Model Fitting

TDV—3D compute vision: labs.

Martin Matoušek

September, 2012[*]

1 Motivation – A Planar Line

Suppose there is an unknown planar line parameterised as a normalised projective line, i.e. $\boldsymbol{l} = [l_1, l_2, l_3]^\top$, where $l_1^2 + l_2^2 = 1$. There is a process generating a set of n planar points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ we are able to observe. Some of the points – the inliers – are generated by a line, the remaining points – the outliers – are placed randomly, not related to the line. Additionally, the inliers are impaired by an isotropic Gaussian noise, i.e.,

$$\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{x}_i; \widehat{\boldsymbol{x}_i}, \sigma),$$

where $\widehat{\boldsymbol{x}_i}$ is the 'correct' point generated by the line, $[\widehat{\boldsymbol{x}_i}^\top\, 1]\, \boldsymbol{l} = 0$. The example situation is shown in Figure 1.

We define a metric $d(\boldsymbol{x}_i, \boldsymbol{l})$ as the orthogonal point-to-line distance,

$$d(\boldsymbol{x}_i, \boldsymbol{l}) = [\boldsymbol{x}_i^\top\, 1]\, \boldsymbol{l}\,.$$

We assume that the points are generated independently, the probability that a point is inlier is p_1 and the probability that a point is outlier is p_0, providing that $p_0 + p_1 = 1$. The outliers are placed uniformly in the image plane and the inliers are placed uniformly along a line (though the infinity of the plane and of the line must be somehow treated). Then it can be shown, that the line likelihood is

$$p(\mathcal{X}|\boldsymbol{l}) = n! \prod_1^n p_0\, c_0 + p_1\, c_1\, \mathcal{N}(d(\boldsymbol{x}_i, \boldsymbol{l}); 0, \sigma)$$
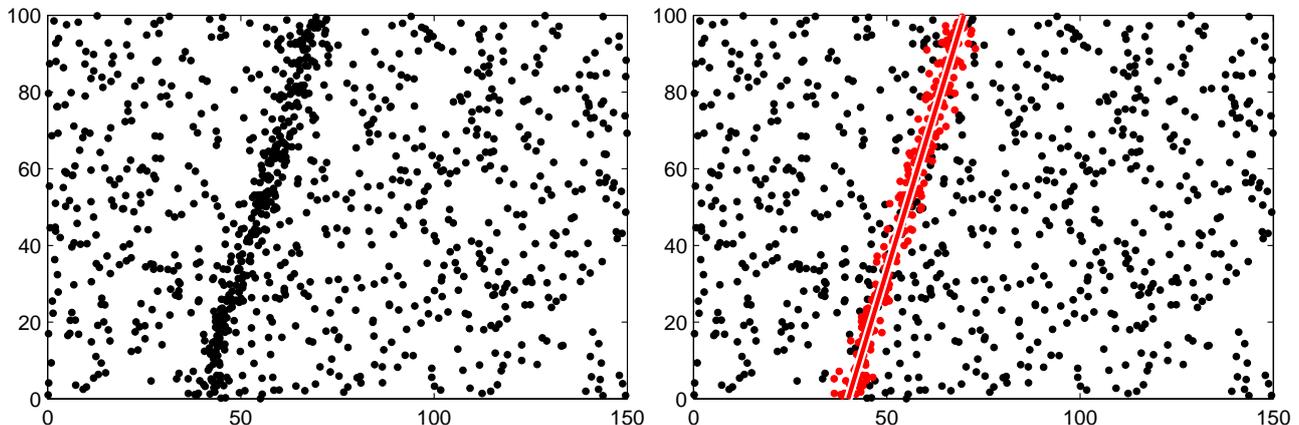
Fig. 1: Single line fitting situation – noisy inliers generated by the line mixed with outliers. The raw data set is on the left, the right figure shows in red the original line and inliers it generates.
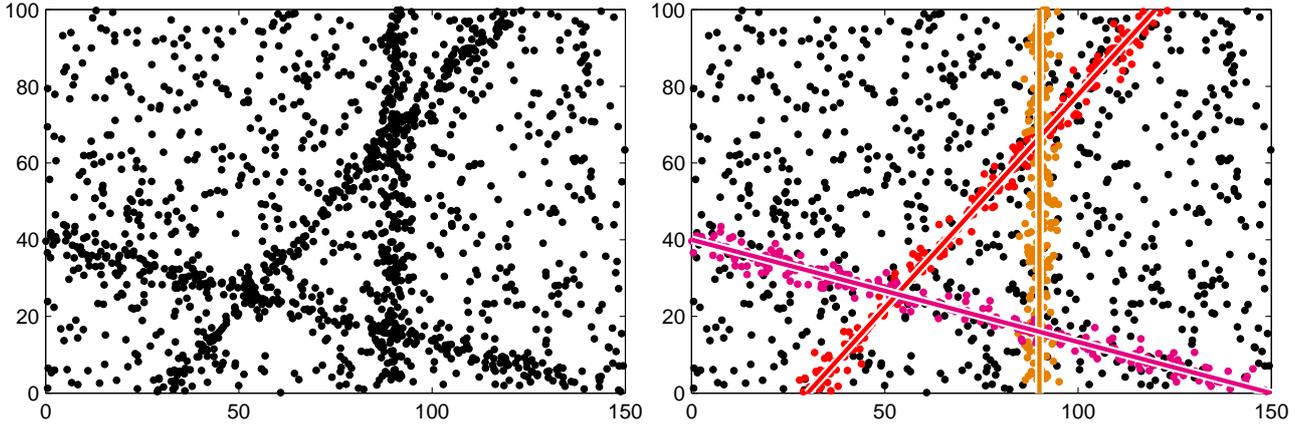
Fig. 2: Multiple lines fitting situation. The raw data set is on the left, the original lines and inliers (red, orange and magenta) are on the right.

where c_0 is a constant given by uniform distribution of outliers in a plane and c_1 is a is constant given by uniform distribution of inliers along a line.

The task. Having the set $\mathcal{X}$ of n planar points measured, we want to estimate the parameters $\boldsymbol{l}^*$ of the unknown line and optionally select the inliers from $\mathcal{X}$.

In the next section, the estimation is formulated more generally as an optimisation problem of fitting a *model* to an observed *sample set*. In the case of planar line estimation, the model is a line and the sample set consists of the measured points.

Multiple models. Imagine another example – a set of points $\mathcal{X}$ generated by two lines $\boldsymbol{l}_1$, $\boldsymbol{l}_2$. Again, there are some outliers and inliers are impaired by a noise. A portion of inliers comes from the first line and a portion from the second one. As in the previous case, the task is to estimate parameters of the two lines and optionally select the points belonging to each.

2 The Problem of Model Fitting

We seek parameters of a model – an unobservable entity, parameterised by a set of parameter vectors $\mathcal{S}$. The entity generates a set of n observable samples $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. The set is polluted by outliers – the erroneous samples not related to the model. Additionally, the inliers – samples related to the model – are impaired by some noise.

The model can be a single indivisible entity parameterised by a single vector, $\mathcal{S} = \{\boldsymbol{l}\}$, or it can be composed from some number of sub-models, each having its own feature vector, $\mathcal{S} = \{\boldsymbol{l}_1, \ldots, \boldsymbol{l}_m\}$. Apparently, the first case is a special case for $m = 1$. We assume, that the number m is fixed and known, though there are more complicated problems violating this.

The fact that a point is either an outlier or an inlier belonging to one of m sub-models is characterised by labels $\Lambda = (\lambda_1, \ldots, \lambda_n)$, where $\lambda_i = 0$ means that the i-th sample is an outlier and $\lambda_i \in \{1, \ldots, m\}$ is for inlier belonging to the λ_i-th sub-model.

The sample-generating process is characterised by probability $p(\mathcal{X}, \mathcal{S}, \Lambda) = p(\mathcal{X}|\mathcal{S}, \Lambda) \, p(\mathcal{S}, \Lambda)$, composed from data-dependent likelihood and data-independent prior. We formulate the model fitting as maximum a-posteriori (MAP) or maximum likelihood (ML) problem. MAP estimate can be computed only if the prior $p(\mathcal{S}, \Lambda)$ is known; otherwise we seek for ML estimate. In some situations the model $\mathcal{S}$ and the labels Λ should be estimated simultaneously, in other situations only the model is estimated. Thus we distinguish four closely related formulations of the optimisation problem.

2

1. MAP estimate of the model and labels.

$$(\mathcal{S}^*, \Lambda^*) = \arg \max_{\mathcal{S}, \Lambda} p(\mathcal{X}|\mathcal{S}, \Lambda)\, p(\mathcal{S}, \Lambda) \tag{1}$$

2. ML estimate of the model and labels.

$$(\mathcal{S}^*, \Lambda^*) = \arg \max_{\mathcal{S}, \Lambda} p(\mathcal{X}|\mathcal{S}, \Lambda) \tag{2}$$

3. MAP estimate of the model.

$$\mathcal{S}^* = \arg \max_{\mathcal{S}} p(\mathcal{X}|\mathcal{S})\, p(\mathcal{S}) \tag{3}$$

4. ML estimate of the model.

$$\mathcal{S}^* = \arg \max_{\mathcal{S}} p(\mathcal{X}|\mathcal{S}) \tag{4}$$

We see that the data-dependent likelihood is a key term of the estimation, either in its full form $p(\mathcal{X}|\mathcal{S}, \Lambda)$ or with labels eliminated by marginalisation, $p(\mathcal{X}|\mathcal{S}) = \sum_{\Lambda} p(\mathcal{X}|\mathcal{S}, \Lambda)p(\Lambda)$.

2.1 Typical Factorisation of Likelihood of a Model

An actual model likelihood $p(\mathcal{X}|\mathcal{S}, \Lambda)$ depends on the actual problem we are solving. Examples include estimating a single line from points, two planar homographies from image-to-image correspondences, epipolar geometry from image-to-image correspondences, or camera resection from image-to-scene correspondences, and many more. However, in many problems the likelihood factorises into the same/similar pieces:

- $p(\mathcal{S})$ – prior distribution of the model.

- $p(\Lambda) = \prod_1^n p(\lambda_i)$ – independent prior distribution of labels.
  We denote $p(\lambda_i = j)$ as p_j, then $p_0 + \ldots + p_m = 1$.

- $p(\mathcal{X}|\mathcal{S}, \Lambda) = n! \prod_{i=1}^n p(x_i|\mathcal{S}, \lambda_i)$ – the samples are independent (the $n!$ term honours the fact that the order of samples is not significant). This leads to $p(\mathcal{X}|\mathcal{S}) = n! \prod_{i=1}^n p(\boldsymbol{x}_i|\mathcal{S})$, i.e., marginalised likelihoods are independent as well.

- $p(x_i|\mathcal{S}, \lambda_i = 0) = c_0$ – distribution of outliers is a constant.

- $p(\boldsymbol{x}_i|\mathcal{S}, \lambda_i = j \neq 0) = c_j\, p\,(d(\boldsymbol{x}_i, \boldsymbol{l}_j)) = c_j\, \mathcal{N}(d(\boldsymbol{x}_i, \boldsymbol{l}_j); 0, \sigma)$ – distribution of inliers is proportional to normal distribution of inlier errors. There must be defined a metric d measuring a 'distance' of a sample from the model. It is reasonable to consider the deviation σ same for all models, since the noise usually comes from some detector of samples (e.g. points in image plane) and is not dependent on a model.

Usually, we assume that the constants c_1 to c_m are the same. The prior probabilities p_1 to p_m are either known or assumed same as well. Thus we express all the unknown constant terms for inliers as $c_1 p_1 = \ldots = c_m p_m = c_x$ and for outliers using the threshold as $p_0 c_0 = c_x \mathcal{N}(\theta; 0, \sigma)$. Note that θ has meaning of threshold separating inliers from outliers, it has no sense when separating inliers for different labels.

The likelihood $p(\mathcal{X}|\mathcal{S}, \Lambda)$ is directly composed using the above factorisation. Note that the task

$$(\mathcal{S}^*, \Lambda^*) = \arg \max_{\mathcal{S}, \Lambda} p(\mathcal{X}|\mathcal{S}, \Lambda)$$

can be decomposed as

$$\begin{aligned}
\mathcal{S}^* &= \arg \max_{\mathcal{S}} \max_{\Lambda} p(\mathcal{X}|\mathcal{S}, \Lambda)\,, \\
\Lambda^* &= \arg \max_{\Lambda} p(\mathcal{X}|\mathcal{S}^*, \Lambda)\,.
\end{aligned}$$

We can write likelihood $p(\mathcal{X}|\mathcal{S})$ by marginalisation over all values of labels,

$$p(\boldsymbol{x}_i|\mathcal{S}) = p_0\,c_0 + p_1\,c_1\,\mathcal{N}(d(\boldsymbol{x}_i,\boldsymbol{l}_1);0,\sigma) + \ldots + p_m\,c_m\,\mathcal{N}(d(\boldsymbol{x}_i,\boldsymbol{l}_m);0,\sigma)\,, \tag{5}$$

$$p(\mathcal{X}|\mathcal{S}) = n!\prod_{i=1}^{n} p(\boldsymbol{x}_i|\mathcal{S})\,. \tag{6}$$

Usually, we assume that the constants c_1 to c_m are same. The prior probabilities p_1 to p_m are either known or assumed same as well. Then (5) can be written as

$$p(\boldsymbol{x}_i|\mathcal{S}) = c_x\Big(\mathcal{N}(\theta;0,\sigma) + \mathcal{N}(d(\boldsymbol{x}_i,\boldsymbol{l}_1);0,\sigma) + \ldots + \mathcal{N}(d(\boldsymbol{x}_i,\boldsymbol{l}_m);0,\sigma)\Big)\,, \tag{7}$$

where the outlier probability is expressed as inlier-outlier separation threshold θ and c_x is some constant, that has no impact to estimation. Note that a case with no outliers ($p_0 = 0$) can be represented by infinite threshold.

3 Model Fitting Algorithms

Depending on an actual shape of model probability/likelihood function we can choose an optimisation algorithm. We will focus on a typical situation where the shape of the function prevents to use a close form solution or some gradient-descent or similar optimisation to find a global optimum.

General Scheme of a Sampling Algorithm Iterate w.r.t some stopping criterion:

1. Make a proposal (hypothesis) of model parameters.

2. Evaluate the proposal using a criterion function (usually based on probability/likelihood).

3. Remember the best proposal (w.r.t criterion) from the proposals generated heretofore.

Algorithms mainly differ in a way the proposals are generated and then how the criterion is evaluated.

3.1 RANSAC and Modifications

Proposal. Randomly select a minimal sample set (MSS) and construct a model from it. E.g, MSS for a line are two points, then the model is the line joining them.

Variants differ by the criterion.

Traditional RANSAC. TODO

MLESAC. TODO

3.2 RANSAC for a Model Composed of Multiple Sub-models.

There are two approaches:

1. Treat as a single (indivisible) model.

2. Apply sequentially for each sub-model, in each step removing inliers (w.r.t. a threshold θ) found heretofore.
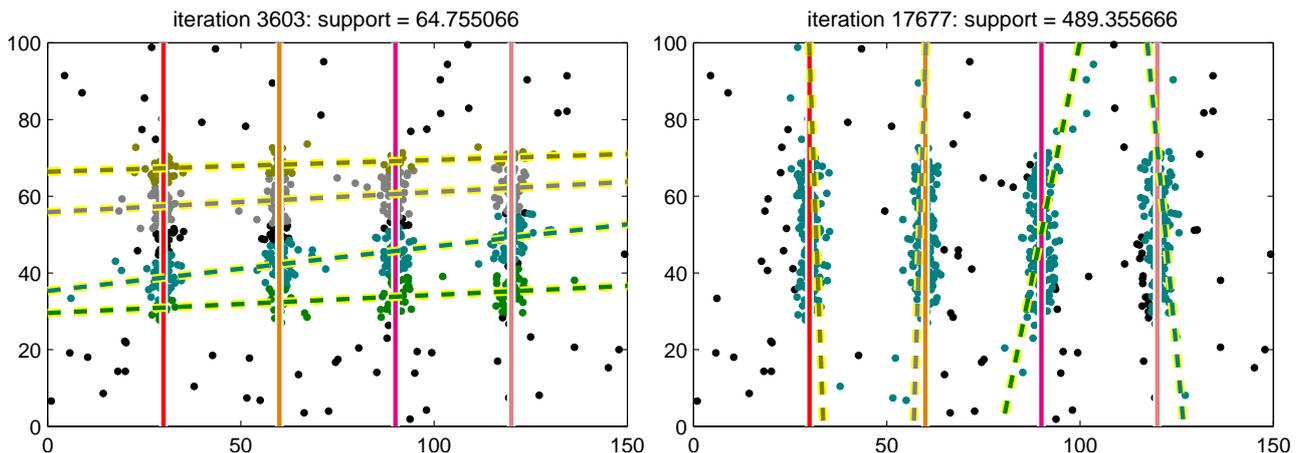
Fig. 3: Four lines fitting situation. Left – result of the sequential RANSAC, right – results of the full RANSAC. Solid lines denote the original lines, dashed lines denote the estimates.

As an example, consider fitting n lines (n is fixed and known) to a set of planar points. Then MSS for n consists of $2n$ points. In the sequential approach, a single line is found by some version of RANSAC in each step and the inlier points are removed.

Depending on a problem, sequential RANSAC can have advantage of faster convergence because of smaller MSS, or it can fall into wrong solution (local optimum). See example in Figure 3.2.

3.3 Metropolis-Hastings Sampler

Appropriate for a model composed of sub-models. We assume, that a new proposal can be generated using some information from the previous proposal, e.g., some parameters of sub-models are kept, some are changed. The question which sub-models are changed is a part of proposal.

Sampling procedure. The procedure generates a sequence of parameters of the model. We also call the set of parameters of the model as a *state*. Let $\mathcal{S}_i$ denote the state (model parameters) generated at i-th step.

Init: Generate random state $\mathcal{S}_1$, e.g., using randomly sampled MSS (of the whole model).

Step $i + 1$:

1. Given $\mathcal{S}_i$, generate a random proposal of the state $\mathcal{S}'$. The generating procedure is a matter of choice.

2. Evaluate the conditional probability $q(\mathcal{S}'|\mathcal{S}_i)$ of this model changing 'forward' step and the probability $q(\mathcal{S}_i|\mathcal{S}')$ of the 'reverse' step. These probabilities are derived for a particular generating procedure.

3. Evaluate probabilities of the proposed state $p(\mathcal{S}'|\mathcal{X})$ and the previous state $p(\mathcal{S}_i|\mathcal{X})$.

4. Compute acceptance ratio a,
$$a = \frac{p(\mathcal{S}'|\mathcal{X})}{p(\mathcal{S}_i|\mathcal{X})} \cdot \frac{q(\mathcal{S}_i|\mathcal{S}')}{q(\mathcal{S}'|\mathcal{S}_i)}.$$

5. Generate random number u from unit-interval uniform distribution $\mathcal{U}_{0,1}$.

6. If $a > u$ then accept the new state, $\mathcal{S}_{i+1} \leftarrow \mathcal{S}'$, else keep the previous state, $\mathcal{S}_{i+1} \leftarrow \mathcal{S}_i$.

7. Iterate and remember the best state $\mathcal{S}^\star$ (the state having $p(\mathcal{S}|\mathcal{X})$ maximal).

3.3.1 M-H Sampler for Four Lines

The possible procedure generating the new proposal $\mathcal{S}'$ can be:

1. Randomly choose the number k of lines changed from uniform distribution, $k \in 1, 2, 3, 4$.

2. Randomly (uniformly) choose which of the four lines are changed; $\binom{4}{k}$ possibilities.

3. Randomly draw sample of $2k$ data points, construct the new lines replacing the lines selected in previous step.

Now the probability of the forward step and the probability of the reverse step is the same, thus terms q are cancelled. Then we evaluate only the posterior probability $p(\mathcal{S}|\mathcal{X})$ in each step, or assuming uniform prior $p(\mathcal{S})$, only the likelihood $p(\mathcal{X}|\mathcal{S})$.

Note that if $k = 4$, the new proposal does not use any information from the previous state. This case is also used for generating the initial state $\mathcal{S}_1$.