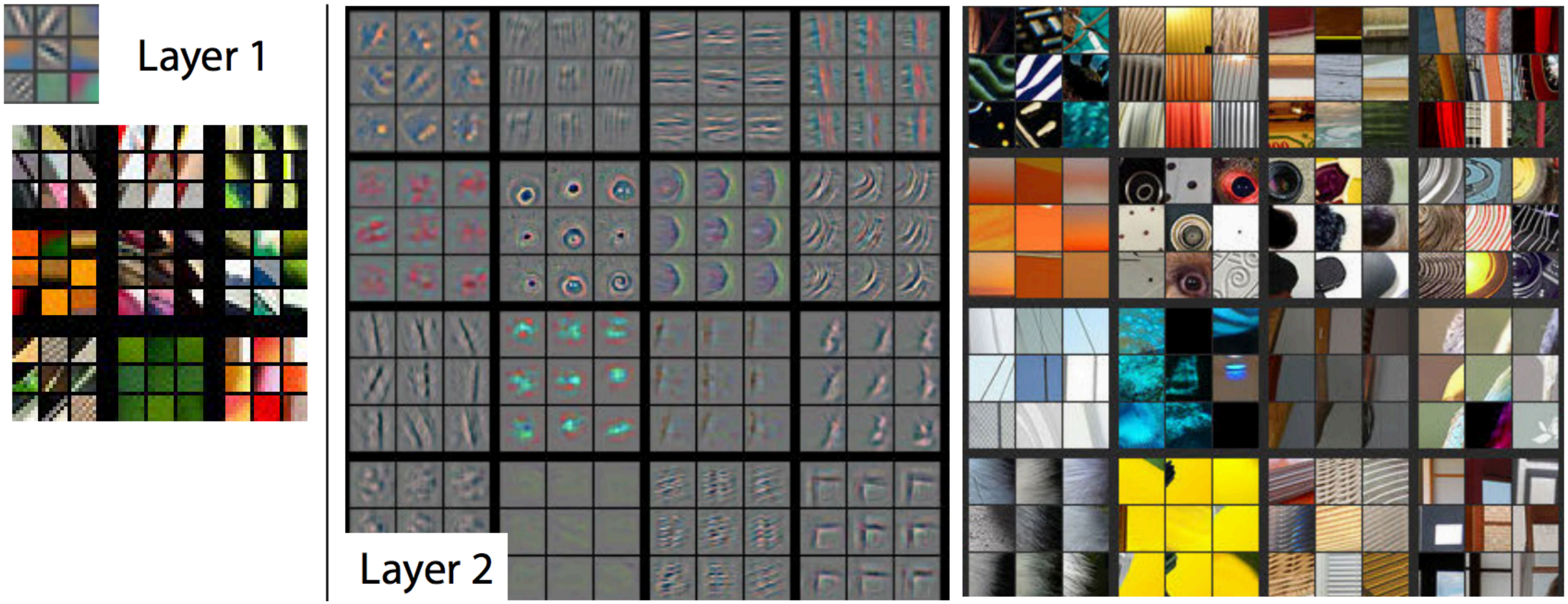
Fully Convolutional Neural Networks

Last Lecture



Preventing overfitting: Dropout

Last Lecture



Low-level features: corners, edges, ...

Transfer Learning

Today

- Model Fitting Basics
- Robust Model Fitting
- RANdom SAmple Consensus (RANSAC)

Model Fitting Basics

SIFT Features

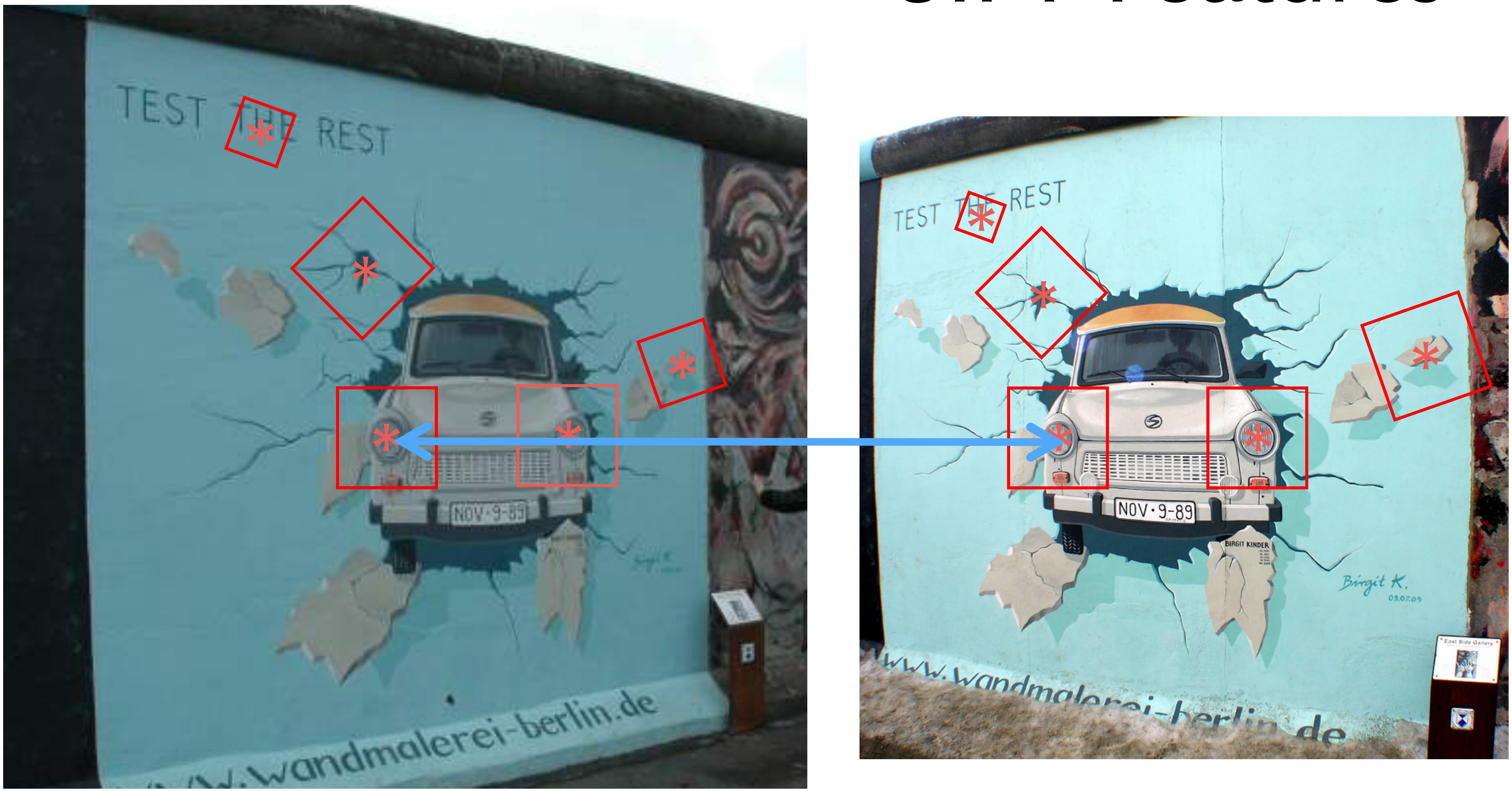


Image Stitching

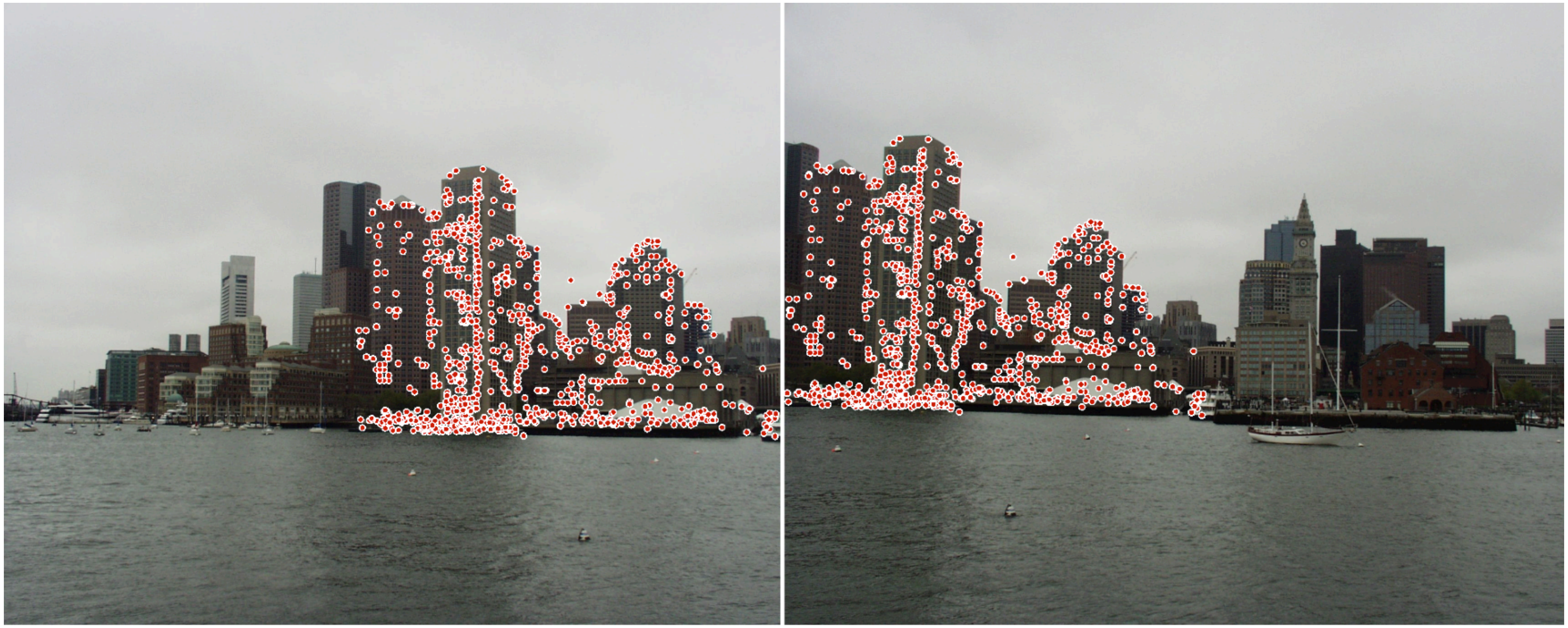
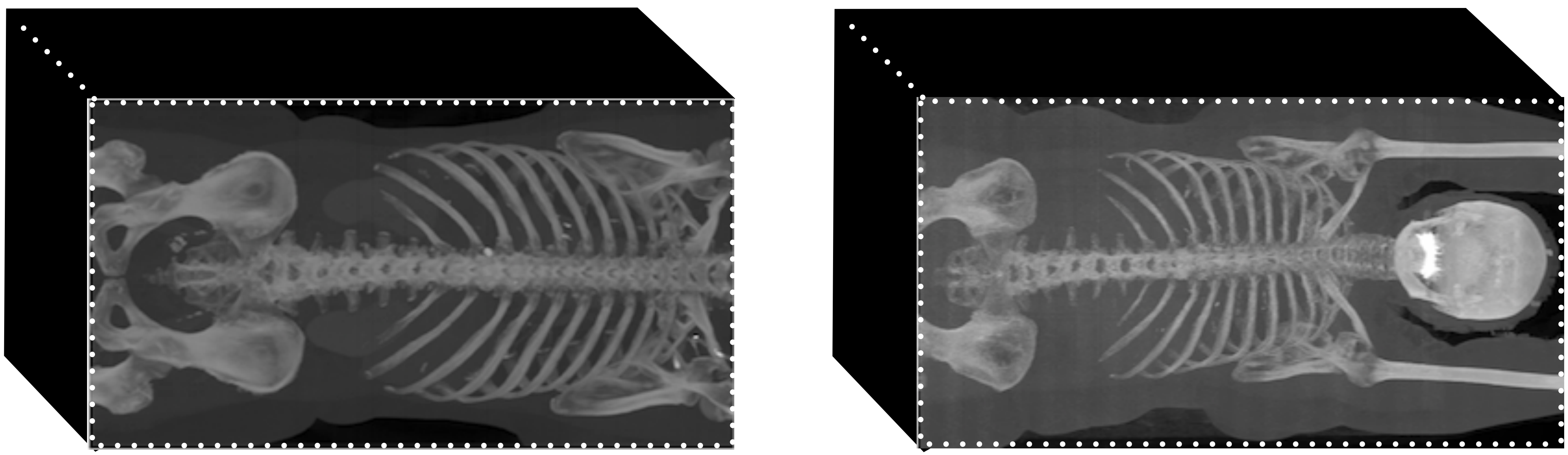


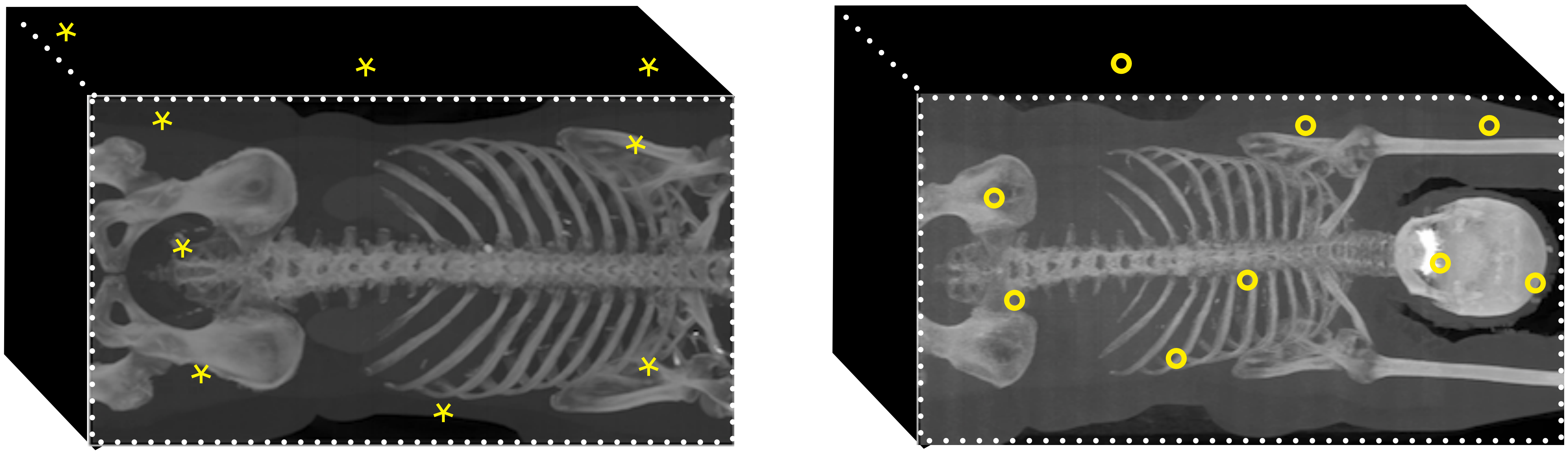
Image Stitching

Registration

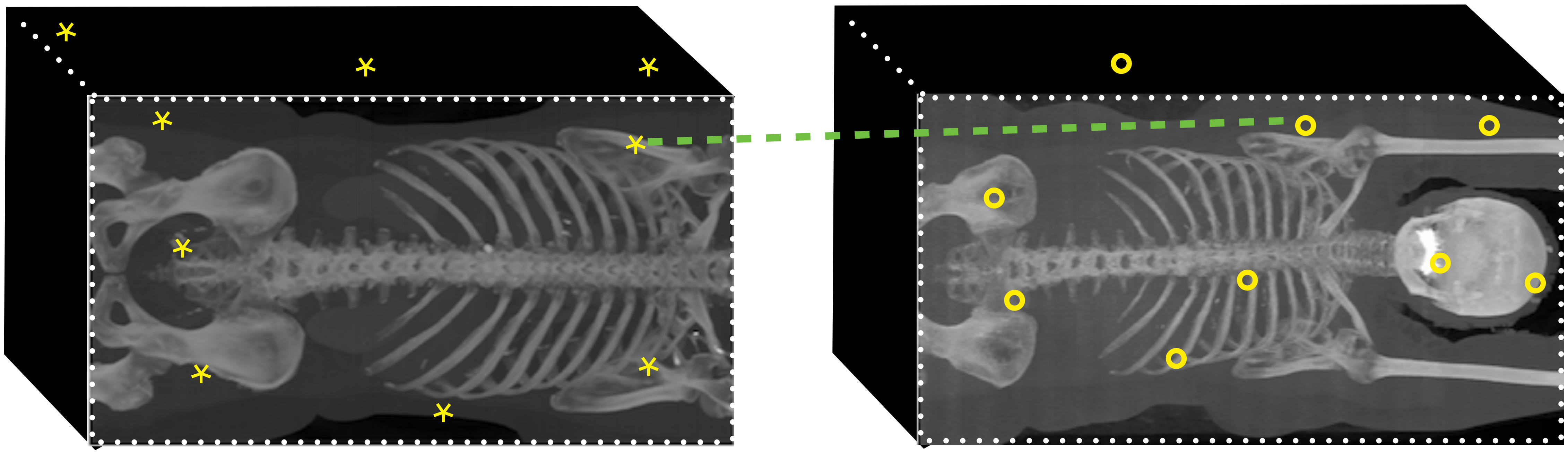


Estimate a transformation

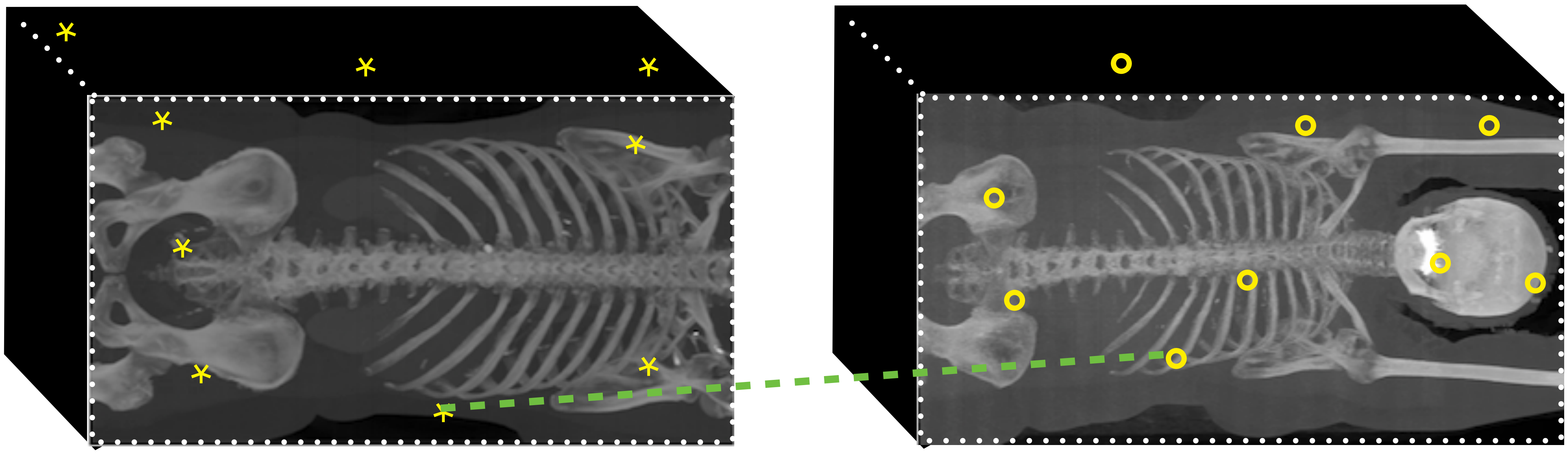
Registration



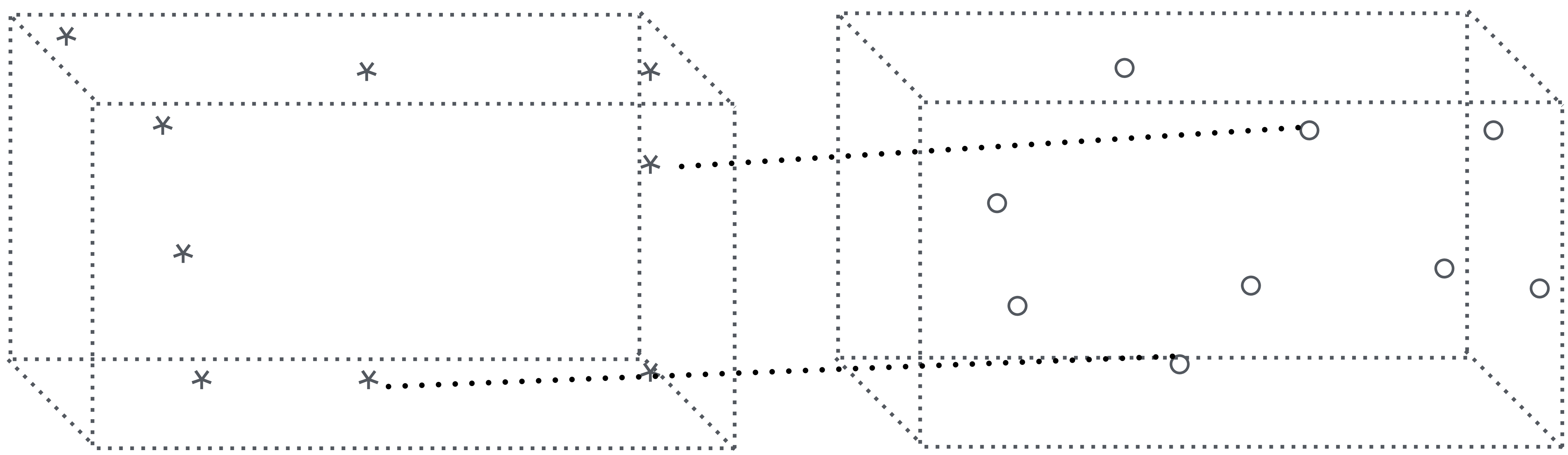
Registration



Registration

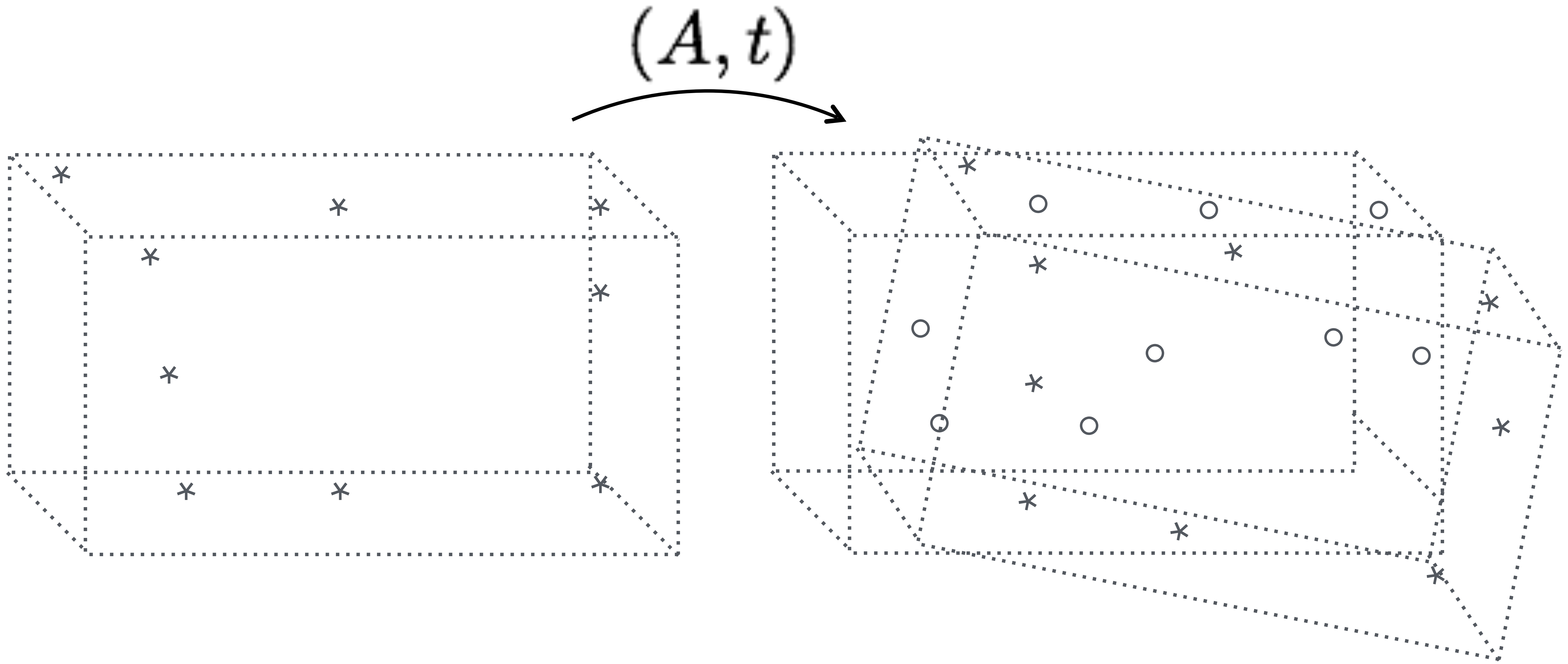


3D Point Set Registration

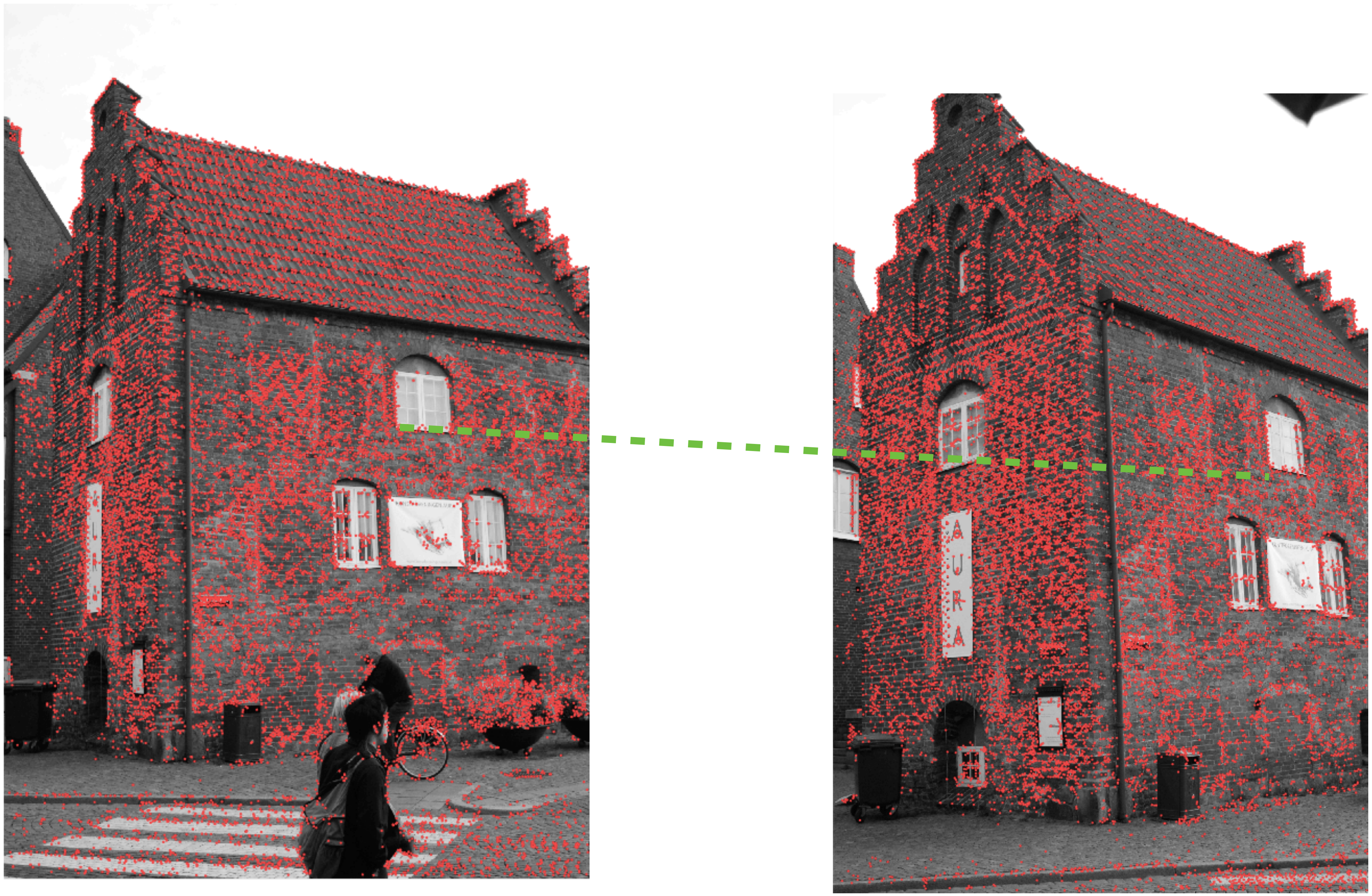


3D Point Set Registration

Find a consistent transformation



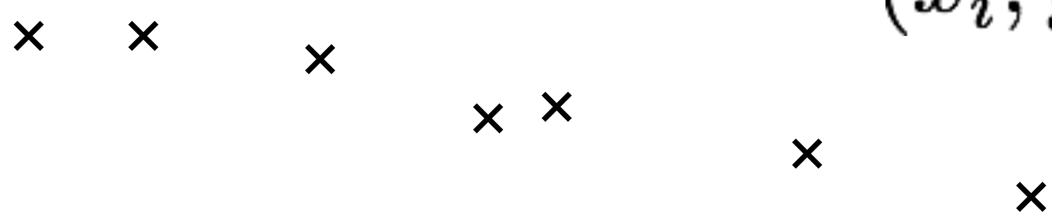
Estimating Camera Motion



Measurements - Models - Parameters

Measurements

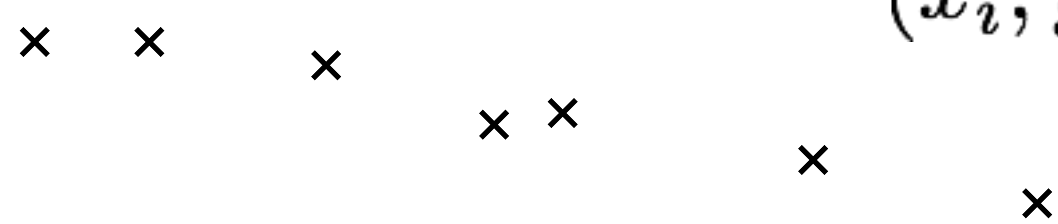
(x_i, y_i)



Measurements - Models - Parameters

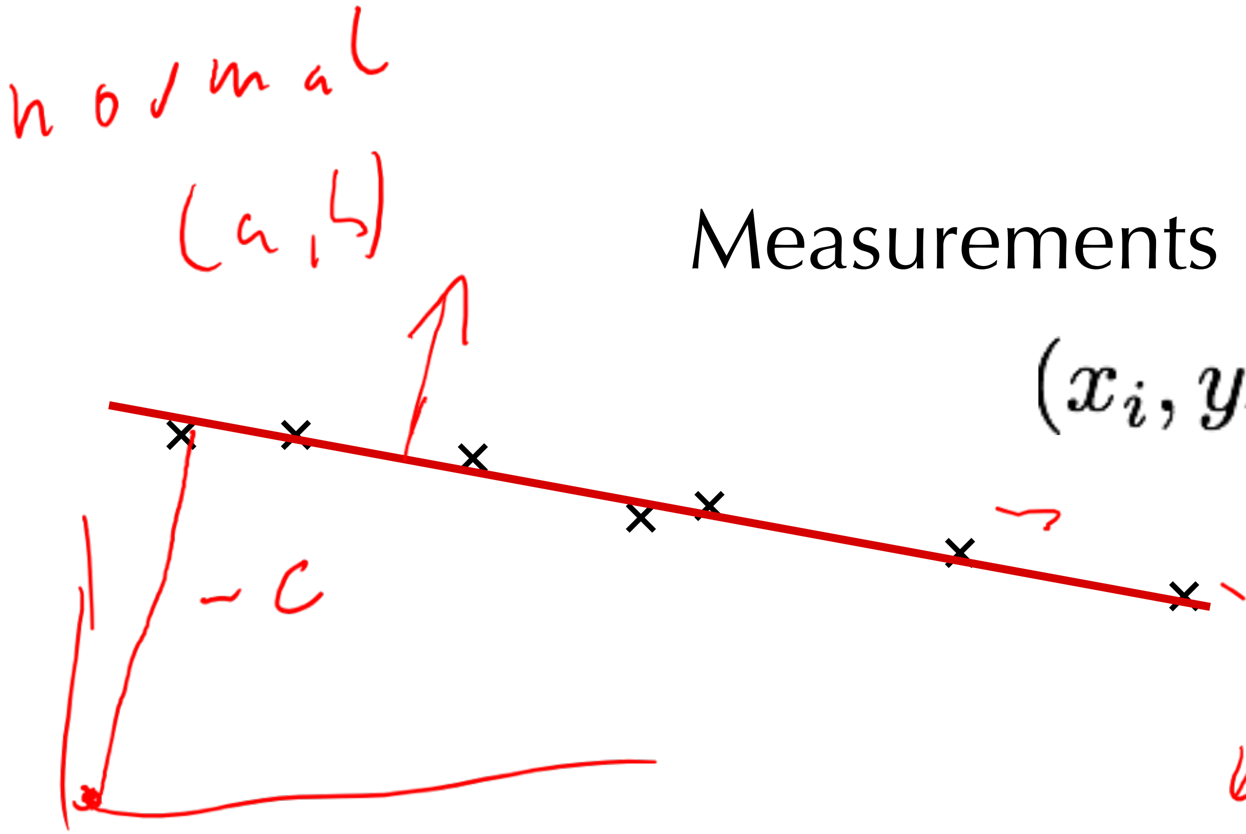
Measurements

(x_i, y_i)



Model: Points on a line: $y_i = kx_i + m$

Measurements - Models - Parameters



Measurements

(x_i, y_i)

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = z$$

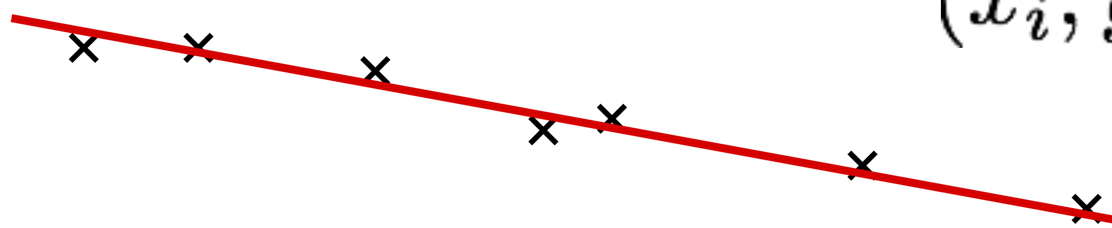
explicit

Model: Points on a line: $y_i = kx_i + m$

implicit representation

Good formulation: $ax_i + by_i + c = 0$

Measurements - Models - Parameters



Model: Points on a line: $y_i = kx_i + m$

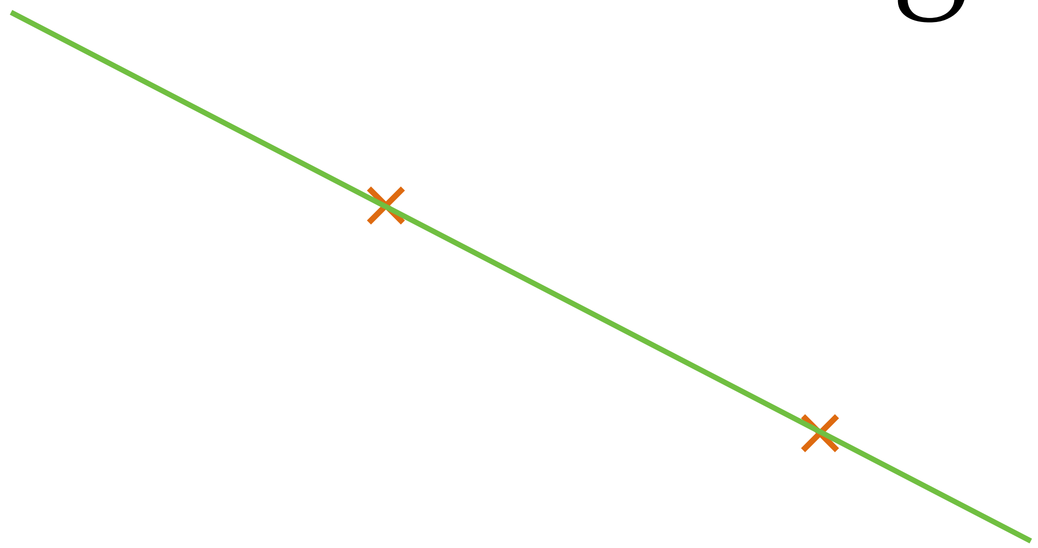
Good formulation: $ax_i + by_i + c = 0$

Model fitting : Estimate a, b and c from measurements

$$\theta = (a, b, c)$$

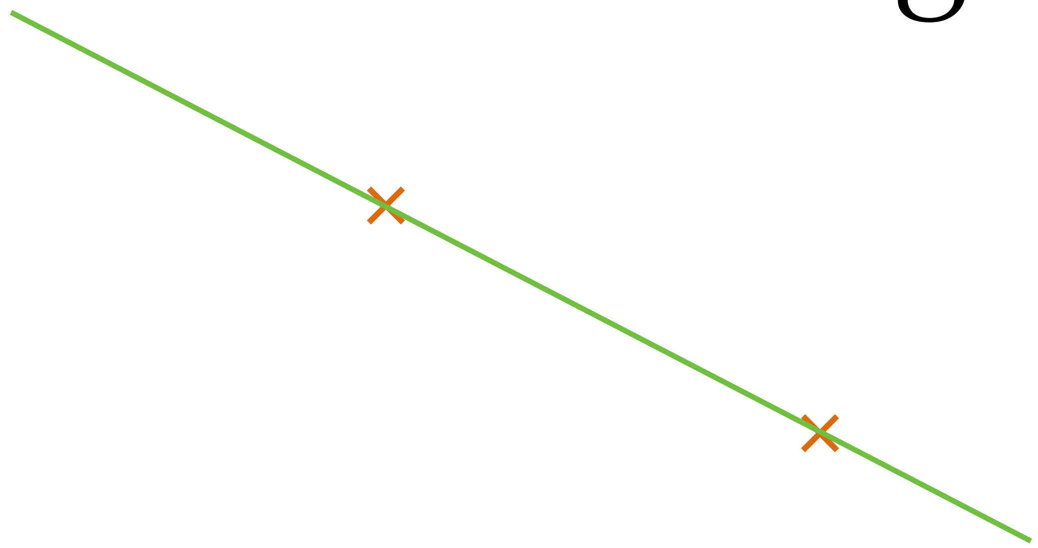
Line Fitting

Two measurements - exact solution

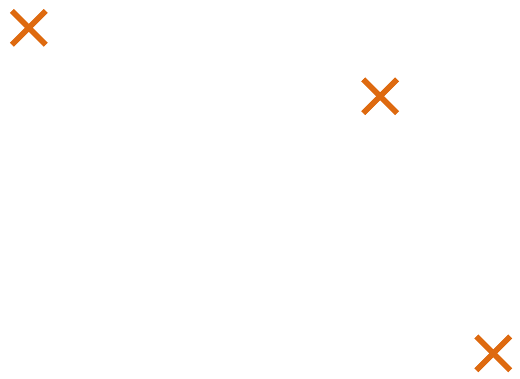


Line Fitting

Two measurements - exact solution

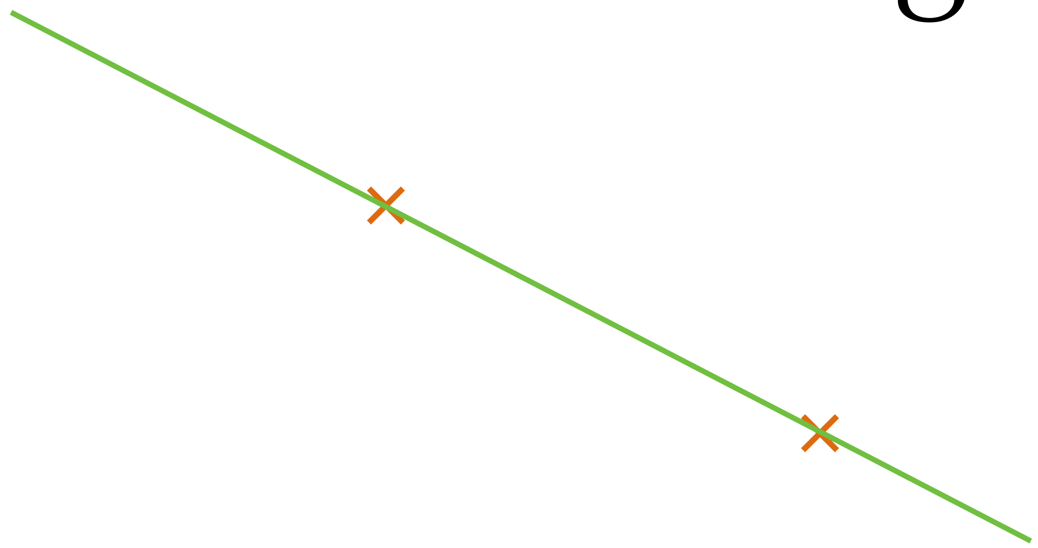


More than two?

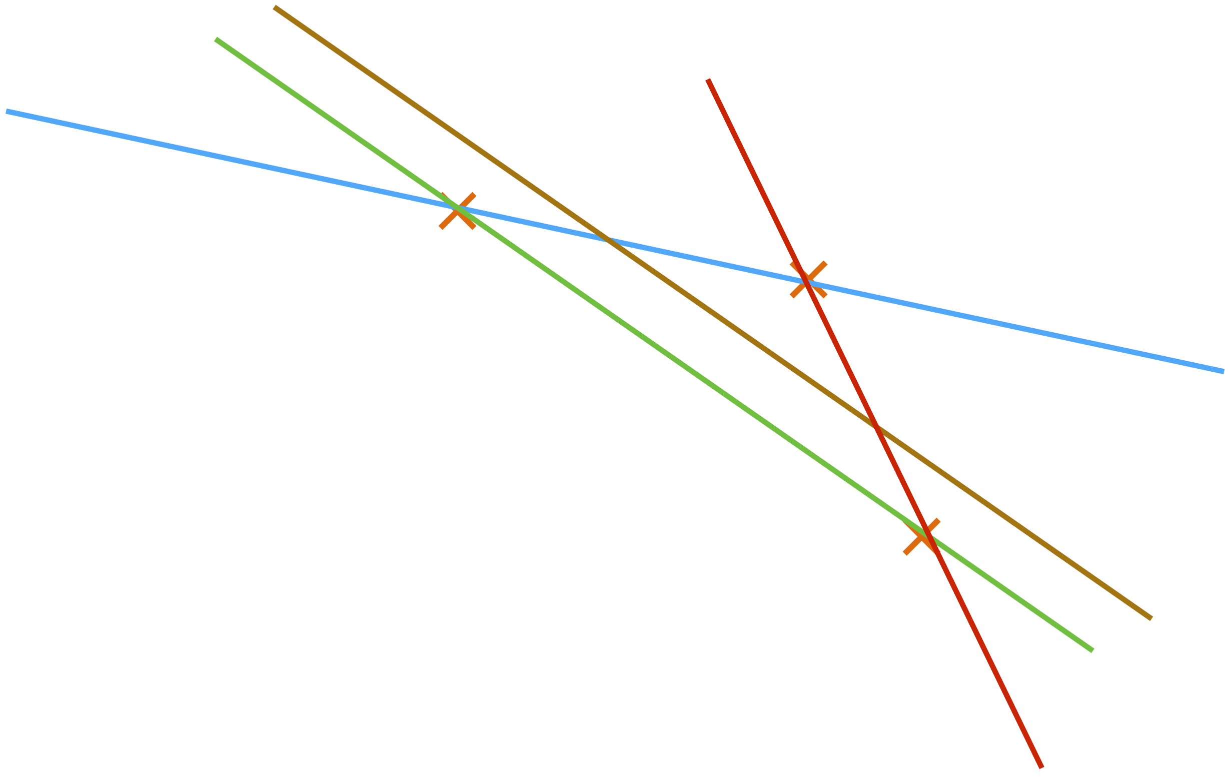


Line Fitting

Two measurements - exact solution



More than two?



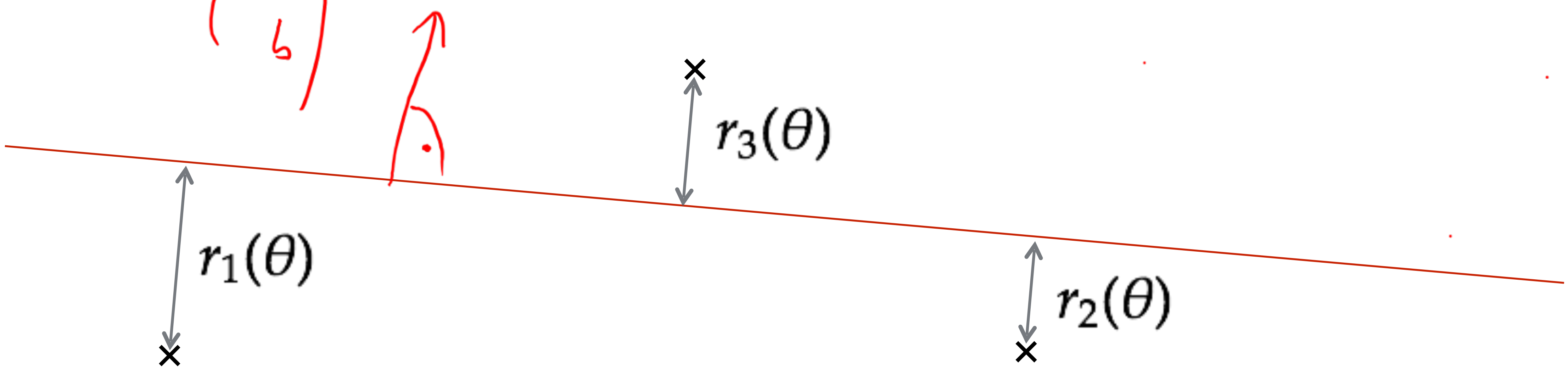
Residuals

$$r_i(\theta) = \left| \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + c \right|$$

scale $\theta = (a, b, c)$ s.t.

$a^2 + b^2 = 1$

$\begin{pmatrix} a \\ b \end{pmatrix}$



affine transformation $A \begin{pmatrix} x \\ y \end{pmatrix} + t$

$$\left\| \begin{pmatrix} x \\ y \end{pmatrix} - \left(A \begin{pmatrix} x \\ y \end{pmatrix} + t \right) \right\|$$

Model Fitting Problem

$$\theta = (a, b, c) \longrightarrow r_1(\theta),\ r_2(\theta),\ r_3(\theta),\ r_4(\theta), \ldots, r_n(\theta)$$

How do we choose an optimal set of parameters?

Why Do We Have Residuals?

Measurements are not exact.
They are affected by Gaussian noise!

Carl Friedrich Gauss

Why Do We Have Residuals?

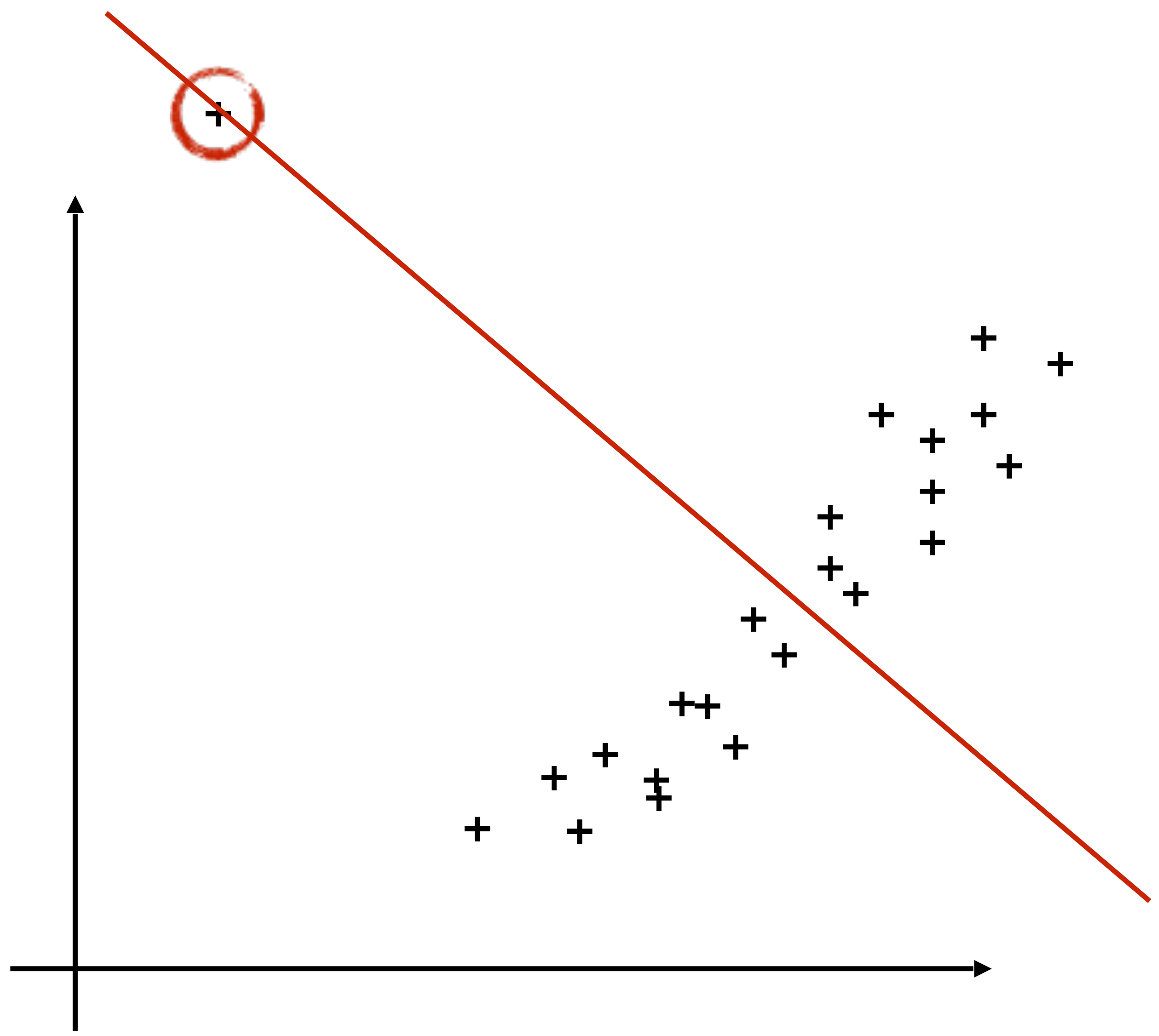
Measurements are not exact.
They are affected by Gaussian noise!

Best thing to do: Least squares!

$$\min_{\theta} \sum r_i(\theta)^2$$

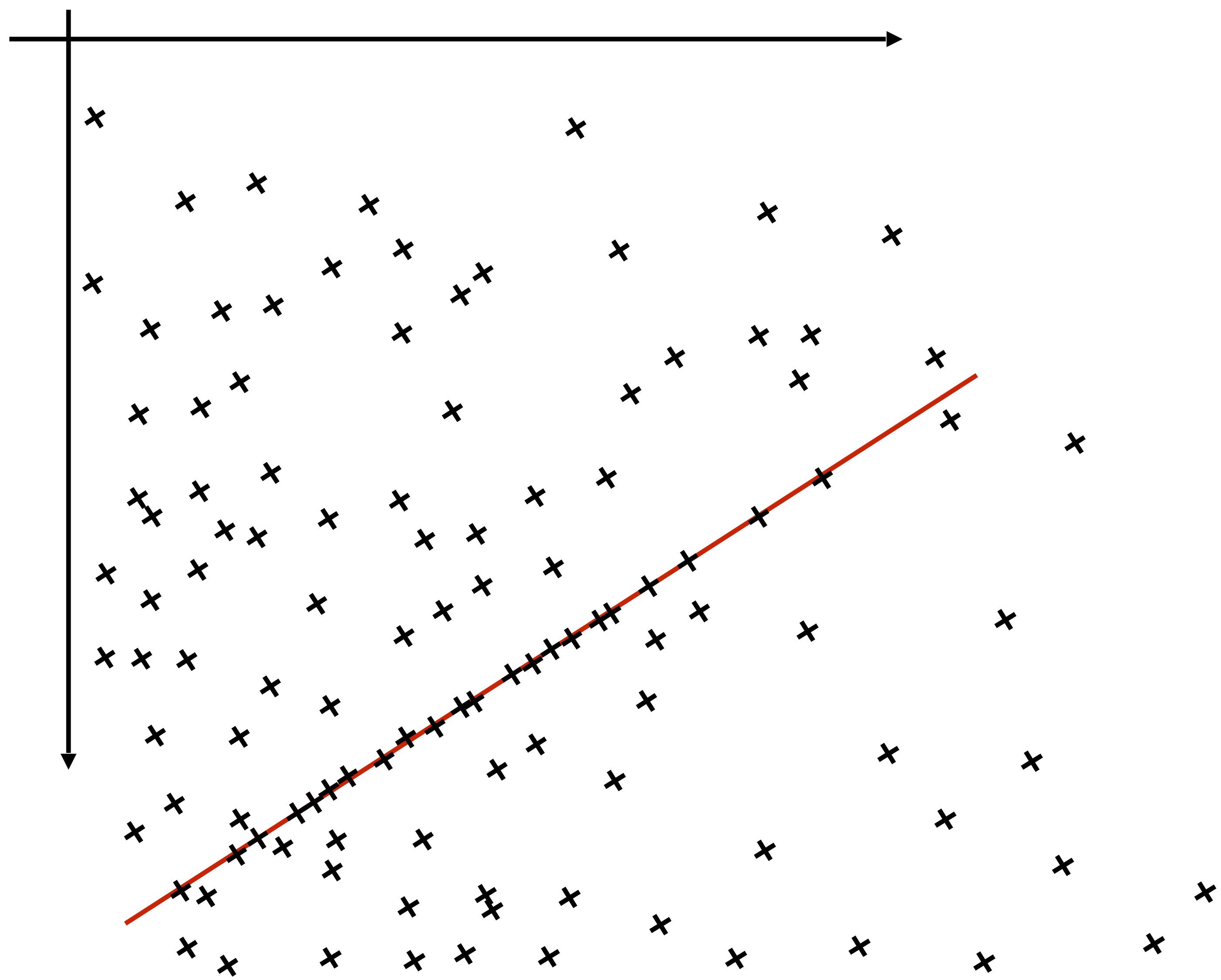
Carl Friedrich Gauss

Outliers



Rare unexpected measurements that don't fit this model.

Outliers in Image Analysis



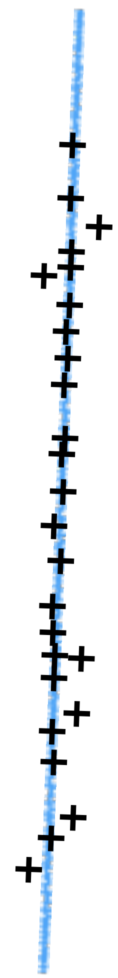
Frequent expected measurements that don't give useful information.

Outliers in Image Analysis

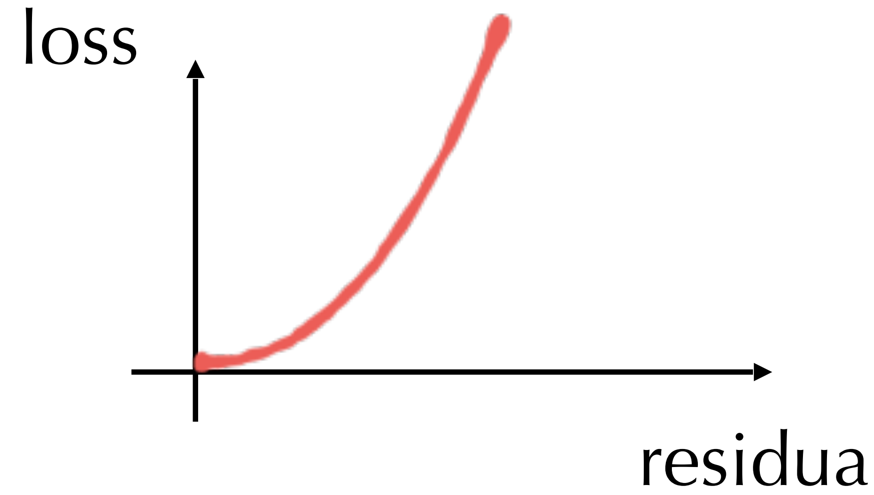
Are We Doing Something Wrong?

Robust Model Fitting

Least-Squares Fitting

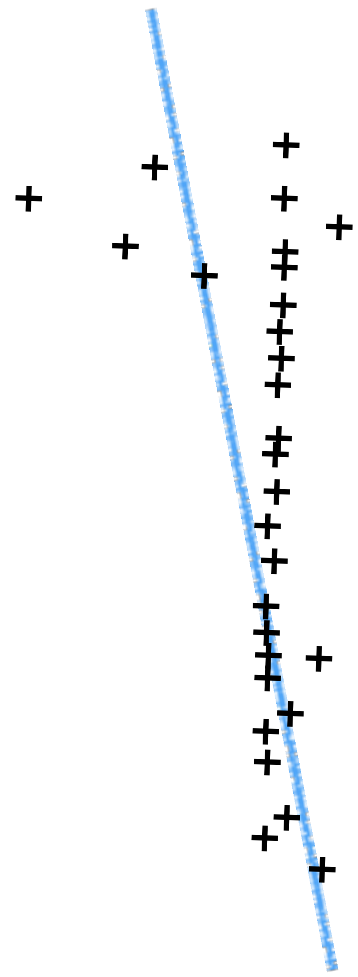


ℓ^2-solution

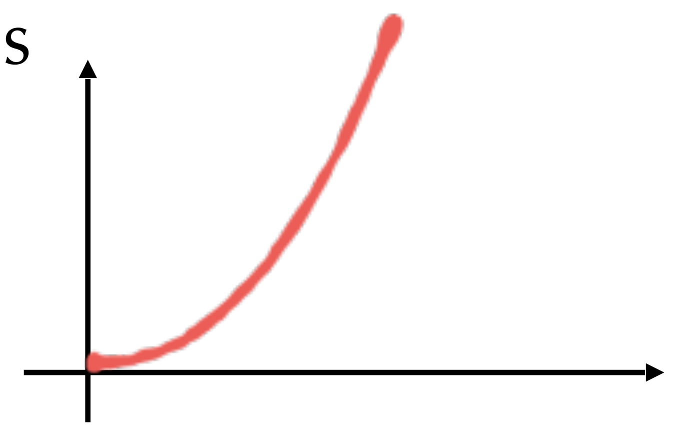


$$\min_{\theta} \sum r_i(\theta)^2$$

Least-Squares Fitting



ℓ^2-solution



$$\min_{\theta} \sum r_i(\theta)^2$$

Minimization under Gaussian Noise

- Assumption: Zero-mean, isotropic Gaussian noise on residuals

$$p(r_i(\theta)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{r_i(\theta)^2}{2\sigma^2}}$$

Minimization under Gaussian Noise

- Assumption: Zero-mean, isotropic Gaussian noise on residuals

$$p(r_i(\theta)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{r_i(\theta)^2}{2\sigma^2}}$$

- Maximum Likelihood Estimate: $\max_\theta \prod_i p(r_i(\theta))$

Minimization under Gaussian Noise

- Assumption: Zero-mean, isotropic Gaussian noise on residuals

$$p(r_i(\theta)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{r_i(\theta)^2}{2\sigma^2}}$$

- Maximum Likelihood Estimate: $\max_\theta \prod_i p(r_i(\theta))$

- Minimizing negative log-likelihood: $\min_\theta - \sum_i \log(p(r_i(\theta)))$

Minimization under Gaussian Noise

- Assumption: Zero-mean, isotropic Gaussian noise on residuals

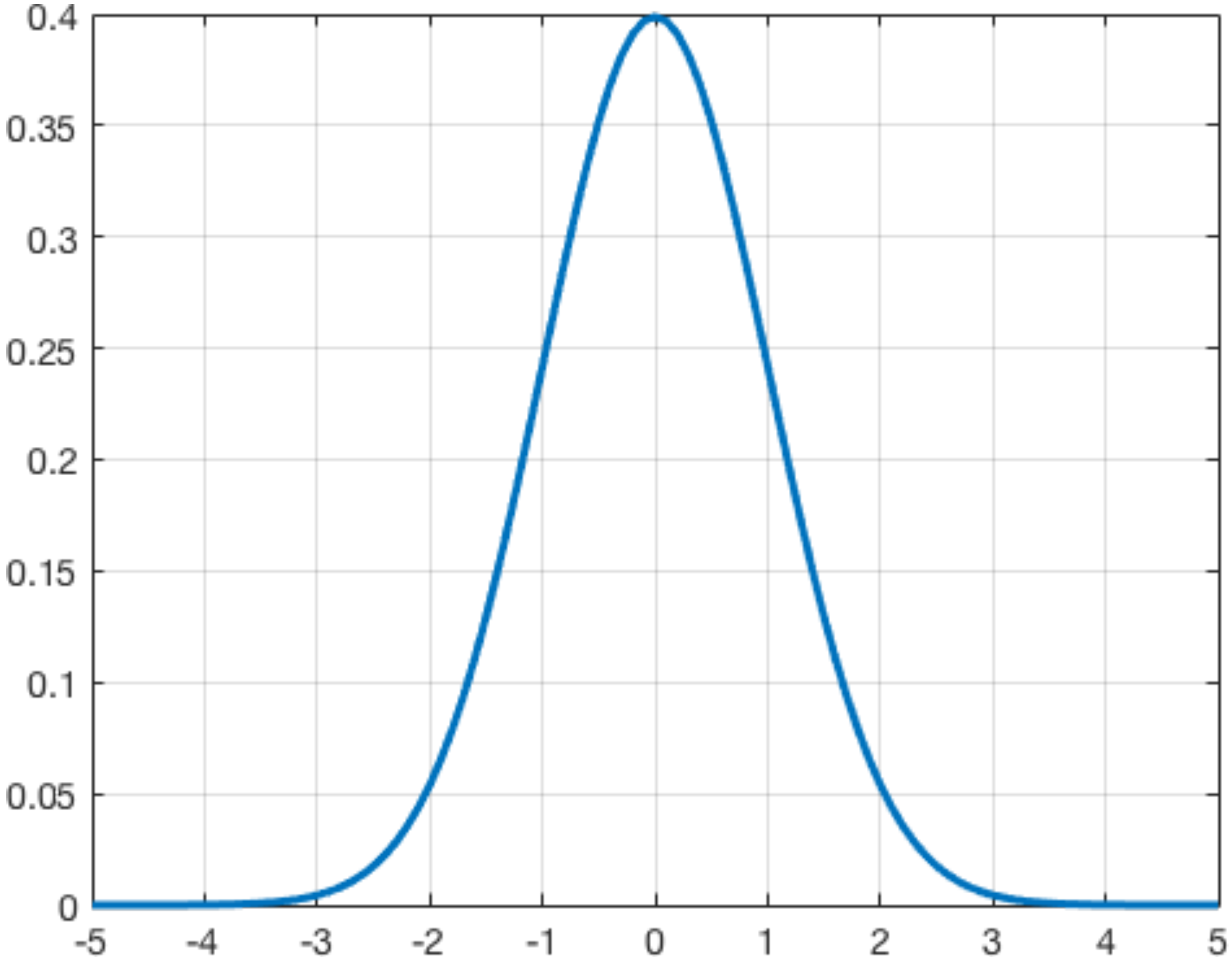
$$p(r_i(\theta)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{r_i(\theta)^2}{2\sigma^2}}$$

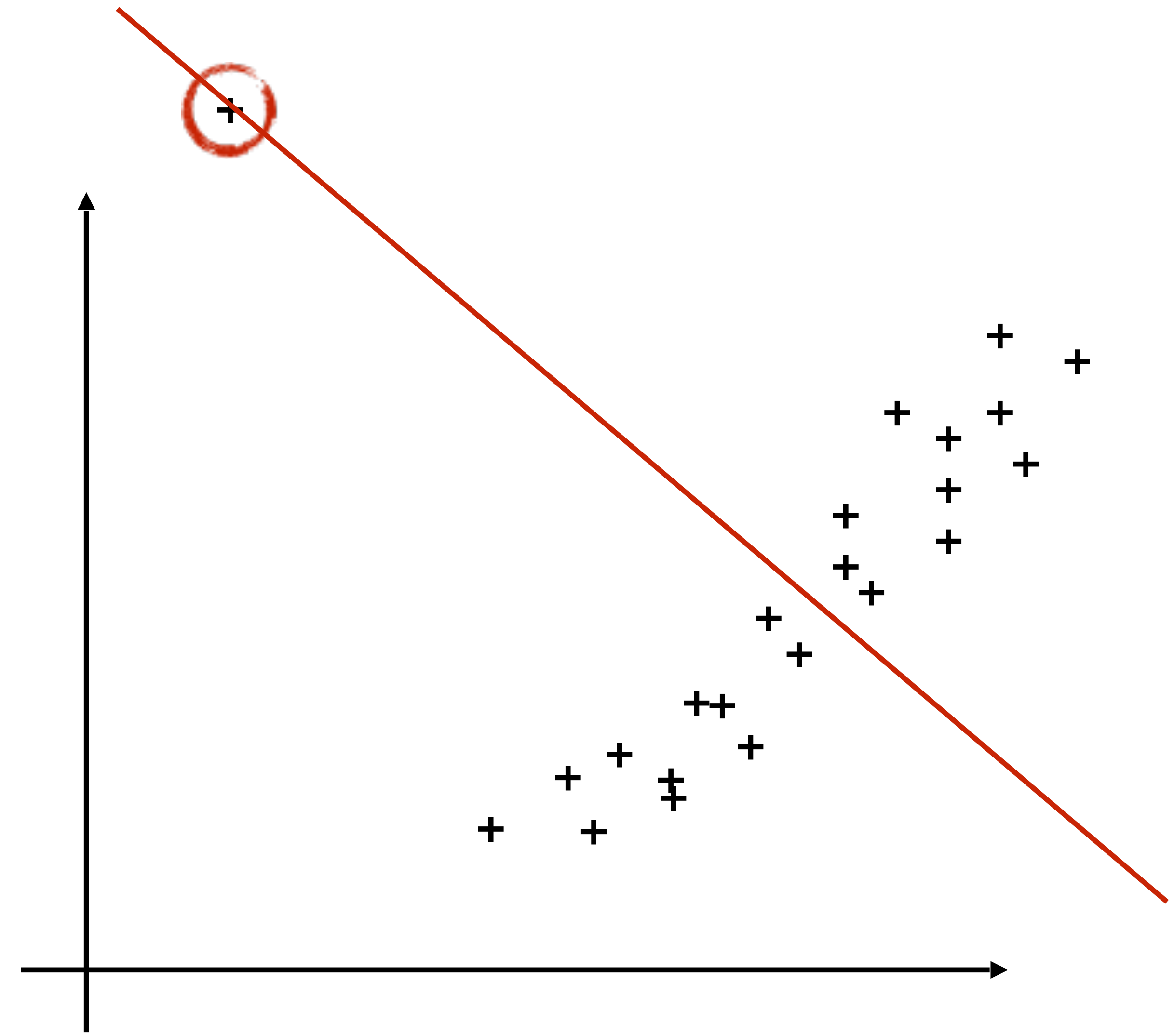
- Maximum Likelihood Estimate: $\max_{\theta} \prod_i p(r_i(\theta))$

- Minimizing negative log-likelihood: $\min_{\theta} - \sum_i \log(p(r_i(\theta)))$

- Equivalent to least-squares fitting: $\min_{\theta} \sum_i r_i(\theta)^2$

Minimization under Gaussian Noise

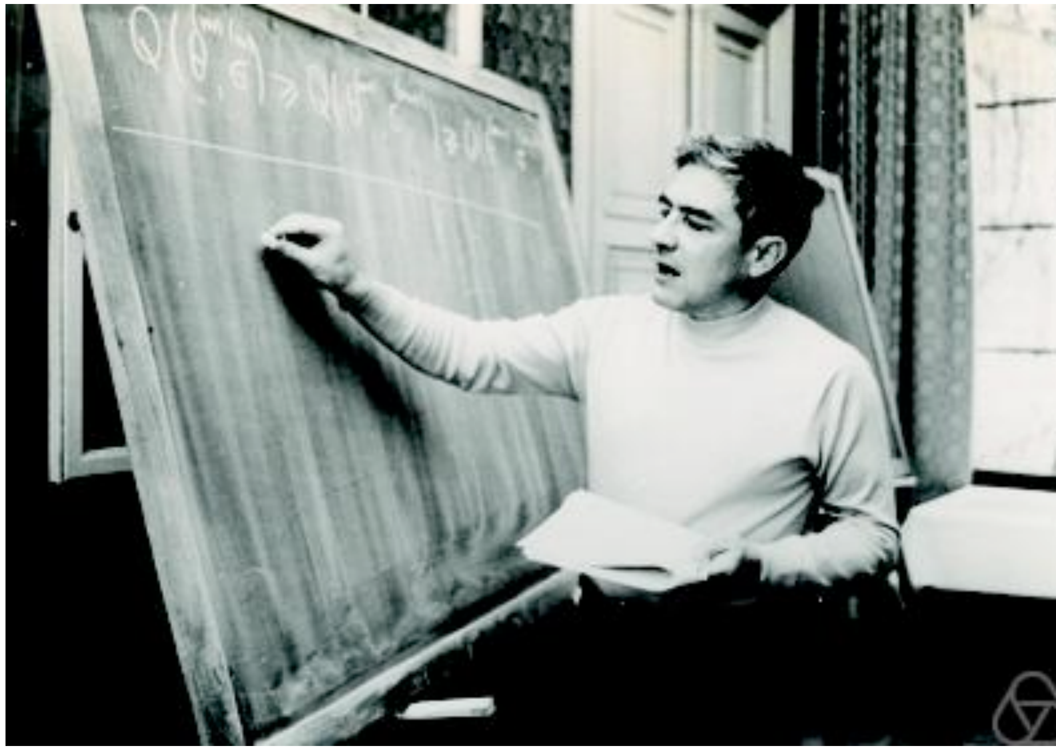




Rare unexpected measurements that don't fit Gaussian noise model.

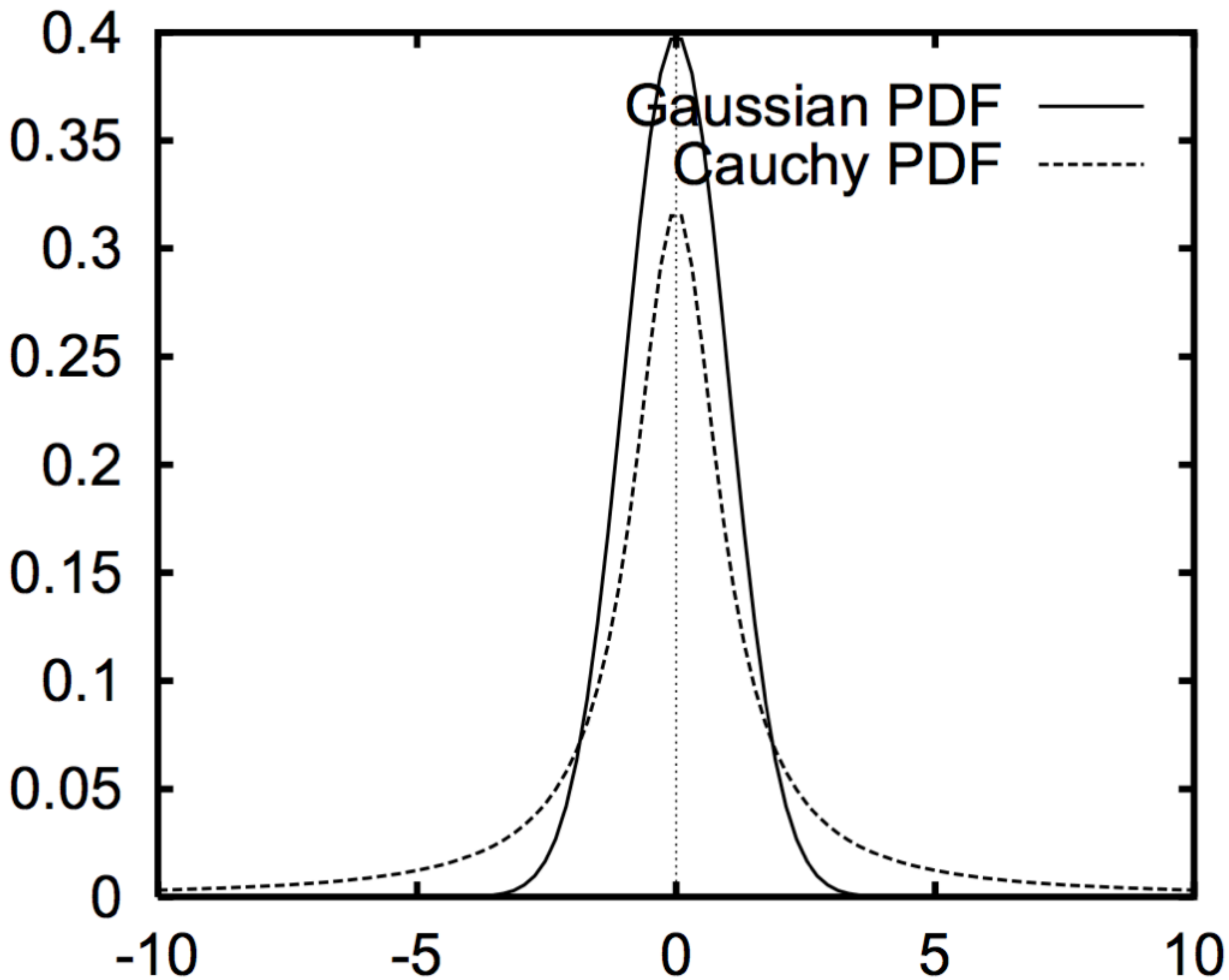
Why Do We Have Residuals?

There are all kinds of noise!
Use robust loss functions!



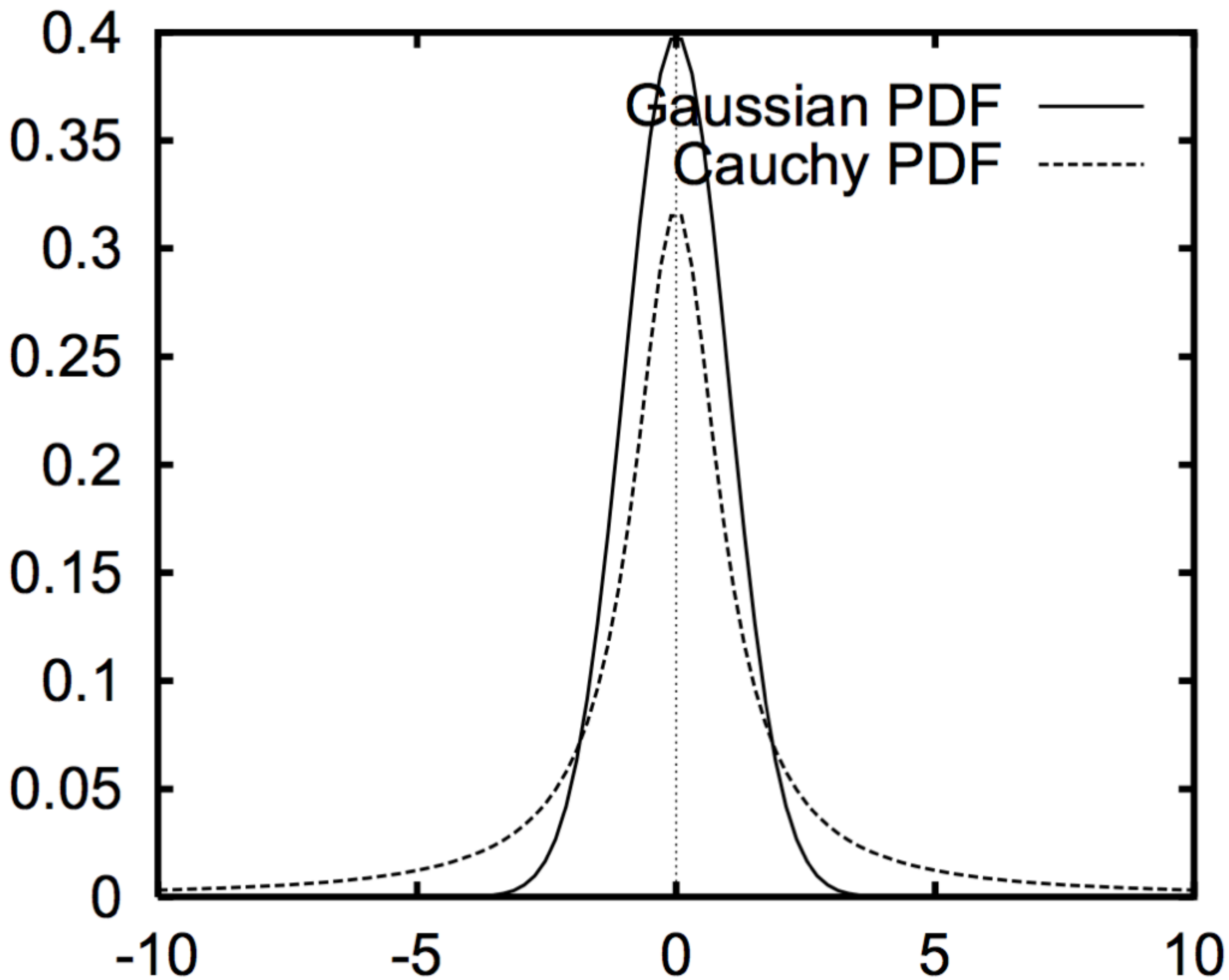
Peter Huber

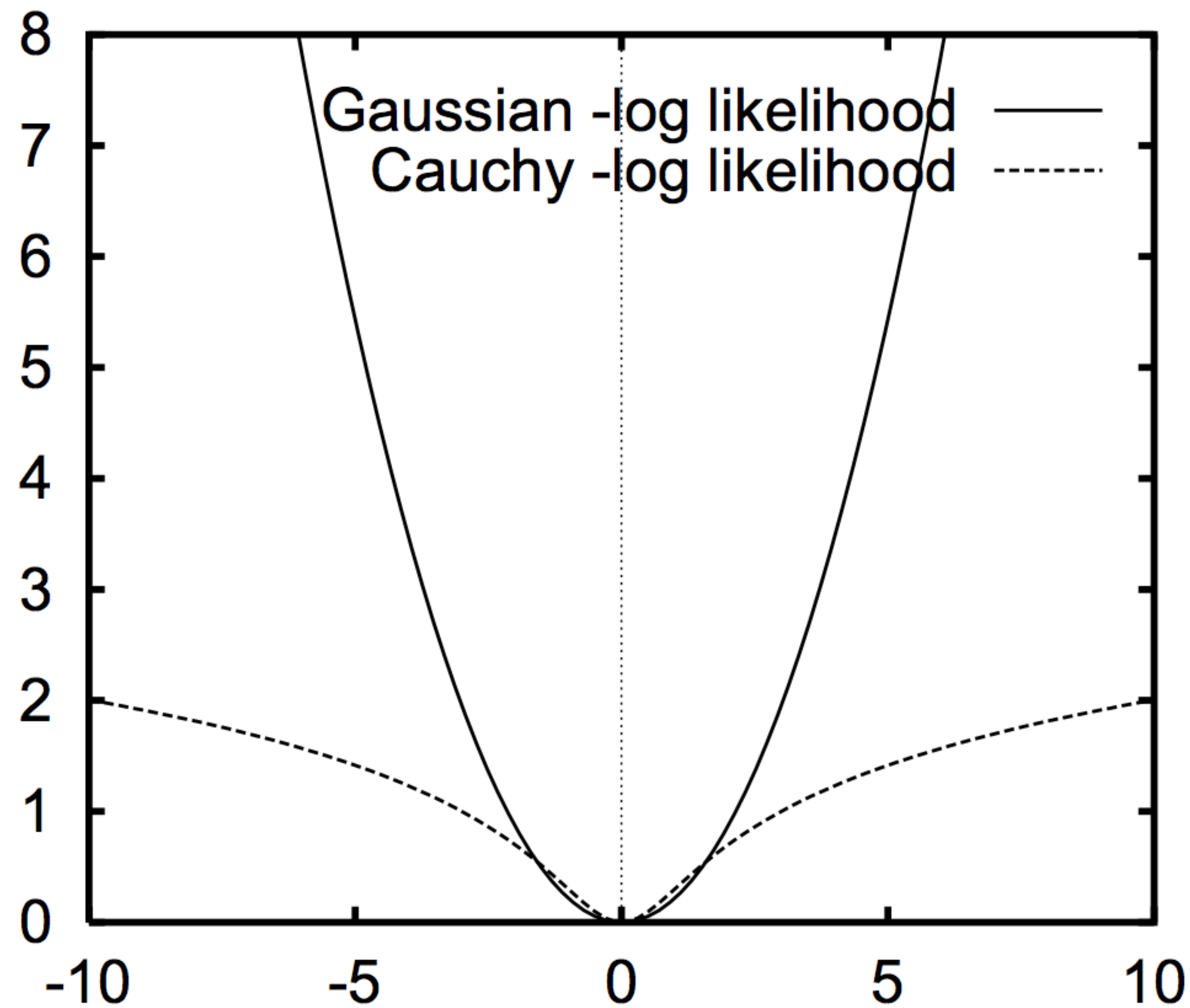
Robust Loss Functions



Cauchy distribution: $p(x) = \dfrac{1}{\pi(1+x^2)}$

Robust Loss Functions





Cauchy distribution: $p(x) = \dfrac{1}{\pi(1+x^2)}$

Minimizing Robust Loss Functions

- Replace

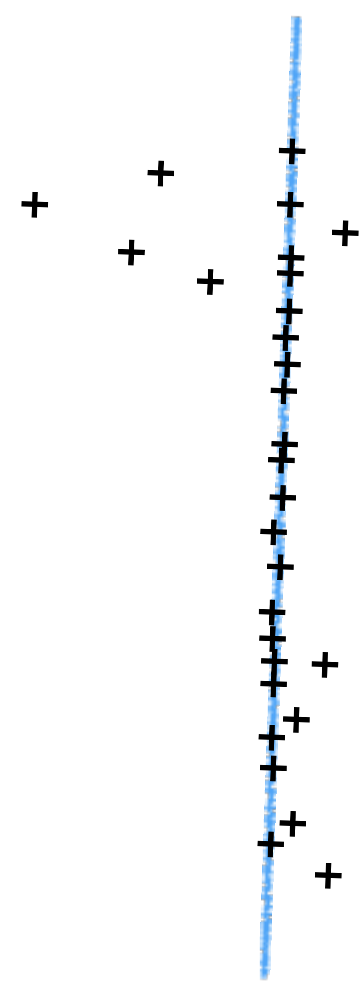
$$\min_{\theta} \sum_i r_i(\theta)^2$$

robust loss function

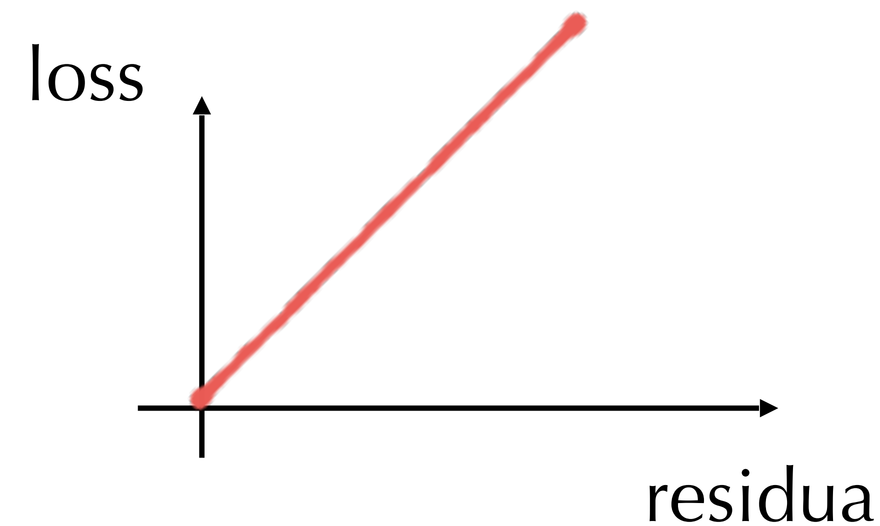
- with

$$\min_{\theta} \sum_i f(|r_i(\theta)|)$$

Least Absolute Residual

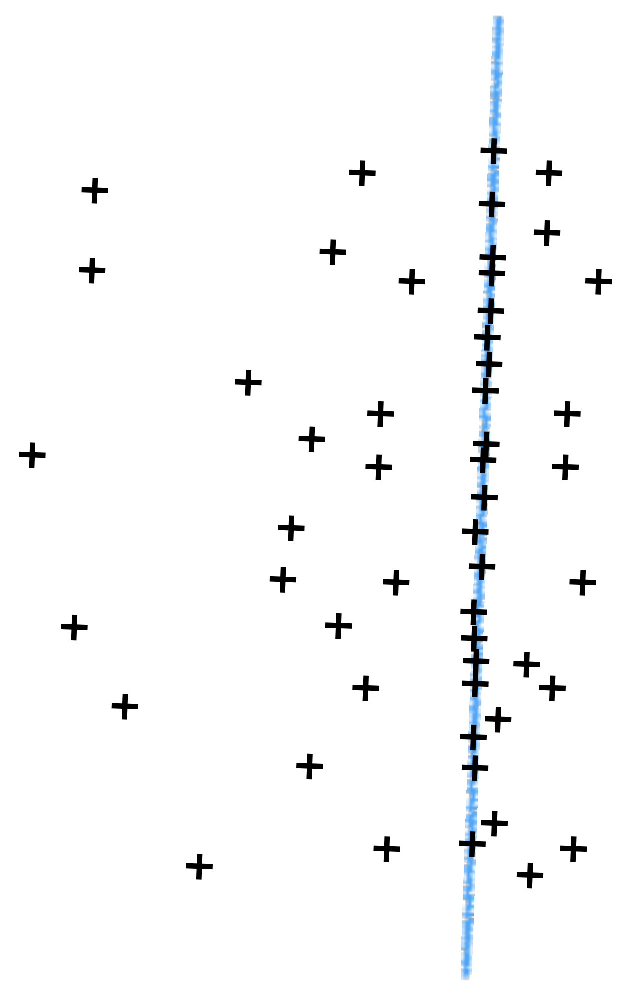


ℓ^1-solution

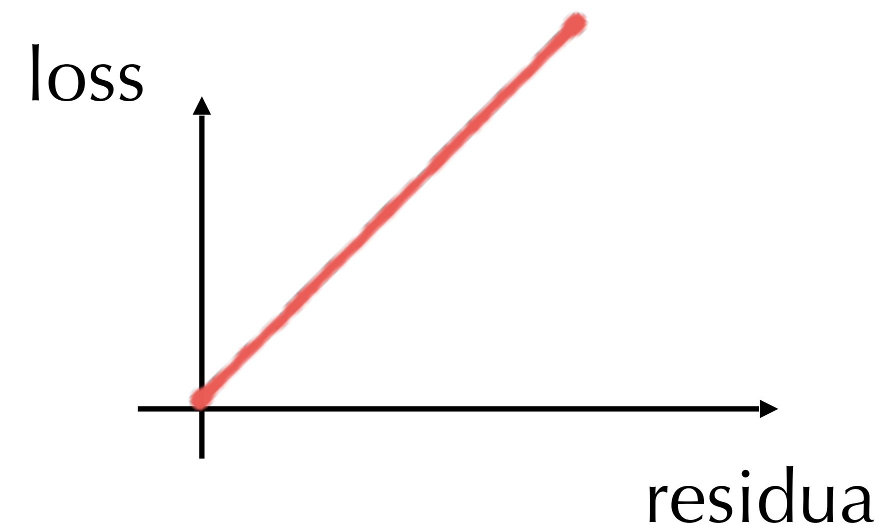


$$\min_{\theta} \sum |r_i(\theta)|$$

Least Absolute Residual

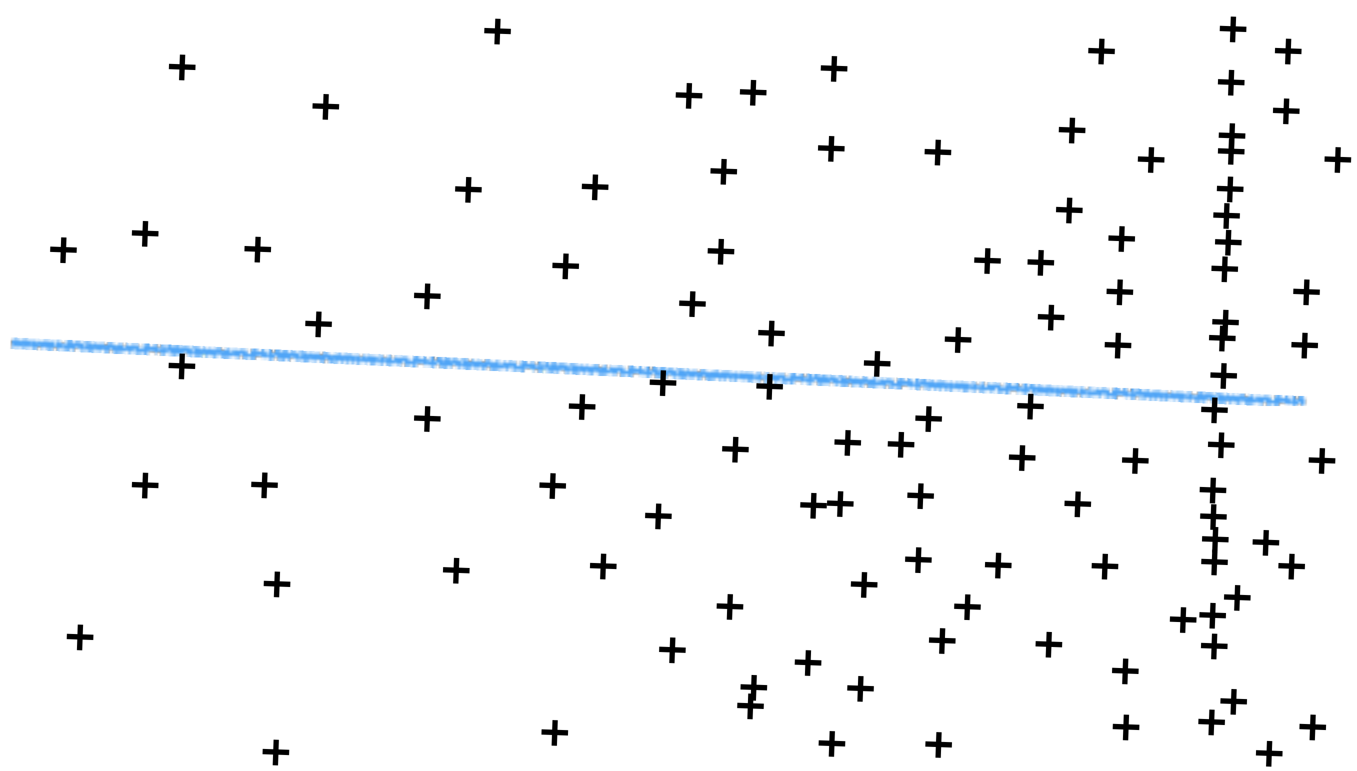


ℓ^1-solution

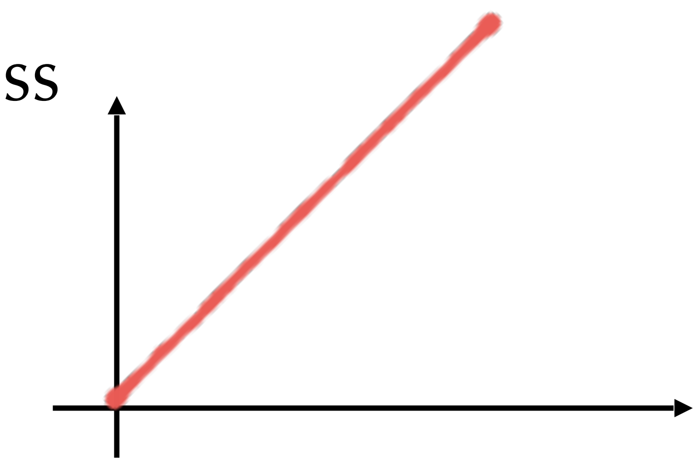


$$\min_\theta \sum |r_i(\theta)|$$

Least Absolute Residual

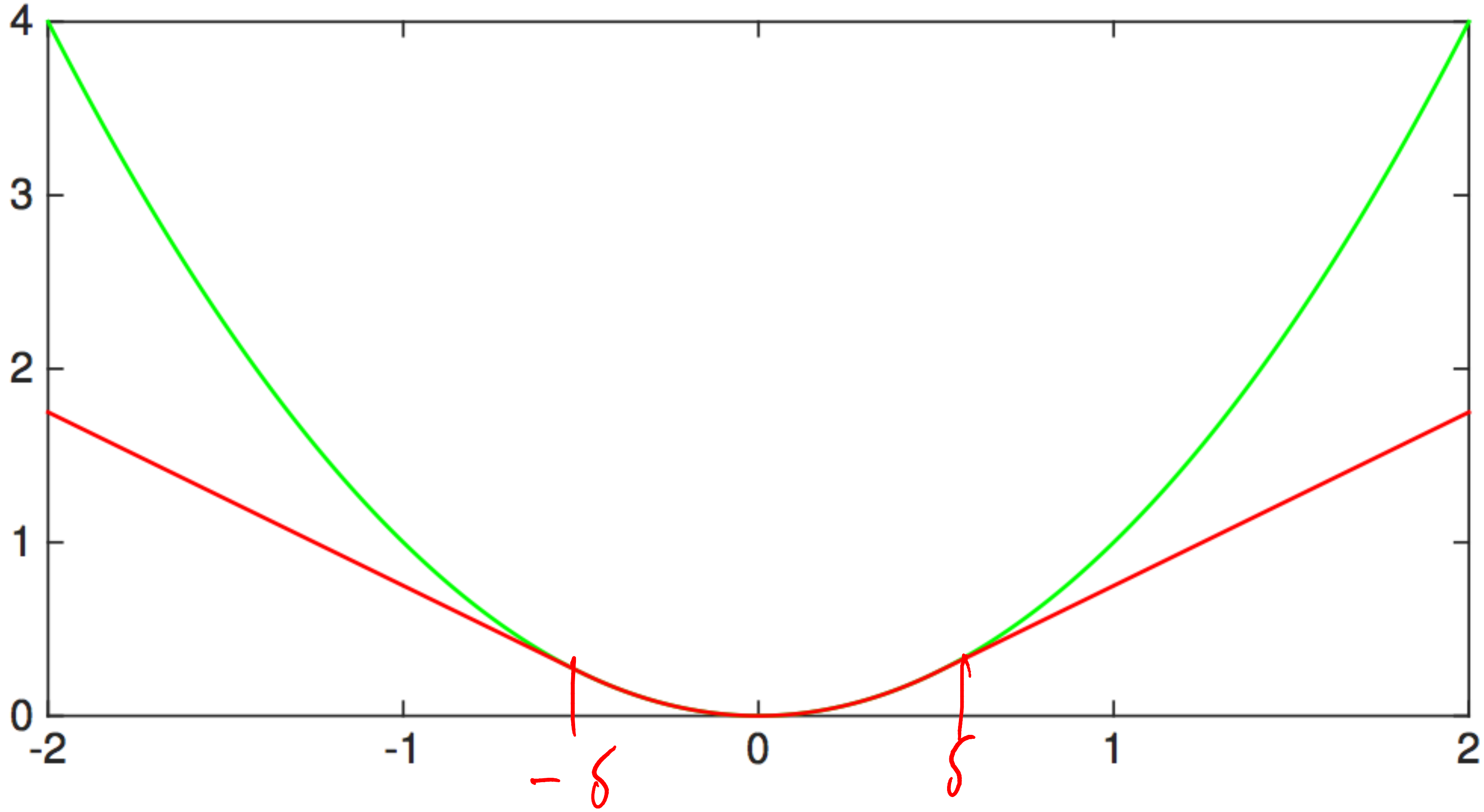


ℓ^1-solution



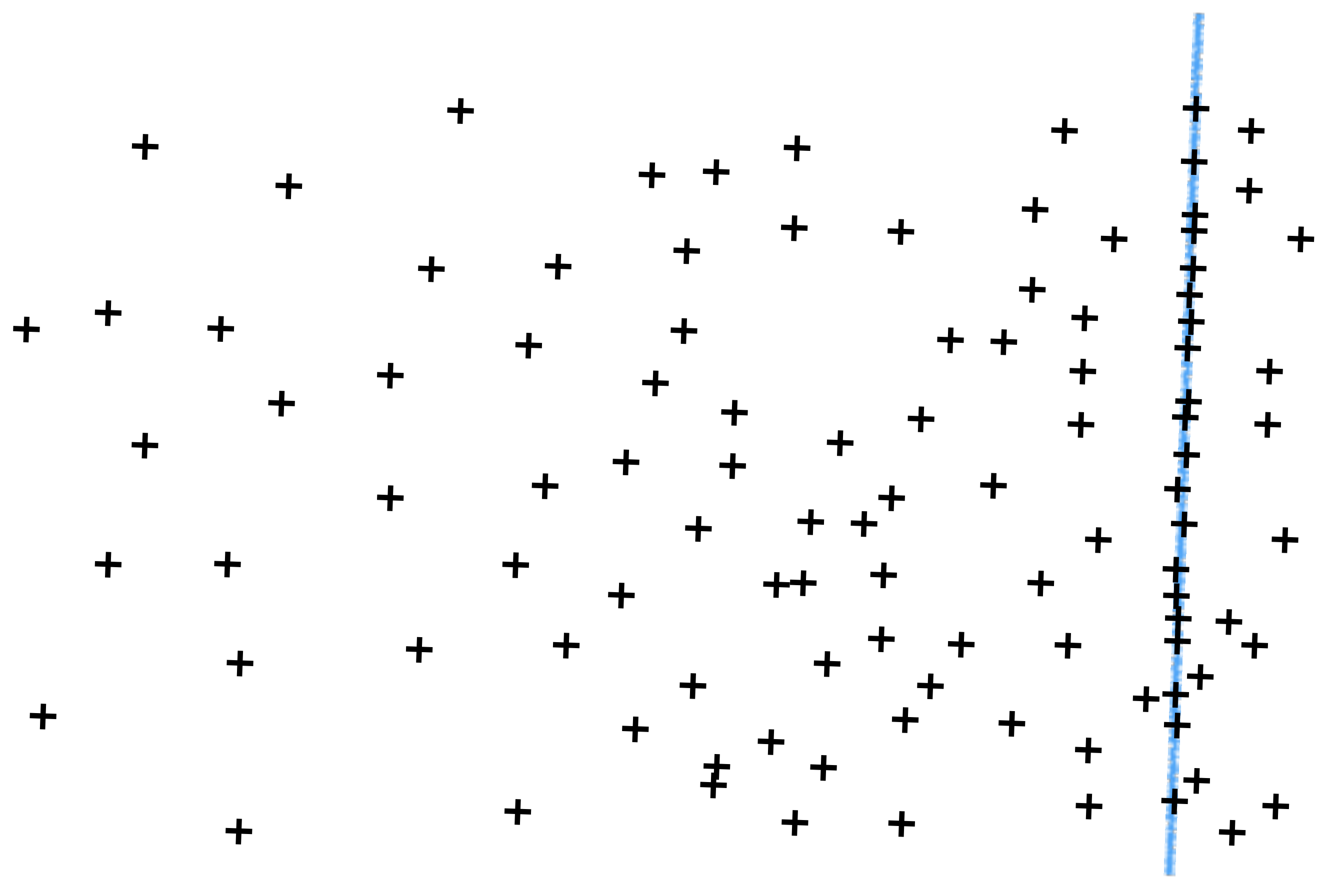
$$\min_{\theta} \sum |r_i(\theta)|$$

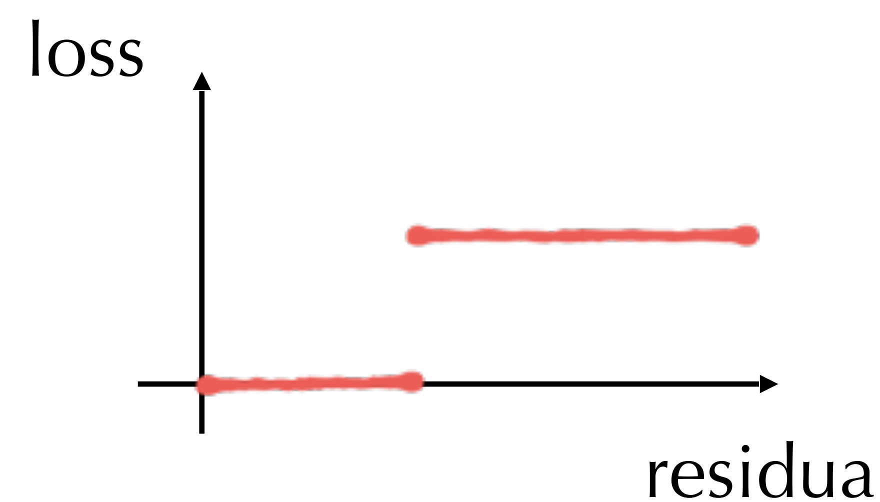
Huber Loss



$$\min_{\theta} \sum_i h(r_i(\theta)^2) \text{ with } h(x) = \begin{cases} x & \text{if } x < \delta \\ 2\delta\sqrt{x} - \delta^2 & \text{if } x \geq \delta \end{cases}$$

Minimizing Number of Outliers

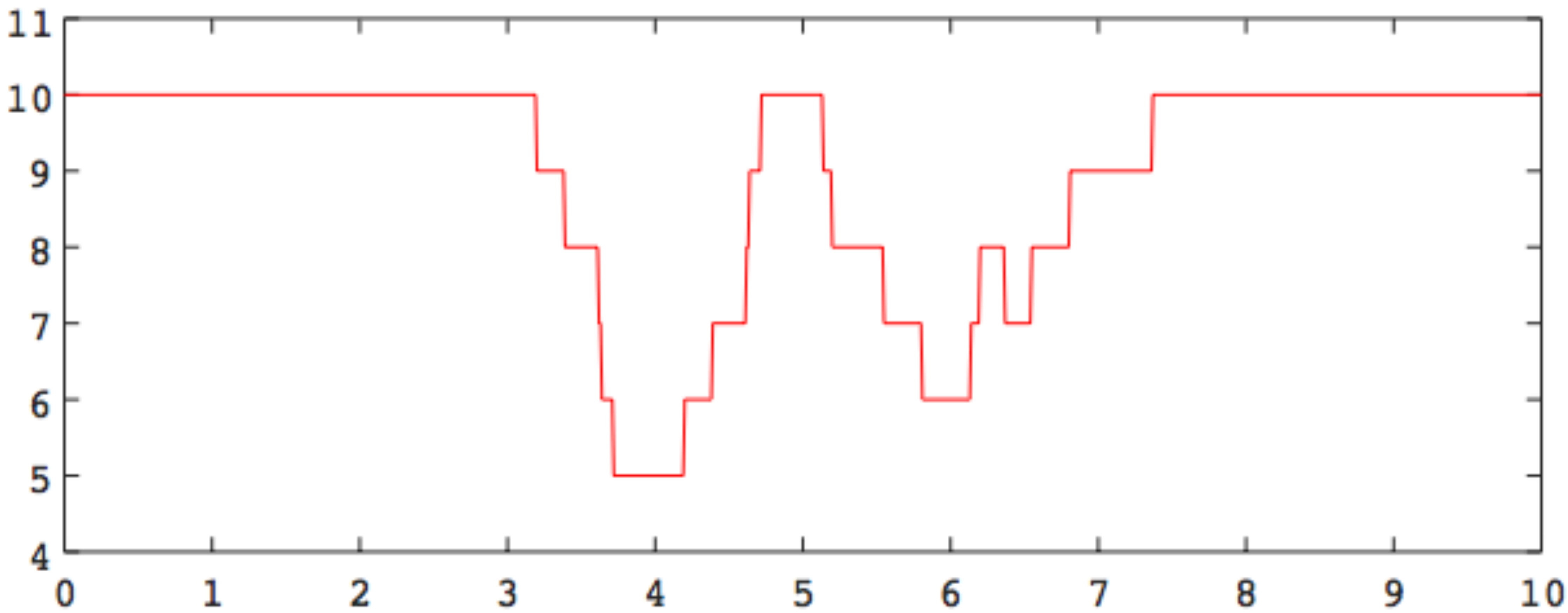




Outlier count

How to Minimize?

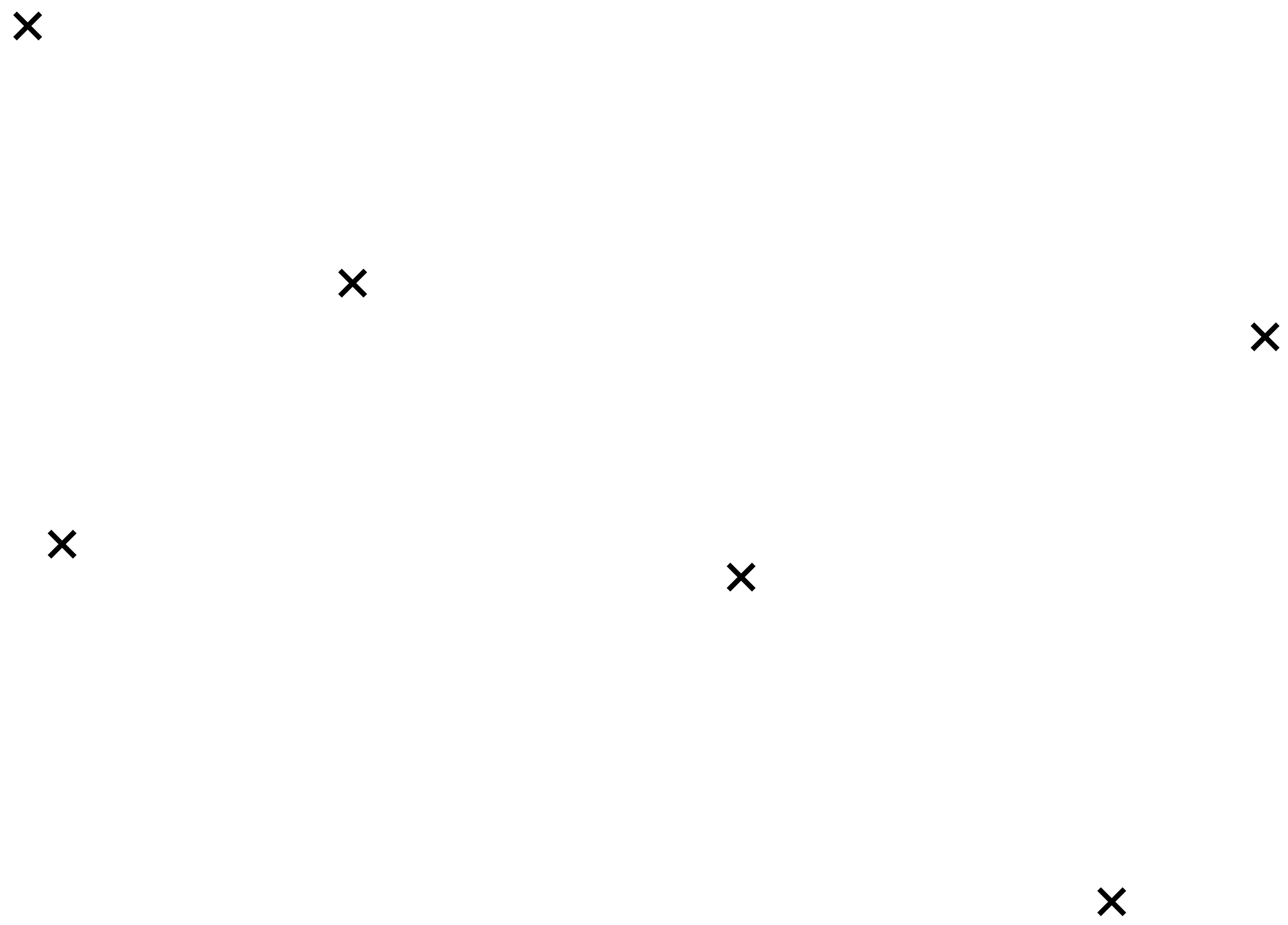
1D example:



RANdom SAmple Consensus (RANSAC)

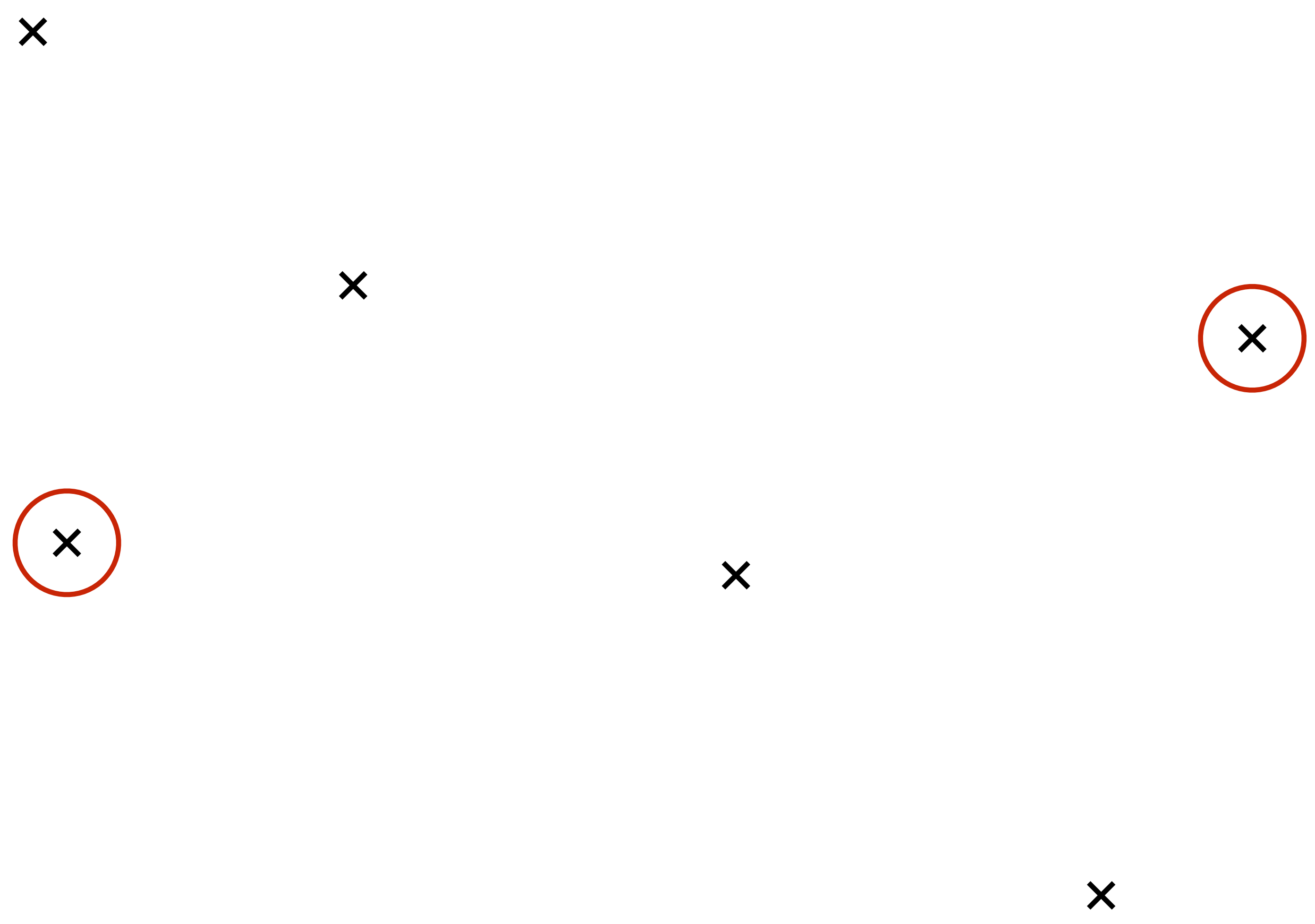
RANDom SAmple Consensus - RANSAC

Line fitting example



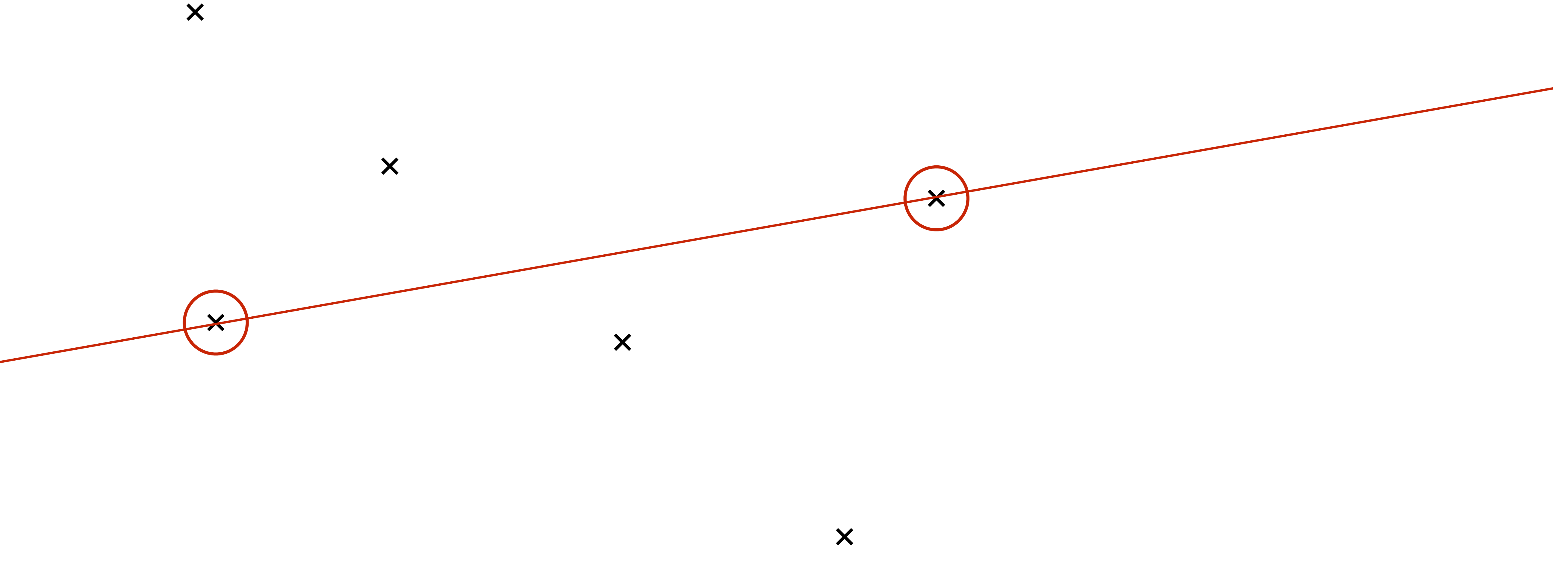
RANdom SAmple Consensus - RANSAC

Line fitting example



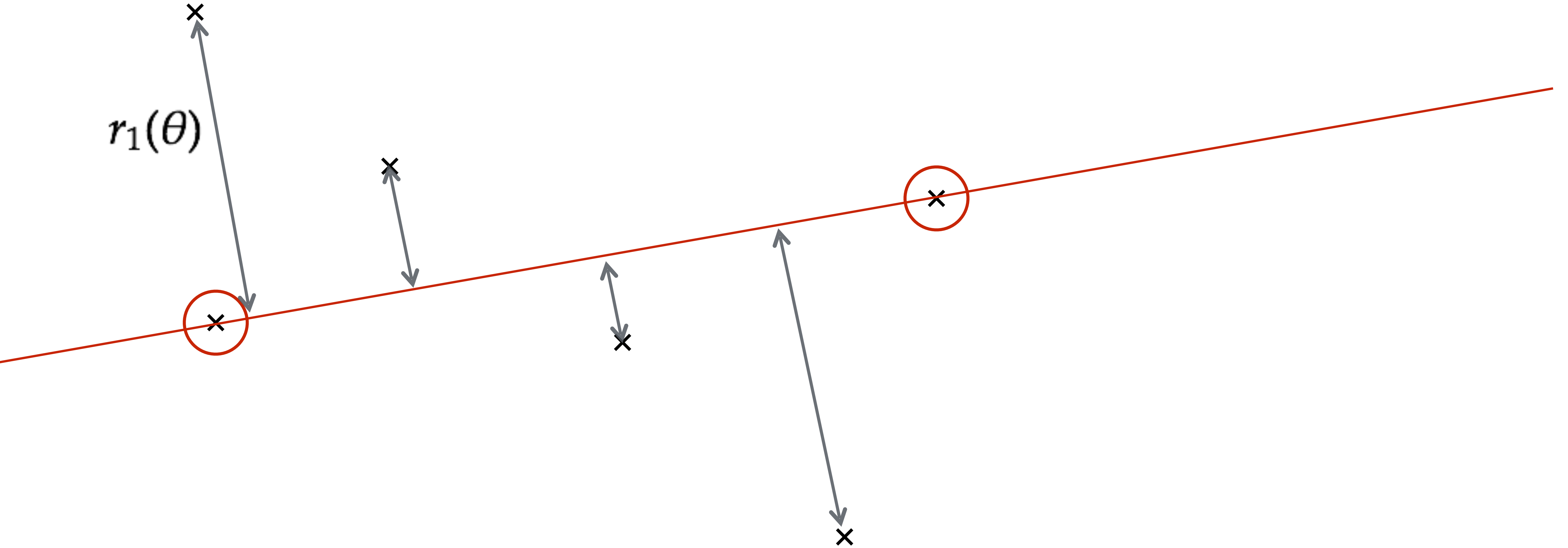
RANdom SAmple Consensus - RANSAC

Line fitting example



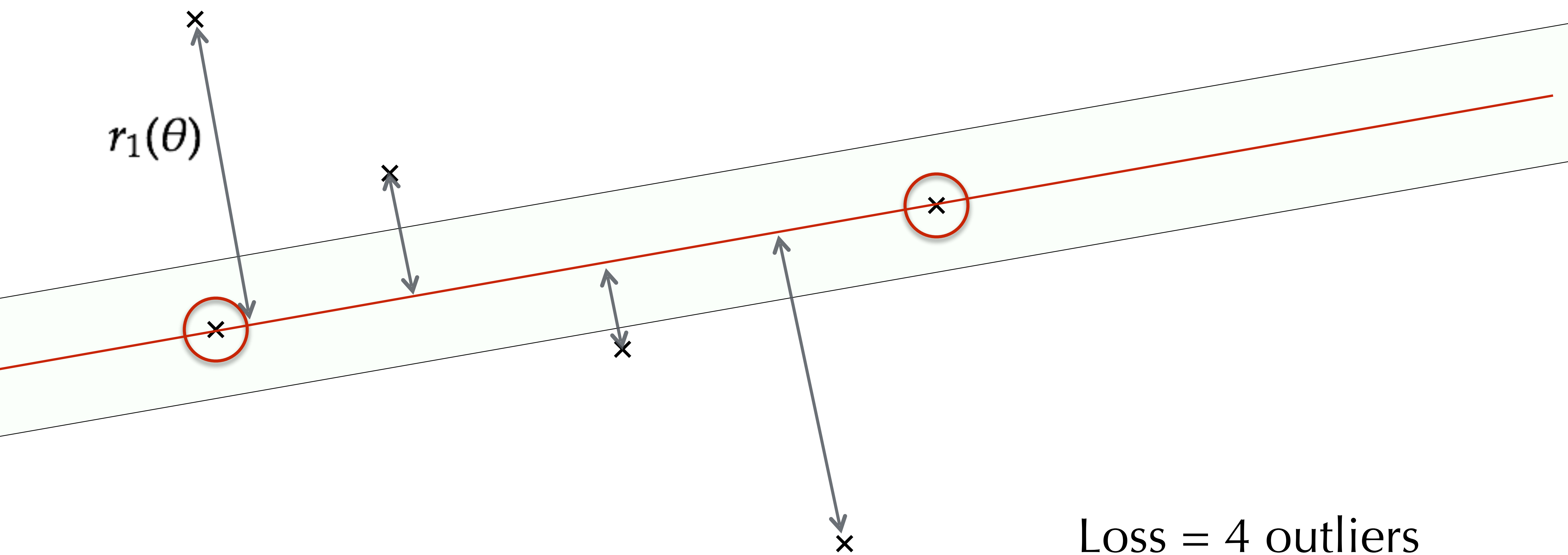
RANdom SAmple Consensus - RANSAC

Line fitting example



RANdom SAmple Consensus - RANSAC

Line fitting example

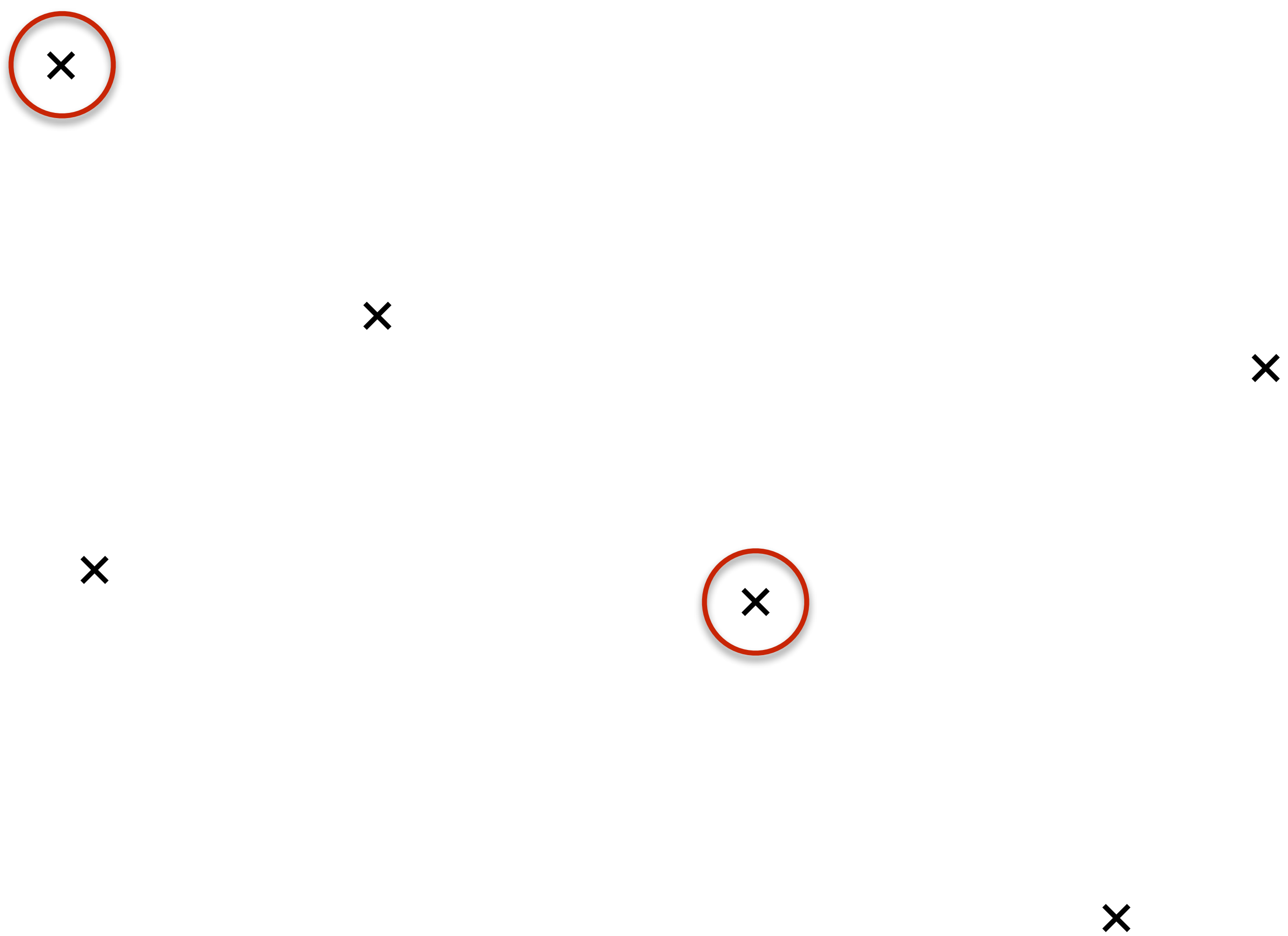


Loss = 4 outliers

RANdom SAmple Consensus - RANSAC

Line fitting example

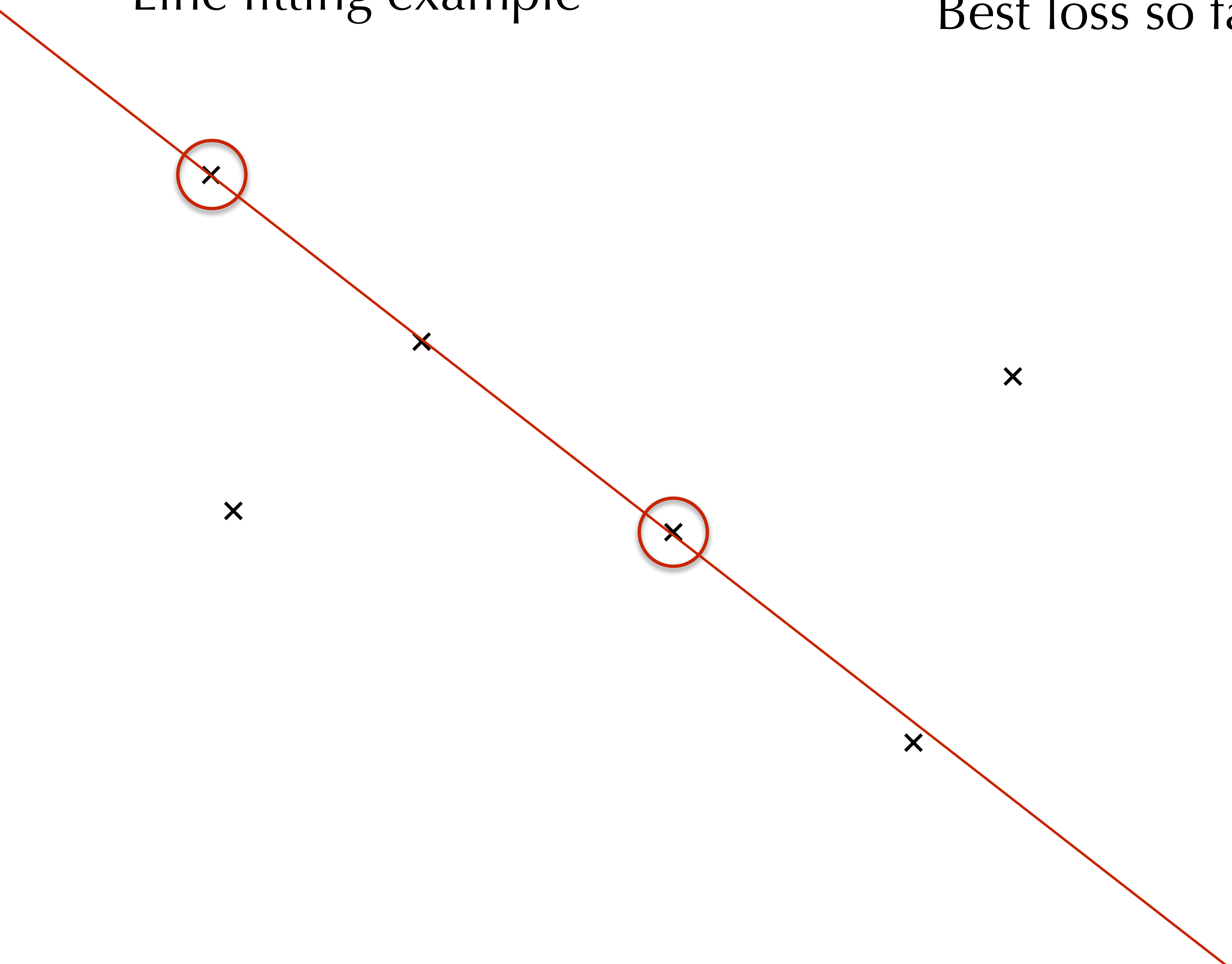
Best loss so far =   4 outliers



RANdom SAmple Consensus - RANSAC

Line fitting example

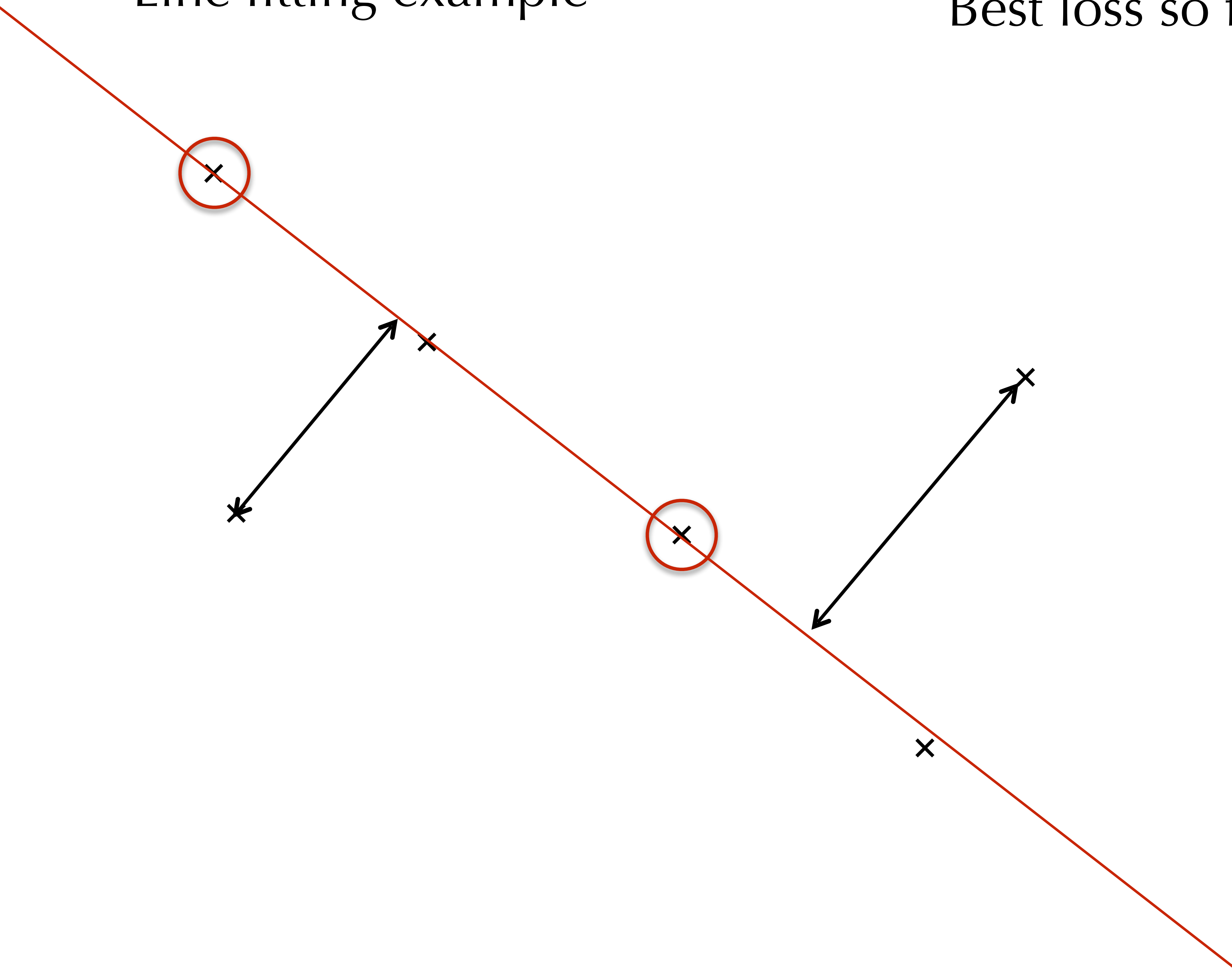
Best loss so far = 4 outliers



RANdom SAmple Consensus - RANSAC

Line fitting example

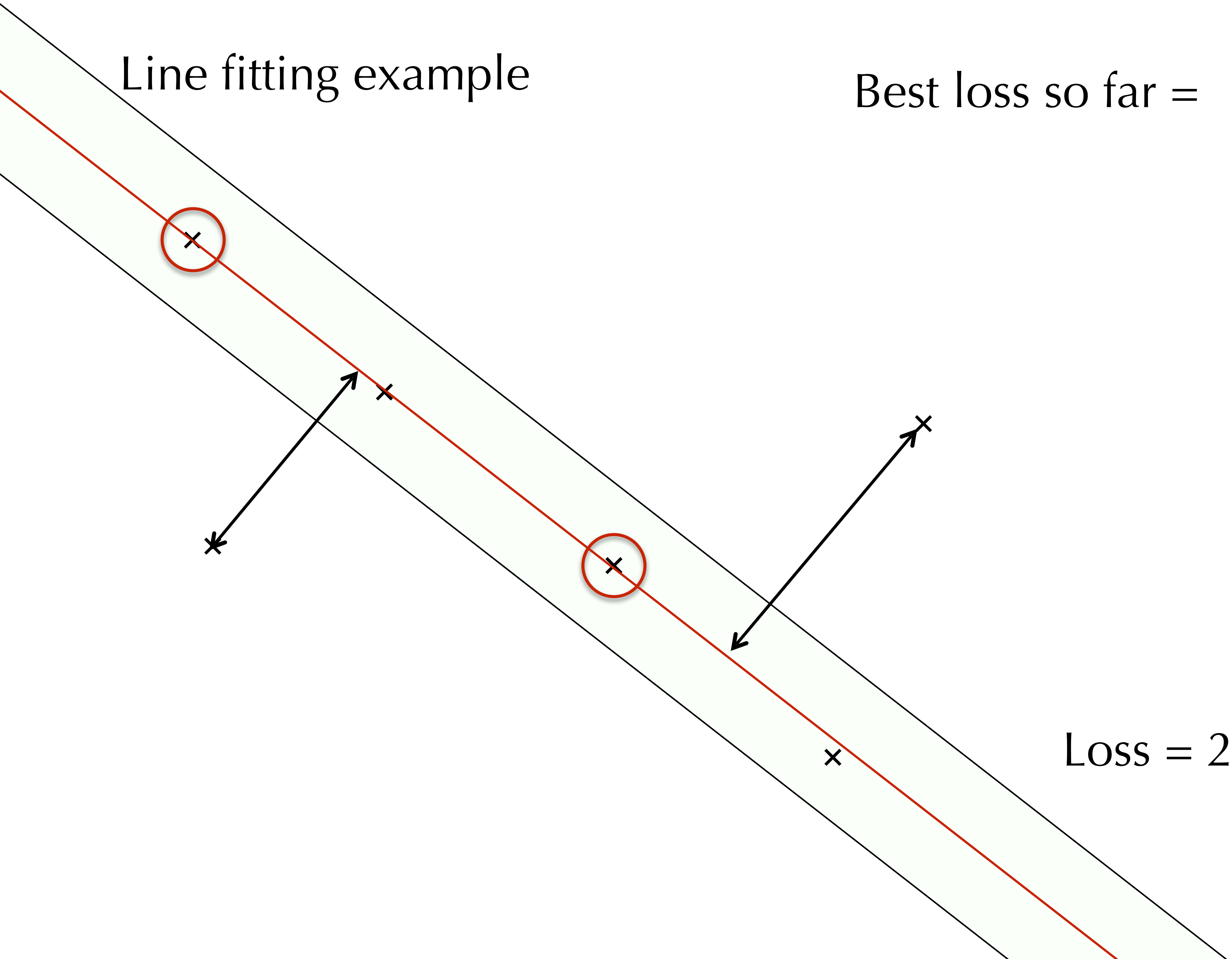
Best loss so far = 4 outliers



RANdom SAmple Consensus - RANSAC

Line fitting example

Best loss so far = 4 outliers

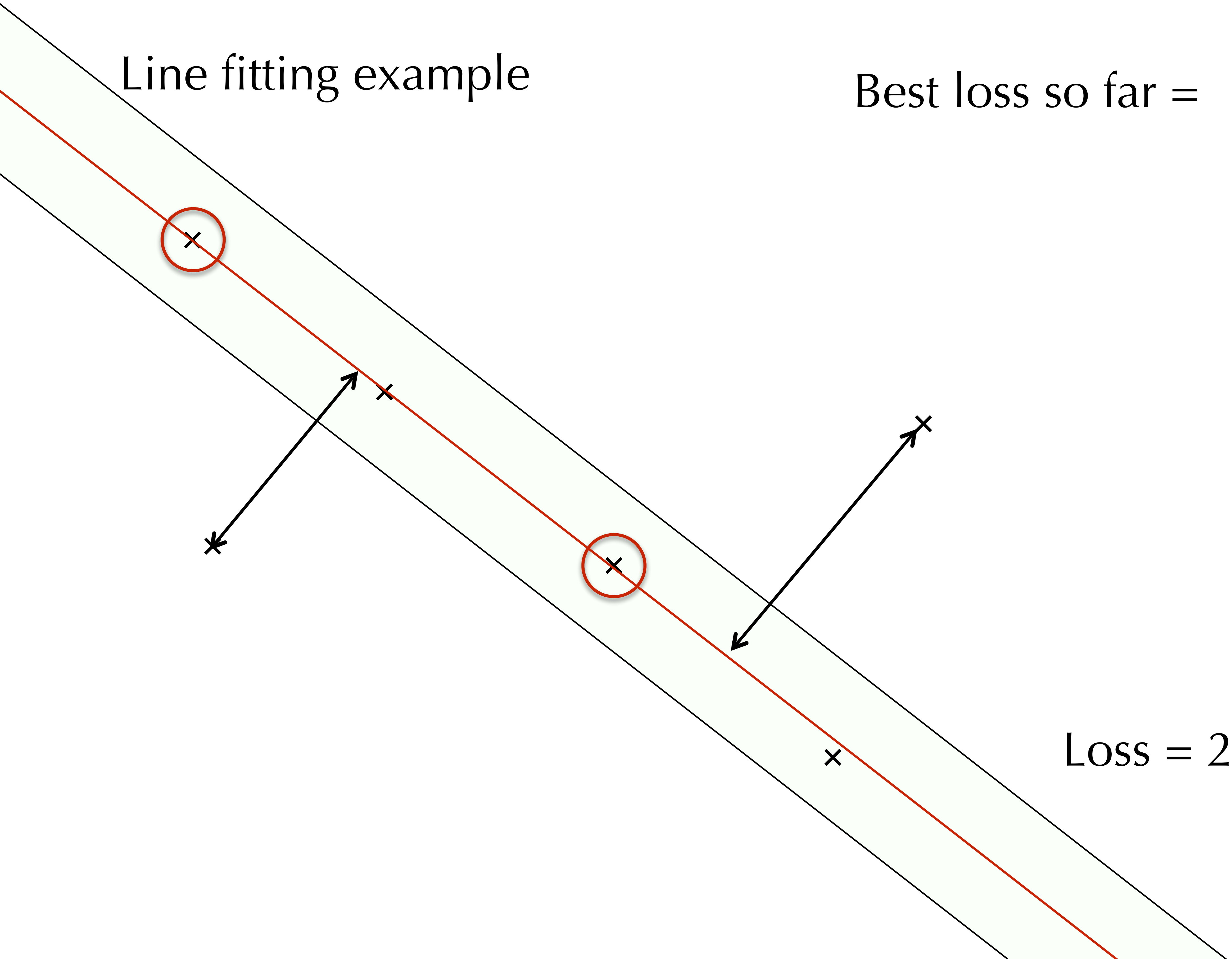


Loss = 2 outliers

RANdom SAmple Consensus - RANSAC

Line fitting example

Best loss so far = 2 outliers

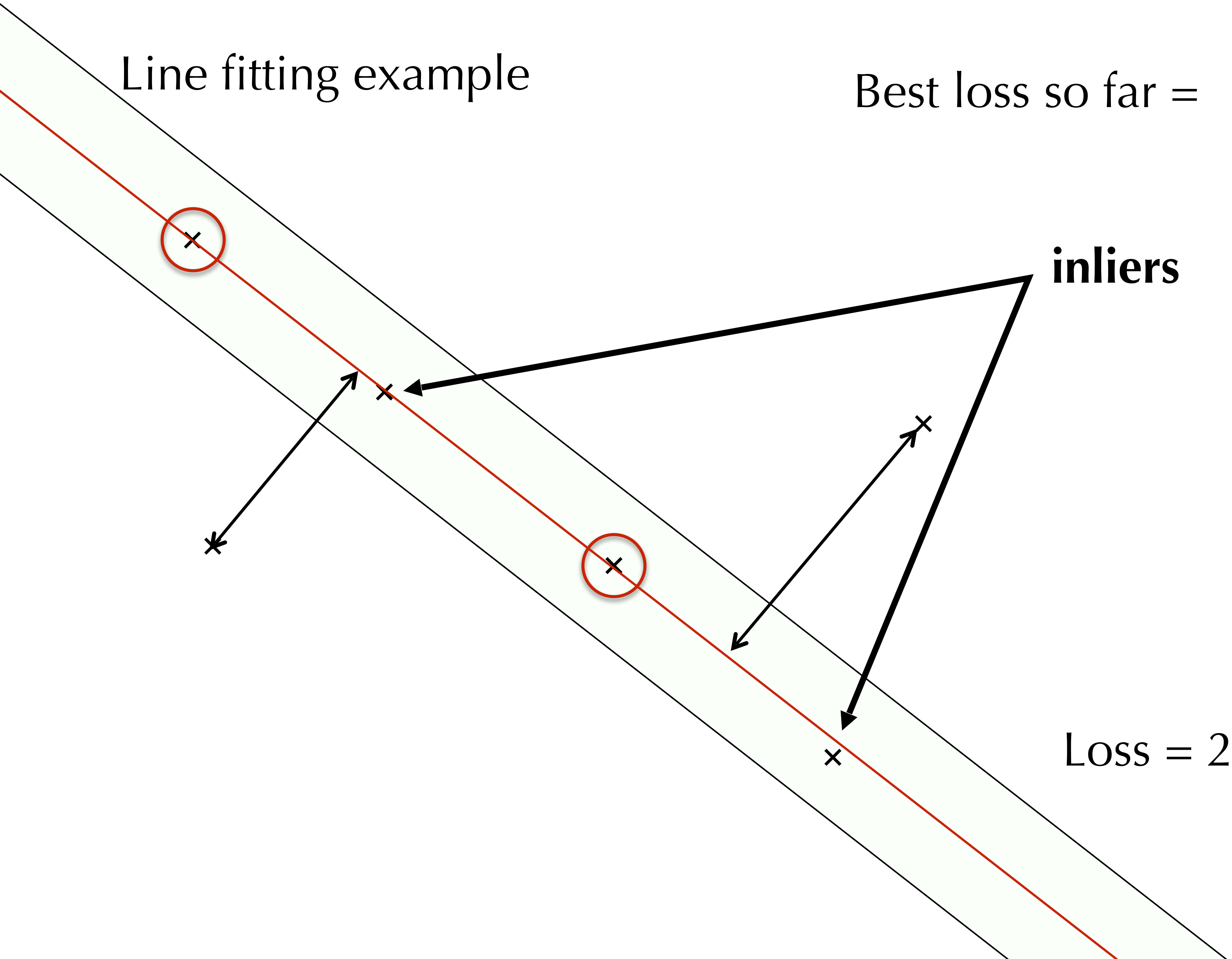


Loss = 2 outliers

RANdom SAmple Consensus - RANSAC

Line fitting example

Best loss so far = 2 outliers



inliers

Loss = 2 outliers

RANdom SAmple Consensus - RANSAC

While probability of missing correct model >η

Estimate model from *n* random data points

Estimate support (= #**inliers**) of model

If more inliers than previous best model

update best model, η

Return: Model with most inliers

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Termination Criterion

- Let's assume we know the **inlier ratio ε** (fraction of inliers)

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Termination Criterion

- Let's assume we know the **inlier ratio ε** (fraction of inliers)
- Probability of picking an inlier randomly: ε

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Termination Criterion

- Let's assume we know the **inlier ratio** ε (fraction of inliers)
- Probability of picking an inlier randomly: ε
- Probability of picking $\mathtt{n}$ inliers randomly: $\varepsilon^{\mathtt{n}}$

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Termination Criterion

- Let's assume we know the **inlier ratio** ε (fraction of inliers)
- Probability of picking an inlier randomly: ε
- Probability of picking n inliers randomly: ε^n
- Probability of non-all inlier sample ($\geq$ 1 outlier): $(1-\varepsilon^n)$

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Termination Criterion

- Let's assume we know the **inlier ratio** ε (fraction of inliers)

- Probability of picking an inlier randomly: ε

- Probability of picking n inliers randomly: ε^n

- Probability of non-all inlier sample ($\geq$ 1 outlier): $(1-\varepsilon^n)$

- Probability of not picking all-inlier sample in k iterations: $(1-\varepsilon^n)^k$

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Termination Criterion

- Terminate if $(1-\varepsilon^n)^k < \eta$

find k_{max} s.t.

$$(1-\varepsilon^n)^{k_{max}} = \eta$$

$$\Leftrightarrow \quad k_{max} \cdot \log(1-\varepsilon^n) = \log \eta$$

$$\Leftrightarrow \quad k_{max} = \frac{\log \eta}{\log(1-\varepsilon^n)}$$

$$\varepsilon' < \varepsilon$$

$$\Rightarrow k_{max}(\varepsilon') \geq k_{max}(\varepsilon)$$

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Termination Criterion

- Terminate if $(1-\varepsilon^n)^k < \eta$
- In practice: Compute maximum number iterations k_{max}

find k_{max} s.t.

$$(1-\varepsilon^n)^{k_{max}} = \eta$$

$$\Leftrightarrow \quad k_{max} \cdot \log(1-\varepsilon^n) = \log \eta$$

$$\Leftrightarrow \quad k_{max} = \frac{\log \eta}{\log(1-\varepsilon^n)}$$

$$\varepsilon' < \varepsilon$$

$$\Rightarrow k_{max}(\varepsilon') \geq k_{max}(\varepsilon)$$

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Termination Criterion

- Terminate if $(1-\varepsilon^n)^k < \eta$
- In practice: Compute maximum number iterations k_{max}

find k_{max} s.t.

$$(1-\varepsilon^n)^{k_{max}} = \eta$$

$$\Leftrightarrow \quad k_{max} \cdot \log(1-\varepsilon^n) = \log \eta$$

$$\Leftrightarrow \quad k_{max} = \frac{\log \eta}{\log(1-\varepsilon^n)}$$

- How do we know inlier ratio ε? $\varepsilon' < \varepsilon$

$$\Rightarrow k_{max}(\varepsilon') \geq k_{max}(\varepsilon)$$

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

RANdom SAmple Consensus - RANSAC

Input: *m* data points

$k = 0$, $\varepsilon = \varepsilon_0$

$k_{max} = \log(\eta)/\log(1-\varepsilon^n)$

While $k < k_{max}$

Estimate model from *n* random data points

Estimate support (= **#inliers**) of model

If more inliers than previous best model

update best model

update ε = #inliers / m

update $k_{max} = \log(\eta)/\log(1-\varepsilon^n)$

Return: Model with most inliers

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Minimal Solvers

- Probability of picking $\mathtt{n}$ inliers randomly: $\varepsilon^{\mathtt{n}}$

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

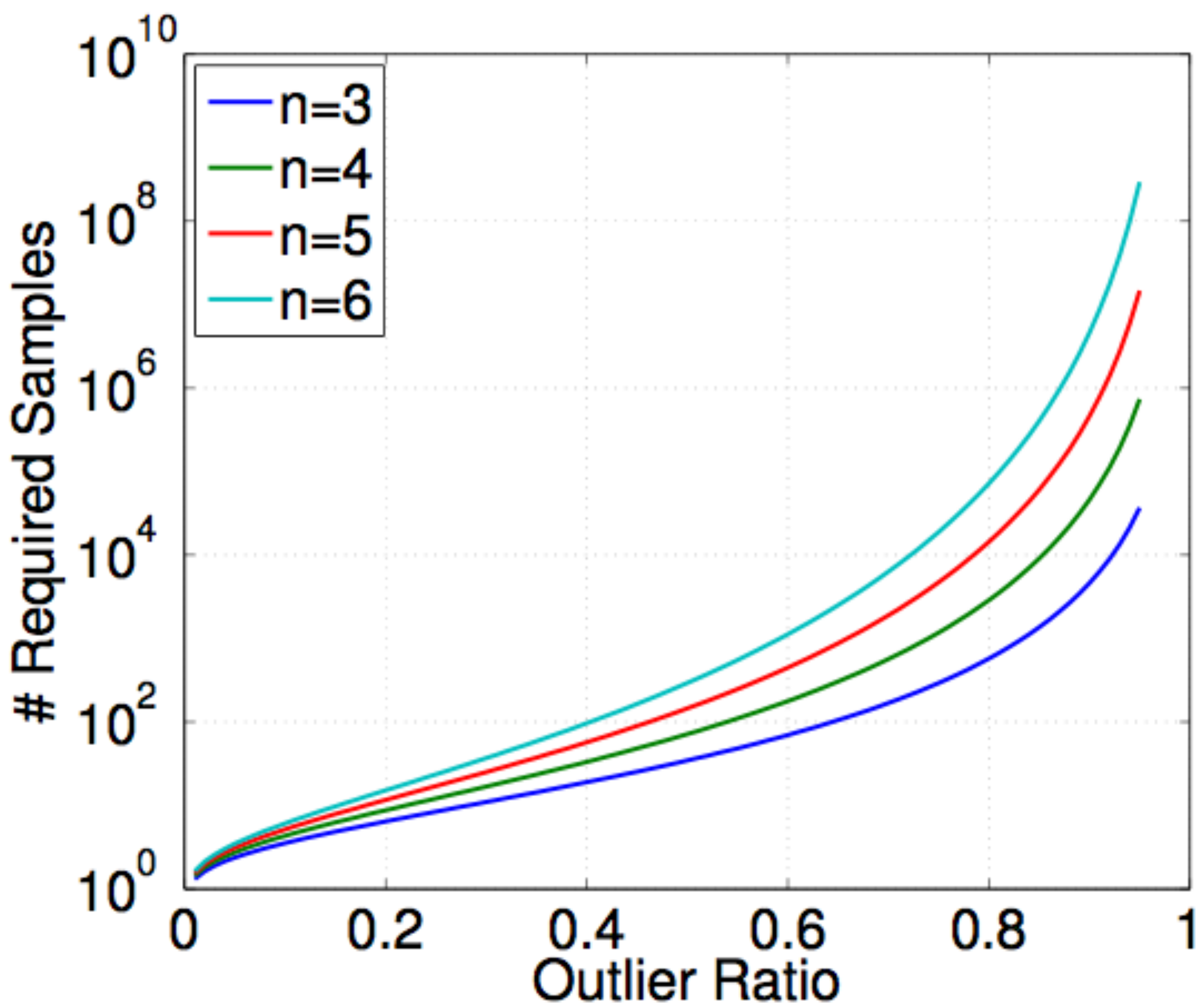
Minimal Solvers

- Probability of picking $\mathtt{n}$ inliers randomly: $\varepsilon^{\mathtt{n}}$
- Maximize $\varepsilon^{\mathtt{n}}$ by minimizing $\mathtt{n}$ → **minimal solver**

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Minimal Solvers

- Probability of picking n inliers randomly: ε^n
- Maximize ε^n by minimizing n → **minimal solver**



[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

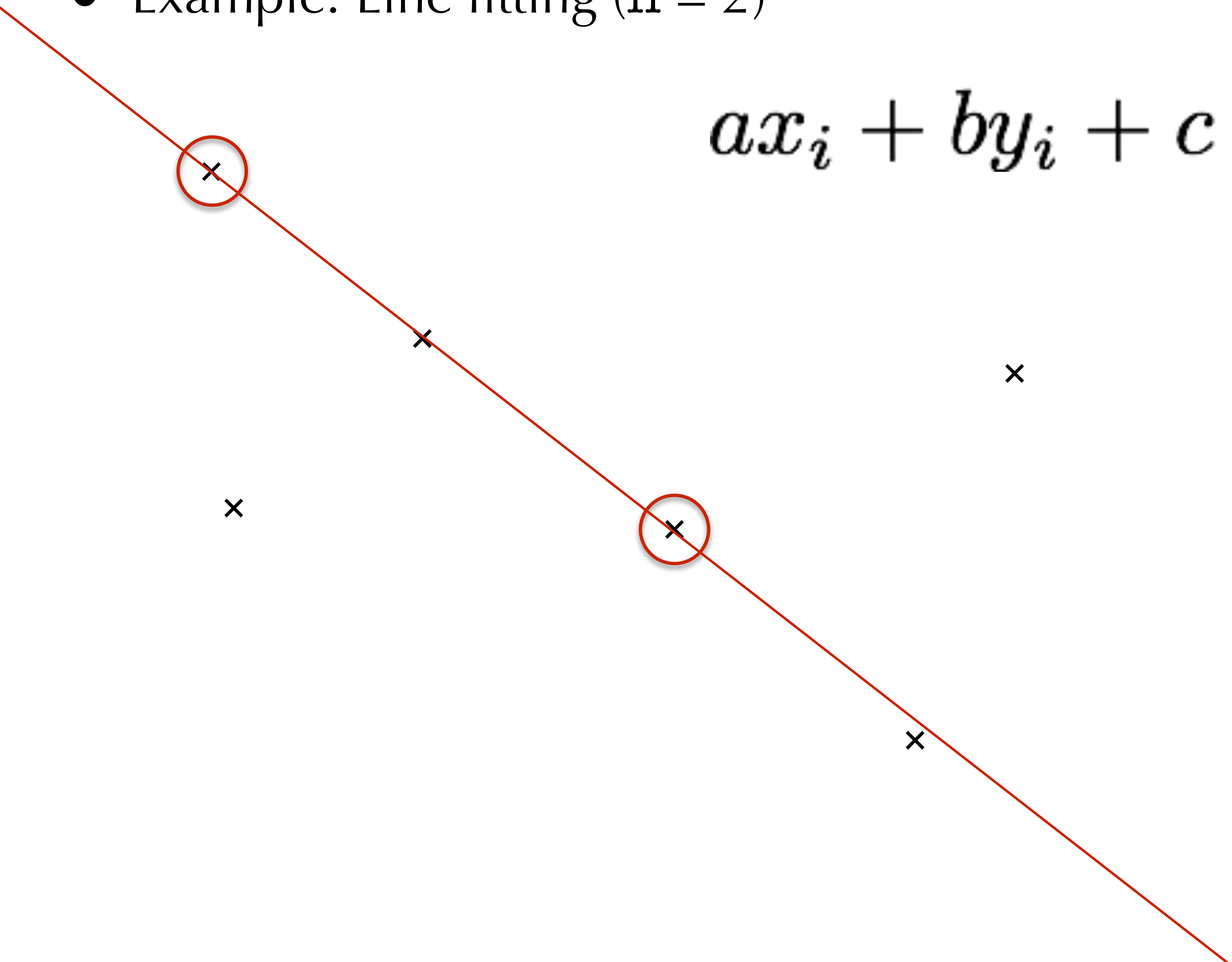
Minimal Solvers

- Solver fits a model into a set of measurements
- **Minimal solver** uses minimum number of measurements

Minimal Solvers

- Example: Line fitting (`n` = 2)

$$ax_i + by_i + c = 0$$



Minimal Solvers

- Lab 3: Create minimal solver for affine transformation

$$\begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

$$\underline{\underline{B}} \begin{pmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \\ t_1 \\ t_2 \end{pmatrix} = 0$$

Minimal Solvers

- Lab 3: Create minimal solver for affine transformation

$$\begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

- **Hint**: Write affine transformation as linear system in the parameters of the affine transformation, solve the system

$$\underline{B} \begin{pmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \\ t_1 \\ t_2 \end{pmatrix} = 0$$

RANSAC in Practice

While probability of missing correct model >η

Estimate model from *n* random data points

Estimate support ~~(= #**inliers**)~~ of model

If new best model

local optimization

update best model, η

Refine best model via least squares on inliers

Return: refined best model

[Chum, Matas, Optimal Randomized RANSAC. PAMI 2008]

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

RANSAC in Practice

While probability of missing correct model >η

Estimate model from *n* random data points

Estimate support ~~(= #**inliers**)~~ of model

[Chum, Matas, Optimal Randomized RANSAC. PAMI 2008]

If new best model

local optimization

[Lebeda, Matas, Chum, Fixing the Locally Optimized RANSAC. BMVC 2012]

update best model, η

Refine best model via least squares on inliers

Return: refined best model

[Fischler & Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. CACM 1981]

Lessons Learned

- Main lessons from this lecture
  - What is model fitting?
  - Impact of outliers on least-squares estimates
  - Robust cost functions
  - RANSAC: Why? How?
  - What are minimal solvers?

Lessons Learned

- Main lessons from this lecture
  - What is model fitting?
  - Impact of outliers on least-squares estimates
  - Robust cost functions
  - RANSAC: Why? How?
  - What are minimal solvers?
- Next lecture: Image Registration