Lecture 8: Model Fitting

1 Noise Models

In Lecture 7 we studied the RANSAC approach for removing incorrect point correspondences. Once these outliers have been removed we still have to deal with noise in the measurements. Since the appearance of a patch changes when the viewpoint changes, exact positioning of corresponding points is not possible, see Figure 1,. Therefore, our point measurements will always be corrupted by noise of various forms and levels.

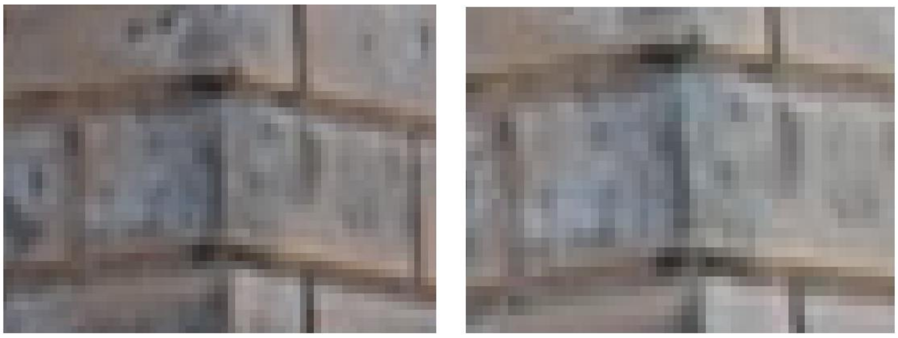


Figure 1: Two patches extracted from images with slightly different viewpoint. Exact localization of corresponding points is made difficult because of slight appearance differences and limited image resolution.

In Lectures 3,4,5 and 6 we solved various problems using linear formulations for approximately solving the governing algebraic equations. While this is an easy approach it does in general not give the "best" possible fit to the data. In this lecture we will derive formulations that gives the "best" fit under the assumption of Gaussian noise. The resulting problems are in general more difficult to solve than the formulations that we have used previously. In many cases they can only be locally optimized. Therefore the linear approaches are still very useful since they provide an easy way of creating a starting solution.

2 Line Fitting

What is meant by the "best" fit depends on the particular noise model. In this section we will consider two different noise models and show that they lead to different optimization criteria. For simplicity we will consider the problem of line fitting since this leads to closed form solutions.

2.1 Linear Least Squares

Suppose that (x_i, y_i) are measurements of 2D-points belonging to a line $y = ax + b$. Furthermore, we assume that y_i is corrupted by Gaussian noise, that is,

$$y_i = \tilde{y}_i + \epsilon_i \tag{1}$$

where $\epsilon_i \in \mathcal{N}(0, 1)$ (Gaussian noise with mean 0 and standard deviation 1) and $\tilde{y}_i$ is the true y-coordinate. Our goal is to estimate the line parameters a and b for which the measurements y_i are most likely. Since $\epsilon_i \in \mathcal{N}(0, 1)$, its probability density function is

$$p(\epsilon_i) = \frac{1}{\sqrt{2\pi}} e^{-\epsilon_i^2/2}. \tag{2}$$

Furthermore, if we assume that the ϵ_i, $i = 1, ..., n$ are independent of each other then their joint distribution is

$$p(\epsilon) = \prod_{i=1}^{n} p(\epsilon_i), \tag{3}$$

where $\epsilon = (\epsilon_1, \epsilon_2, ..., \epsilon_n)$. Since $\epsilon_i = y_i - \tilde{y}_i$ we can compute the likelihood of the measurements by

$$p(\epsilon) = \prod_{i=1}^{n} p(\epsilon_i) = \prod_{i=1}^{n} p(y_i - \tilde{y}_i) = \prod_{i=1}^{n} p(y_i - (ax_i + b)) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-(y_i - (ax_i + b))^2/2}. \tag{4}$$

We now want to find the the line parameters a and b that make these measurements most likely. To simplify the maximization we maximize the logarithm of the likelihood

$$\log \left(\prod_{i=1}^{n} p(\epsilon_i) \right) = -\sum_{i=1}^{n} \frac{(y_i - (ax_i + b))^2}{2} + \sum_{i=1}^{n} \log \left(\frac{1}{\sqrt{2\pi}} \right). \tag{5}$$

Since the second term does not depend on a of b this is the same as minimizing

$$\sum_{i=1}^{n} (y_i - (ax_i + b))^2. \tag{6}$$

In matrix form we can write this as

$$\left\| \underbrace{\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}}_{=A} \begin{pmatrix} a \\ b \end{pmatrix} - \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{=B} \right\|^2. \tag{7}$$

The minimum of this expression can be computed using the normal equations

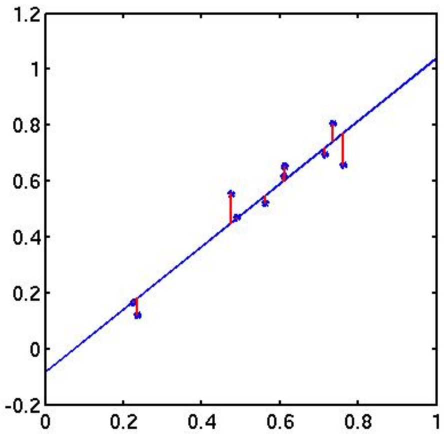
$$\begin{pmatrix} a \\ b \end{pmatrix} = (A^T A)^{-1} A^T B, \tag{8}$$

which we will derive in Lecture 9. The geometric interpretation of (6) is that under this noise model the vertical distance between the line and the measurement should be minimized, see Figure 2.

2.2 Total Linear Least Squares

Next we will assume that we have noise in both coordinates, that is,

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} \tilde{x}_i \\ \tilde{y}_i \end{pmatrix} + \delta_i, \tag{9}$$



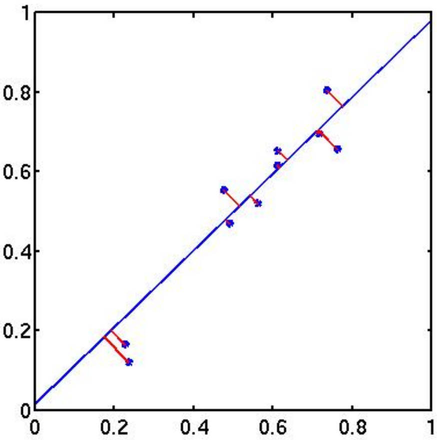


Figure 2: Left: The vertical distances between the line and the measured points are minimized in (6). In contrast, the minimal distances between the line and the measured points are minimized in (12).

where $\delta_i \in \mathcal{N}(0, I)$ and $a\tilde{x}_i + b\tilde{y}_i = c$. The δ_i now belong to a two dimensional normal distribution with probability density function

$$p(\delta_i) = \frac{1}{2\pi} e^{-\|\delta_i\|^2/2}. \tag{10}$$

The log likelihood function is

$$\sum_{i=1}^{n} \log(p(\delta_i)) = -\sum_{i=1}^{n} \frac{(x_i - \tilde{x}_i)^2 + (y_i - \tilde{y}_i)^2}{2} + \sum_{i=1}^{n} \log(\frac{1}{2\pi}). \tag{11}$$

Therefore, to maximize the likelihood we need to minimize

$$\sum_{i=1}^{n} ((x_i - \tilde{x}_i)^2 + (y_i - \tilde{y}_i)^2), \tag{12}$$

where $a\tilde{x}_i + b\tilde{y}_i = c$. The point $(\tilde{x}_i, \tilde{y}_i)$ can be any point on the line, however since we are minimizing (12) we can restrict it to be the closest point on the line. The expression (12) then becomes the distance between (x_i, y_i) and the line. This distance can be expressed using the distance formula as

$$\frac{|ax_i + by_i + c|}{a^2 + b^2}. \tag{13}$$

Without loss of generality we ca assume that $a^2 + b^2 = 1$, and therefore we need to solve

$$\min \quad \sum_{i=1}^{n} (ax_i + by_i + c)^2 \tag{14}$$

$$s.t. \quad a^2 + b^2 = 1. \tag{15}$$

This problem is often referred to as the total linear least squares problem.

2.2.1 Solving the Total Least Squares Problem

To solve (14),(15) we first take derivatives with respect to c. This shows that the the optimal solution must fulfill

$$c = -(a\bar{x} + b\bar{y}), \tag{16}$$

where $\bar{x}$ and $\bar{y}$ are the mean values

$$(\bar{x}, \bar{y}) = \frac{1}{n}\sum_{i=1}^{n}(x_i, y_i). \tag{17}$$

Substituting into (14) we get

$$\begin{aligned} \min \quad & \sum_{i=1}^{m}(a(x_i - \bar{x}) + b(y_i - \bar{y}))^2 & (18)\\ \text{such that} \quad & 1 - (a^2 + b^2) = 0. & (19) \end{aligned}$$

By forming the matrix

$$M = \sum_{i=1}^{m}\begin{pmatrix} (x_i - \bar{x})^2 & (x_i - \bar{x})(y_i - \bar{y}) \\ (x_i - \bar{x})(y_i - \bar{y}) & (y_i - \bar{y})^2 \end{pmatrix}, \tag{20}$$

we can write this problem as

$$\begin{aligned} \min \quad & t^T M t & (21)\\ \text{such that} \quad & 1 - t^T t = 0, & (22) \end{aligned}$$

where t is a 2×1 vector containing a and b. This is a constrained optimization problem of the type

$$\begin{aligned} \min \quad & f(t) & (23)\\ \text{such that} \quad & g(t) = 0. & (24) \end{aligned}$$

According to Persson-Böiers, "Analys i flera variabler" and the method of Lagrange multipliers the solution of such a system has to fulfill

$$\nabla f(t) + \lambda \nabla g(t) = 0. \tag{25}$$

Therefore the solution of (21)-(22) must fulfill

$$2Mt + \lambda(-2t) = 0 \Leftrightarrow Mt = \lambda t. \tag{26}$$

That is, the solution t has to be an eigenvector of the matrix M. Furthermore, inserting into (21), and using that $t^T t = 1$ we see that it has to be the eigenvector corresponding to the smallest eigenvalue.

3 The Maximum Likelihood Solution for Camera Systems

In this section we derive the maximum likelihood estimator for our class projection problems. Suppose the 2D-point $x_{ij} = (x_{ij}^1, x_{ij}^2)$ is a projection in regular Cartesian coordinates of the 3D-point $\mathbf{X}_j$ in camera P_i. The projection in regular coordinates can be written

$$\left(\frac{P_i^1\mathbf{X}_j}{P_i^3\mathbf{X}_j}, \frac{P_i^2\mathbf{X}_j}{P_i^3\mathbf{X}_j}\right), \tag{27}$$

where P_i^1, P_i^2, P_i^3 are the rows of the camera matrix P_i. Also we assume that the observations are corrupted by Gaussian noise, that is,

$$(x_{ij}^1, x_{ij}^2) = \left(\frac{P_i^1\mathbf{X}_j}{P_i^3\mathbf{X}_j}, \frac{P_i^2\mathbf{X}_j}{P_i^3\mathbf{X}_j}\right) + \delta_{ij}, \tag{28}$$

and δ_{ij} is normally distributed with covariance I. The probability density function is then

$$p(\delta_{ij}) = \frac{1}{2\pi}e^{-\frac{1}{2}||\delta_{ij}||^2}. \tag{29}$$

Similarly to Section 2.2 we now see that the model configuration that maximizes the likelihood of the obtaining the observations $x_{ij} = (x_{ij}^1, x_{ij}^2)$ is obtained by solving

$$\min \sum_{i=1}^{n} \sum_{j=1}^{m} \left\| \left(x_{ij}^1 - \frac{P_i^1 \mathbf{X}_j}{P_i^3 \mathbf{X}_j},\ x_{ij}^2 - \frac{P_i^2 \mathbf{X}_j}{P_i^3 \mathbf{X}_j} \right) \right\|^2. \tag{30}$$

where n is the number of cameras and m is the number of scene points. The geometric interpretation of the above expression is that the distance between the projection and the measured point in the image should be minimized, see Figure 3. Note that it does not matter which of the variables P_i and X_i we consider as unknowns, it is always the reprojection error that should be minimized.

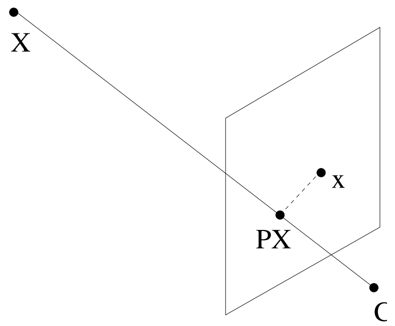


Figure 3: Geometric interpretation of the maximum likelihood estimate for projection problems. The dashed distance should be minimized.

3.1 Affine Cameras

In general the maximum likelihood estimator (30) can only be solved using local iterative methods. However in the special case of affine cameras there is a closed form solution. An affine camera is a camera where the third row of P, $P^3 = [0\ 0\ 0\ t_3]$. Since the scale of the camera matrix is arbitrary we may assume that $t_3 = 1$, and therefore the camera matrix has the form

$$P = \begin{bmatrix} A & t \\ 0 & 1 \end{bmatrix}, \tag{31}$$

where A is a 2×3 matrix and t is a 2×1 vector. If we use regular Cartesian coordinates for both image points and scene points the camera equations can be simplified. If x_{ij} is the projection of the scene point X_j in the affine cameras P_i then the projection can be written

$$x_{ij} = A_i X_j + t_i. \tag{32}$$

To find he maximum likelihood estimate we therefore need to solve

$$\min \sum_{i=1}^{n} \sum_{j=1}^{m} \|x_{ij} - A_i X_j + t_i\|^2. \tag{33}$$

Dy differentiating with respect to t_i it can be seen that the optimal t_i is given by

$$t_i = \bar{x}_i - A_i \bar{X},$$

where $\bar{X} = \frac{1}{m} \sum_j X_j$ and $\bar{x}_i = \frac{1}{m} \sum_j x_{ij}$. To simplify the problem we therefore change coordinates so that all these mean values are zero by translating all image points and scene points. Using $\tilde{x}_{ij} = x_{ij} - \bar{x}_i$ and $\tilde{X}_i = X_i - \bar{X}$, gives the simplified problem

$$\min \sum_{ij} ||\tilde{x}_{ij} - A_i \tilde{X}_j||^2. \tag{34}$$

In matrix form we can write this as

$$\min \left\| \underbrace{\begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \dots & \tilde{x}_{1m} \\ \tilde{x}_{21} & \tilde{x}_{22} & \dots & \tilde{x}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{n1} & \tilde{x}_{n2} & \dots & \tilde{x}_{nm} \end{bmatrix}}_{M} - \underbrace{\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} \begin{bmatrix} \tilde{X}_1 & \tilde{X}_2 & \dots & \tilde{X}_m \end{bmatrix}}_{\text{rank 3 matrix}} \right\|^2 . \tag{35}$$

Since the A_i has only 3 columns the second term will be rank 3 matrix. Thus our problem is to find the matrix of rank 3 that best approximates M. The best approximating matrix can be found by computing the SVD of M and setting all but the first 3 singular values to zero.

We summarize the algorithm for affine cameras here:

1. Re-center all images so that the center of mass of the image points is zero in each image.
2. Form the measurement matrix M.
3. Compute the SVD:
$$M = USV^T. \tag{36}$$
4. A solution can be found by extracting the cameras from $U(:, 1 : 3)$ and the structure from $S(1 : 3, 1 : 3) * V(:, 1 : 3)'$.
5. Transform back the solution to the original image coordinates.

Note that the approach only works when all points are visible in all images. Furthermore, the camera model is affine, which is a simplification. This is often a good approximation when the scene point have roughly the same depth.