OPPA European Social Fund
Prague & EU: We invest in your future.

Fitting of a Planar Line

A[E]4M33TDV—3D compute vision: labs.

Martin Matoušek

October, 2011*

1 Maximum Likelihood Model Fitting

Given a set of measurements $\mathcal{X}$, a model $\boldsymbol{l}$ is estimated by maximising likelihood function

$$\boldsymbol{l}^* = \arg\max_{\boldsymbol{l}} p(\mathcal{X}|\boldsymbol{l})\,. \tag{1}$$

Hence the conditional probability of the measurements given the model $p(\mathcal{X}|\boldsymbol{l})$ must be known. Should also a prior probability $p(\boldsymbol{l})$ of the model be known, the maximum a posteriori estimate can be found by maximising $p(\mathcal{X}|\boldsymbol{l})p(\boldsymbol{l})$ instead.

In the case of fitting a line to a set of points, the measurements consists of a set of n point locations, $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ and the model $\boldsymbol{l}$ is an appropriate line representation. The conditional probability of the points given the line is derived in the next sections.

2 Trivial Example: Single Point Fitting

For the purpose of demonstration, a trivial example is considered first. The model $\boldsymbol{l}$ in (1) is a single point $\boldsymbol{x}$. Given the point (Euclidean) coordinates, the measurement is modelled as this very point polluted by an isotropic Gaussian noise,

$$\boldsymbol{x}_i = \boldsymbol{x} + \boldsymbol{e}_i\,, \tag{2}$$

$$\boldsymbol{e}_i \sim p(\boldsymbol{e}_i) = \mathcal{N}(\boldsymbol{e}_i; \boldsymbol{0}, \boldsymbol{\Sigma}_{\mathbf{e}})\,, \quad \boldsymbol{\Sigma}_{\mathbf{e}} = \begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix})\,. \tag{3}$$

Thus the conditional probability of a single measurement is

$$p(\boldsymbol{x}_i|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}_i - \boldsymbol{x}; \boldsymbol{0}, \boldsymbol{\Sigma}_{\mathbf{e}}) = \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{x}, \boldsymbol{\Sigma}_{\mathbf{e}}) = \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}_{\mathbf{e}}|}} e^{0.5(\boldsymbol{x}_i-\boldsymbol{x})^\top \boldsymbol{\Sigma}_{\mathbf{e}}{}^{1}(\boldsymbol{x}_i-\boldsymbol{x})} = \frac{1}{2\pi\sigma_e^2} e^{\frac{1}{2\sigma_e^2}||\boldsymbol{x}_i-\boldsymbol{x}||^2} \tag{4}$$

The measurements are assumed to be independent, so the overall probability of the measurements is

$$p(\mathcal{X}|\boldsymbol{x}) = n! \prod_{i=1}^{n} p(\boldsymbol{x}_i|\boldsymbol{x}) = n! \prod_{i=1}^{n} \mathcal{N}(\boldsymbol{x}_i - \boldsymbol{x}; \boldsymbol{0}, \boldsymbol{\Sigma}_{\mathbf{e}})\,, \tag{5}$$

where the $n!$ term is added since the order of the measurements does not matter.

The ML estimate of the point from the measurements is then

$$\begin{aligned} \boldsymbol{x}^* &= \arg\max_{\boldsymbol{x}} p(\mathcal{X}|\boldsymbol{x}) = \arg\min_{\boldsymbol{x}} \left(-\log p(\mathcal{X}|\boldsymbol{x})\right) = \\ &= \arg\min_{\boldsymbol{x}} \left\{ -\log(n!) + \sum_{i=1}^{n} \left(-\log \frac{1}{2\pi\sigma_e^2} + \frac{||\boldsymbol{x}_i - \boldsymbol{x}||^2}{2\sigma_e^2} \right) \right\} = \arg\min_{\boldsymbol{x}} \sum_{i=1}^{n} ||\boldsymbol{x}_i - \boldsymbol{x}||^2\,, \end{aligned} \tag{6}$$

*Last revision: October 13, 2011

where a monotonous function (log) was applied and some constants were omitted. This simplification obviously does not affect the result, and transforms the original problem of likelihood maximisation into the sum-of-squared-errors (SSE) minimisation.

Note, that in this simple case the coordinate components $x, y, \ldots$ are independent. Hence the minimiser of (6) is found in closed-form simply by laying the derivation according to each component of $\boldsymbol{x}$ equal zero,

$$\boldsymbol{0} = \frac{\partial}{\partial \boldsymbol{x}} \sum_{i=1}^{n} ||\boldsymbol{x}_i - \boldsymbol{x}^*||^2 = \sum_{i=1}^{n} 2(\boldsymbol{x}^* - \boldsymbol{x}_i) = 2n\boldsymbol{x}^* - 2\sum_{i=1}^{n} \boldsymbol{x}_i \Rightarrow \boldsymbol{x}^* = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \,. \tag{7}$$

As expected, the ML estimate (in arbitrary number of dimensions) of the point under isotropic Gaussian noise is simply the mean of the set $\mathcal{X}$.

3 Model for Noisy Planar Line and ML Estimate

A normalised planar line $\boldsymbol{l} = (\boldsymbol{n}^\top, d)$ is given, where $\boldsymbol{n}$ is normal vector. There is also direction vector $\boldsymbol{u}$, such that $\boldsymbol{u} \perp \boldsymbol{n}$. These two vectors forms a basis of line coordinates. A point $\tilde{\boldsymbol{x}}$ (exactly) lying on the line is created as

$$\tilde{\boldsymbol{x}}_i = \begin{bmatrix} \boldsymbol{u} & \boldsymbol{n} \end{bmatrix} \begin{bmatrix} t_i \\ -d \end{bmatrix}, \tag{8}$$

where t is a parameter along a line. Note, that the Euclidean coordinates of points are used. The distribution of the parameter is chosen as

$$t_i \sim p(t_i) = \mathcal{N}(t_i; \mu_t, \sigma_t)\,. \tag{9}$$

The measurement $\boldsymbol{x}_i$ is the line point polluted by an isotropic Gaussian noise,

$$\boldsymbol{x}_i = \tilde{\boldsymbol{x}}_i + \boldsymbol{e}_i\,, \tag{10}$$

$$\boldsymbol{e}_i \sim p(\boldsymbol{e}_i) = \mathcal{N}(\boldsymbol{e}_i; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_e & 0 \\ 0 & \sigma_e \end{bmatrix})\,. \tag{11}$$

The noise can be expressed in the line coordinates as

$$\boldsymbol{e}_i = \begin{bmatrix} \boldsymbol{u} & \boldsymbol{n} \end{bmatrix} \begin{bmatrix} t_i^e \\ d_i^e \end{bmatrix}, \quad p_e\left(\begin{bmatrix} d_i^e \\ t_i^e \end{bmatrix}\right) = \mathcal{N}(d_i^e; 0, \sigma_e)\,\mathcal{N}(t_i^e; 0, \sigma_e)\,, \tag{12}$$

and the measurement expressed in the line coordinates is then

$$\boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{u} & \boldsymbol{n} \end{bmatrix} \begin{bmatrix} t_i + t_i^e \\ -d + d_i^e \end{bmatrix}. \tag{13}$$

This leads to (using transformation of probability by linear transformation with unit Jacobian)

$$p(\boldsymbol{x}_i|\boldsymbol{l}, t_i) = p_e\left(\begin{bmatrix} \boldsymbol{u}^\top \\ \boldsymbol{n}^\top \end{bmatrix} \boldsymbol{x}_i - \begin{bmatrix} t_i \\ -d \end{bmatrix}\right) = \mathcal{N}(\boldsymbol{u}^\top \boldsymbol{x}_i - t_i; 0, \sigma_e)\,\mathcal{N}(\underbrace{\boldsymbol{n}^\top \boldsymbol{x}_i + d}_{d_i}; 0, \sigma_e)\,. \tag{14}$$

Here the term $d_i = \boldsymbol{n}^\top \boldsymbol{x}_i + d$ represents the orthogonal distance of the point $\boldsymbol{x}_i$ to the line $\boldsymbol{l}$.

Now the probability $p(\boldsymbol{x}_i|\boldsymbol{l})$ is derived from the joint p.d.f. $p(\boldsymbol{x}_i, t_i|\boldsymbol{l})$ by marginalisation over t_i.

$$\begin{aligned} p(\boldsymbol{x}_i|\boldsymbol{l}) &= \int_{-\infty}^{\infty} p(\boldsymbol{x}_i, t_i|\boldsymbol{l})\,\mathbf{d}t_i = \int_{-\infty}^{\infty} p(\boldsymbol{x}_i|\boldsymbol{l}, t_i)\,p(t_i)\,\mathbf{d}t_i = \\ &= \int_{-\infty}^{\infty} \mathcal{N}(\boldsymbol{u}^\top \boldsymbol{x}_i - t_i; 0, \sigma_e)\,\mathcal{N}(d_i; 0, \sigma_e)\,\mathcal{N}(t_i; \mu_t, \sigma_t)\,\mathbf{d}t_i = \end{aligned}$$

$$= \mathcal{N}(d_i;0,\sigma_e) \int_{-\infty}^{\infty} \mathcal{N}(t_i;-\boldsymbol{u}^\top \boldsymbol{x}_i,\sigma_e)\,\mathcal{N}(t_i;\mu_t,\sigma_t)\,\mathbf{d}t_i =$$

$$= \mathcal{N}(d_i;0,\sigma_e)\,\frac{1}{\sqrt{2\pi(\sigma_e^2+\sigma_t^2)}} e^{\frac{-(-\boldsymbol{u}^\top \boldsymbol{x}_i-\mu_t)^2}{2(\sigma_e^2+\sigma_t^2)}} \tag{15}$$

(The last step with fruitful help of Maple solver.)

Now the variance σ_t is chosen much larger than the size of image, where the points are observed, the mean μ_t is chosen e.g. zero (assuming that origin is in the image centre). Then the exponent is close to zero, and the probability is approximately

$$p(\boldsymbol{x}_i|\boldsymbol{l}) \approx \gamma\,\mathcal{N}(d_i;0,\sigma_e)\,, \tag{16}$$

where γ is some constant.

Again, the measurements are assumed to be independent, so the overall conditional probability of the measurements is

$$p(\mathcal{X}|\boldsymbol{l}) = n!\prod_{i=1}^{n} p(\boldsymbol{x}_i|\boldsymbol{l}) = n!\,\gamma^n \prod_{i=1}^{n} \mathcal{N}(d_i;0,\sigma_e)\,. \tag{17}$$

Finally, log is applied and unnecessary constants are omitted to form a log-likelihood to minimise. The ML estimate of the line from the measurements is then

$$\begin{aligned} \boldsymbol{l}^* &= \arg\max_{\boldsymbol{l}} p(\mathcal{X}|\boldsymbol{l}) = \arg\min_{\boldsymbol{l}} (-\log p(\mathcal{X}|\boldsymbol{l})) = \\ &= \arg\min_{\boldsymbol{l}} \left\{ -\log(n!\gamma^n) + \sum_{i=1}^{n} \left(-\log\frac{1}{\sqrt{2\pi\sigma_e^2}} + \frac{d_i^2}{2\sigma_e^2} \right) \right\} = \arg\min_{\boldsymbol{x}} \sum_{i=1}^{n} d_i^2\,. \end{aligned} \tag{18}$$

Again, the original problem of likelihood maximisation is transformed into the sum-of-squared-errors (SSE) minimisation. There is no closed form solution of (18), some numerical approach must be used.

4 Model for Noisy Planar Line Points with Outliers.

The process that creates set of points for a given line $\boldsymbol{l}$ is modelled by following three random processes that participates on generating each planar point $\boldsymbol{x}_i$.

1. Label generator. A label $L_i \in \{I, O\}$ determining if a point is inlier or outlier is randomly drawn.

$$L_i \sim p_L(L_i)\,,\quad p_L(L_i = I) = \alpha\,(\mathrm{const})\,,\quad p_L(L_i = 0) = 1-\alpha$$

2. Inlier generator. If $L_i = I$, a point location $\boldsymbol{x}_i$ belonging to the line is generated as described in Section 3.

$$\boldsymbol{x}_i \sim p(\boldsymbol{x}_i|\boldsymbol{l})$$

3. Outlier generator. If $L_i = O$, an outlier (not dependent on the line) is drawn from uniform distribution, assuming finite image of area $1/\beta$.

$$\boldsymbol{x}_i \sim p_O(\boldsymbol{x}_i) = \beta(\mathrm{const.})$$

This leads the robust joint probability of point locations and labels, given the line, to be

$$p_R(\boldsymbol{x}_i, L_i|\boldsymbol{l}) = \begin{cases} \alpha\,p(\boldsymbol{x}_i|\boldsymbol{l}) & \text{if } L_i = I \\ (1-\alpha)\,\beta & \text{if } L_i = O \end{cases} \tag{19}$$

The resulting robust probability $p_R(\boldsymbol{x}_i|\boldsymbol{l})$ is obtained by marginalisation (over all values of L_i) as

$$p_R(\boldsymbol{x}_i|\boldsymbol{l}) = \sum_{L_i \in \{I,O\}} p_R(\boldsymbol{x}_i, L_i|\boldsymbol{l}) = \alpha\, p(\boldsymbol{x}_i|\boldsymbol{l}) + (1-\alpha)\,\beta\,. \tag{20}$$

The points are assumed to be independent, so using (16) the overall robust probability of measurements is

$$\begin{aligned}
p_R(\mathcal{X}|\boldsymbol{l}) &= n!\prod_{i=1}^{n} p_R(\boldsymbol{x}_i|\boldsymbol{l}) = n!\prod_{i=1}^{n}\left(\alpha\, p(\boldsymbol{x}_i|\boldsymbol{l}) + (1-\alpha)\,\beta\right) = \\
&= n!\prod_{i=1}^{n}\left(\frac{\alpha\gamma}{\sqrt{2\pi\sigma_e^2}}e^{-\frac{d_i^2}{2\sigma_e^2}} + (1-\alpha)\beta\right). \\
&= c_1\prod_{i=1}^{n}\left(e^{-\frac{d_i^2}{2\sigma_e^2}+c_2}\right),
\end{aligned} \tag{21}$$

where c_1, c_2 are some constants. Again, cost function from minus log-likelihood is constructed prior to optimisation

$$-\log\left(p_R(\mathcal{X}|\boldsymbol{l})\right) = -c_1 - \sum_{i=1}^{n}\log\left(e^{\frac{-d_i^2}{2\sigma_e^2}} + c_2\right)$$

$$C(\boldsymbol{l}) = \sum_{i=1}^{n} -e^{2\sigma_e^2}\log\left(e^{-\frac{d_i^2}{2\sigma_e^2}} + e^{\frac{\theta^2}{2\sigma_e^2}}\right) = \sum_{i=1}^{n} C_1(d_i)\,. \tag{22}$$

Here the constant c_2 was replaced by a threshold θ. The cost function (robust penalty) $C1$ is analysed in the next section.

4.1 Robust Penalty

The previous section reveals a typical property of the robust model fitting problem under Gaussian noise: the error (d_i in the case of line fitting) is modelled using a mixture of normal and constant probability density. The mixture p.d.f and its negative logarithm is

$$p(d_i) = \alpha\mathcal{N}(d_i;0,\sigma_e) + (1-\alpha)\beta\,, \quad -\log\left(p(d_i)\right) = -\log\left(\frac{\alpha}{\sqrt{2\pi\sigma_e^2}}e^{\frac{-d_i^2}{2\sigma_e^2}} + (1-\alpha)\beta\right) \tag{23}$$

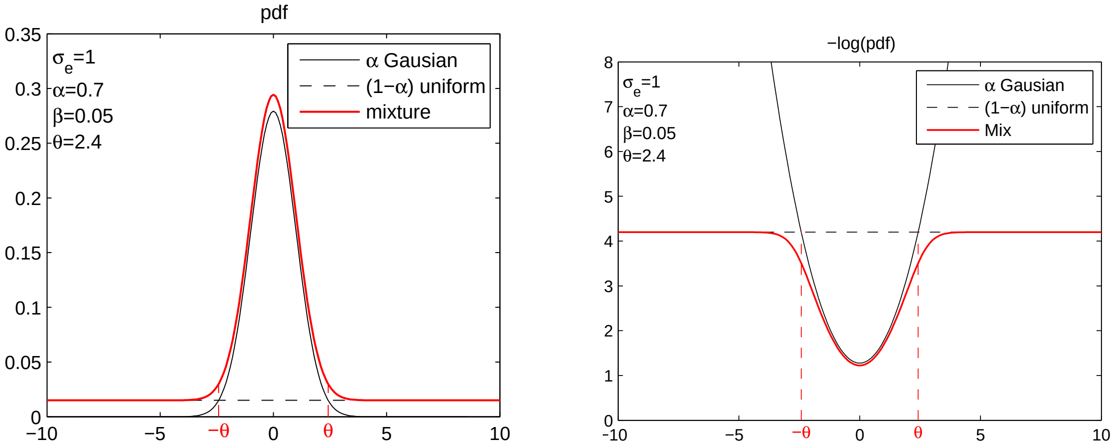
where α is a mixing coefficients. Example is in Figure 1.

There is an important value of d_i on the intersection of both densities (i.e. the probability of both processes is the same), denoted as a threshold θ.

$$\frac{\alpha}{\sqrt{2\pi\sigma_e^2}}e^{\frac{-\theta^2}{2\sigma_e^2}} = (1-\alpha)\beta \quad \Rightarrow \quad \theta = \sqrt{-\log\left(\frac{(1-\alpha)\beta}{\alpha}\sqrt{2\pi\sigma_e^2}\right)2\sigma_e^2} \tag{24}$$

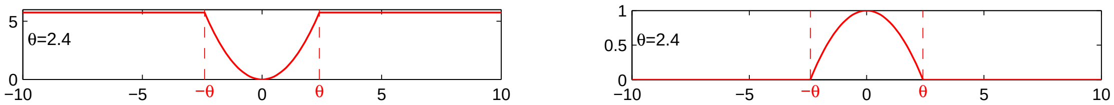
Usually, the threshold is used to parametrise the mixture, and the Gaussian variance σ_e is assumed approx $\sigma_e \in (0.1\theta, 0.5\theta)$ (the variance affects only the curvature of the penalty function near the threshold). Then the robust penalty function after removing some constants is

$$C_1(d_i) = -2\sigma_e^2\log\left(e^{\frac{-d_i^2}{2\sigma_e^2}} + e^{\frac{-\theta^2}{2\sigma_e^2}}\right). \tag{25}$$



(a) (b)

Fig. 1: Mixture of a Gaussian and a constant probability density. (a) Probability density function and (b) its negative logarithm.



(a) (b)

Fig. 2: Approximation of a robust penalty. (a) penalty function C_1 (26), (b) likelihood function L_1 (28).

This function can be approximated by two segments, one quadratic and one constant,

$$C_1(d_i) \approx \begin{cases} d_i^2 & \text{if } |d_i| < \theta \\ \theta^2 & \text{if } |d_i| > \theta \end{cases} . \tag{26}$$

The original problem is then solved by a numeric minimisation of $\sum C_1(d_i)$. Alternatively, the negative of $C_1(d_i)$ can be used for maximisation, i.e. (with a constant shift that do not affect optimisation),

$$L_1(d_i) = 1 - \frac{C_1(d_i)}{\theta^2} , \tag{27}$$

which is then approximated as

$$L_1(d_i) \approx \begin{cases} 1 - \frac{d_i}{\theta}^2 & \text{if } |d_i| < \theta \\ 0 & \text{if } |d_i| > \theta \end{cases} . \tag{28}$$

Both approximations are demonstrated in Figure 2. Then the original problem is solved by a numeric maximisation of $\sum L_1(d_i)$.

OPPA European Social Fund
Prague & EU: We invest in your future.