CSci 6974 and ECSE 6966 Math. Tech. for Vision, Graphics and Robotics
Lectures 19&20, April 6&13, 2006
Background for Estimation Techniques

Planned Schedule for Remainder of the Semester

- April 6 & 13: Introduction to estimation
- April 10: Test 2
- April 17: Estimation of planar homographies
- April 20: Robust estimation
- April 24 & May 1: Non-linear estimation
- April 27: Linear programming (Prof. Trinkle)

We will discuss the date and format of the 3rd exam later in April.

Major Issues in Estimation

Each time an estimation problem is addressed several questions must be answered:

Data: What is the form of the data? What is the uncertainty in the measurements? Might there be outliers (gross errors)? Are there multiple structures (populations) in the data?

Model and Parameters to be Estimated: What is the model and what is the parameter set? Are there constraints on the parameter set? We have already dealt with points, lines and planes, and found several choices of parameterizations.

Error Distance Metric: What is the "distance" between a point and an instance of the model.

Objective Function: What function of the error distances for the individual points is to be minimized? The model instance (parameter set) that minimizes this objective function over all possible instances is the solution to the estimation problem.

Minimization Technique: What method/algorithm will be used to search for the minimum of the objective function? Is data normalization necessary? Is the minimization linear or non-linear?

Robustness and Outliers: Are the objective function and minimization technique "robust" to outliers? In other words, is the estimate substantially altered by the addition of outliers?

Stability: How much does the estimated model change when the data is subject to small random perturbations?

A good choice reading reference is Hartley and Zisserman, Chapter 3 (1st edition) or Chapter 4 (2nd edition).

Preliminary Stuff

Before starting, we need to spend time on the mathematical and statistical background. This will include:

- Random variables.
- Probability density functions and cumulative density functions.
- Covariance matrices.
- Constrained minimization and Lagrange multipliers.
- Constrained minimization via the SVD.

Many of you will already know this material, but for others it will be relatively new. Please be patient.

Measurement Error

Each datum will be a measurement vector, $\mathbf{x}$. This may be written as

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Delta x} \tag{1}$$

where

- $\overline{\mathbf{x}}$ is the 'true" or "unperturbed" measurement vector, and
- $\mathbf{\Delta x}$ is the measurement error vector.

Random Vectors and Random Variables

The measurement error will be treated as a "random vector" — a vector of "random variables":

- Each random variable is a quantity governed by a probability. The value of the random variable changes each time it is "sampled", "tested" or "queried".
- Example random variables include:

- the result of a coin toss,
- the sum of the values on two dice when they are rolled,
- the intensity measured at an image pixel,
- the $(x, y)^\top$ location of a corner or edge in an image.

- In the case of the intensity measurement, the error is the direct result of a physical process, whereas for the corner or edge location, the error may be thought of as imperfections in the behavior of an algorithm.

Distributions

- The *cumulative distribution function* ("distribution" or "cdf") of a random variable, x_0, is a function $F(x)$ giving the probability that $x_0 \leq x$.

- Since probabilities are bounded in the range $[0, 1]$, we have

$$\lim_{x \to -\infty} F(x) = 0 \qquad \text{and}$$
$$\lim_{x \to \infty} F(x) = 1.$$

- $F(x)$ is clearly monotonically non-decreasing.

Densities

- The *probability density function* ("density" or "pdf") associated with a distribution is

$$f(x) = \frac{\partial F}{\partial x}. \tag{2}$$

- The *expected value* or *first moment* of a probability density function is

$$\mu = \mathrm{E}[x] = \int_{-\infty}^{\infty} x f(x) dx. \tag{3}$$

This equates to our intuitive notion of an average.

- The *variance* or *central second moment* of f is

$$\sigma^2 = \mathrm{E}[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx. \tag{4}$$

This measures the spread of the distribution. The value σ is the *standard deviation* of the distribution.

Examples: Normal, Uniform and Exponential Distributions

- The normal or Gaussian distribution is defined by the probability density function
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(\frac{-(x-\mu)^2}{2 * \sigma^2}\Big). \tag{5}$$

- The cumulative distribution is
$$F(x) = \int_{-\infty}^{x} f(u)du. \tag{6}$$

- The mean of this distribution is μ and the variance is σ^2. Proving this is a homework question.

- The density for a uniform distribution over an interval $[a, b]$, is simply
$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

- The density of the exponential distribution for $x \geq 0$ is
$$f(x) = \frac{1}{c} e^{-x/c} \tag{8}$$

  In class we will show that this is indeed a density and then compute the mean of this distribution.

Discrete Random Variables

- When the random variable x can only take on a discrete set of values, $\mathcal{X} = \{x_1, \ldots, x_k\}$, then we consider the probability density function as a "probability mass function" and its domain becomes the set $\mathcal{X}$.

- The integrals in the above expressions are replaced by sums.

- We must then have
$$\sum_{x_i \in \mathcal{X}} f(x_i) = 1 \tag{9}$$
  and the
$$F(x) = \sum_{x_i \in \mathcal{X}, x_i < x} f(x_i) \tag{10}$$

- The example we will consider in class is outcome of a role of fair dice.

Joint Probabilities

Consider two random variables x_0 and y_0.

- The joint distribution is a function $F(x, y)$ giving the probability that $x_0 \leq x$ and $y_0 \leq y$.

- If the two variables are independent then

$$F(x, y) = F(x)F(y). \tag{11}$$

- The joint density is

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}. \tag{12}$$

- The mean is now a vector. Defining $\mathbf{x} = (x, y)^T$, the mean is now

$$\mu = \mathrm{E}[\mathbf{x}] = \int_{\infty}^{\infty} \int_{\infty}^{\infty} \mathbf{x} f(x, y) dx dy. \tag{13}$$

  We write μ component-wise as $\mu = (\mu_x, \mu_y)^T$.

- The variance becomes a matrix called the covariance matrix:

$$\mathbf{\Sigma} = \mathrm{E}[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T] = \int_{\infty}^{\infty} \int_{\infty}^{\infty} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T f(x, y) dx dy. \tag{14}$$

  The integral of the matrix here looks intimidating, but it can be handled component-wise.

- The covariance matrix is positive semi-definite, and for two variables is written:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \tag{15}$$

  where σ_x^2 and σ_y^2 are the variances of the independent variables and σ_{xy} is the covariance of x and y.

- If x and y are independent, then $\sigma_{xy} = 0$. The converse is not necessarily true.

- The *marginal density* of one of the random variables is obtained by integrating out the other variable (or variables):

$$f(x) = \int_{\infty}^{\infty} f(x, y) dy \tag{16}$$

Back to Random Vectors: Multivariate Gaussian

For mean vector μ and covariance $\mathbf{\Sigma}$ the multivariable normal distribution is

$$f(\mathbf{x}) = (2\pi)^{-n/2} \det(\mathbf{\Sigma})^{-1/2} \exp(-(\mathbf{x}-\mu)^T \mathbf{\Sigma}^{-1} (\mathbf{x}-\mu)/2), \tag{17}$$

where n is the dimension of the vector.

- When the variables are independent, this reduces to the product of individual normal distributions.
- The main random vector we will consider is the calculated x and y locations of image points such as edges and corners.
- The simplest assumption we will make is that the errors in these values are independent random variables governed by the same distribution.

Samples, Means and Covariance Matrices

Suppose you are given N vectors $\mathbf{x}_i$ all from the same, unknown distribution.

- The estimated mean of the distribution is

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i. \tag{18}$$

- The estimated covariance matrix of the distribution is

$$\hat{\mathbf{\Sigma}} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})^T. \tag{19}$$

- These are called the *sample mean* and *sample covariance*, whereas the corresponding values in the actual distribution are called the *population mean* and *population covariance*.
- These measures provide a 2nd order (essentially Gaussian) approximation to the probability density function (pdf). There are many other methods to approximate the pdf.

Constrained Minimization Via Lagrange Multipliers

Goal: minimize a function $f(\mathbf{x})$ subject to a constraint $g(\mathbf{x}) = 0$. The teschnique of Lagrange multipliers is used to solve this problem as follows:

- Form the function
$$F(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x}). \tag{20}$$

- Calculate the partial derivatives and equate them to 0:
$$\frac{\partial F}{\partial \mathbf{x}} = 0$$
$$\frac{\partial F}{\partial \lambda} = 0$$

- Eliminating λ and solving yields the desired value of $\mathbf{x}$.

Example: Constrained Minimization

Minimize $\|\mathbf{A}\mathbf{x}\|$ subject to $\|x\| = 1$.

- The minimization is equivalent to minimizing $\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}$, while the constraint is equivalent to $\mathbf{x}^T\mathbf{x} - 1 = 0$.

- Form,
$$F(\mathbf{x}, \lambda) = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \lambda(\mathbf{x}^T\mathbf{x} - 1). \tag{21}$$

- Calculate partial derivatives:
$$\frac{\partial F}{\partial \mathbf{x}} = 2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\lambda\mathbf{x} = \mathbf{0},$$
$$\frac{\partial F}{\partial \lambda} = \mathbf{x}^T\mathbf{x} - 1 = 0$$

- The first constraint ensures that the solution $\mathbf{x}$ is an eigenvector of $\mathbf{A}^T\mathbf{A}$. The second constraint ensures that $\mathbf{x}$ is a unit vector.

- Looking at the spectral decomposition,
$$\mathbf{A}^T\mathbf{A} = \sum_{i=1}^{n} \sigma_i \mathbf{v}_i \mathbf{v}_i^T \tag{22}$$
tells us that $\mathbf{x}$ is the unit eigenvector corresponding to the smallest eigenvalue.

- **Important note:** The SVD solution to this minimization problem led to the same result. The lesson is that Lagrange multipliers are **not** the only technique that may be used.

Lagrange Multipliers: Second Example

What is the distance of a point, (u, v), to a conic?

- We want to find the point (x, y) on the conic that minimizes the distance to (u, v).
- Thus, our minimization problem is

$$F(x, y) = (x - u)^2 + (y - v)^2 + \lambda(ax^2 + bxy + cy^2 + dx + ey + f) \quad (23)$$

- Taking the partial derivatives with respect to x, y, and λ yields the system of equations

$$\begin{aligned} 2(x - u) + \lambda(2ax + by + d) &= 0 \\ 2(y - v) + \lambda(bx + 2cy + e) &= 0 \\ ax^2 + bxy + cy^2 + dx + ey + f &= 0 \end{aligned}$$

  Notice that the constraint always appears in exact form in the last term.

- Solving this leads to the problem of finding the roots of a 4th order polynomial!
  - We will discuss the geometric intuition about why this is true.
- This example shows that
  - Distance calculations aren't always trivial.
  - Root finding has significant practical applications.

Aside: Homogeneous vs. Inhomogeneous Forms of Estimation

Students often have trouble distinguishing between when to write a problem in the form

$$\|\mathbf{Ax}\|^2, \quad (24)$$

where $\mathbf{x}$ is the set of parameters to be solved, and and when to write it in the form

$$\|\mathbf{By} - \mathbf{c}\|^2 \quad (25)$$

where $\mathbf{y}$ is the set of parameters to be solved. (Of course $\mathbf{x}$ and $\mathbf{y}$ are closely related.) Here are some rules-of-thumb:

- Write down the distance metric.
- Write down the constraints on the parameters.
- Write down the minimization problem.

- Try to separate terms. Are there any data terms that are independent of the parameters? (Do not do this artificially!!)
  - If so, they can be used to form the vector **c**, and the inhomogeneous form should be used.
  - Otherwise, the homogeneous form should be used.
- We will consider this by reconsidering the two problems of finding the line closest to a set of points and the point closest to a set of lines.

Practice Problems

1. Using the given fact that

$$\int_{-\infty}^{\infty} \exp(-t^2/2)dt = \sqrt{2 * \pi},$$

   prove that the mean and variance of a univariate Gaussian distribution are μ and σ^2 respectively.

2. Derive the mean and variance of a uniform distribution taken over the interval $[a, b]$.

3. Compute the variance of the exponential distribution.

4. Consider the outcomes of rolling a pair of dice.

   (a) Write this as a bivariate distribution, where the first variable, x_1, is the outcome of rolling the first di and the second variable, x_2, is the outcome of rolling the second. Write the 2d probability mass function.

   (b) Use this to compute the probability that neither di is 6 (i.e. both dice have values less than 6).

   (c) Derive the mean of (x_1, x_2).

   (d) Derive the covariance matrix.

5. Use Lagrange multipliers to derive the minimum distance from a point to a plane in $\mathbb{R}^n$.

Homework Problems; Due Monday April 17, 2006

1. Use Lagrange multipliers to determine the closest point on the surface of the sphere with center $\mathbf{x}_c$ and radius r to a given point $\mathbf{p}$.

2. Given the function

$$f(x) = \begin{cases} c(1-u^2)^2 & |u| \leq 1 \\ 0 & |u| > 1 \end{cases}.$$

   (a) Derive the value of c necessary to make f a density.

   (b) Derive the mean and variance of f.