RANSAC
Robust Model Estimation
From Data Contaminated By Outliers

Lecturer: Jiří Matas

Authors: Ondřej Chum, Jiří Matas, Ondřej Drbohlav

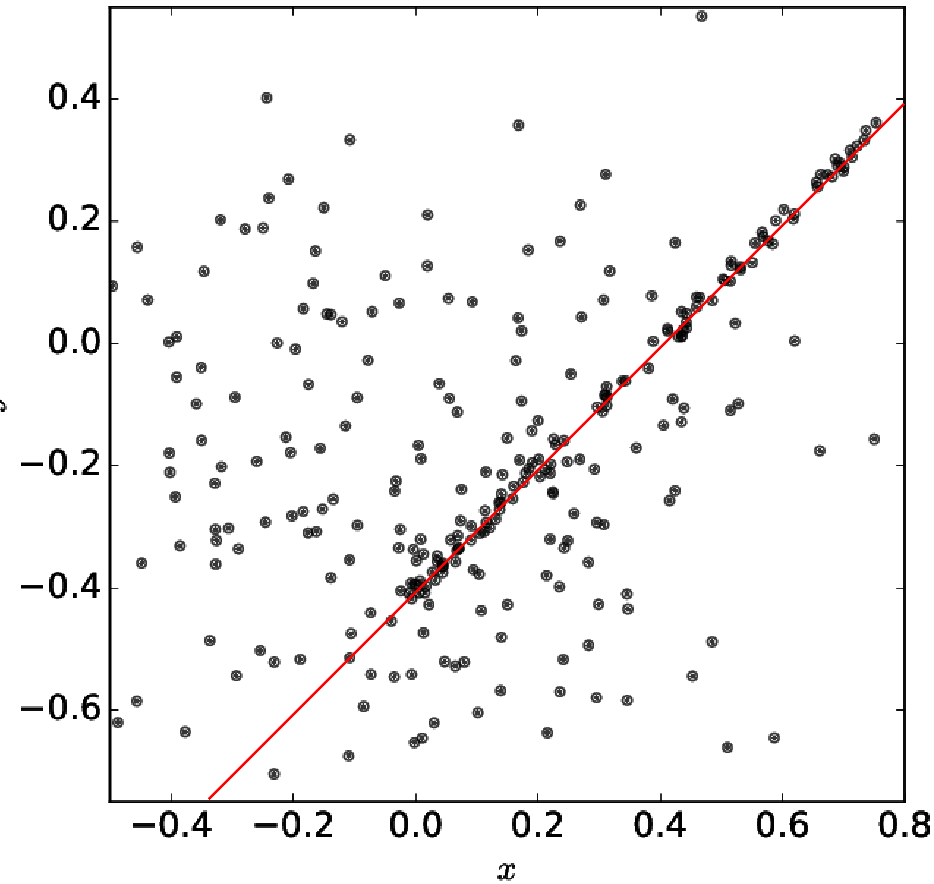
Czech Technical University in Prague

http://cmp.felk.cvut.cz

What is RANSAC?

- RANSAC = RANdom SAmple Consensus
- M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. CACM, 24(6):381–395, June 1981.

- **Example**: Finding a line in 2D data.
  - Not all input points are on a line.
  - Finding a line also implicitly divides points to **inliers** (=those on a line) and **outliers** (=those not on a line).



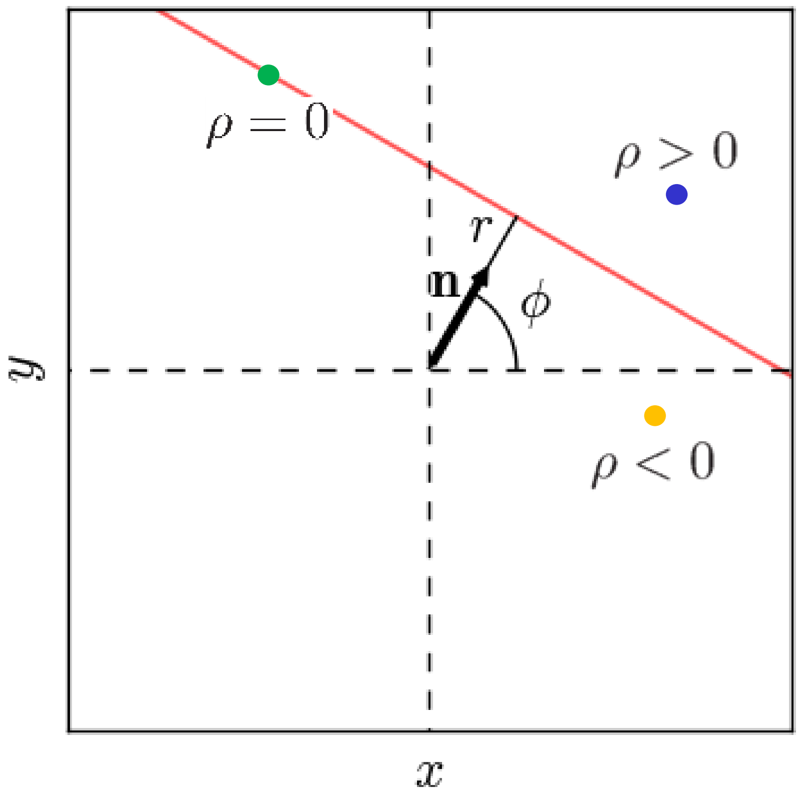
Line Fitting: Line Parametrization

- Line parametrization (homogeneous):

$$ax + by + c = 0, \qquad (a \neq 0 \vee b \neq 0) \tag{1}$$
$$a, b, c \in \mathbb{R} : \text{ line parameters} \tag{2}$$
$$(x, y) : \text{ point coordinates} \tag{3}$$

- Line parametrization (radial):

$$x \cos\phi + y \sin\phi = r, \tag{4}$$
$$\phi \in [0, \pi[, \ r \in \mathbb{R} : \text{ line parameters} \tag{5}$$



Note: $\mathbf{n} = (\cos\phi, \sin\phi)$ (thus $\|\mathbf{n}\| = 1$)

- Line parameters: $\phi \in [0, \pi[,\ r \in \mathbb{R}$
- Point $\mathbf{x} = (x, y)$ on the line:

$$x \cos\phi + y \sin\phi = r$$
$$\Leftrightarrow \mathbf{x} \cdot (\cos\phi, \sin\phi) = r$$

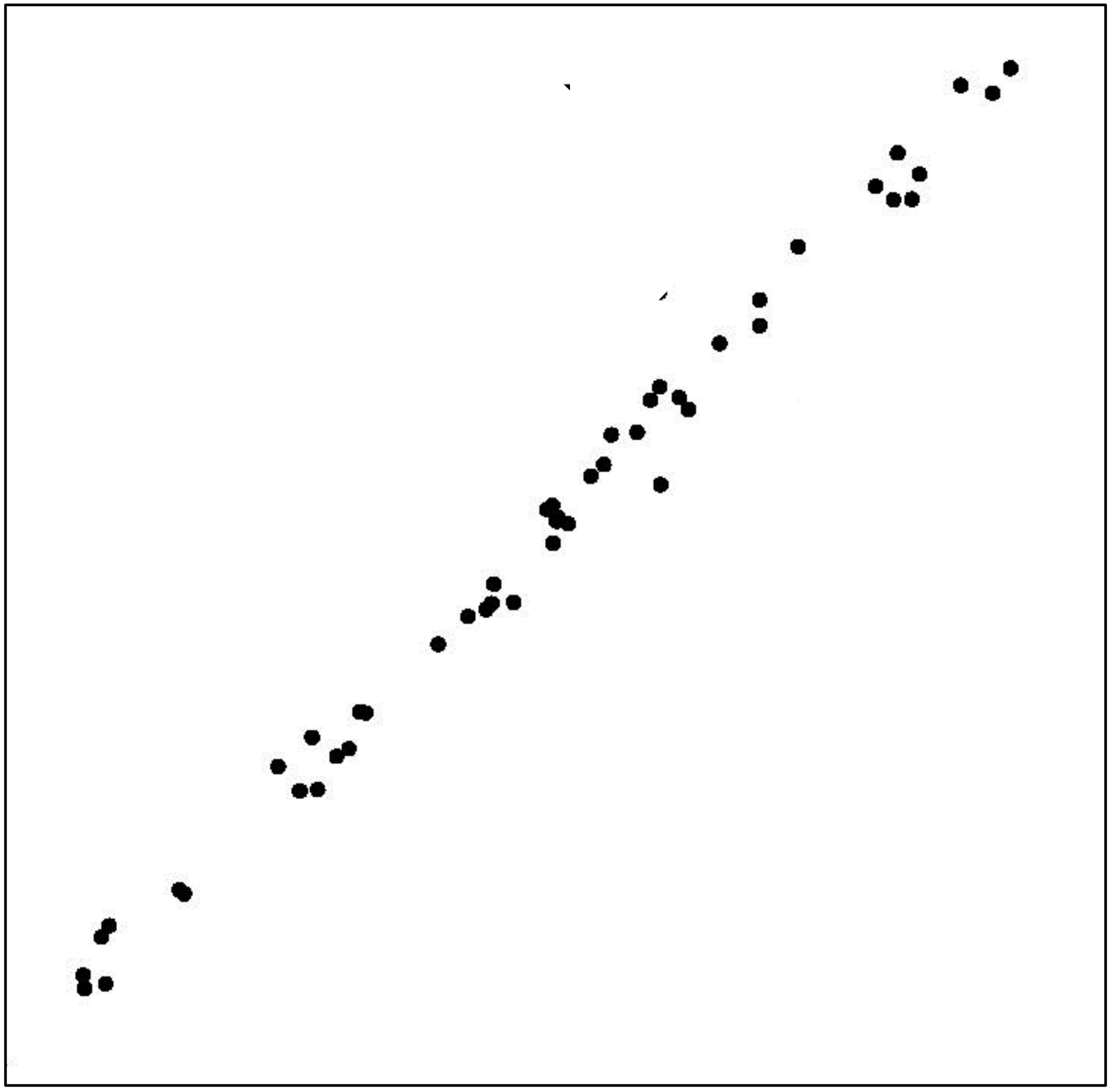
- Point $\mathbf{x} = (x, y)$ not on the line:

$$\mathbf{x} \cdot (\cos\phi, \sin\phi) \neq r$$

  – Signed distance $\rho(\mathbf{x})$ from line:

$$\rho(\mathbf{x}) = \mathbf{x} \cdot (\cos\phi, \sin\phi) - r$$

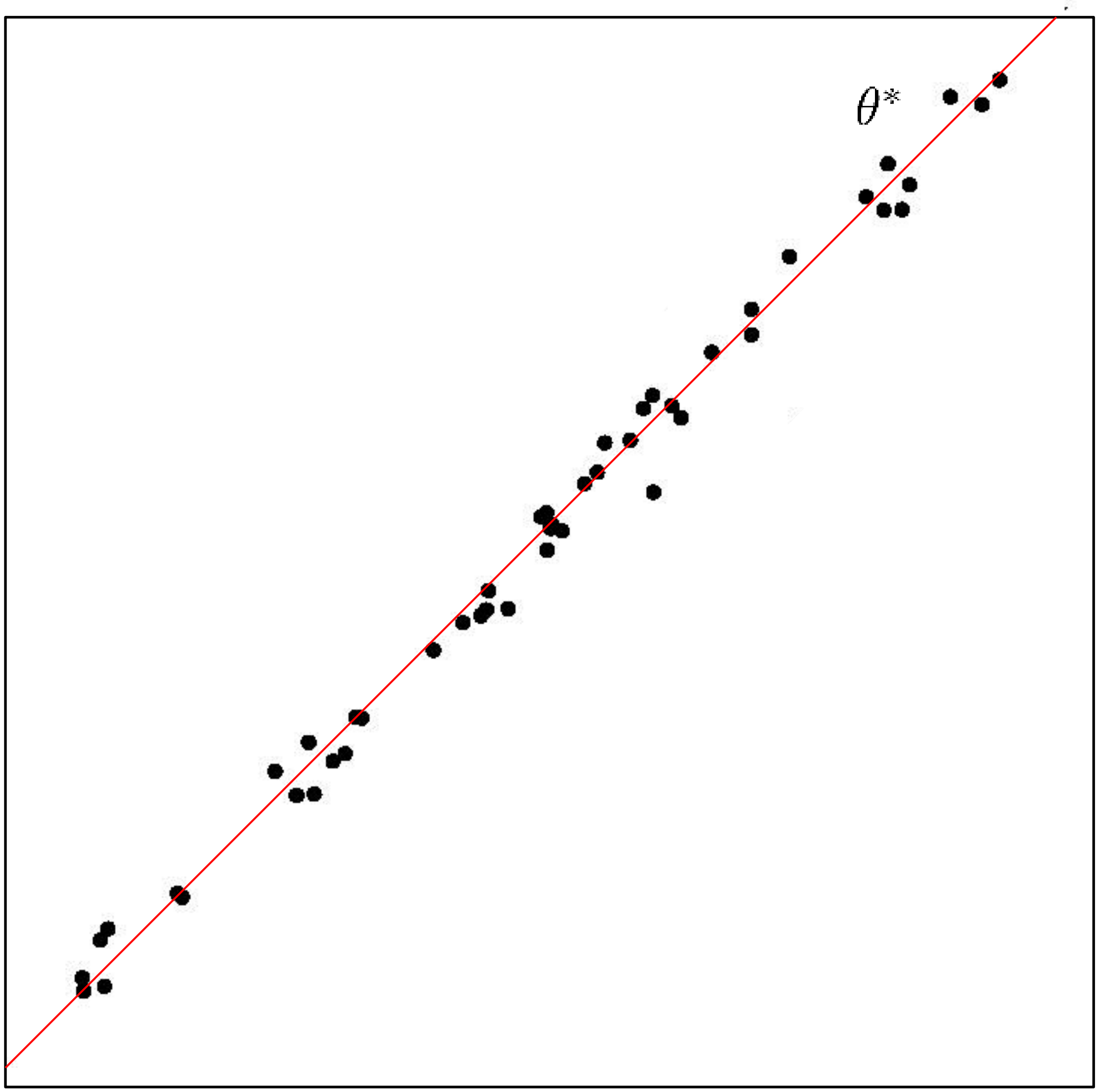
Line Fitting, Inliers Only: Easy!



Data points

$$\mathcal{X} = \{\mathbf{x}_j, j = 1, 2, ..., N_p\}$$
$$(\mathbf{x}_j \in \mathbb{R}^2)$$

Find the line which "best fits" these points.



Data points

$$\mathcal{X} = \{\mathbf{x}_j, j = 1, 2, ..., N_p\}$$
$$(\mathbf{x}_j \in \mathbb{R}^2)$$

Find the line which "best fits" the points.

As optimization: Find best line with parameters θ^* as

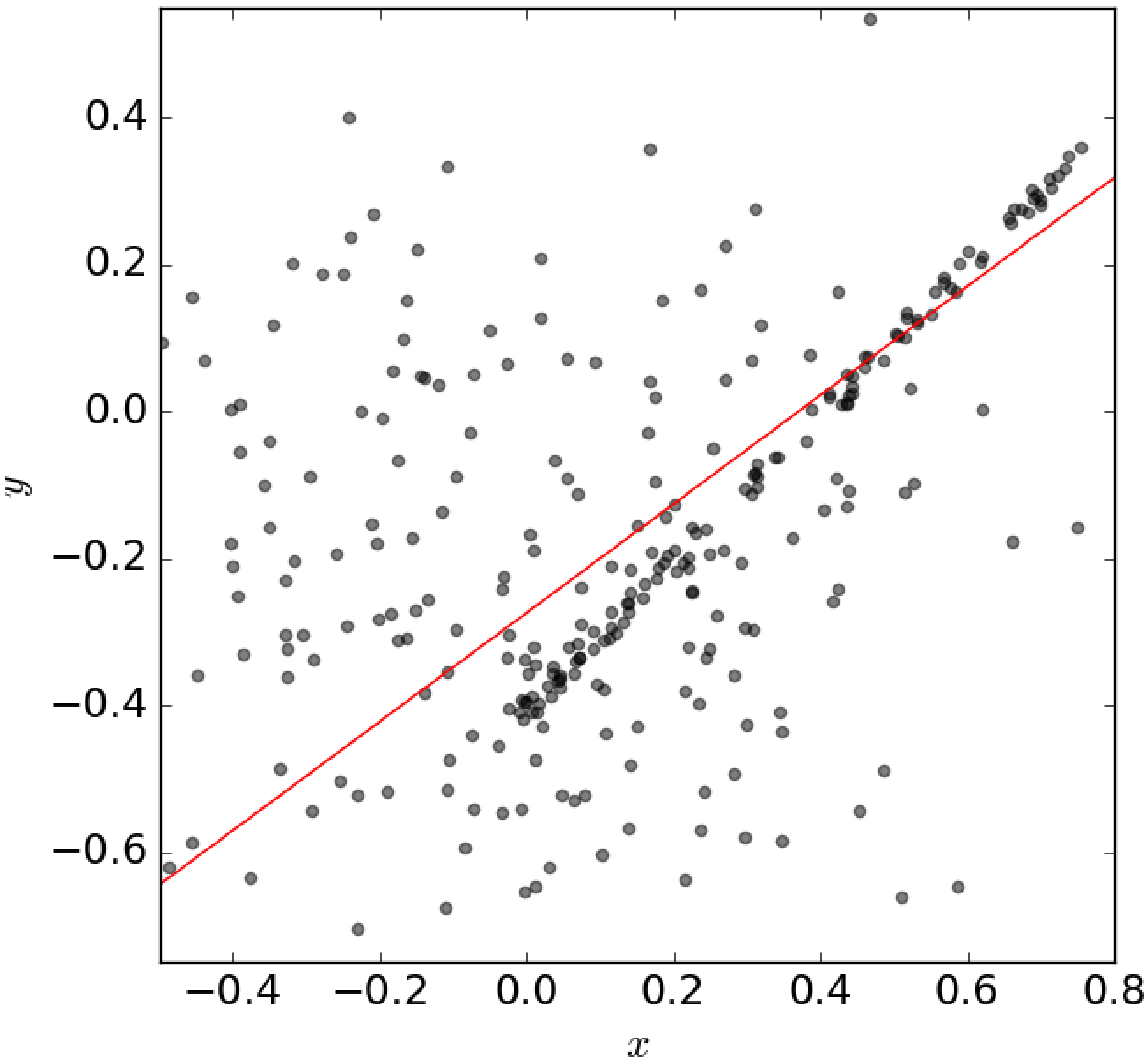
$$\theta^* = \underset{\theta}{\text{argmin}} \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$$

For $f_{LSQ}(\mathbf{x}, \theta) = [\rho(\mathbf{x})]^2$

this is easily solvable by Singular Value Decomposition (SVD).

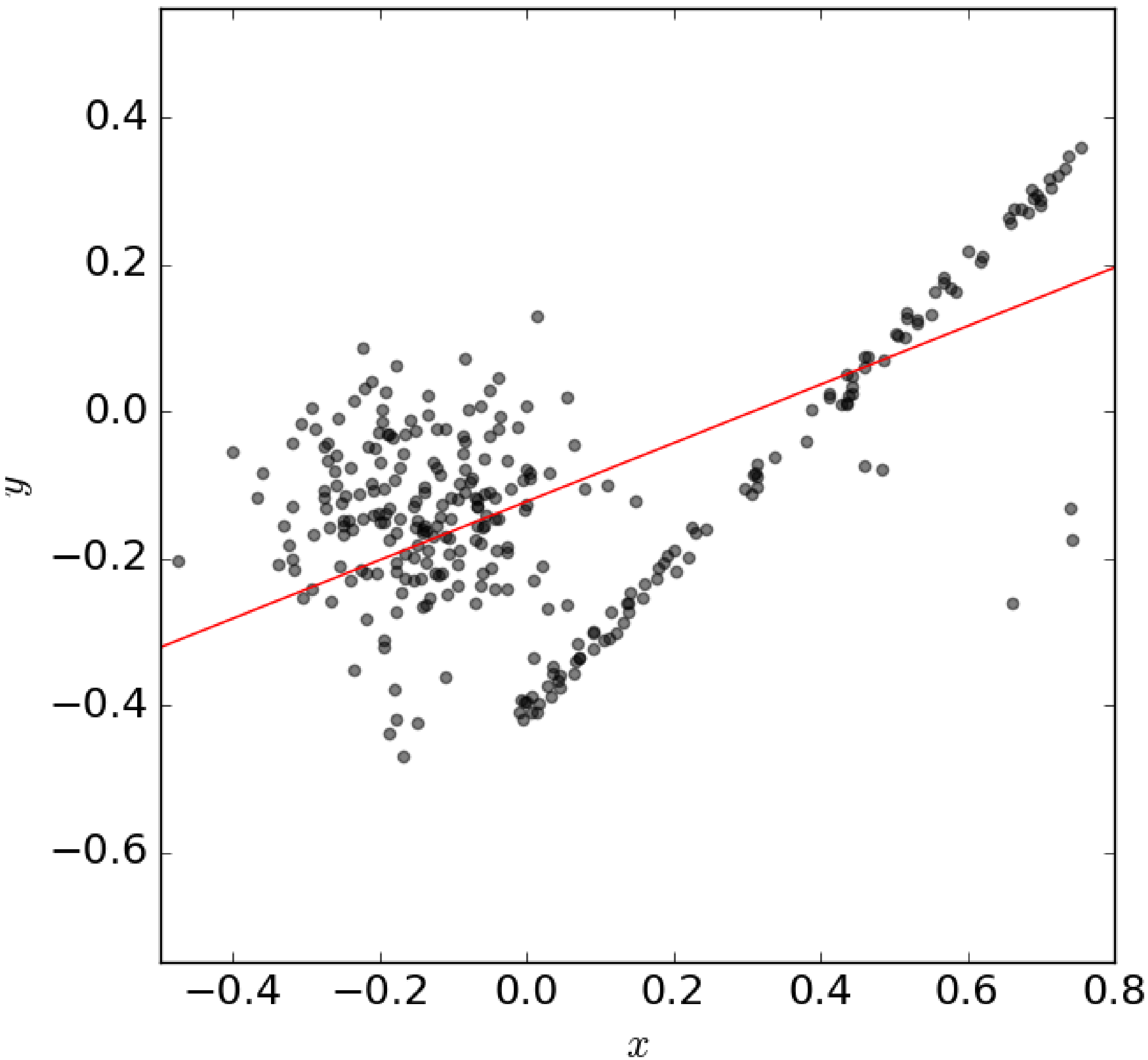
General Case with Outliers, Example 1

Example 1



Least squares fit

Example 2



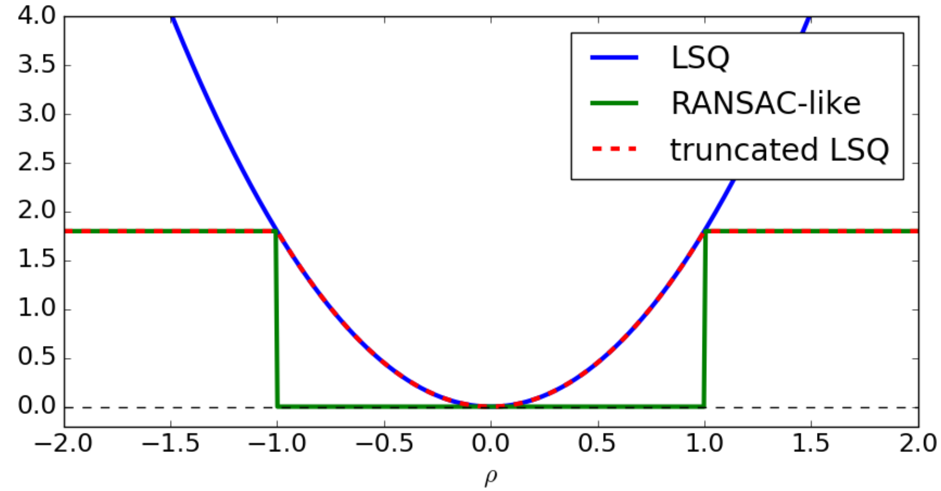
Least squares fit

- $\mathcal{X} = \{\mathbf{x}_j\}_{j=1}^{N_p}$ set of data points

<u>Find</u>:

$$\theta^* = \arg\min_{\theta} \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$$

$$\theta = (r, \phi)$$



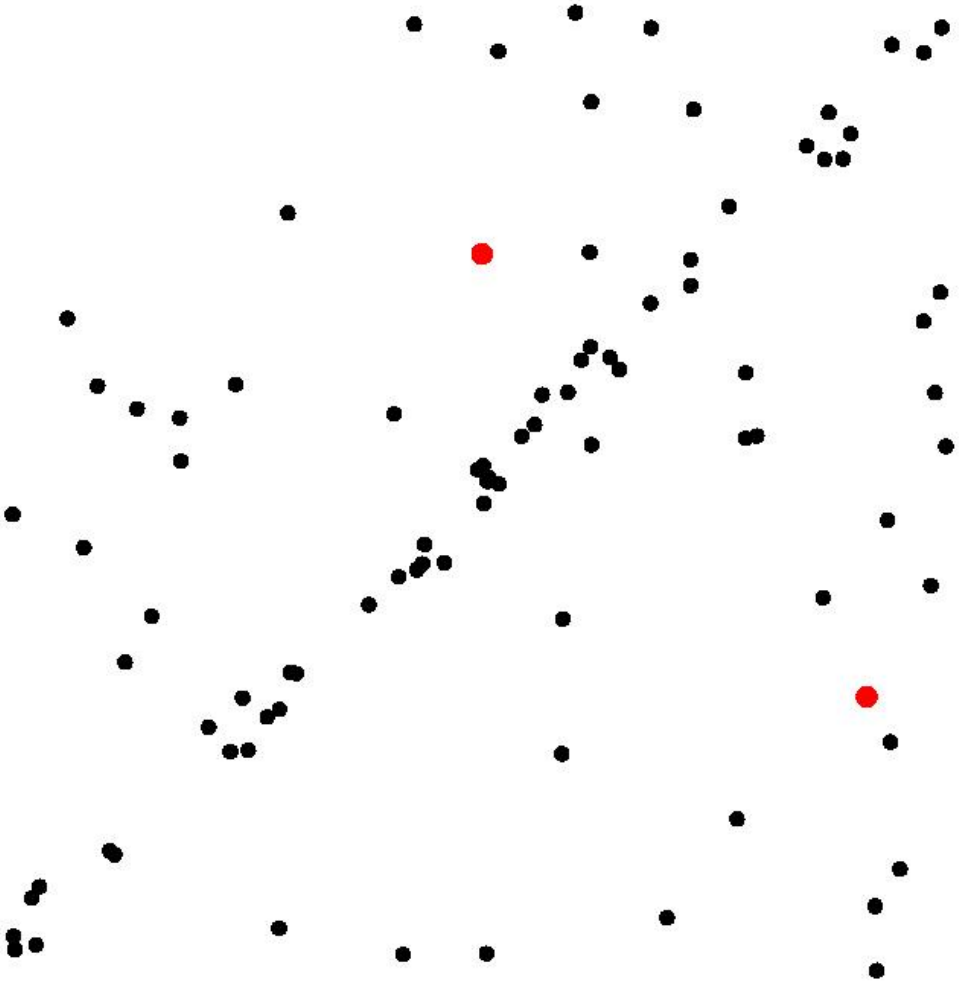
- No outliers: $f_{LSQ}(\mathbf{x}, \theta) = [\rho(\mathbf{x})]^2$

- **Use instead**:

$$f_{\text{RANSAC}}(\mathbf{x}, \theta) = \begin{cases} 0, & \text{if } \rho(\mathbf{x}) \leq \text{threshold } \sigma \\ \text{const}, & \text{otherwise} \end{cases}$$

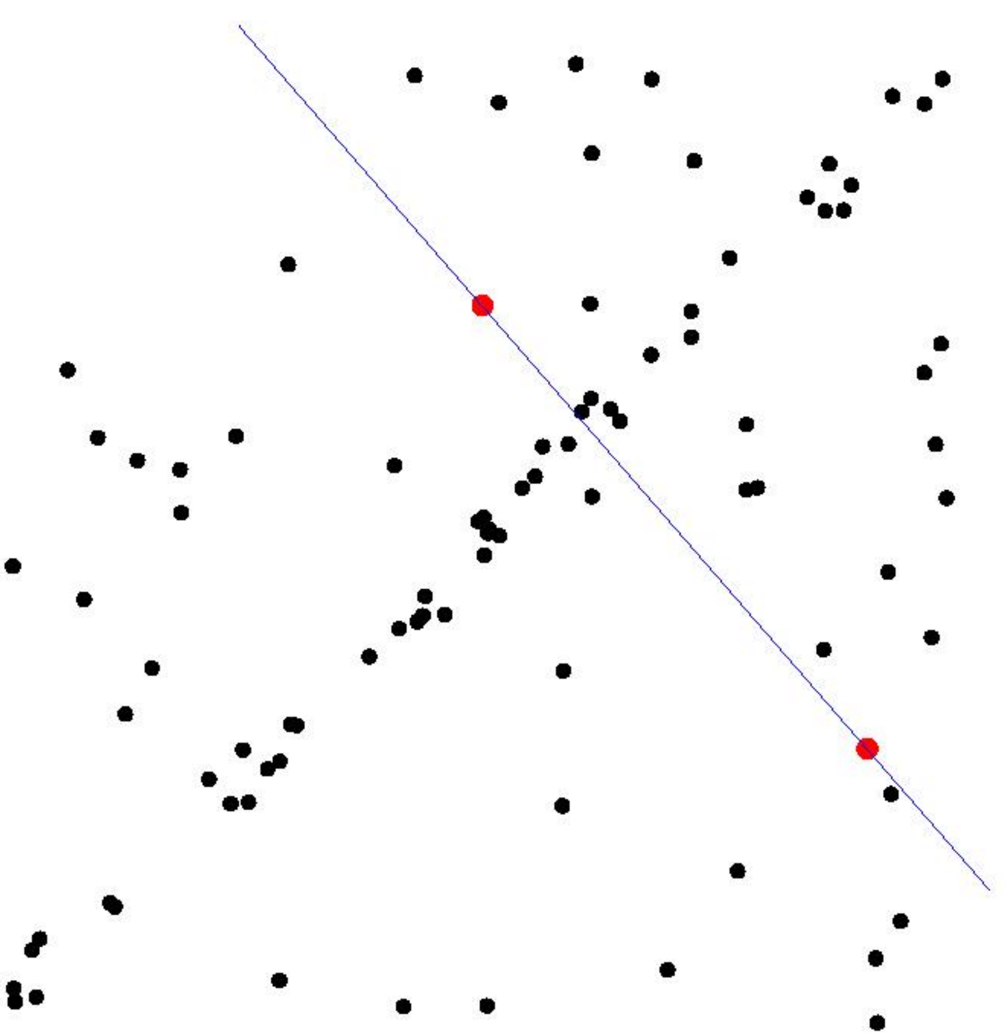
- Such cost function is non-convex

- How to find optimal line parameters?



Select sample of m points at random (here m=2)

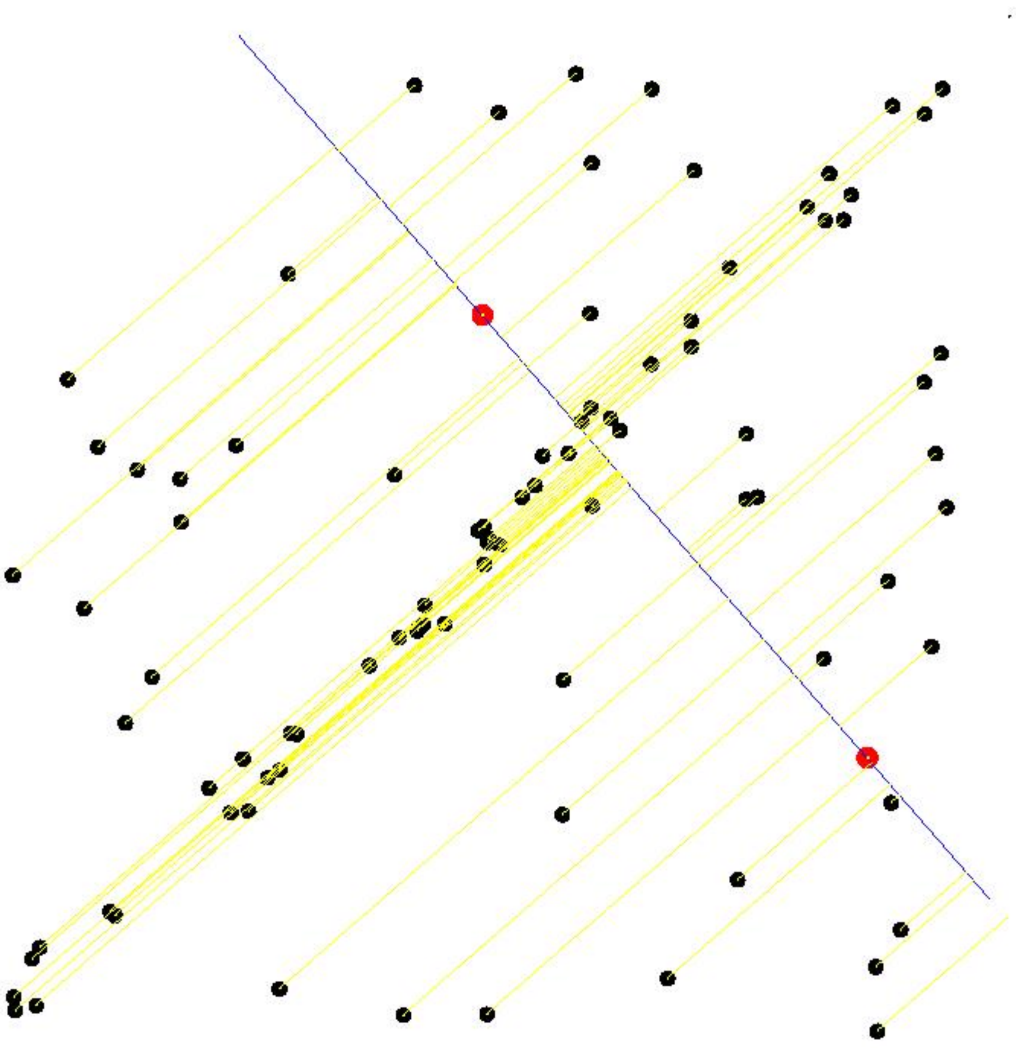
RANSAC



Select sample of m points at random

Estimate model parameters from the data in the sample

RANSAC

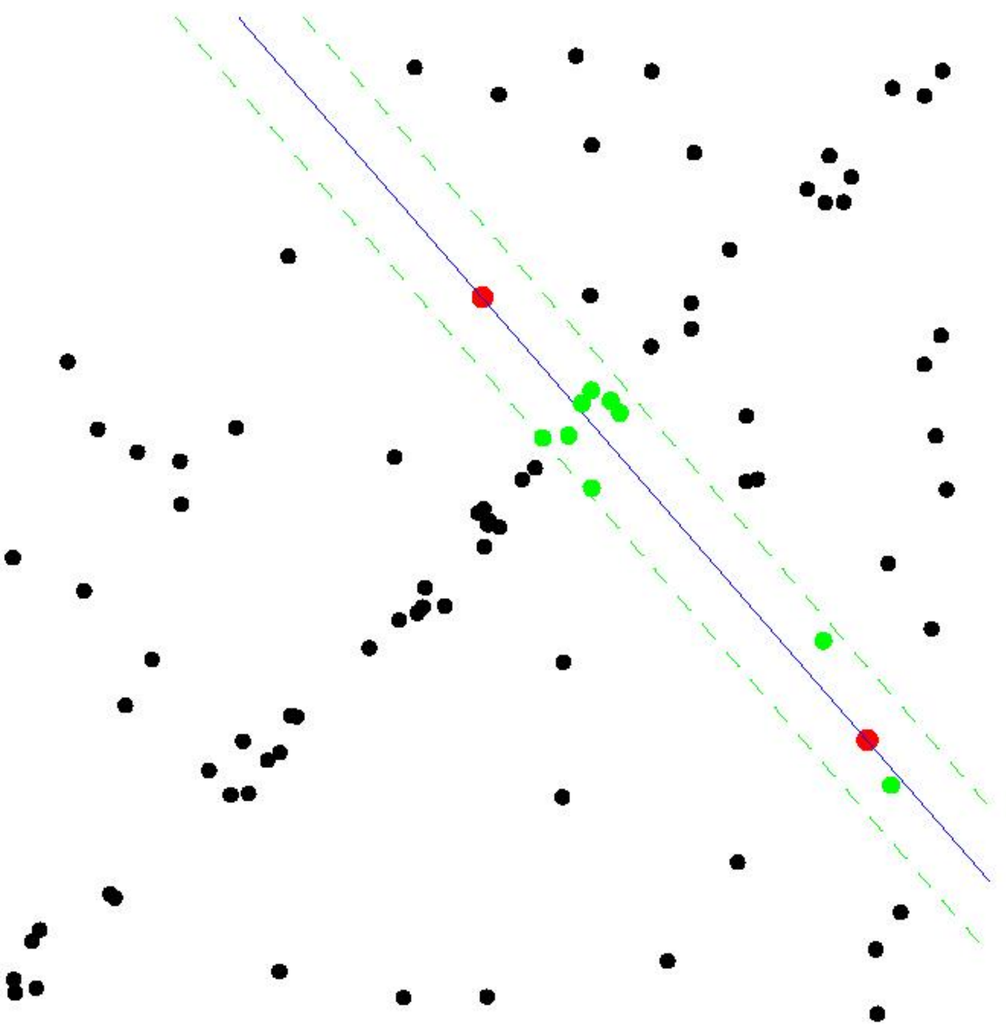


Select sample of m points at random

Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point

RANSAC

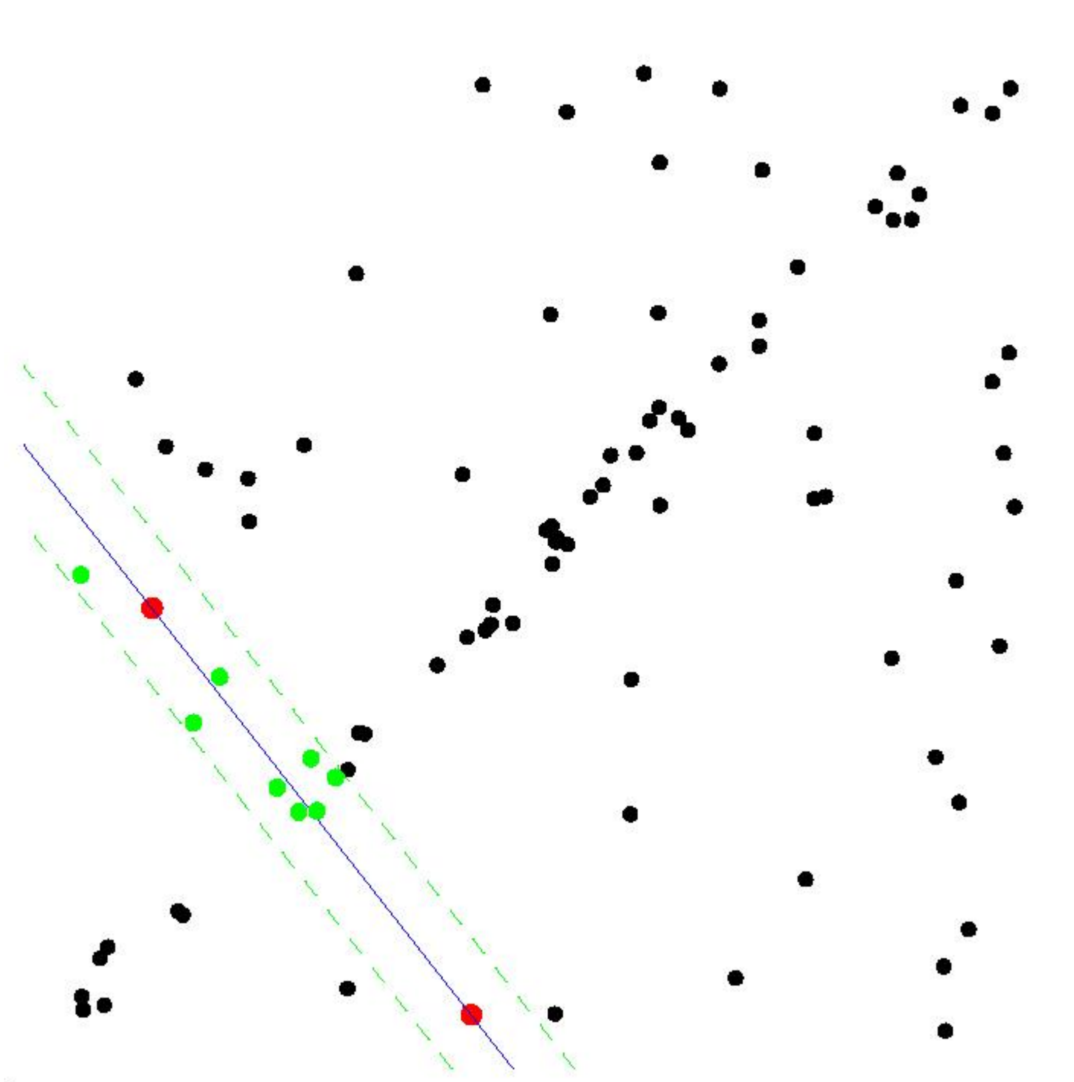


Select sample of m points at random

Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point

Select data that support the current hypothesis



Select sample of m points at random

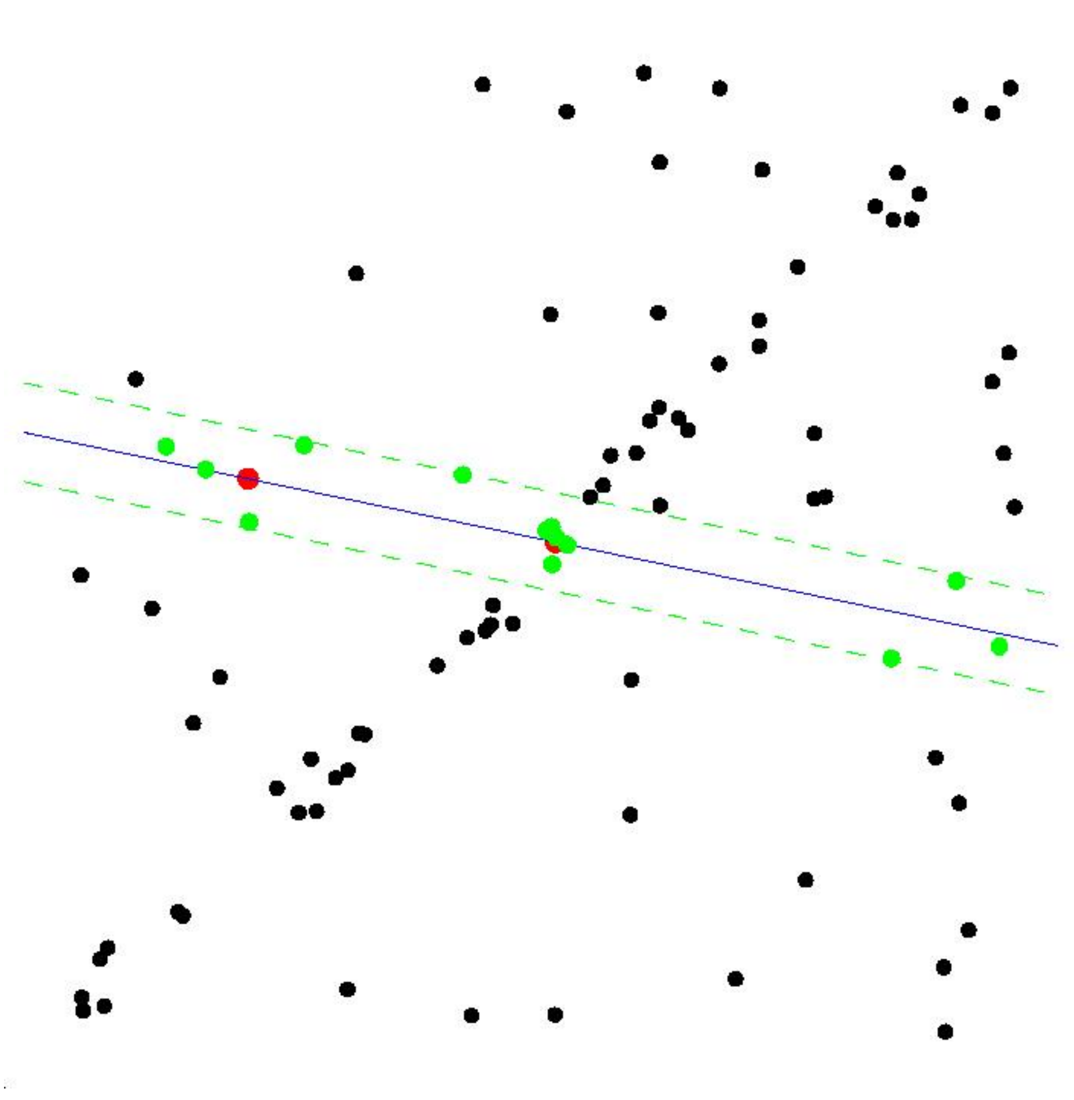
Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point

Select data that support the current hypothesis

Repeat sampling

RANSAC



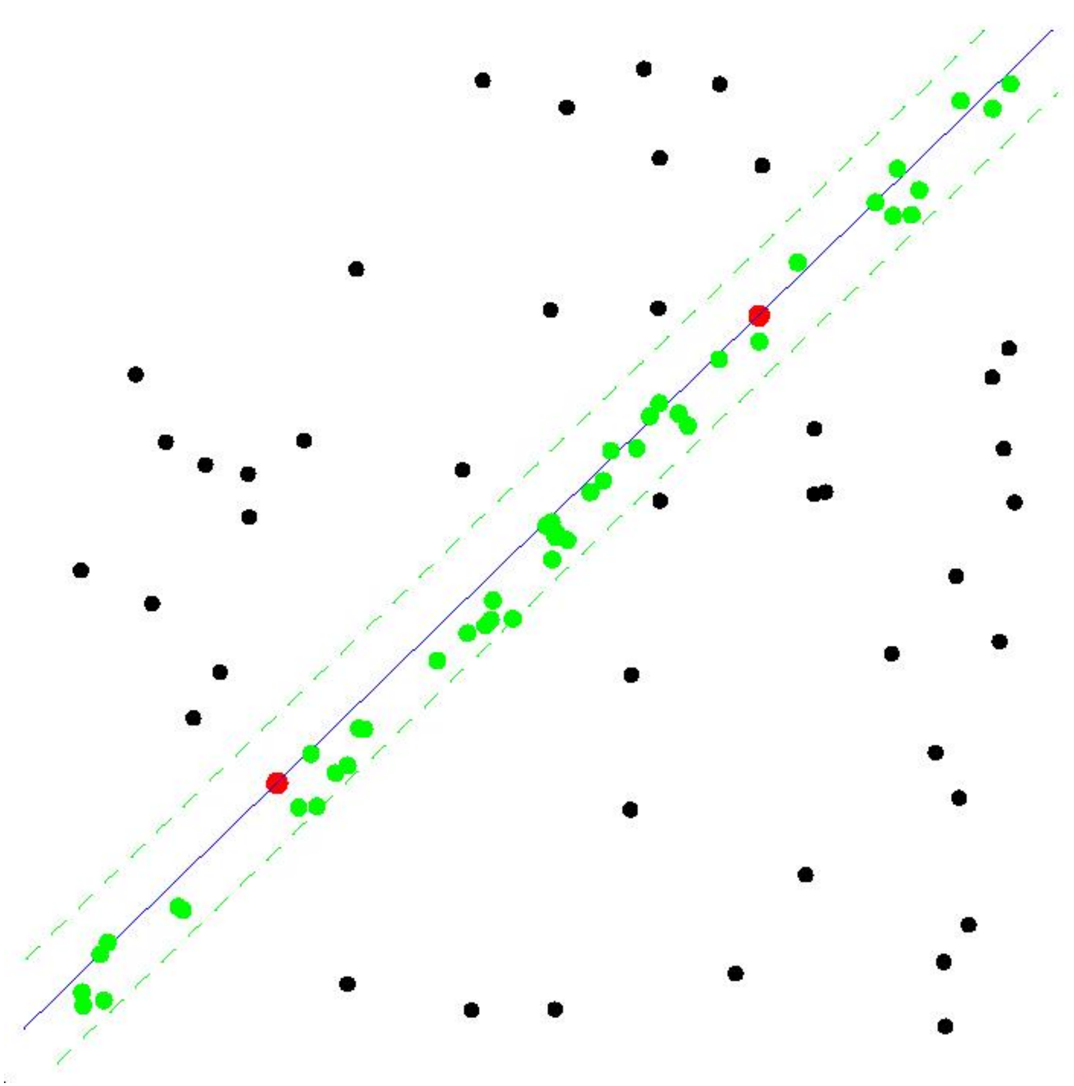
Select sample of m points at random

Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point

Select data that support the current hypothesis

Repeat sampling



Select sample of m points at random

Estimate model parameters from the data in the sample

Evaluate the error (residual) for each data point

Select data that support the current hypothesis

Repeat sampling

RANSAC [Fischler and Bolles 1981]

Input: $\mathcal{X} = \{\mathbf{x}_j\}_{j=1}^{N}$ data points

$e(S) = \theta$ estimates *model parameters* θ given sample $S \subseteq \mathcal{X}$

$$f(\mathbf{x}, \theta) = \begin{cases} 0, & \text{if distance to model} \leq \text{threshold } \sigma \\ 1, & \text{otherwise} \end{cases}$$

Cost function for single data point $\mathbf{x}$

$\Rightarrow J(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$ is #outliers

η – required confidence in the solution, σ – outlier threshold

Output: θ^* parameter of the model minimizing the cost function

$iter \leftarrow 0,\ J^* \leftarrow \infty$
repeat
    Select *random* $S \subseteq \mathcal{X}$ (sample size $m = |S|$) SAMPLING
    Estimate parameters $\theta = e(S)$
    Evaluate $J(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$ VERIFICATION
    If $J(\theta) < J^*$ then SO-FAR-THE-BEST
        $\theta^* \leftarrow \theta,\ J^* \leftarrow J(\theta)$
    $iter \leftarrow iter + 1$
until P(better solution exists) $= f(|\mathcal{X}|, J^*, iter) < \eta$
Compute θ^* from all inliers $\mathcal{X}_{in}$: $\theta^* \leftarrow$ LocalOptimization$(\mathcal{X}_{in}, \theta^*)$

RANSAC – how many samples?

- N Number of points
- Q Number of inliers, $Q = N - J^*$
- m Size of sample
- $\epsilon = Q/N$ Inlier ratio

Probability of all-inlier (uncontaminated) sample:

$$P(\text{inlier sample}) = \frac{\binom{Q}{m}}{\binom{N}{m}} = \frac{Q(Q-1)...(Q-m+1)}{N(N-1)...(N-m+1)} \approx \epsilon^m$$

Mean time for hitting all-inliers sample is proportional to $1/P$.

RANSAC – how many samples?

- How about this formulation:
  - Set the number of samples k such that ***at least one*** pair of points from the line has been hit with probability larger than η
  - Equivalently ... such that ***no*** pair of points from the line has been hit with probability lower than 1 - η
- Q Number of inliers, $Q = N - J^*$
- m Size of sample
- $\epsilon = Q/N$ Inlier ratio

Probability of all-inlier (uncontaminated) sample:

$$P(\text{inlier sample}) = \frac{\binom{Q}{m}}{\binom{N}{m}} = \frac{Q(Q-1)...(Q-m+1)}{N(N-1)...(N-m+1)} \approx \epsilon^m$$

We require:

$$P(\text{bad pair } k \text{ times}) = (1\text{-}P(\text{inlier sample}))^k < 1 - \eta$$

Finding the solution with confidence η therefore requires at least:

$$k \geq \log(1-\eta) \;/\; \log\left(1-\epsilon^m\right)$$

RANSAC termination - How many samples?

Inlier ratio $\epsilon = Q/N$ [%]

Size of the sample m

	15%	20%	30%	40%	50%	70%
2	132	73	32	17	10	4
4	5916	1871	368	116	46	11
7	$1.75 \cdot 10^6$	$2.34 \cdot 10^5$	$1.37 \cdot 10^4$	1827	382	35
8	$1.17 \cdot 10^7$	$1.17 \cdot 10^6$	$4.57 \cdot 10^4$	4570	765	50
12	$2.31 \cdot 10^{10}$	$7.31 \cdot 10^8$	$5.64 \cdot 10^6$	$1.79 \cdot 10^5$	$1.23 \cdot 10^4$	215
18	$2.08 \cdot 10^{15}$	$1.14 \cdot 10^{13}$	$7.73 \cdot 10^9$	$4.36 \cdot 10^7$	$7.85 \cdot 10^5$	1838
30	∞	∞	$1.35 \cdot 10^{16}$	$2.60 \cdot 10^{12}$	$3.22 \cdot 10^9$	$1.33 \cdot 10^5$
40	∞	∞	∞	$2.70 \cdot 10^{16}$	$3.29 \cdot 10^{12}$	$4.71 \cdot 10^6$

computed for $\eta = 0.95$

RANSAC Notes

Pros:

- extremely popular (>17000 citations in Google Scholar)
- used in many applications
- percentage of inliers not needed and not limited
- a probabilistic guarantee for the solution
- mild assumptions: σ known

Cons:

- slow if inlier ratio low
- It was observed experimentally that RANSAC takes several times longer than theoretically expected. This is due to noise – not every all-inlier sample generates a good hypothesis:

$$P(\text{inlier sample}) \neq P(\text{good model estimate})$$

RANSAC Issues, Variants

- **Cost function:** MLESAC, Huber loss, ...
- **Outlier threshold** σ: Least median of Squares, MINPRAN, ...

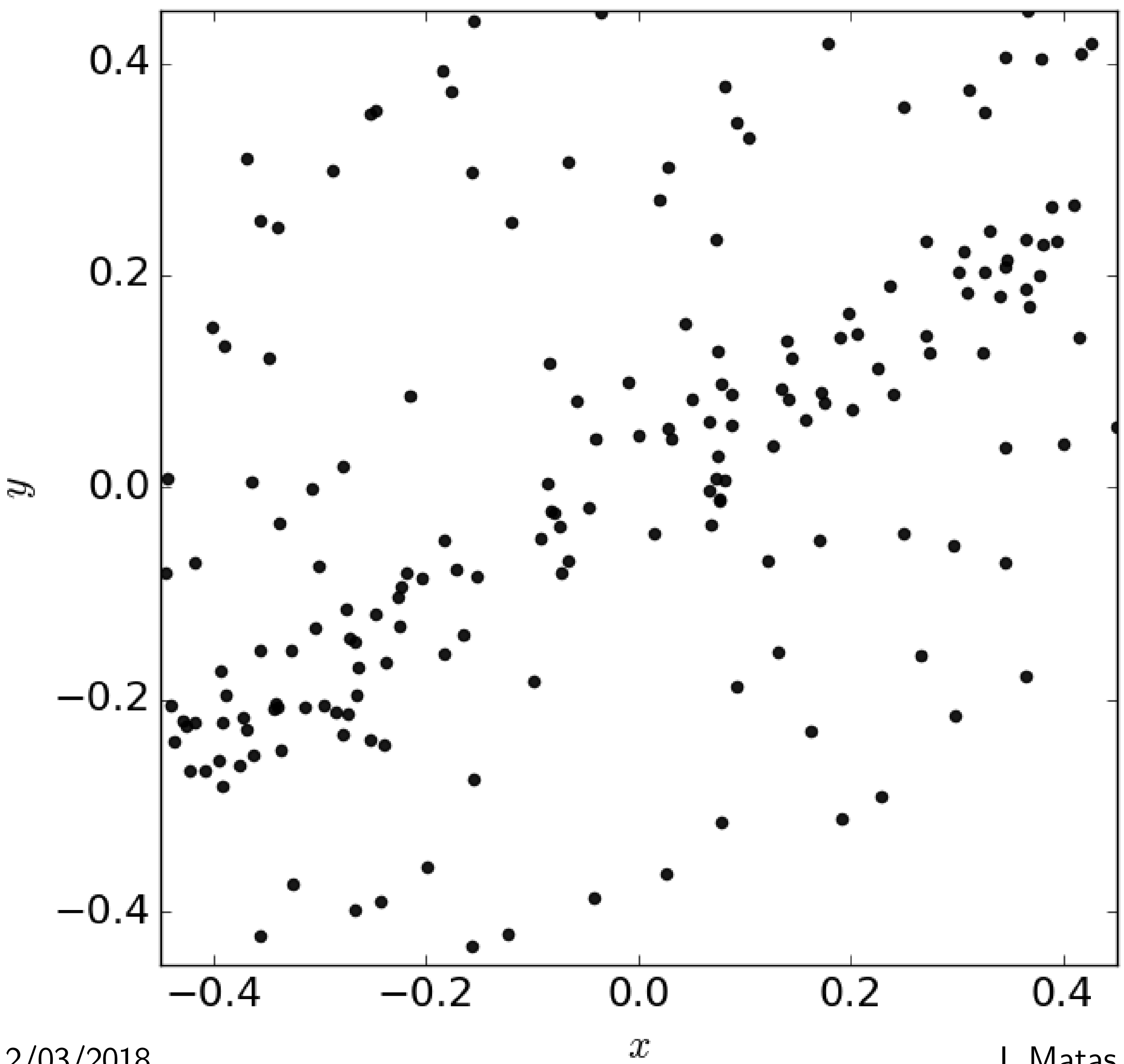
- **Correctness of the results. Degeneracy.**
  Solution: DegenSAC.

- **Accuracy** (parameters are estimated from minimal samples).
  Solution: Locally Optimized RANSAC

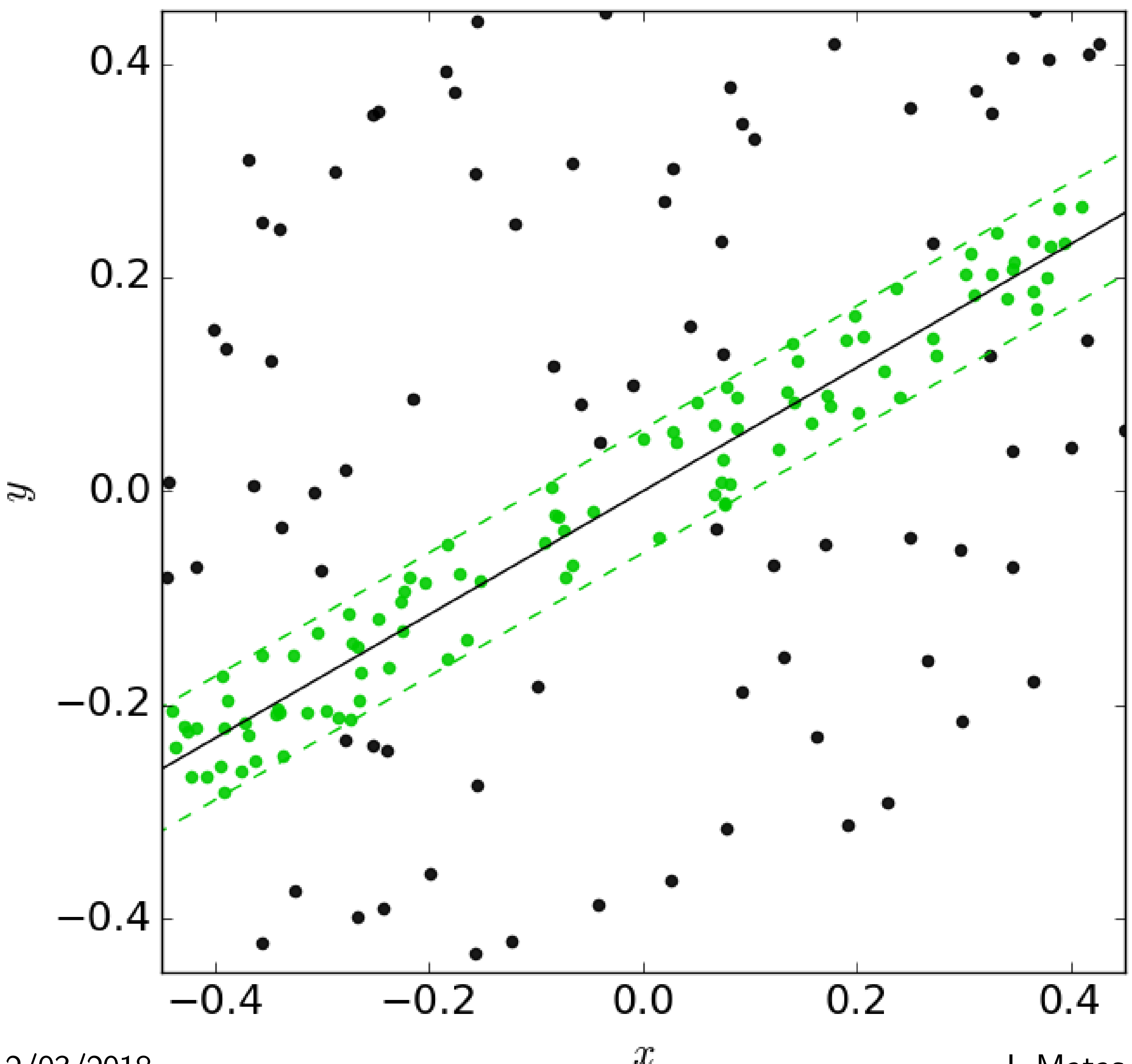
- **Speed:** Running time grows with
  1. number of data points,
  2. number of iterations (polynomial in the inlier ratio)

  Addressing the problem:
  RANSAC with SPRT (WaldSAC), PROSAC

Locally Optimized RANSAC (LO-RANSAC): Problem Intro



Data: 200 points

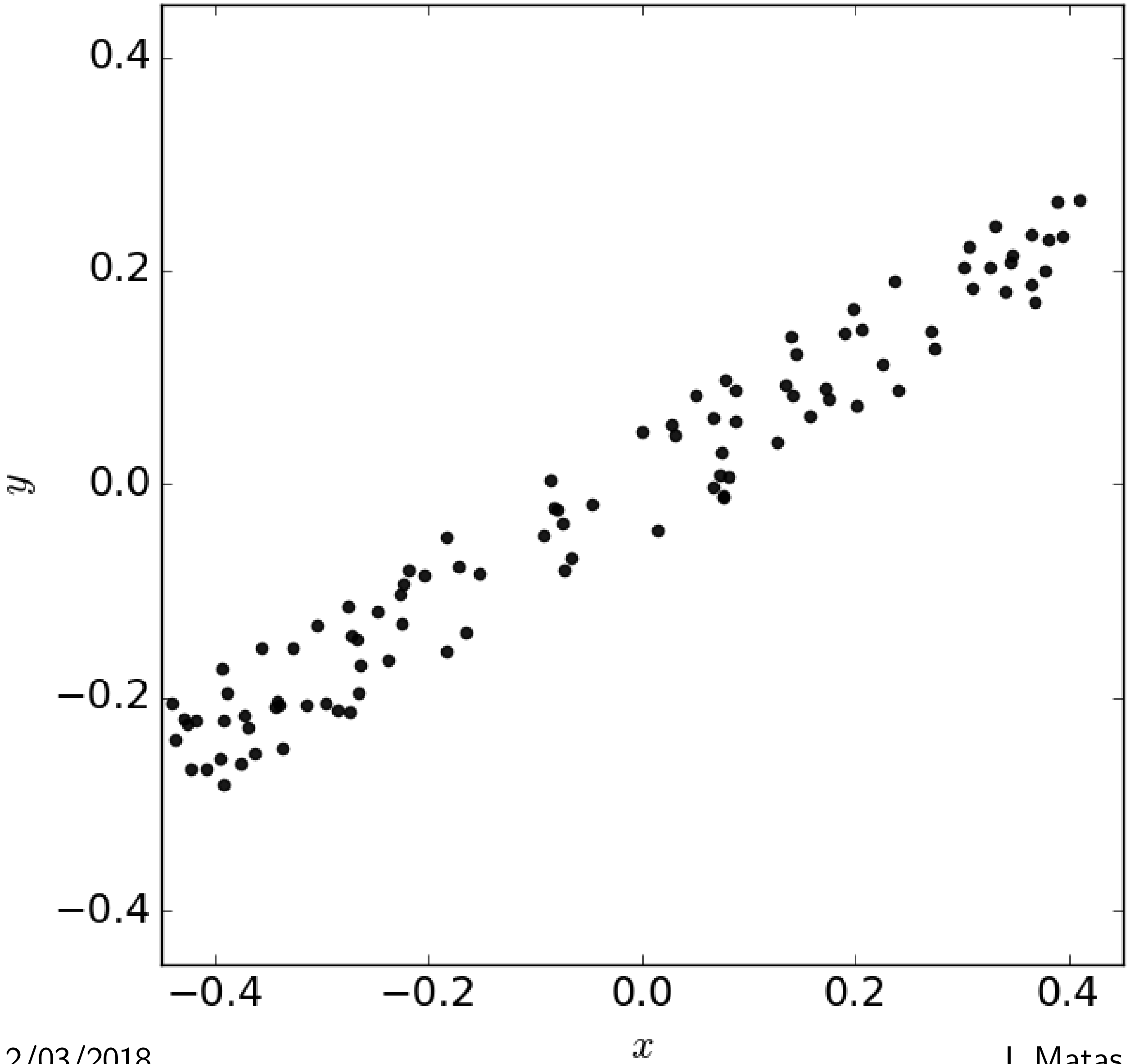


Data: 200 points

Model, 100 inliers

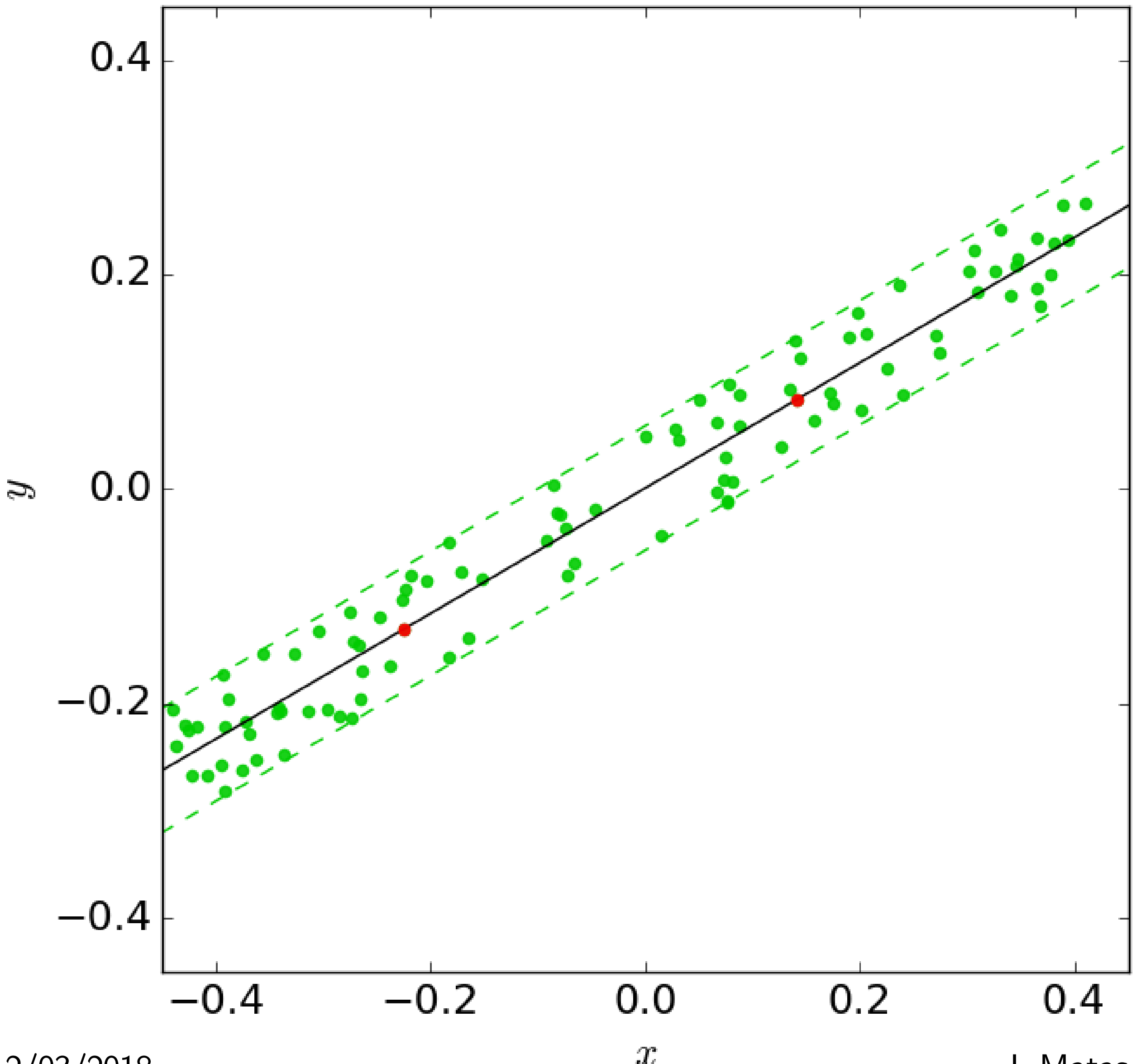
LO-RANSAC: Problem Introduction

For simplicity, consider only points belonging to the model (100 points)



LO-RANSAC: Problem Introduction

For simplicity, consider only points belonging to the model (100 points)



RANSAC

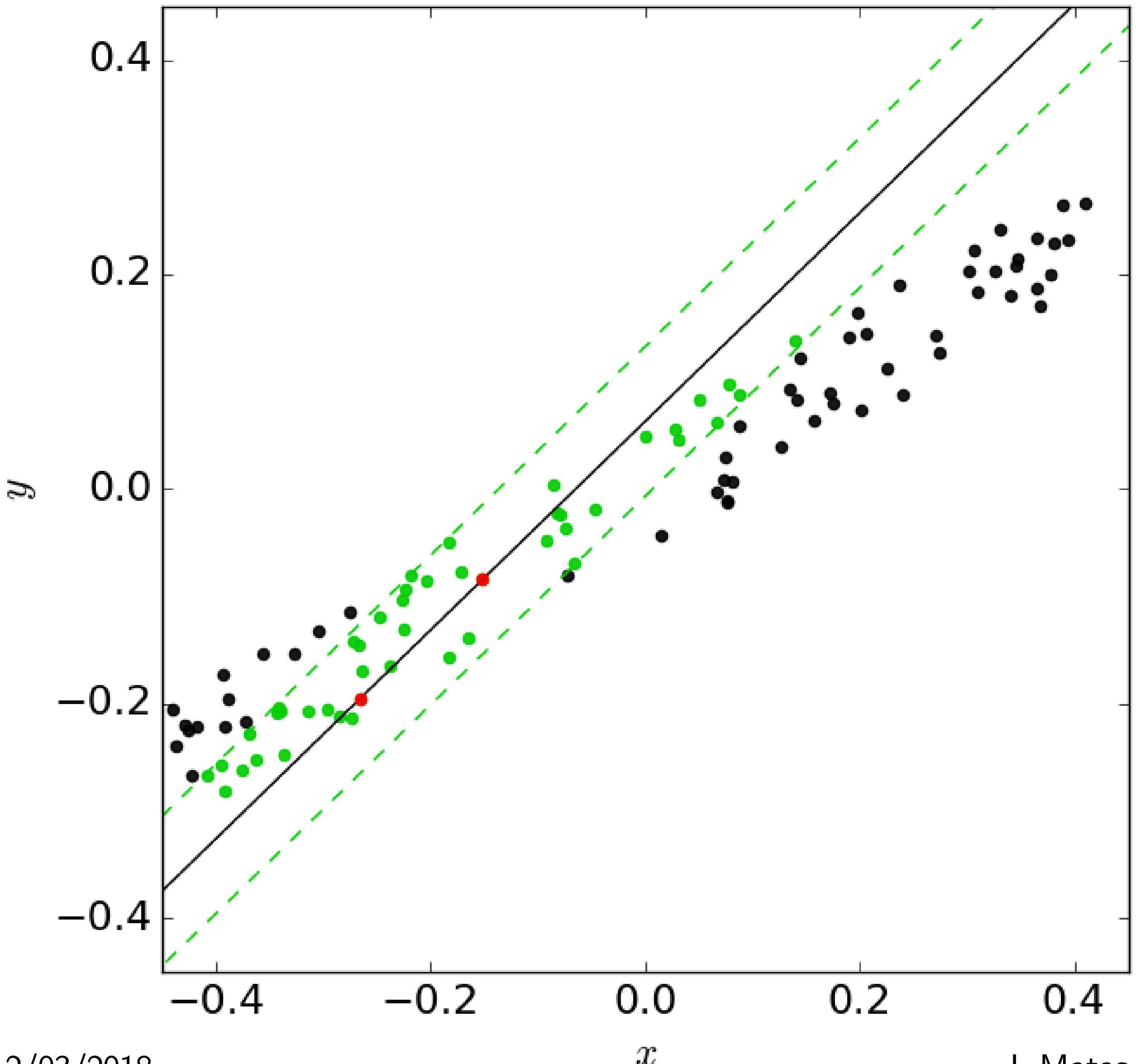
Hypothesis generation from 2 points

Will every two points generate the whole inlier set?

This sample: YES. 100 inliers.

LO-RANSAC: Problem Introduction

For simplicity, consider only points belonging to the model (100 points)



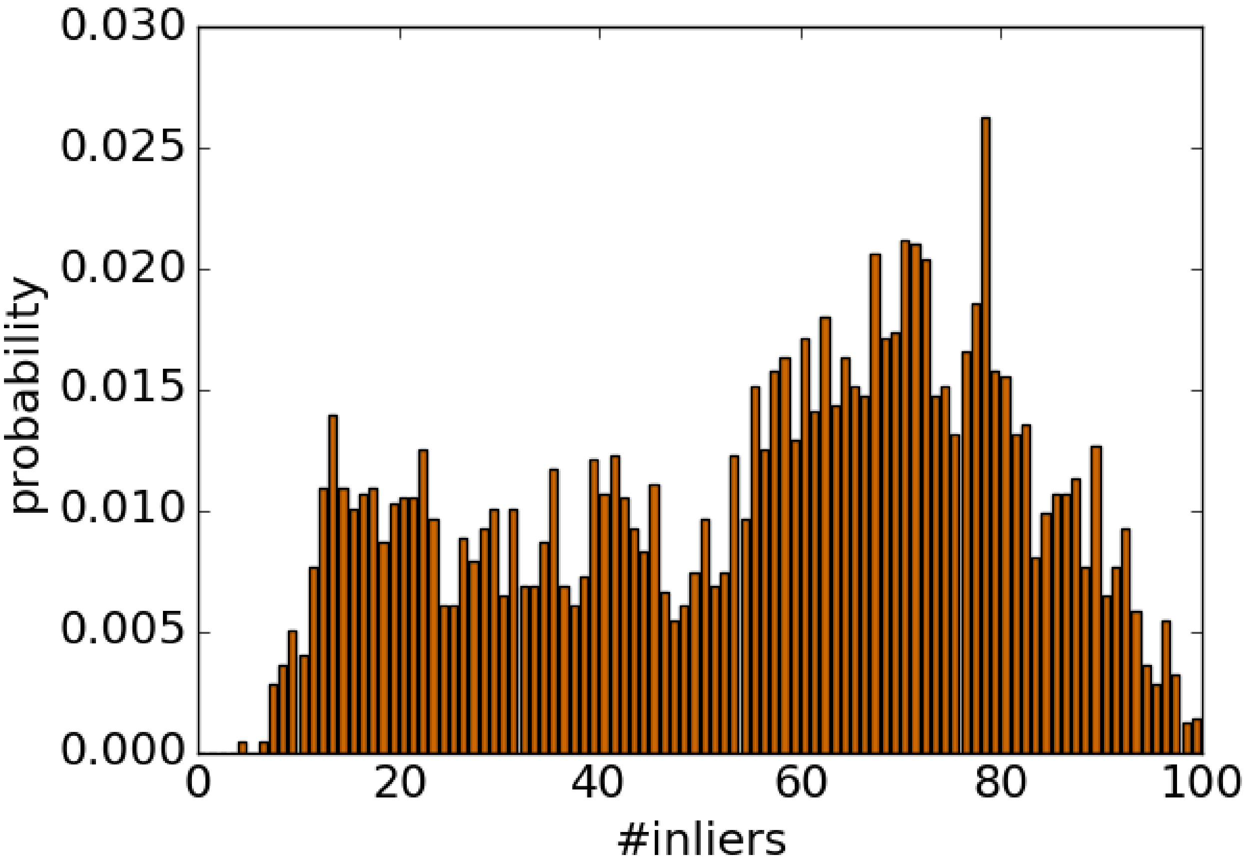
RANSAC

Hypothesis generation from 2 points

Will every two points generate the whole inlier set?

This sample:
NO. 45 inliers.

For simplicity, consider only points belonging to model (100 points)



The distribution of the number of inliers obtained while randomly sampling points pairs

RANSAC

Hypothesis generation from 2 points

Will every two points generate the whole inlier set?

Input: $\mathcal{X} = \{\mathbf{x}_j\}_{j=1}^N$ data points

$e(S) = \theta$ estimates *model parameters* θ given sample $S \subseteq \mathcal{X}$

$$f(\mathbf{x}, \theta) = \begin{cases} 0, & \text{if distance to model } \leq \text{ threshold } \sigma \\ 1, & \text{otherwise} \end{cases}$$

Cost function for single data point $\mathbf{x}$

$\Rightarrow J(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$ is #outliers

η – required confidence in the solution, σ – outlier threshold

Output: θ^* parameter of the model minimizing the cost function

$iter \leftarrow 0,\ J^* \leftarrow \infty$
repeat
  Select *random* $S \subseteq \mathcal{X}$ (sample size $m = |S|$) SAMPLING
  Estimate parameters $\theta = e(S)$
  Evaluate $J(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$ VERIFICATION
  If $J(\theta) < J^*$ then SO-FAR-THE-BEST
    $\theta^* \leftarrow \theta,\ J^* \leftarrow J(\theta)$
  $iter \leftarrow iter + 1$
until $P(\text{better solution exists}) = f(|\mathcal{X}|, J^*, iter) < \eta$
Compute θ^* from all inliers $\mathcal{X}_{in}$: $\theta^* \leftarrow$ LocalOptimization$(\mathcal{X}_{in}, \theta^*)$

LO-RANSAC

Input: $\mathcal{X} = \{\mathbf{x}_j\}_{j=1}^{N}$ data points

$e(S) = \theta$ estimates *model parameters* θ given sample $S \subseteq \mathcal{X}$

$$f(\mathbf{x}, \theta) = \begin{cases} 0, & \text{if distance to model} \leq \text{threshold } \sigma \\ 1, & \text{otherwise} \end{cases}$$

Cost function for single data point $\mathbf{x}$

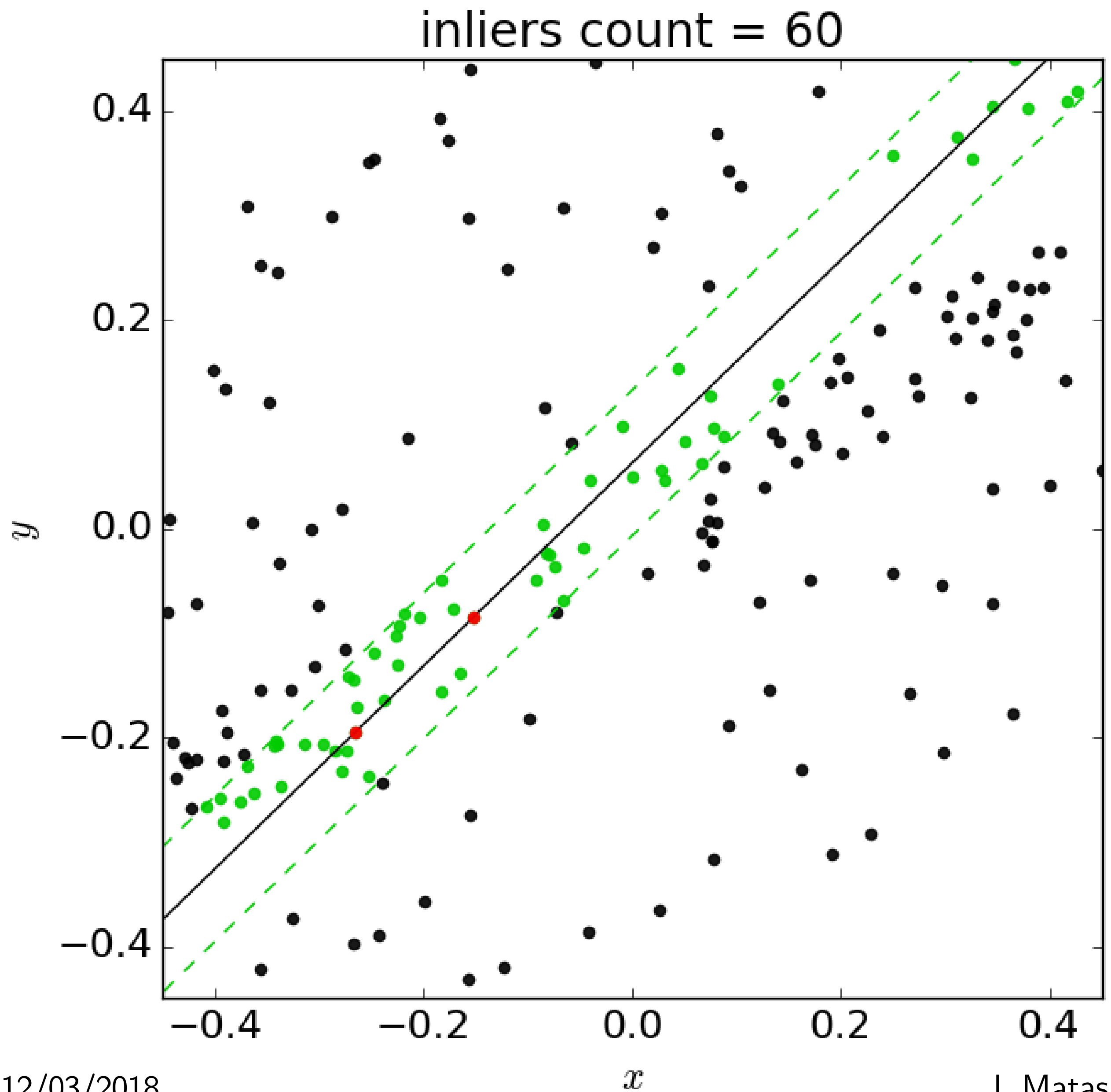
$\Rightarrow J(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$ is #outliers

η – required confidence in the solution, σ – outlier threshold

Output: θ^* parameter of the model minimizing the cost function

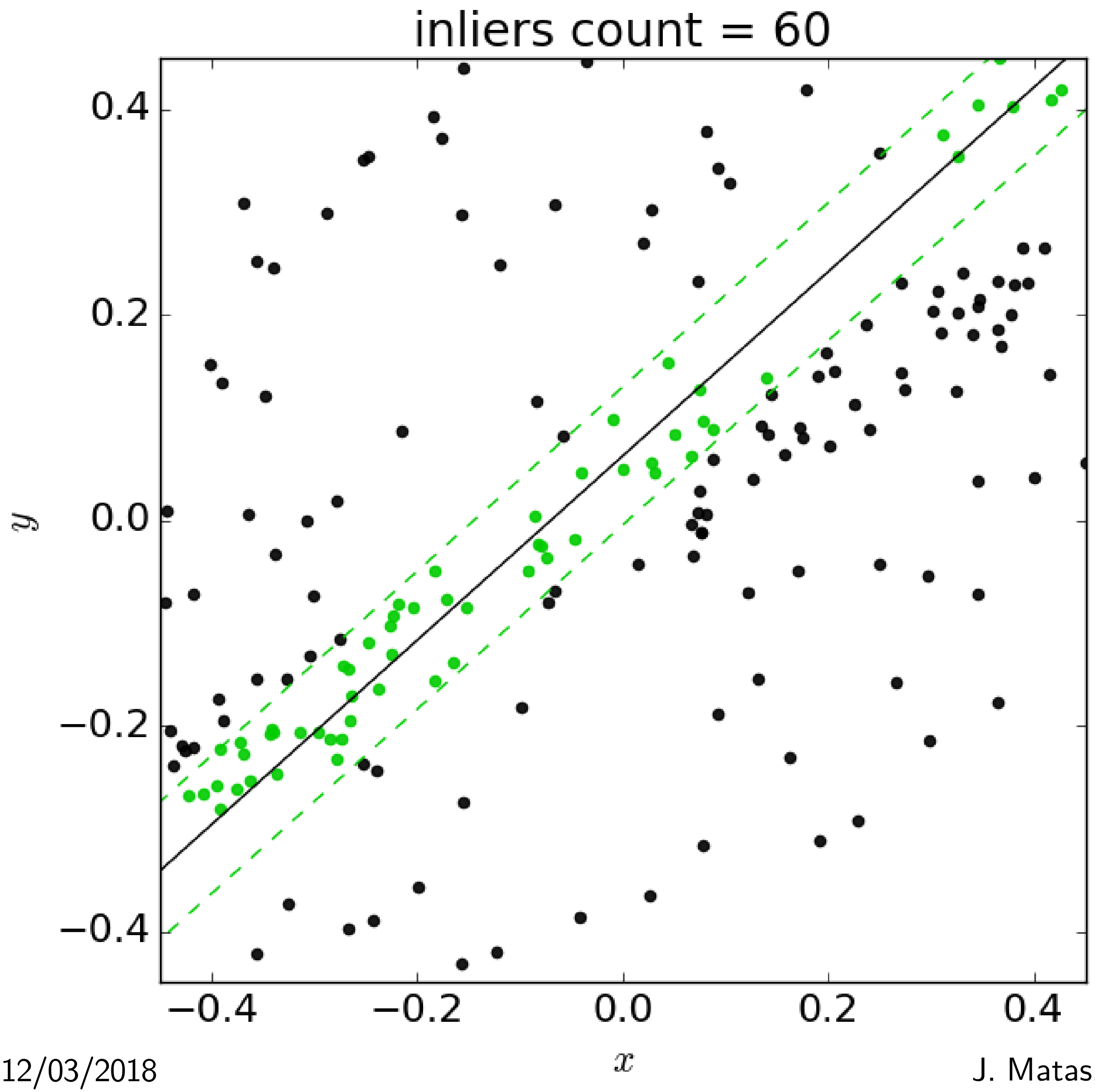
$iter \leftarrow 0,\ J^* \leftarrow \infty$
repeat
    Select *random* $S \subseteq \mathcal{X}$ (sample size $m = |S|$) SAMPLING
    Estimate parameters $\theta = e(S)$
    Evaluate $J(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \theta)$ VERIFICATION
    If $J(\theta) < J^*$ then SO-FAR-THE-BEST
        $\theta^* \leftarrow$ LocalOptimization$(\mathcal{X}_{in}, \theta)$, $J^* \leftarrow J(\theta^*)$
    $iter \leftarrow iter + 1$
until $P(\text{better solution exists}) = f(|\mathcal{X}|, J^*, iter) < \eta$
gone

LO-RANSAC: Example



Init

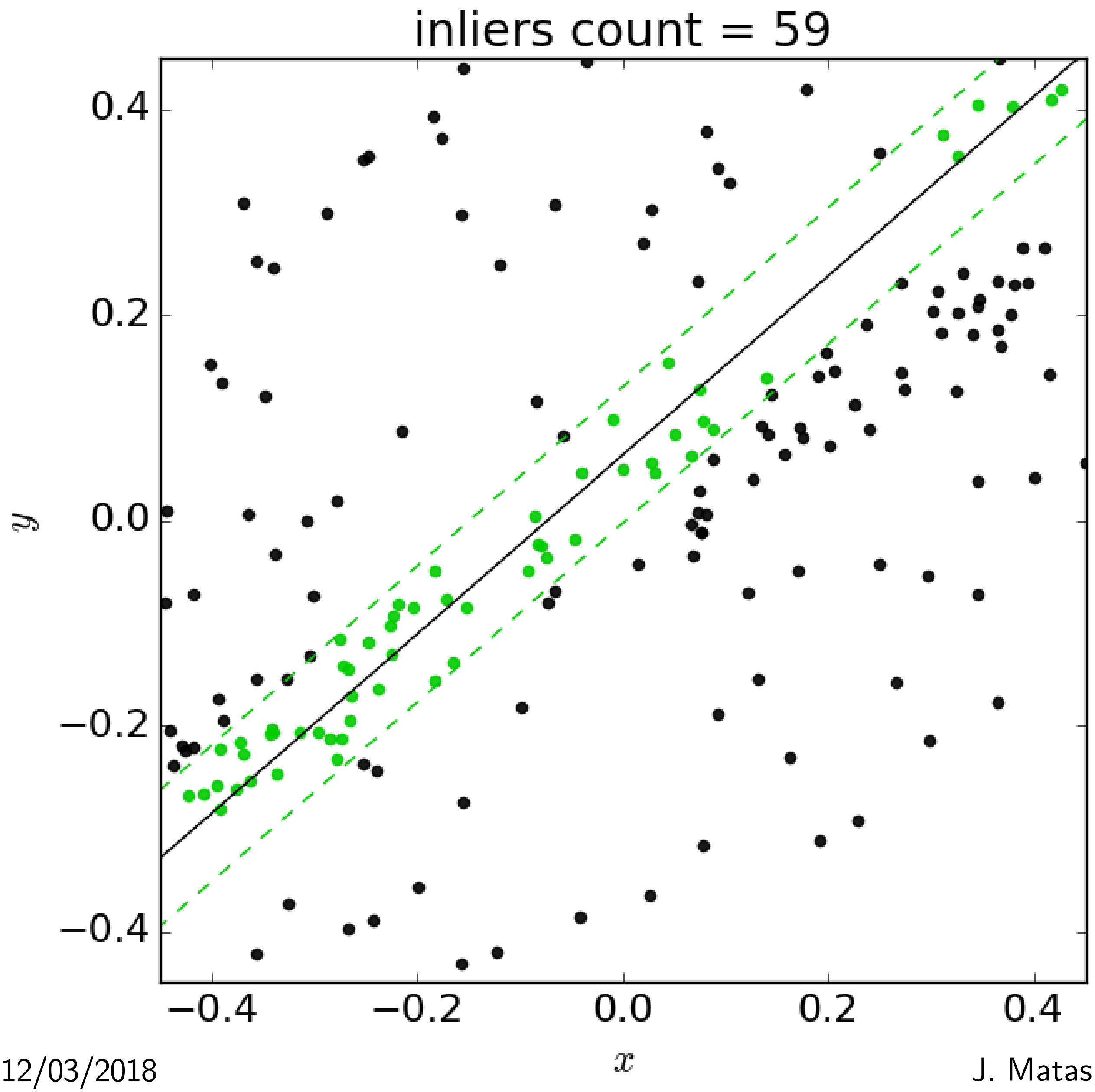
LO-RANSAC: Example



Init

Iteration 1

LO-RANSAC: Example

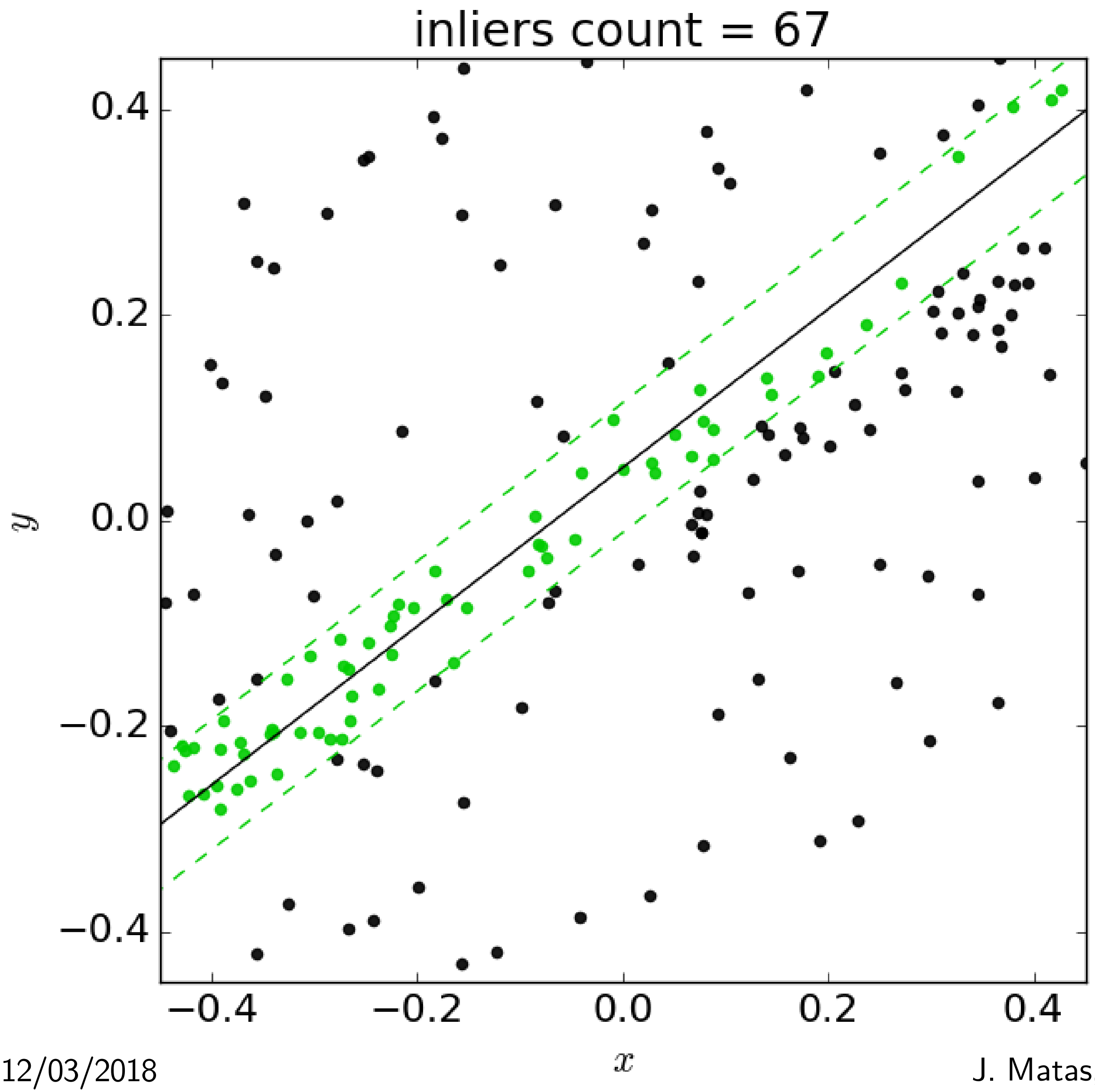


Init

Iteration 1

Iteration 2

LO-RANSAC: Example



Init

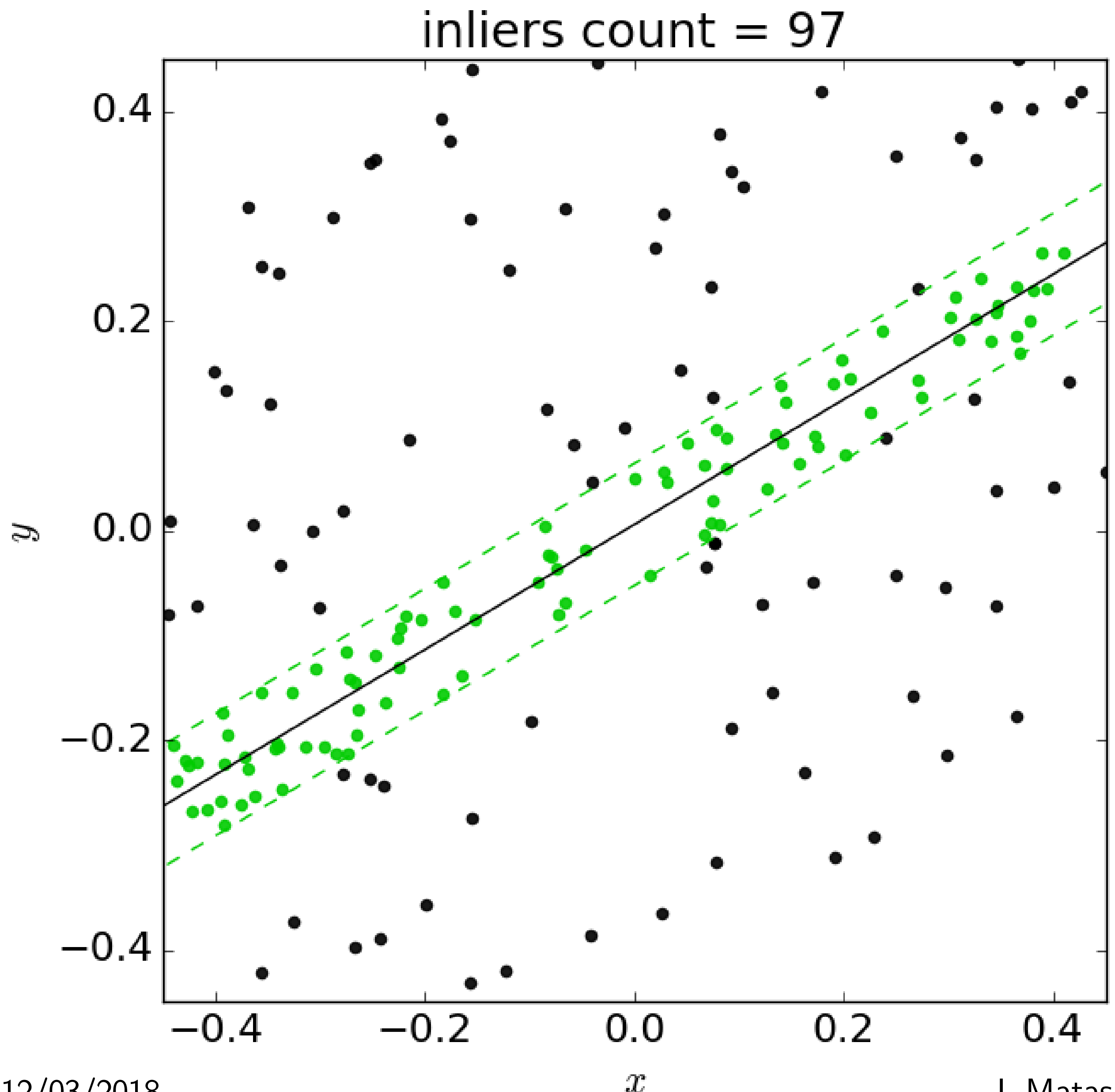
Iteration 1

Iteration 2

...

Iteration 7

LO-RANSAC: Example



Init

Iteration 1

Iteration 2

...

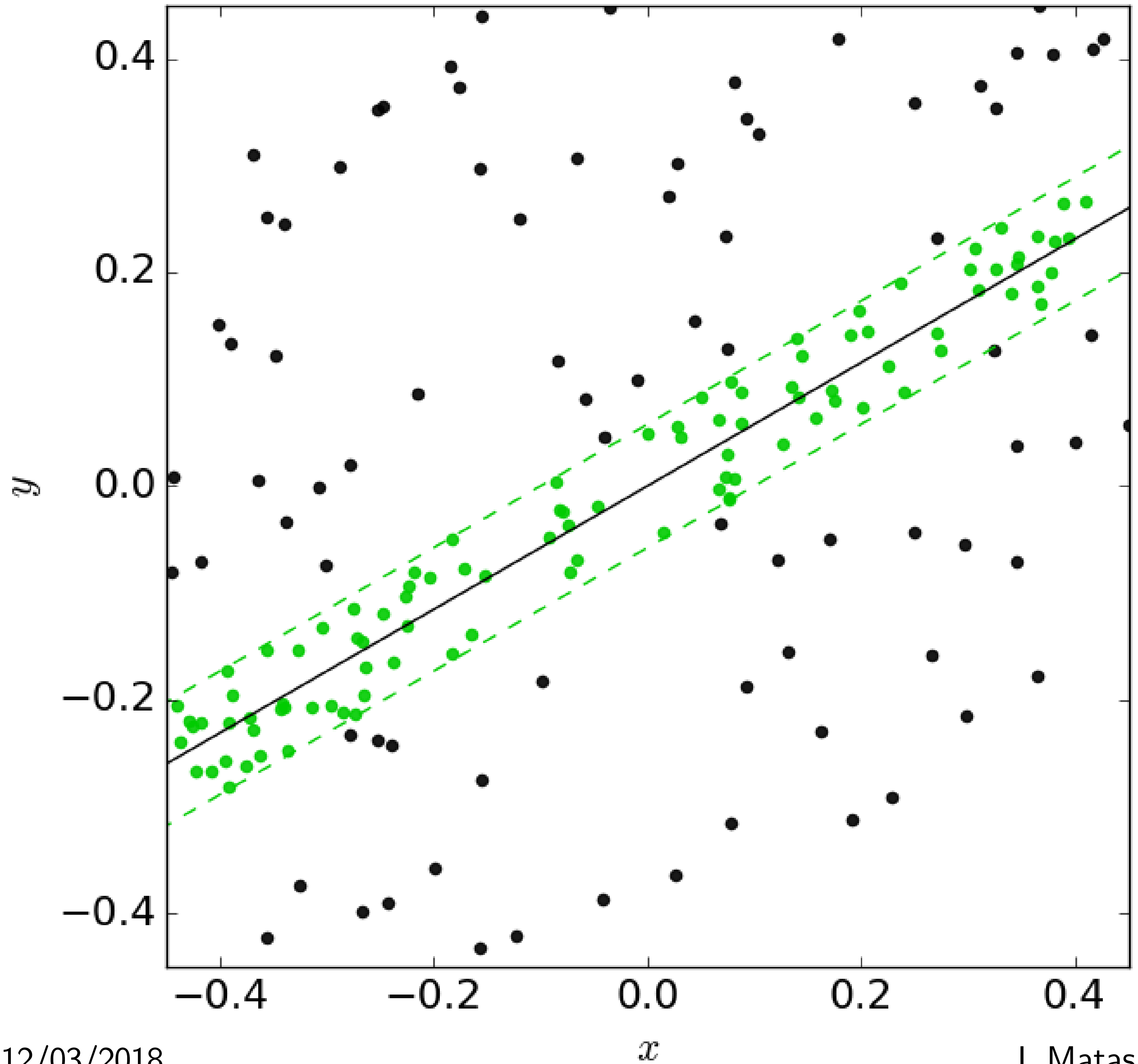
Iteration 7

...

Itration 15

LO-RANSAC: Example

Comparison with model (100 inliers):

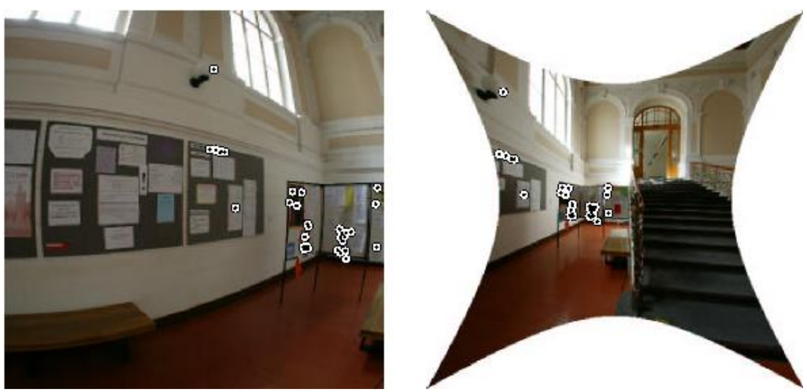


Locally Optimized RANSAC

Estimation of (approximate) models with lower complexity (less data points in the sample) followed by LO step estimating the desired model speeds the estimation up significantly.

The estimation of epipolar geometry is up to 10000 times faster when using 3 region-to-region correspondences rather than 7 point-to-point correspondences.

Simultaneous estimation of radial distortion and epipolar geometry with LO is superior to the state-of the art in both speed a precision of the model.



Fish-eye images by Braňo Mičušík

Chum, Matas, Obdržálek: Enhancing RANSAC by Generalized Model Optimization, *ACCV* 2004

LO-RANSAC: Problem Summary

It was observed experimentally that RANSAC takes several times longer than theoretically expected. This is due to the noise – not every all-inlier sample generates a good hypothesis.

By applying local optimization (LO) to the-best-so-far hypotheses:

(i) a near perfect agreement with theoretical performance

(ii) lower sensitivity to noise and poor conditioning.

The LO is shown to be executed so rarely, log(*iter*) times, that it has minimal impact on the execution time.

RANSAC – Time Complexity

Repeat *k* times (*k* is a function of η, *Q*, *N*)

1. Hypothesis generation
   - Select a sample of *m* data points
   - Calculate parameters of the model(s)
2. Model verification
   - Find the support (consensus set) by
   - verifying all *N* data points

t_M – time needed to draw a sample

$\overline{m}_s$ – average number of models per sample

Total **running** time:

$$t = k(t_M + \overline{m}_s N)$$

Randomised RANSAC [Matas, Chum 02]

Repeat $k/(1-\alpha)$ times

1. Hypothesis generation
2. Model pre-verification $T_{d,d}$ test

- Verify d << N data points, reject
- the model if not all d data points
- are consistent with the model

3. Model verification

   Verify the rest of the data points

V – average number of data points verified

α – probability that a good model is rejected by $T_{d,d}$ test

$$t = \frac{k}{1-\alpha}(t_M + \overline{m}_s V)$$

Optimal Randomised Strategy

Model Verification is Sequential Decision Making

$$H_g\colon\ P(x_i = 1|H_g) \geq \varepsilon$$

$$H_b\colon\ P(x_i = 1|H_b) = \delta$$

$x_i = 1$ $\quad$ x_i is consistent with the model

where
H_g - hypothesis of a `good` model (≈ from an uncontaminated sample)
H_b - hypothesis of a `bad` model, (≈ from a contaminated sample)
δ - probability of a data point being consistent with an arbitrary model

Optimal (the fastest) test that ensures with probability α that that H_g is not incorrectly rejected is the Sequential probability ratio test (SPRT) [Wald47]

SPRT [simplified from Wald 47]

Compute the likelihood ratio $\lambda_i = \prod_{j=1}^{i} \frac{P(x_j|H_b)}{P(x_j|H_g)}$

if $\lambda_i > A$ reject the model

if $i = N$ accept model as 'good'

Two important properties of SPRT:

1. probability of rejecting a \good\ model $\alpha < 1/A$
2. average number of verifications $V = C \log(A)$

$$C \approx \left(P(0|H_b) \log \frac{P(0|H_b)}{P(0|H_g)} + P(1|H_b) \log \frac{P(1|H_b)}{P(1|H_g)} \right)^{-1}$$

SPRT properties

1. Probability of rejecting a \good\ model $\alpha=1/A$

$$\lambda_i = \prod_{j=1}^{i} \frac{P(x_j|H_b)}{P(x_j|H_g)} = \frac{P(x|H_b)}{P(x|H_g)}, x = (x_1, \ldots, x_i)$$

If $\lambda_i > A$ then $P(x|H_g) < P(x|H_b)/A$, therefore

$$\alpha = \int_{\lambda_i > A} P(x|H_g)dx < \int_{\lambda_i > A} P(x|H_b)/A dx =$$

$$= \frac{1}{A}\int_{\lambda_i > A} P(x|H_b)dx \leq \frac{1}{A}\int P(x|H_b)dx = \frac{1}{A}$$

WaldSAC

Repeat k/(1-1/A) times

1. Hypothesis generation
2. Model verification, use SPRT

$$\overline{m}_S \cdot C \log A$$

$$C \approx ((1-\delta)\log\frac{1-\delta}{1-\varepsilon} + \delta \log\frac{\delta}{\varepsilon})^{-1}$$

$$t(A) = \frac{k}{(1-1/A)}(t_M + \overline{m}_S C \log A)$$

In sequential statistical decision problem decision errors are traded off for time. These are two incomparable quantities, hence the constrained optimization.

In WaldSAC, decision errors cost time (more samples) and there is a single minimised quantity, time t(A), a function of a single parameter A.

Optimal test (optimal A) given ε and δ

Optimal A*

$$A^* = \arg\min_A t(A)$$

Optimal A* found by solving

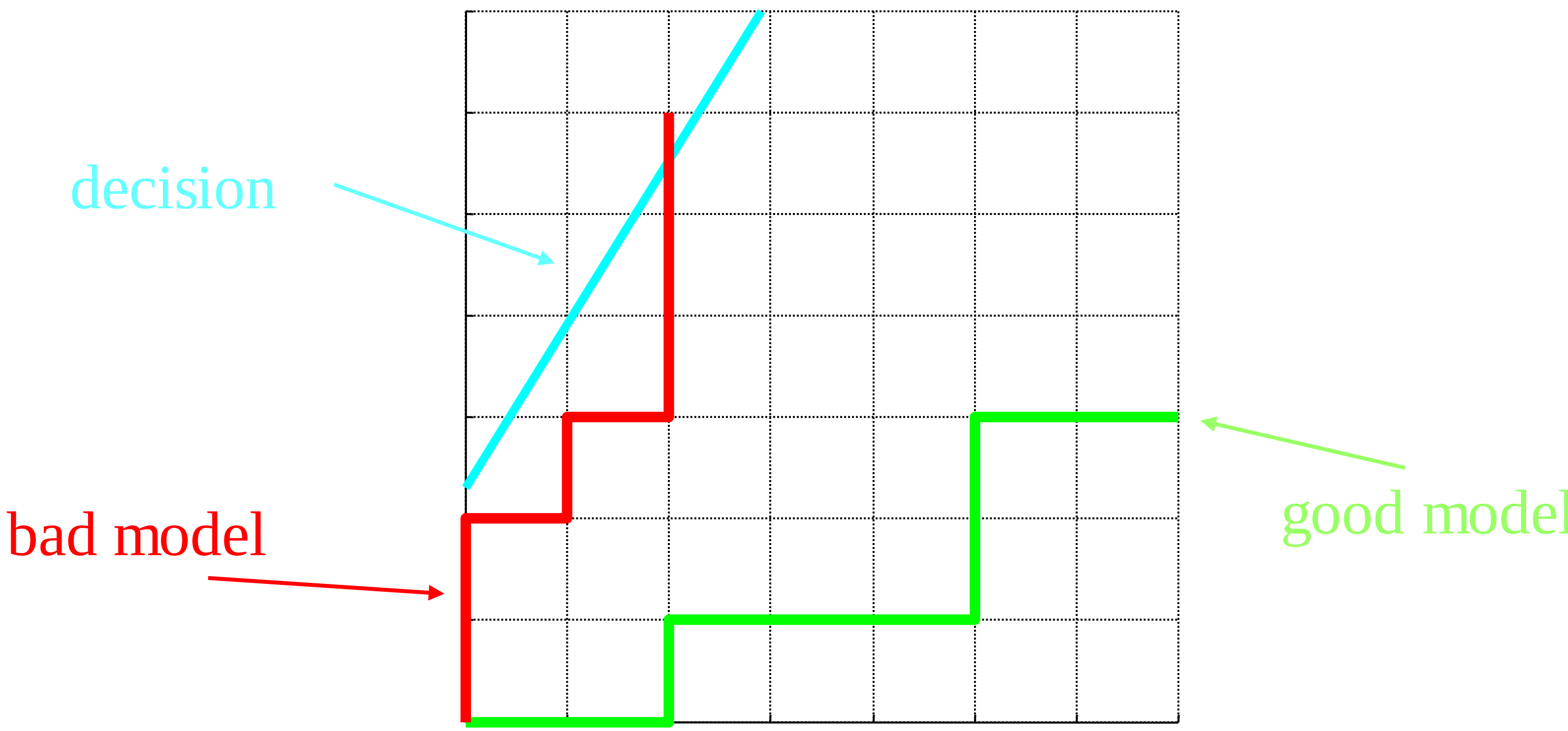
$$\frac{\partial t}{\partial A} = 0$$

$$A^* = \frac{t_M}{\overline{m}_s C} + 1 + \log A^*$$

$$A^* = \lim_{n \to \infty} A_n$$

$$A_0 = \frac{t_M}{\overline{m}_s C} + 1, \quad A_{n+1} = \frac{t_M}{\overline{m}_s C} + 1 + \log A_n$$

SPRT

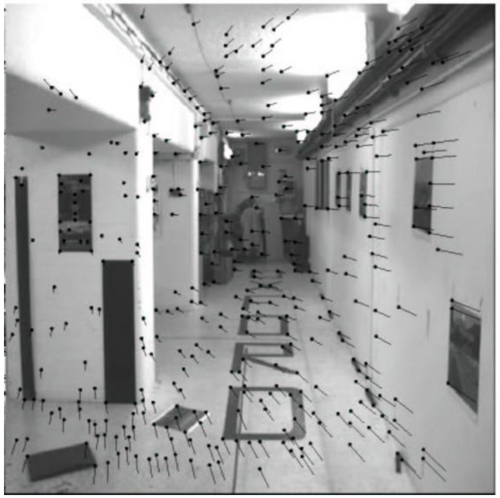


Note: the Wald's test is equivalent to series of $T(d,c)$, where $c = \lceil (\log A - d \log \lambda_1) / \log \lambda_0 \rceil$

Exp. 1: Wide-baseline matching

	samples	models	V	time	spd-up
R	2914	7347	110.0	1099504	1.0
R-R	7825	19737	3.0	841983	1.3
Wald	3426	8648	8.2	413227	2.7

Exp. 2 Narrow-baseline stereo



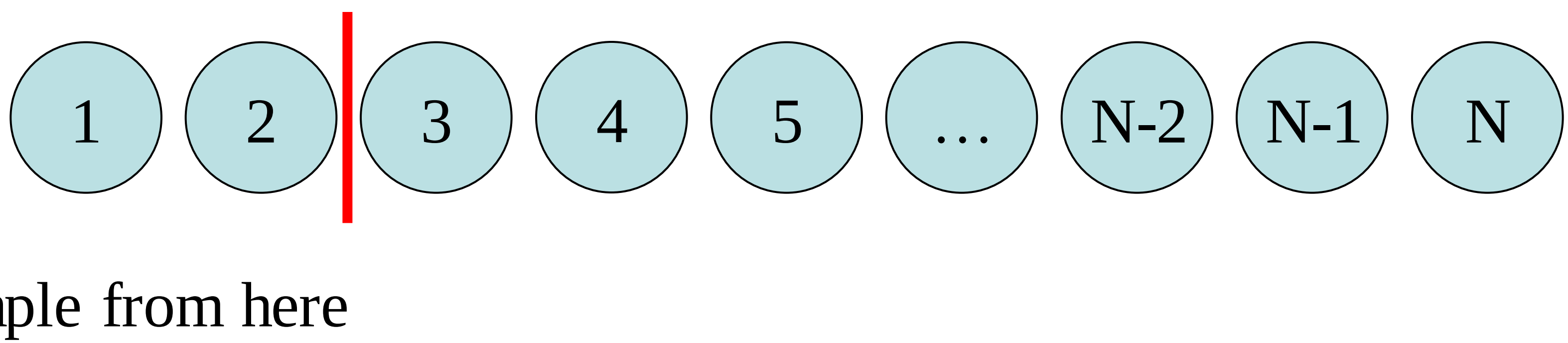
	samples	models	V	time	spd-up
R	155	367	600.0	235904	1.0
R-R	247	587	86.6	75539	3.1
Wald	162	384	23.1	25032	9.4

Randomised Verification in RANSAC: Conclusions

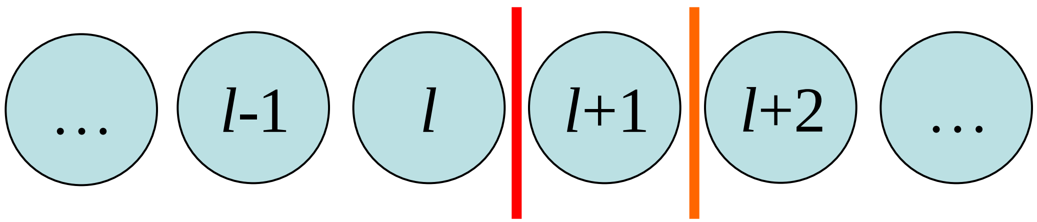
- The same confidence η in the solution reached faster (data dependent, $\approx$ 10x)
- No change in the character of the algorithm, it was randomised anyway.
- Optimal strategy derived using Wald`s theory for known ε and δ.
- Results with ε **and** δ estimated during the course of RANSAC are not significantly different. Performance of SPRT is insensitive to errors in the estimate.
- δ can be learnt, an initial estimate can be obtained by geometric consideration
- Lower bound on e is given by the best-so-far support
- Note that the properties of WaldSAC are quite different from preemptive RANSAC!

PROSAC – PROgressive SAmple Consensus

- Not all correspondences are created equally
- Some are better than others
- Sample from the best candidates first



PROSAC Samples



Draw T_l samples from $(1 \ldots l)$

Draw T_{l+1} samples from $(1 \ldots l+1)$

Samples from $(1 \ldots l)$ that are not from $(1 \ldots l+1)$ contain $l+1$

Draw $T_{l+1} - T_l$ samples of size m-1 and add $l+1$

Degenerate Configurations

The presence of degenerate configuration causes RANSAC to fail in estimating a correct model, instead a model consistent with the degenerate configuration and some outliers is found.

The DEGENSAC algorithm handles scenes with:
• all points in a single plane
• majority of the points in a single plane and the rest off the plane
• no dominant plane present

No a-priori knowledge of the type of the scene is required

Chum, Werner, Matas: Epipolar Geometry Estimation unaffected by dominant plane, *CVPR* 2005