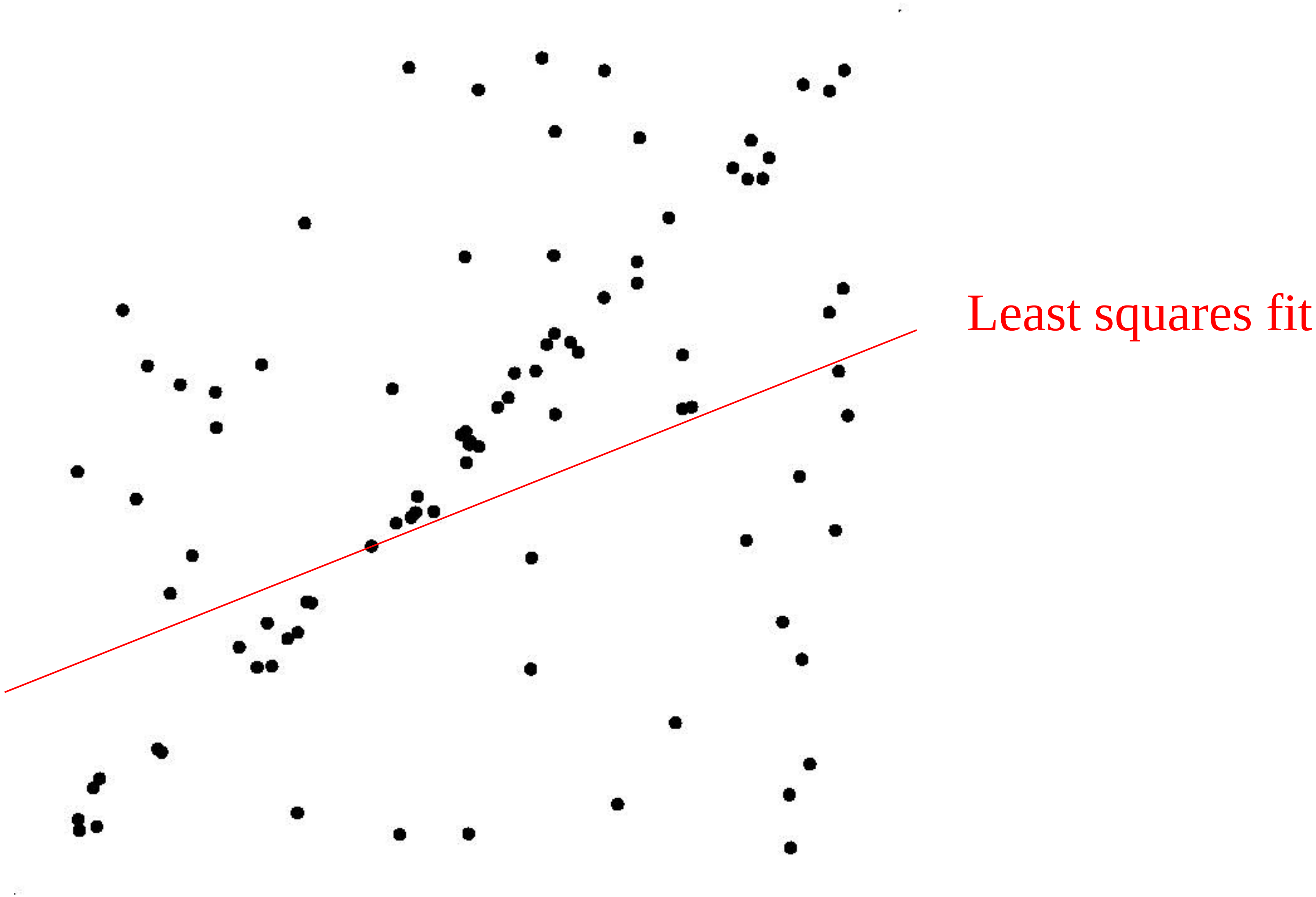
RANSAC

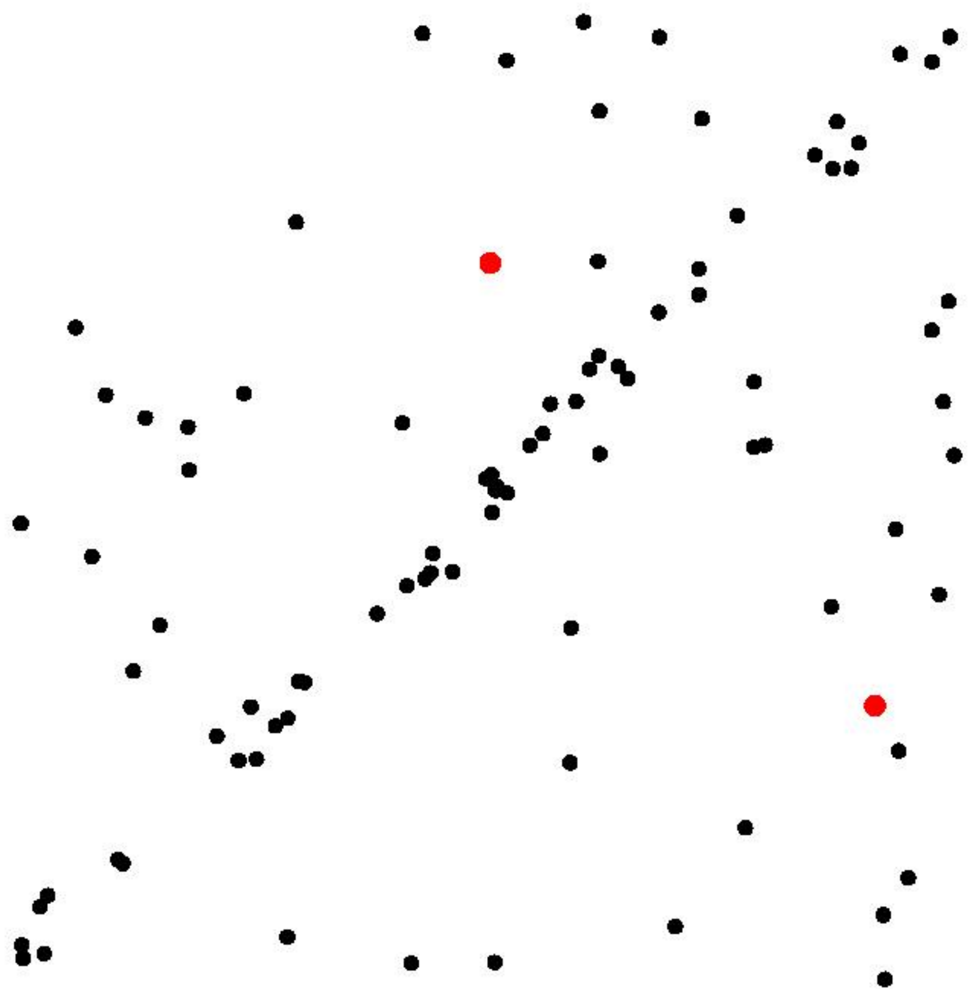
Robust model estimation from data contaminated by outliers

Ondřej Chum

Fitting a Line

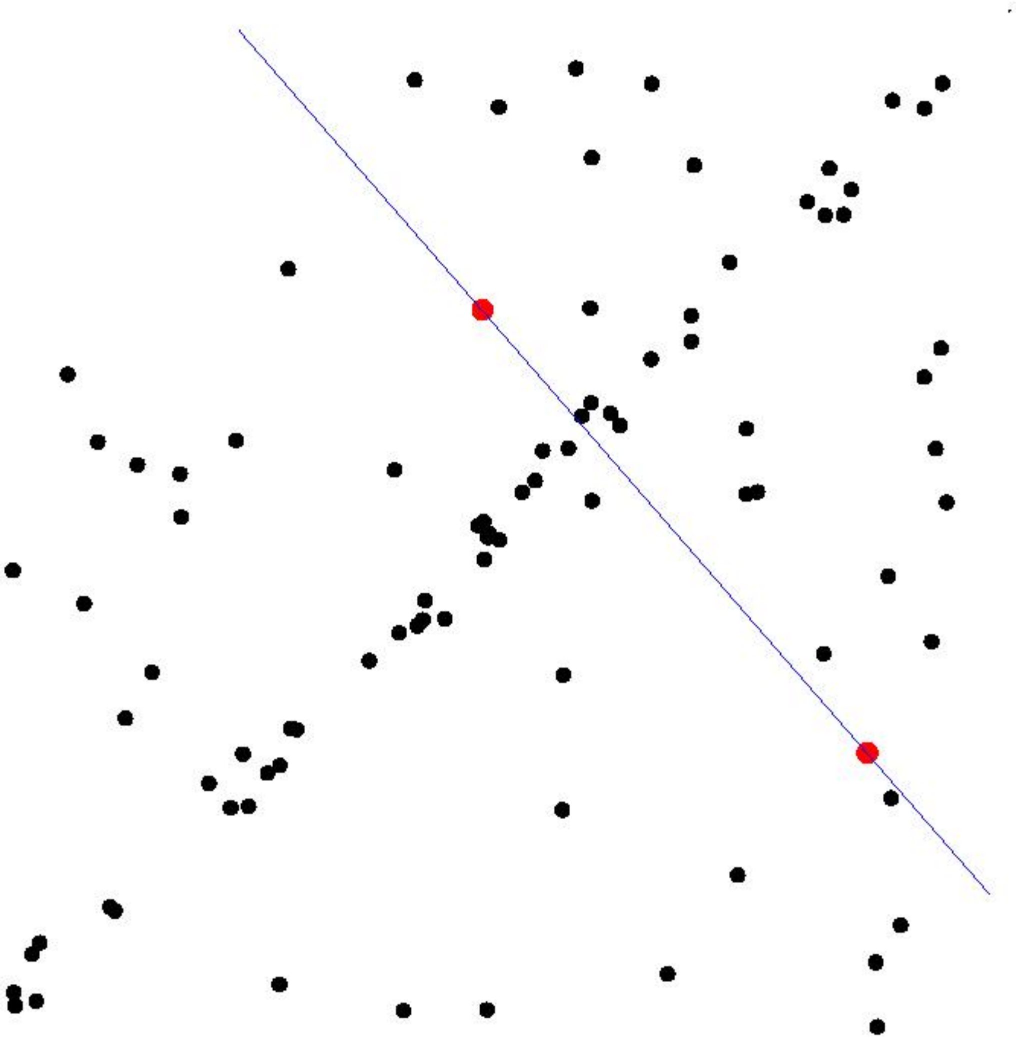


RANSAC



• **Select sample of m points at random**

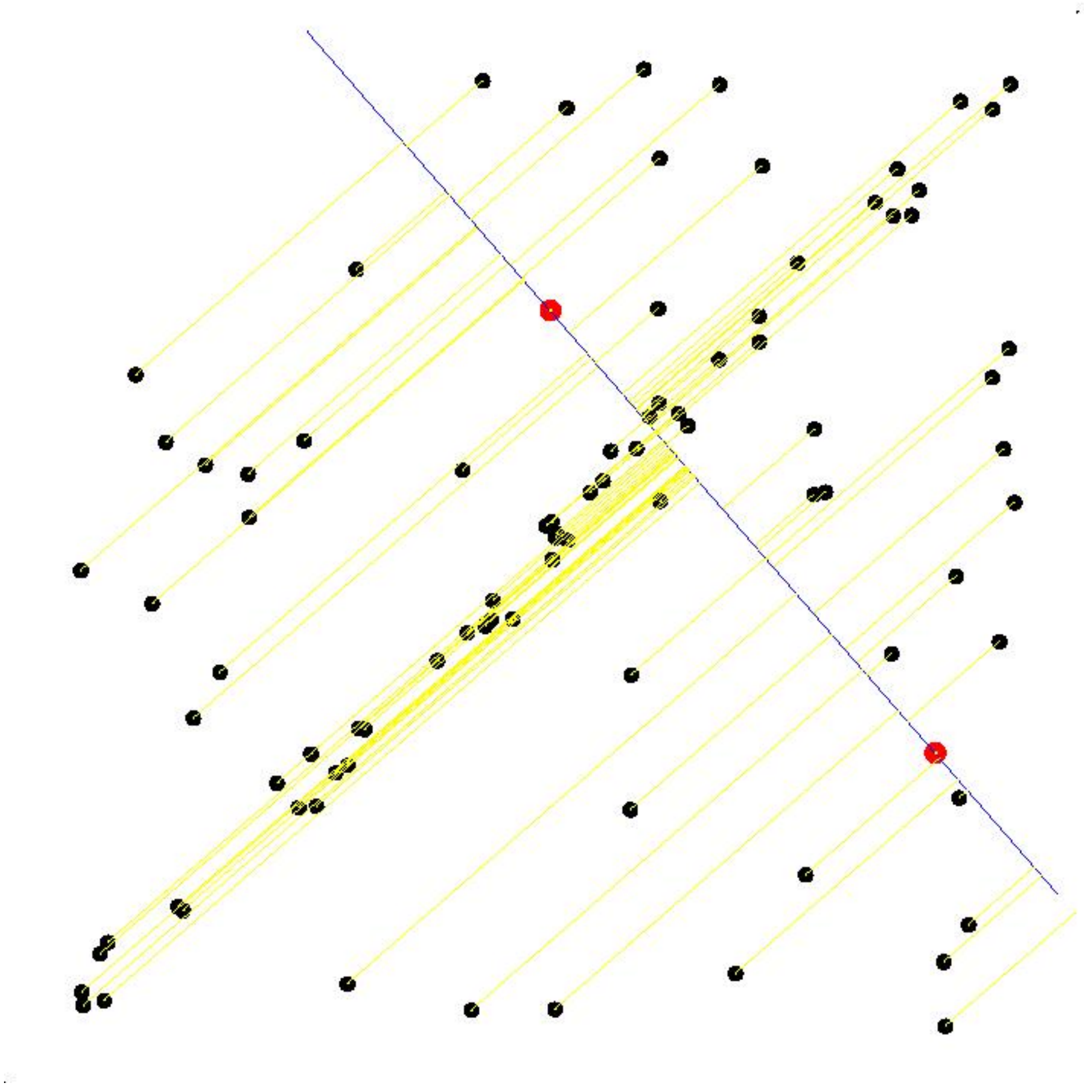
RANSAC



- Select sample of m points at random

- **Calculate model parameters that fit the data in the sample**

RANSAC

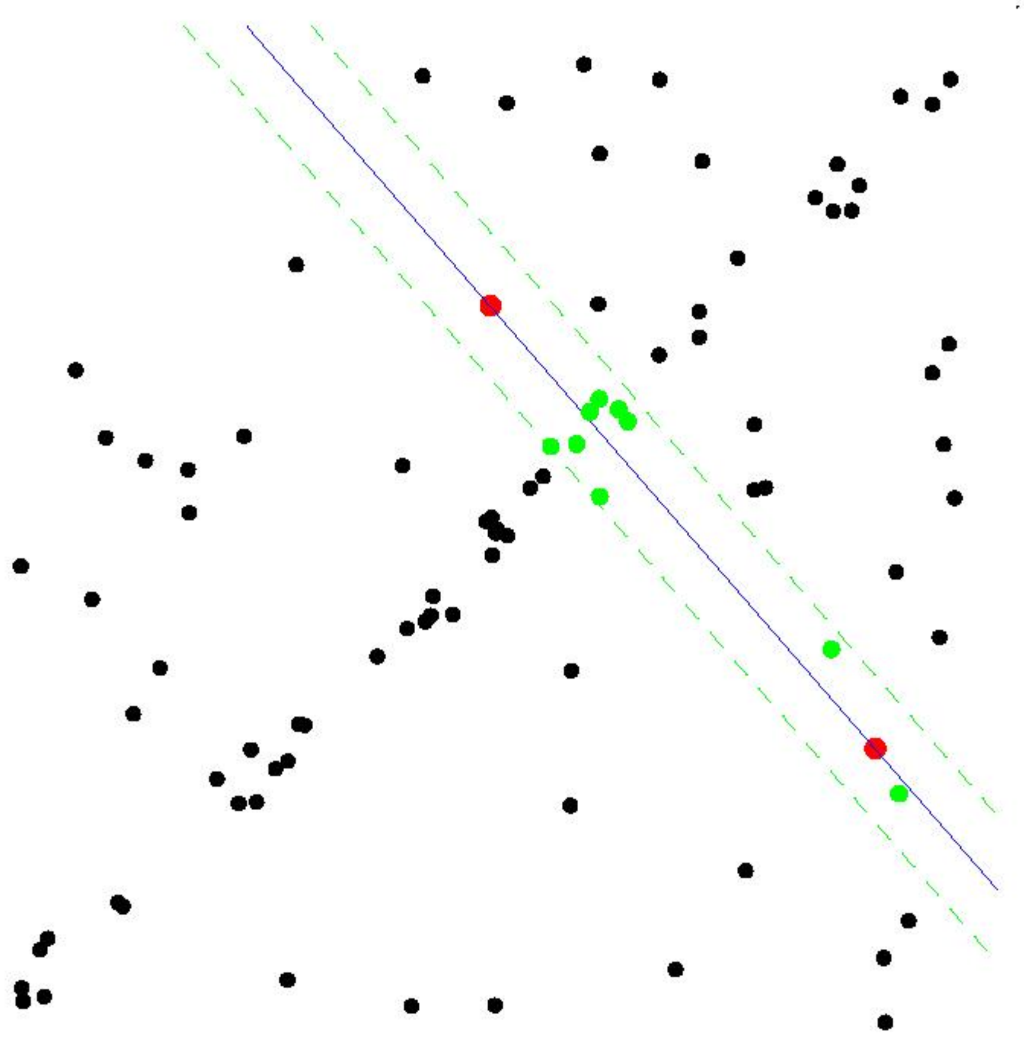


• Select sample of m points at random

• Calculate model parameters that fit the data in the sample

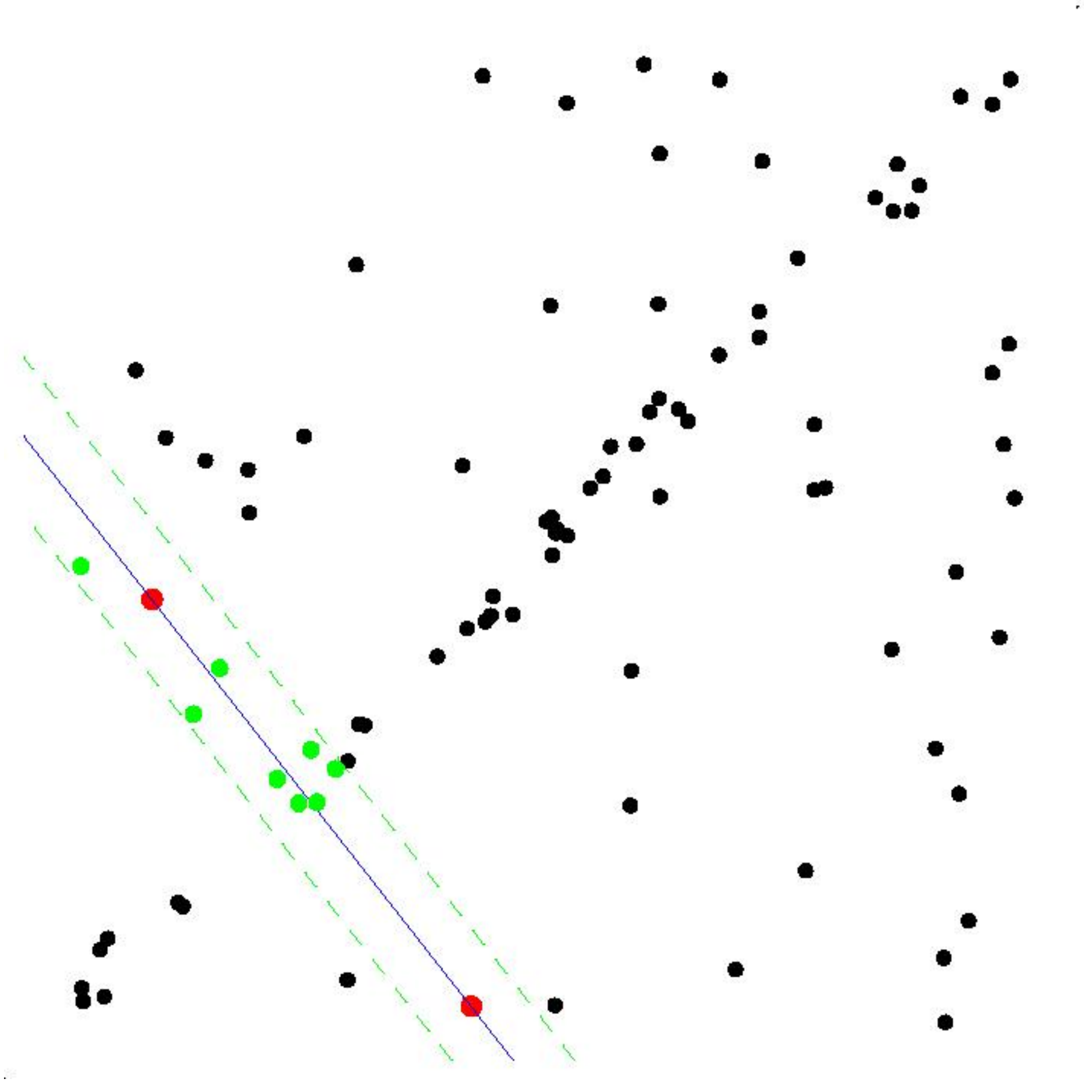
• Calculate error function for each data point

RANSAC



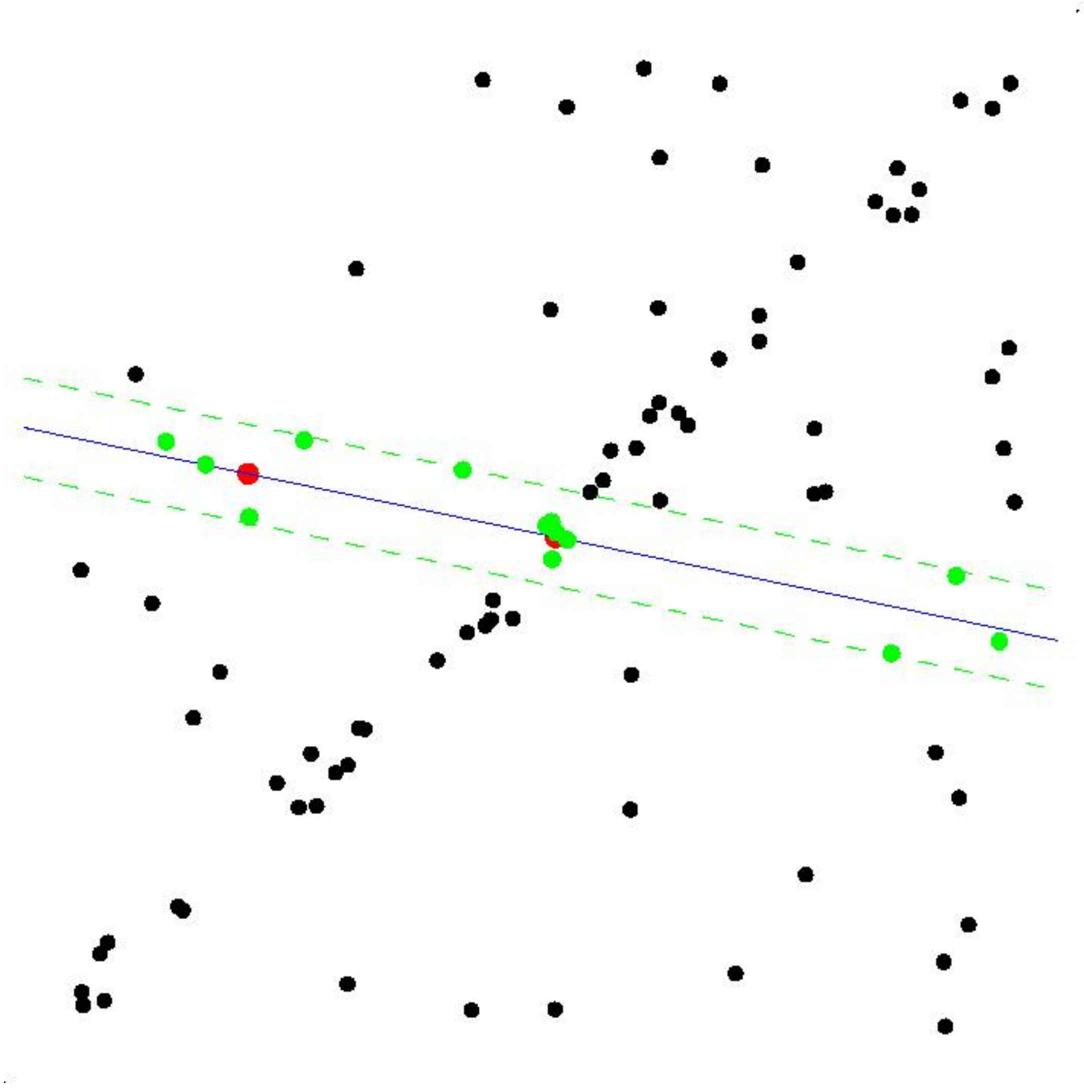
- Select sample of m points at random
- Calculate model parameters that fit the data in the sample
- Calculate error function for each data point
- **Select data that support current hypothesis**

RANSAC



- Select sample of m points at random
- Calculate model parameters that fit the data in the sample
- Calculate error function for each data point
- Select data that support current hypothesis
- **Repeat sampling**

RANSAC



- Select sample of m points at random
- Calculate model parameters that fit the data in the sample
- Calculate error function for each data point
- Select data that support current hypothesis
- **Repeat sampling**

How Many Samples?

On average

N	… number of point
I	… number of inliers
m	… size of the sample

$$P(\text{good}) = \frac{\binom{I}{m}}{\binom{N}{m}} = \prod_{j=0}^{m-1} \frac{I-j}{N-j}$$

mean time before the success

$E(k) = 1 / P(\text{good})$

How Many Samples?

With confidence p

How large k?

. . . to hit at least one pair of points on the line l with probability larger than p (0.95)

Equivalently

. . . the probability of not hitting any pair of points on l is $\leq 1 - p$

How Many Samples?

With confidence *p*

N	… number of point
I	… number of inliers
m	… size of the sample

$$P(\text{good}) = \frac{\binom{I}{m}}{\binom{N}{m}} = \prod_{j=0}^{m-1} \frac{I-j}{N-j}$$

P(bad) = 1 – P(good)

$$P(\text{bad } k \text{ times}) = \big(1 - P(\text{good})\big)^{k}$$

How Many Samples?

With confidence *p*

$$P(\text{bad } k \text{ times}) = \left(1 - P(\text{good})\right)^k \leq 1 - p$$

$$k \log \left(1 - P(\text{good})\right) \leq \log(1 - p)$$

$$k \geq \log(1 - p) \,/\, \log \left(1 - P(\text{good})\right)$$

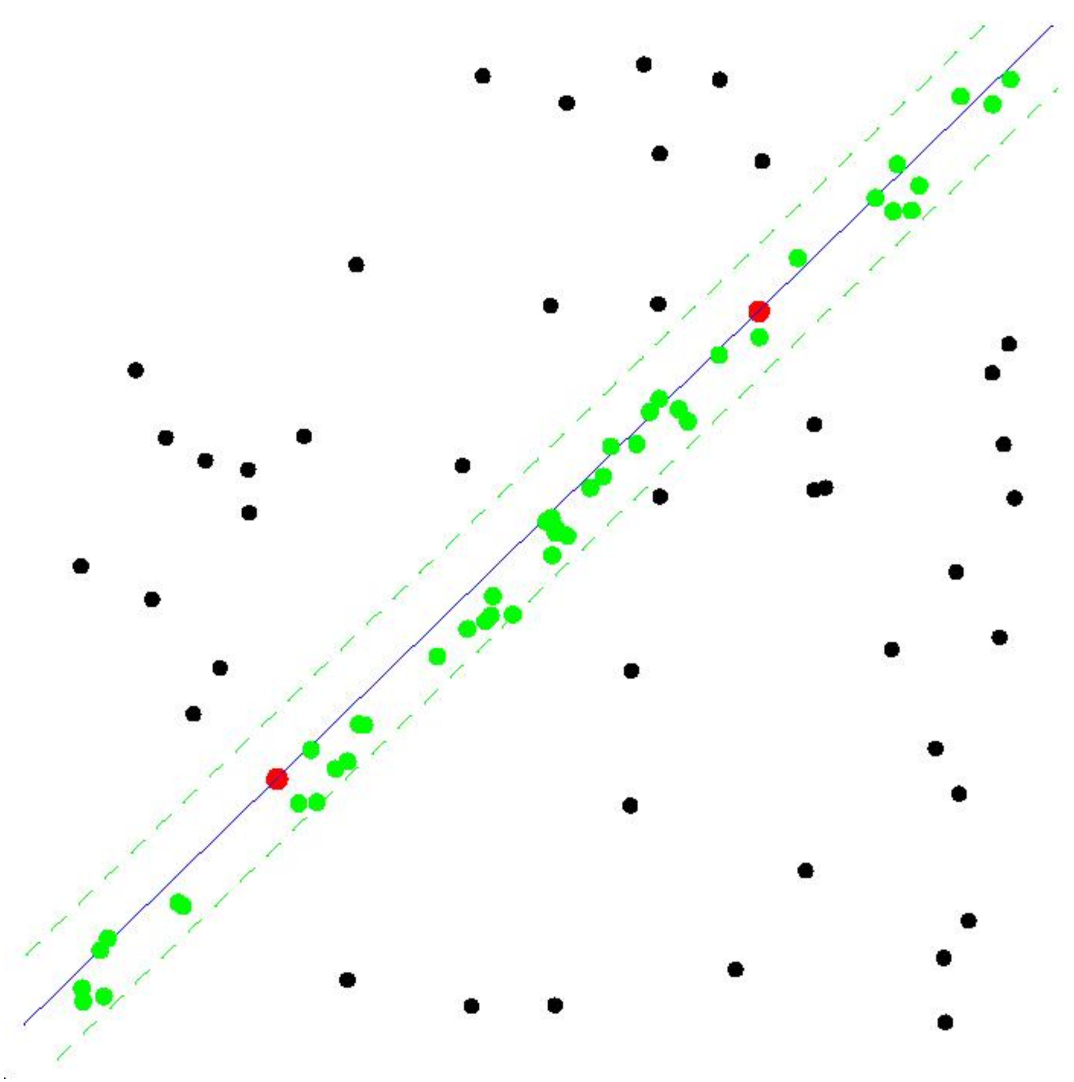
How Many Samples

I / 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	∞	∞	∞	$2.70 \cdot 10^{16}$	$3.29 \cdot 10^{12}$	$4.71 \cdot 10^6$

RANSAC



$$k = \frac{\log(1 - p)}{\log\left(1 - \frac{I}{N}\frac{I-1}{N-1}\right)}$$

k … number of samples drawn

N … number of data points

I … time to compute a single model

p … confidence in the solution (.95)

RANSAC [Fischler, Bolles '81]

In: $U = \{x_i\}$ set of data points, $|U| = N$

$f(S) : S \to p$ function f computes model parameters p given a sample S from U

$\rho(p, x)$ the cost function for a single data point x

Out: p^* p^*, parameters of the model maximizing the cost function

k := 0

Repeat until P{better solution exists} < η (a function of C^* and no. of steps k)

k := k + 1

I. Hypothesis

(1) select randomly set $S_k \subset U$, sample size $|S_k| = m$

(2) compute parameters $p_k = f(S_k)$

II. Verification

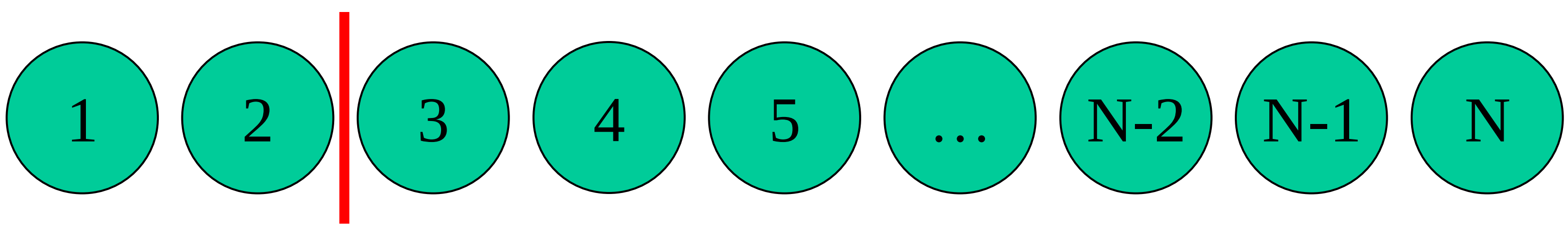
(3) compute cost $C_k = \sum_{x \in U} \rho(p_k, x)$

(4) if $C^* < C_k$ then $C^* := C_k$, $p^* := p_k$

end

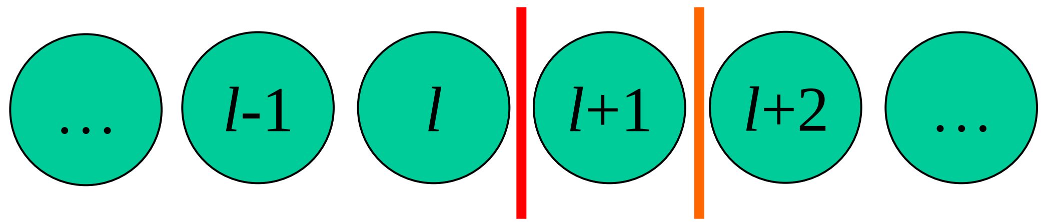
PROSAC – PROgressive SAmple Consensus

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



Sample from here

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 $l+1$