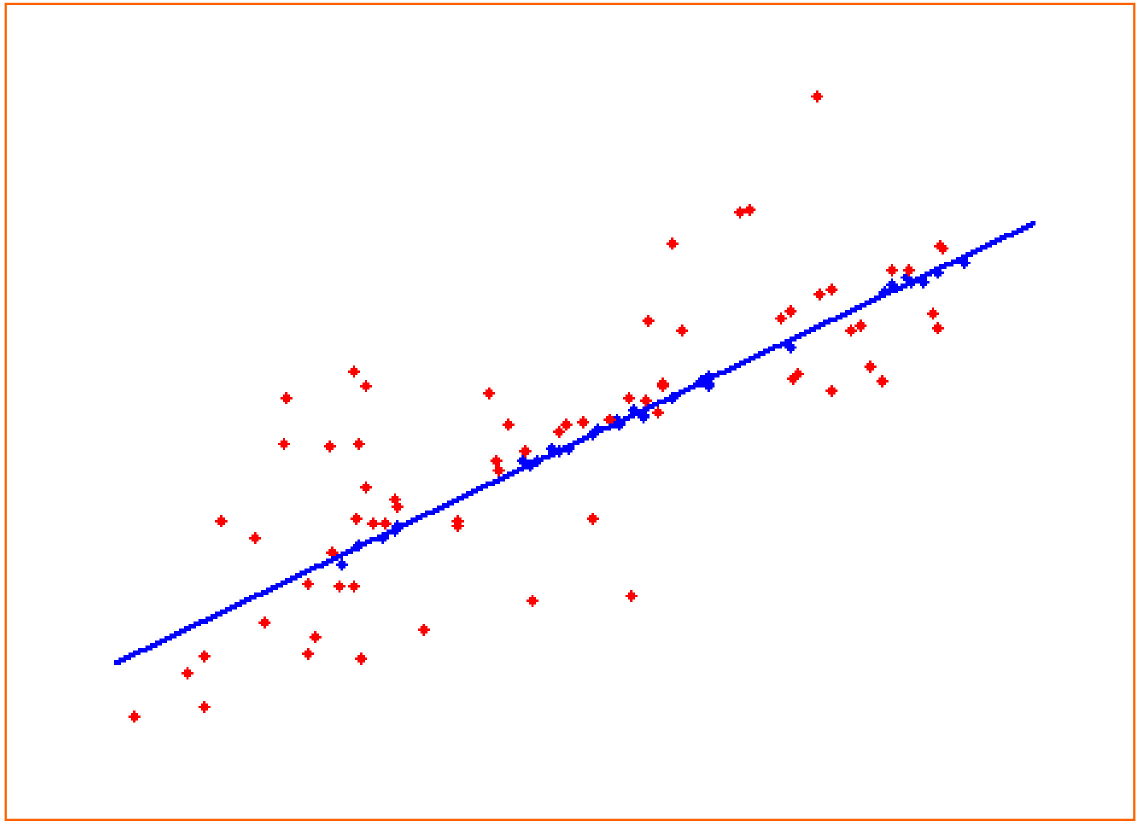
RANSAC – Robust Fitting

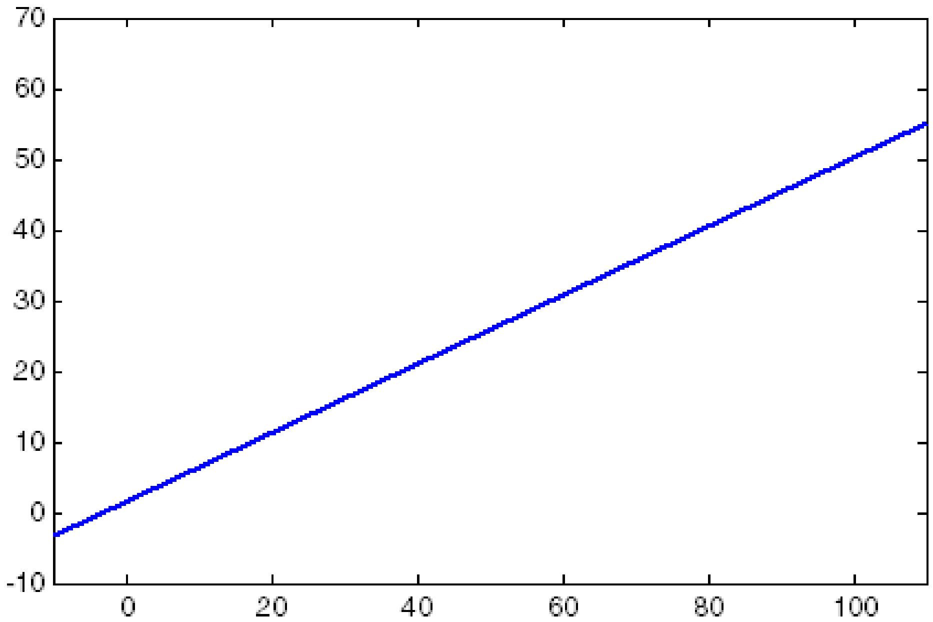


Tomáš Pajdla

21 April 2007

Example

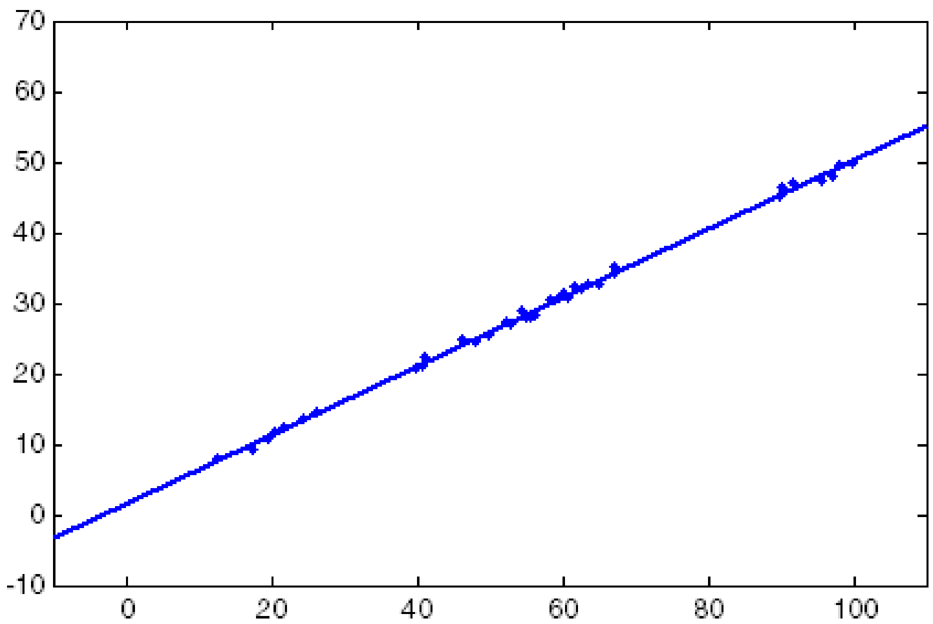
There is a line ...



Example

There is a line ...

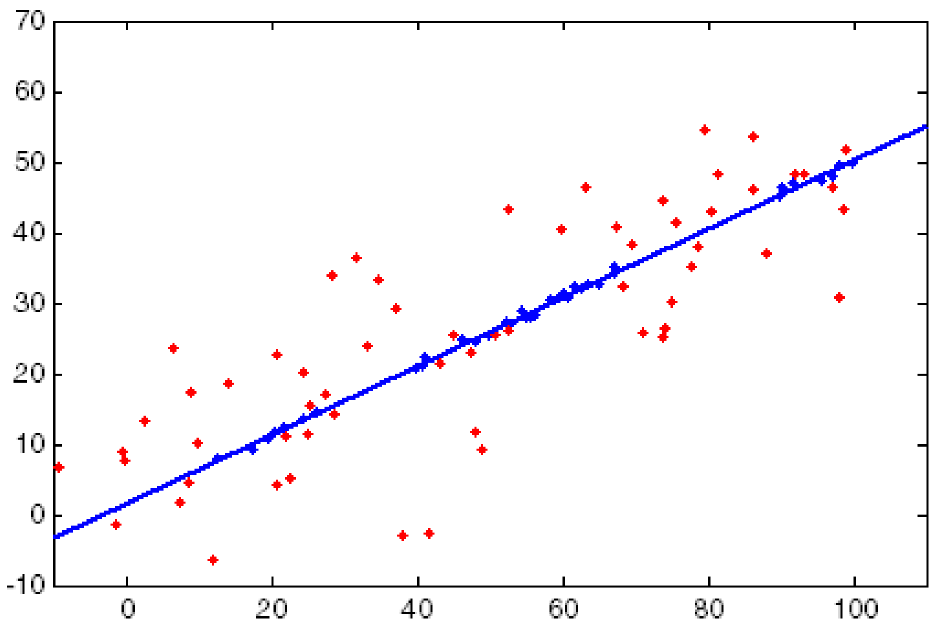
$Y = \{\mathbf{x}_i\}_{i=1}^{M}$ ... a set of points on the line l is measured with Gaussian noise $N(\mathbf{0}, \sigma)$



Example

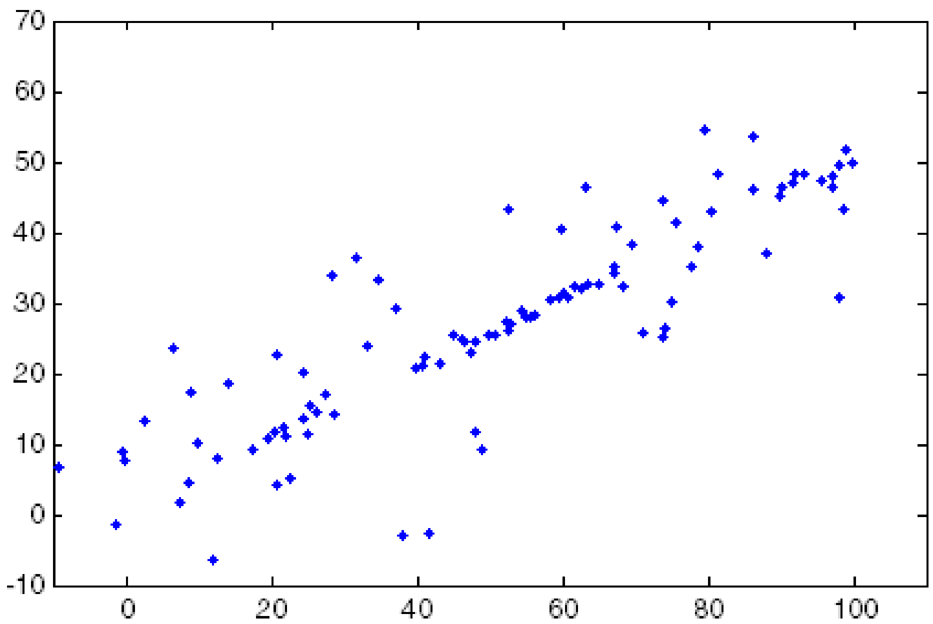
There is a line ...

$X = \{\mathbf{x}_i\}_{i=1}^{N}$ ... other points, unrelated to the line.



Task

Find a line and its points from data contaminated by mesurements unrelated to the line.

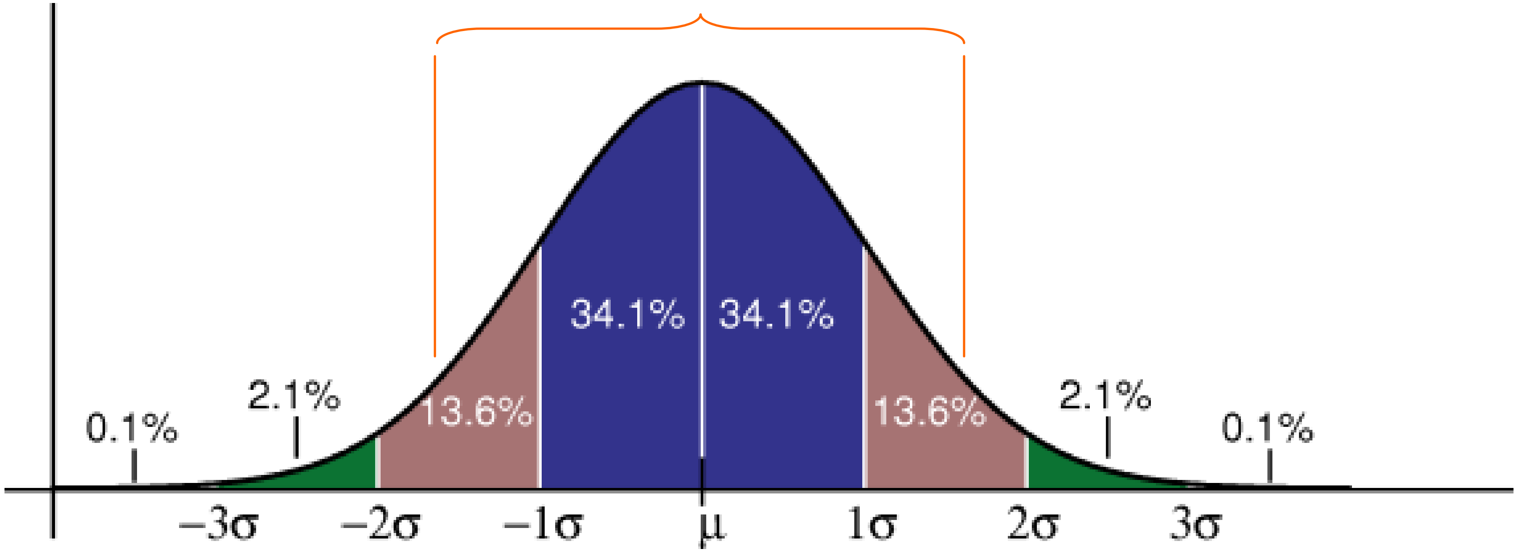


Assumptions

1. σ is known

2. The largest subset of X of points for which there is a line which is closer than 1.65 σ to all the points contains points that were measured on the line l.

$$2 \times 1.65\,\sigma \sim 95\%$$

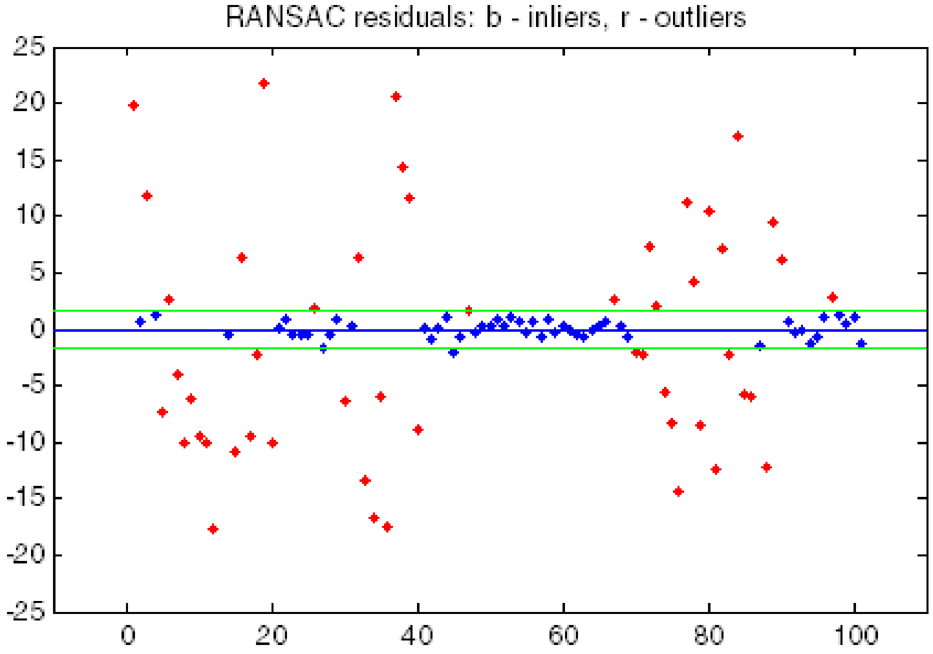


Problem formulation

Find the maximal subset I of the given set X

such that

there is a line which is closer than 1.65σ to all points in I.



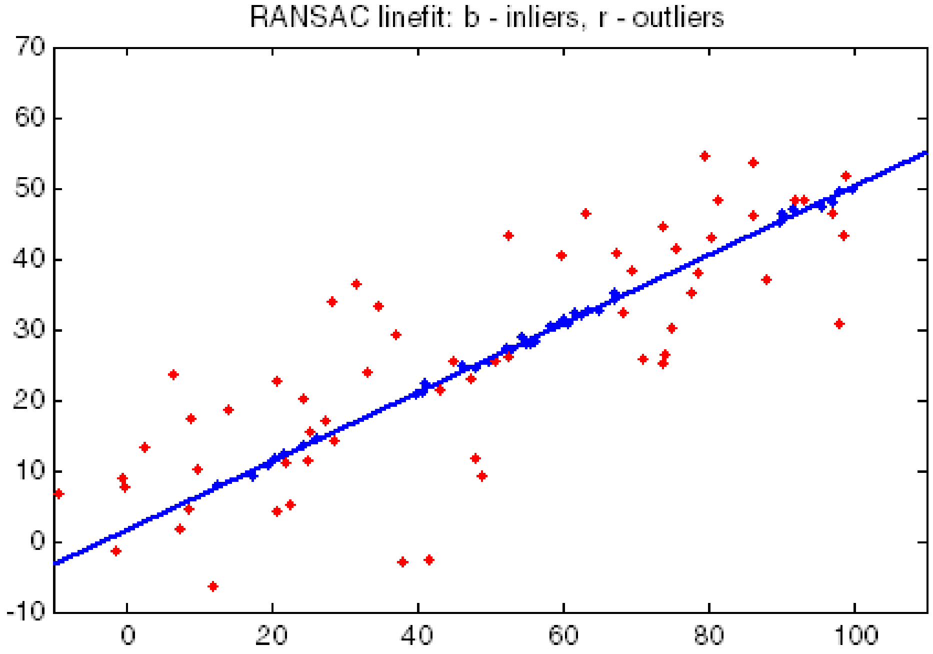
Notation

l . . . line

X . . . all data

I . . . inliers = points closer to l than a threshold

$O = X \setminus I$ . . . outliers



Exhaustive Search

Exhaustive Search

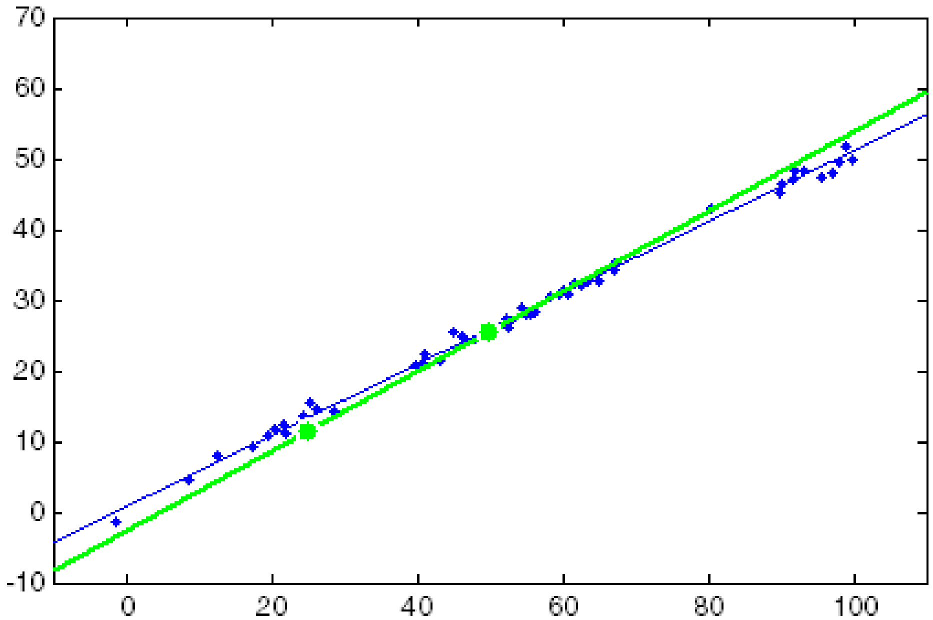
$$I = \arg\max_{S \subseteq X} (\text{the number of inliers in } X \text{ for the best line fit to } S)$$

does not work:

There is 2^N candidate subsets S to be tested

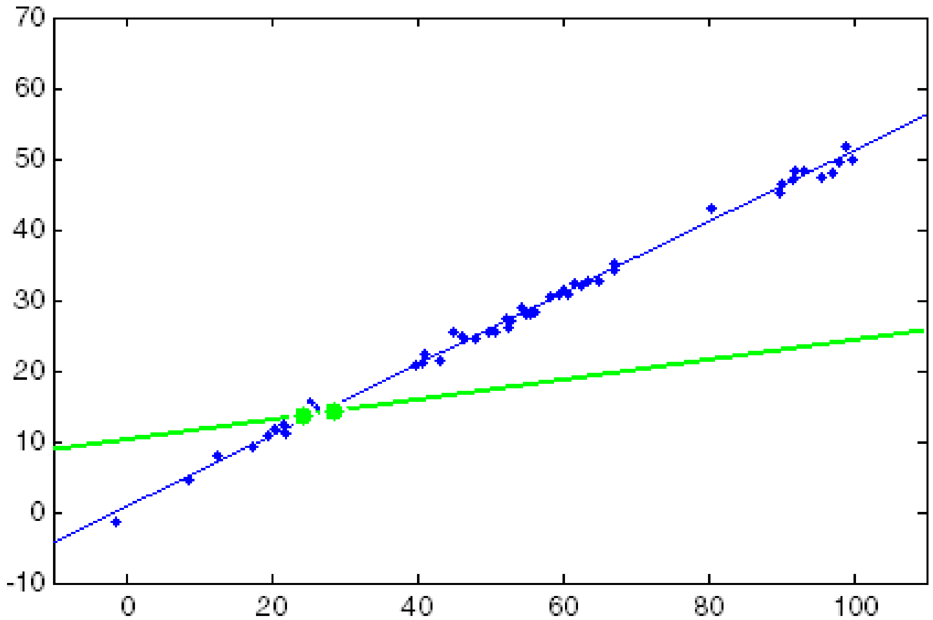
... infeasible for useful N's

Observation



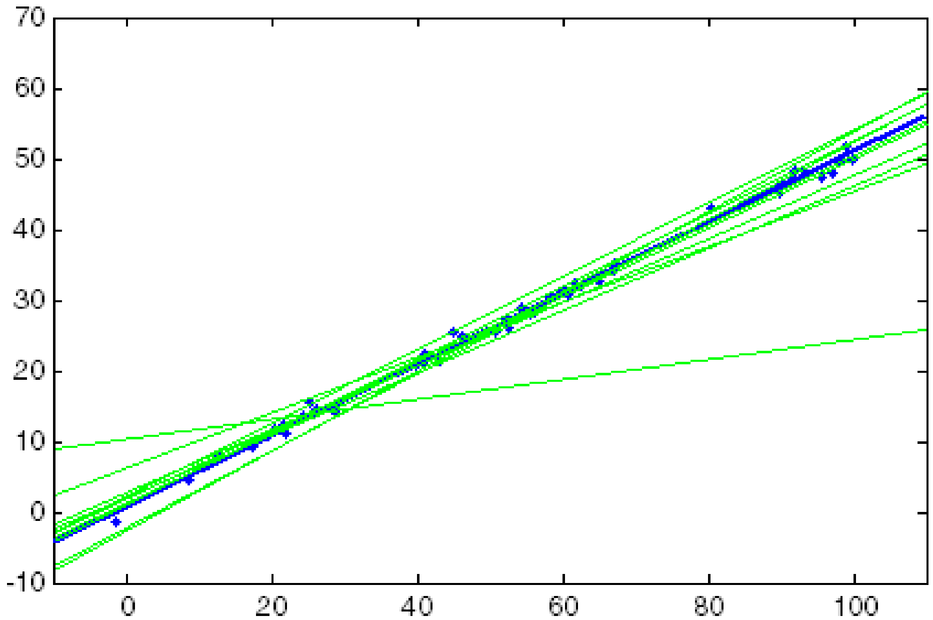
Two (a minimal sample of) *good* points (points measured on the line l) generate a line which is close to l.

Observation



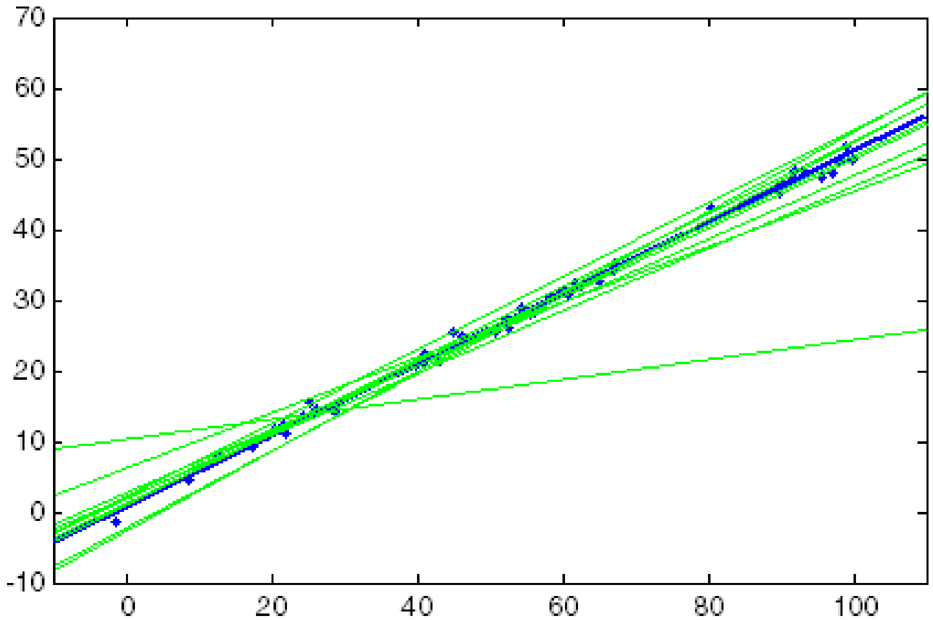
Not all pairs of "good" points are good due to noise ...

Observation



… but many are.

Observation



There is "only" $\binom{N}{2} = \dfrac{N(N-1)}{2}$ pairs of distinct points.

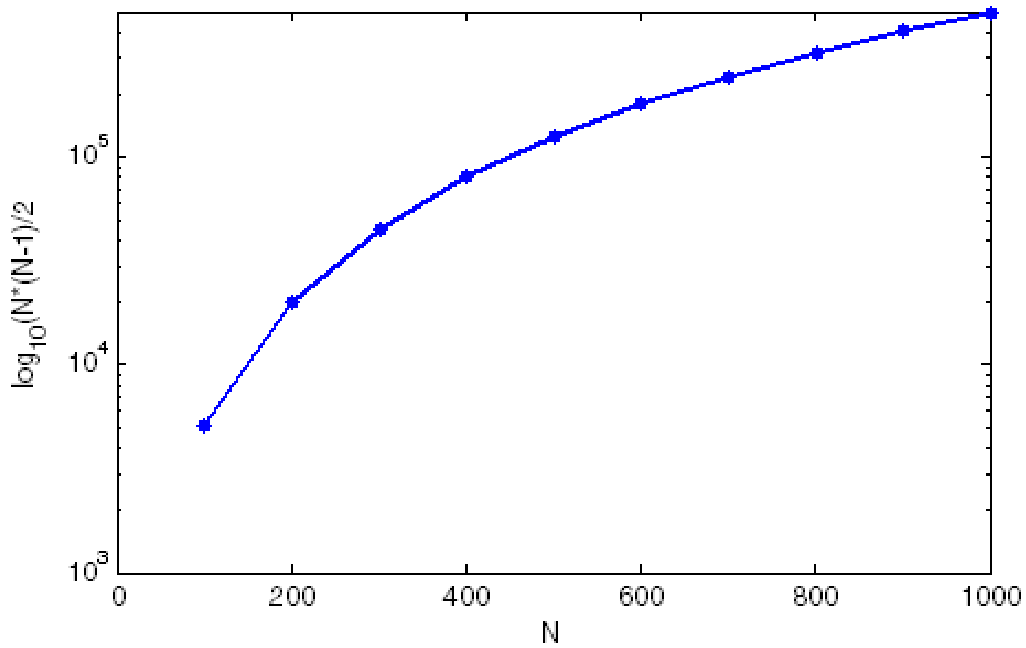
Exhaustive Minimal Sampling

Exhaustive Minimal Sampling

$$I = \arg \max_{\{x_1, x_2\} \subseteq X} (\ \#\ \text{inliers in}\ X\ \text{for the line through}\ \{x_1, x_2\})$$

needs to examine "only" $\frac{N(N-1)}{2} \ll 2^N$ pairs of distinct points.

Exhaustive Minimal Sampling



The number of samples: $\binom{N}{2} = \frac{N(N-1)}{2}$

... is often still too high.

Exhaustive Minimal Sampling with zero noise

Simplified analysis: assume no noise, i.e. $\sigma = 0$.

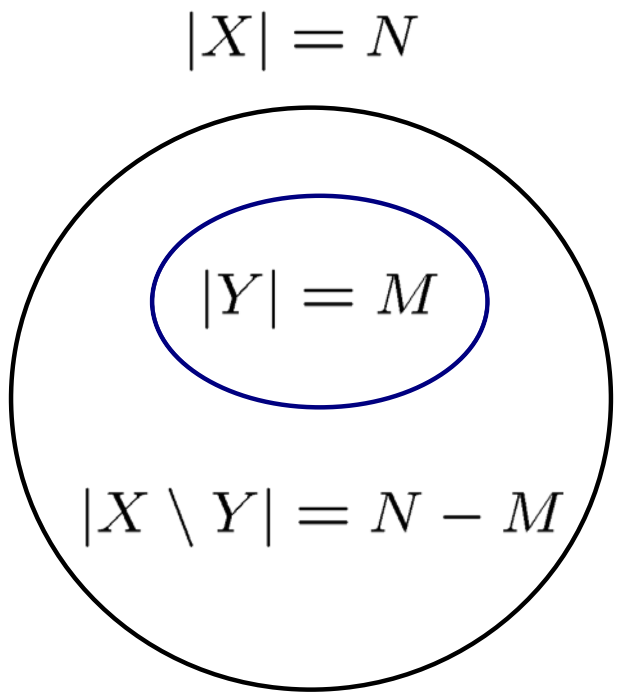
The goal is to make as many samples as to be sure not to miss the set Y of points on the line l.

Assume that there is M points in Y on the line l among the total number N of points in X.

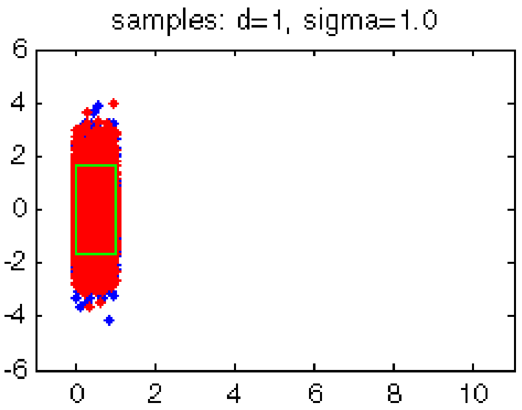
One has to try more than

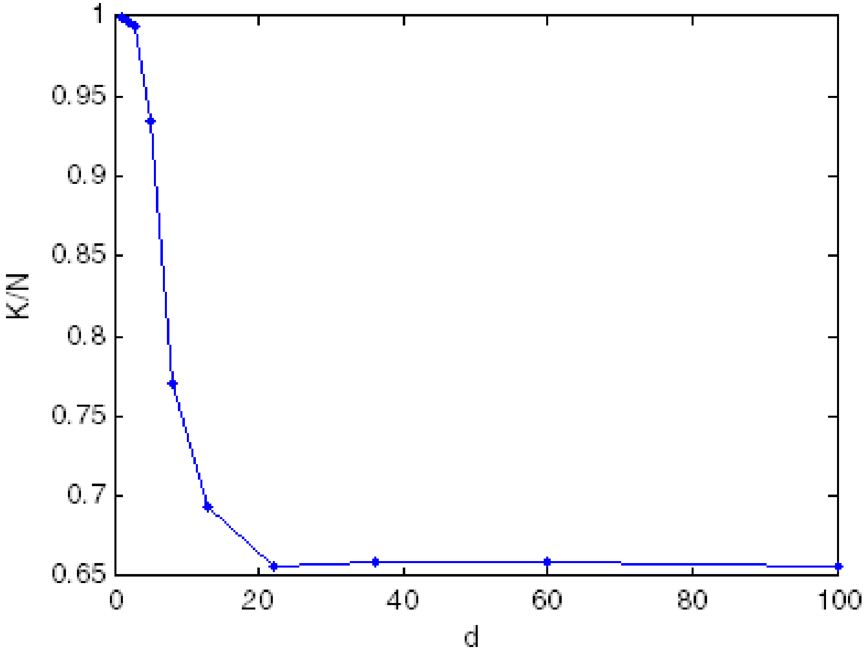
$$\binom{N}{2} - \binom{M}{2}$$

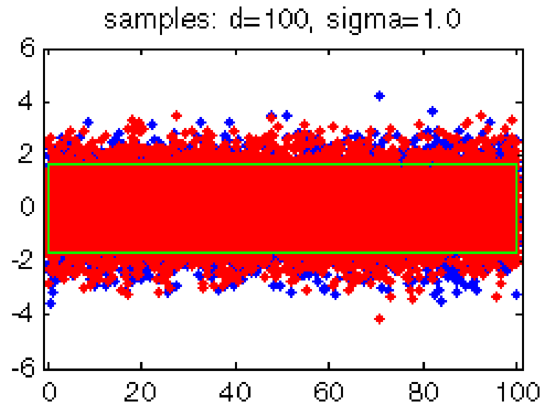
point pairs when drawing samples without repetition, which goest to ∞ for $N \to \infty$.



Observation







1. d . . . the length of the line segment
2. K . . . the number of pairs generating a line that has more that 95 % of points of X closer than 1.65 σ.
3. the shorter the line, the higher the chance to generate a good sample
4. for $d \gg \sigma$ . . . $\frac{K}{N} \approx 0.65$

pajdla@cmp.felk.cvut.cz

Relaxation

It is enough to have a high (95 %) chance of finding a good estimate of the true line segment l.

65 % of samples generate a line that is closer to more than 95 % of points in X

Relaxation

Using sufficient number of randomly chosen pairs of points guarrantees average succeess.

Random Minimal Sampling

Random Minimal Sampling

$$I = \arg \max_{\{x_1, x_2\} \subseteq R \subset X} (\text{\# inliers in } X \text{ for the line through } \{x_1, x_2\})$$

needs to examine "only" $|R|$ pairs of distinct points in a randomly chosen subset R of X.

How many random samples should be tried?

pajdla@cmp.felk.cvut.cz

Random Minimal Sampling for Robust Line Fitting

The goal is to make k samples to hit at leat one pair of points on the line l with probability larger than p.

Equivalently, we look for k such that the probability of not hitting any pair of points on l is smaller or equal to $1-p$

$$\left(1-\frac{M(M-1)}{N(N-1)}\right)^k \leq 1-p$$

$\frac{M}{N}$	...	the probability if drawing a good data point
$\frac{M(M-1)}{N(N-1)}$	...	the probability of drawing (without repetition) a good pair of data points
$1-\frac{M(M-1)}{N(N-1)}$	...	the probability of drawing a bad pair of data points
$(1-\frac{M(M-1)}{N(N-1)})^k$	...	the probability of drawing (with repetitions) k bad pairs of data points in a row

Random Minimal Sampling for General Models

For $N \to \infty$ and fixed fraction $\epsilon = \frac{M}{N}$ of good points

$$\lim_{N\to\infty} \frac{M(M-1)}{N(N-1)} = \lim_{N\to\infty} \frac{\epsilon N(\epsilon N-1)}{N(N-1)} = \epsilon^2$$

which depends only on the fraction of good points and for large N leads to the necessary number of samples

$$k \geq \frac{\log(1-p)}{\log\left(1-\epsilon^2\right)}$$

Random Minimal Sampling for General Models

Generalization for the samples with m points:

$$\lim_{N\to\infty} \frac{M(M-1)\dots(M-m)}{N(N-1)\dots(N-m)} = \lim_{N\to\infty} \frac{\epsilon N(\epsilon N-1)\dots(\epsilon N-m)}{N(N-1)\dots(N-m)} = \epsilon^m$$

and thus for large N we get

$$(1-\epsilon^m)^k \leq 1-p$$

which leads to

$$k \geq \frac{\log(1-p)}{\log(1-\epsilon^m)}$$

How many samples?