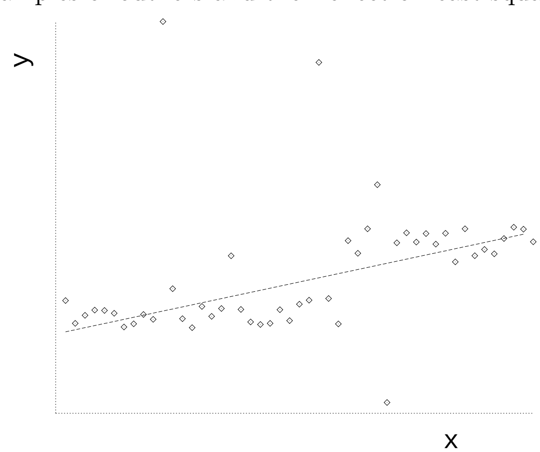
CSci 6974 and ECSE 6966 Math. Tech. for Vision, Graphics and Robotics
Lecture 22, April 20, 2006
Robust Estimation

Today's Lecture

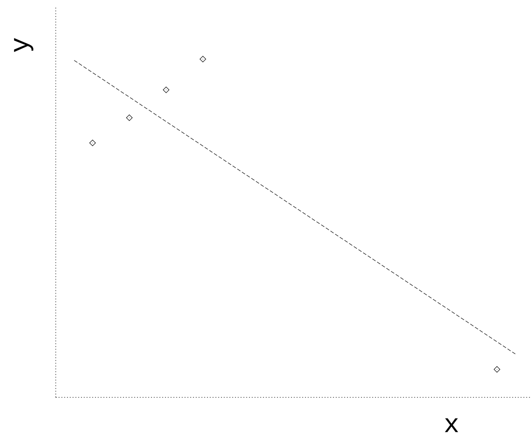
- Summarize different objective functions for homography estimation
- Normalization methods to improve numerical stability in general and the performance of techniques based on algebraic distance in particular.
- The following notes on robust estimation.

Robust Estimation Overview

Recall that "outliers" are extreme errors in measurement or are data from a different "population", other than the population of interest. The following two figures illustrate examples of outliers and their effect of least-squares line fits.



and



In the latter case, a single point — called a *leverage point* has radically skewed the least-squares fit.

Why This Happens

Consider the simple case of ordinary regression.

- The objective function
$$\sum_i (y_i - \mathbf{a}^\top \mathbf{x_i})^2 \tag{1}$$
gives a heavy penalty for points with large errors.
- The estimates are therefore skewed to reduce these large errors.
- In fact, one point out of n can skew the fit arbitrarily, regardless of n.
  - The minimum fraction of points that can arbitrarily corrupt an estimate is called the "breakdown point".
  - The breakdown point of least-squares is 0.
- There are two common solution approaches. The first is to reduce the "penalty" for large residuals to less than the quadratic function used in least-squares estimation. The second is to eliminate the largest residuals from consideration altogether.
- We will explore both types of solutions.

M-Estimators

M-Estimators estimate parameters by minimizing an objective function of the form
$$E(\mathbf{a}) = \sum_i \rho(u_i) \tag{2}$$

where ρ is a function whose growth as a function of $|u_i|$ is less than quadratic, and

$$u_i = \frac{y_i - \mathbf{a}^\top \mathbf{x}}{\sigma_i} \tag{3}$$

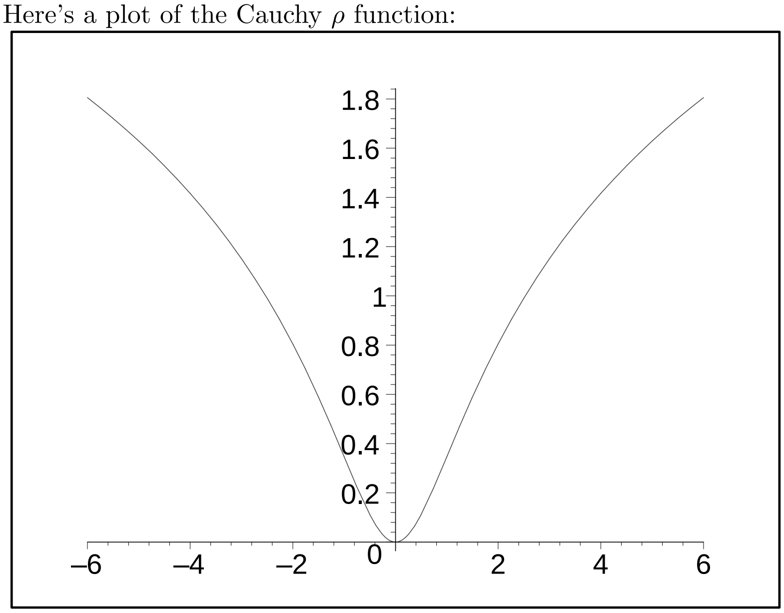
with "scale" value (error standard deviation) σ_i.

- Many robust ρ functions have been devised.
- One example is the Cauchy function, derived from the maximum-likelihood estimator of the Cauchy distribution:

$$\rho_C(u) = \frac{c^2}{2} \log\bigl[1 + \bigl(\frac{u}{c}\bigr)^2\bigr] \tag{4}$$

  where c is a constant.

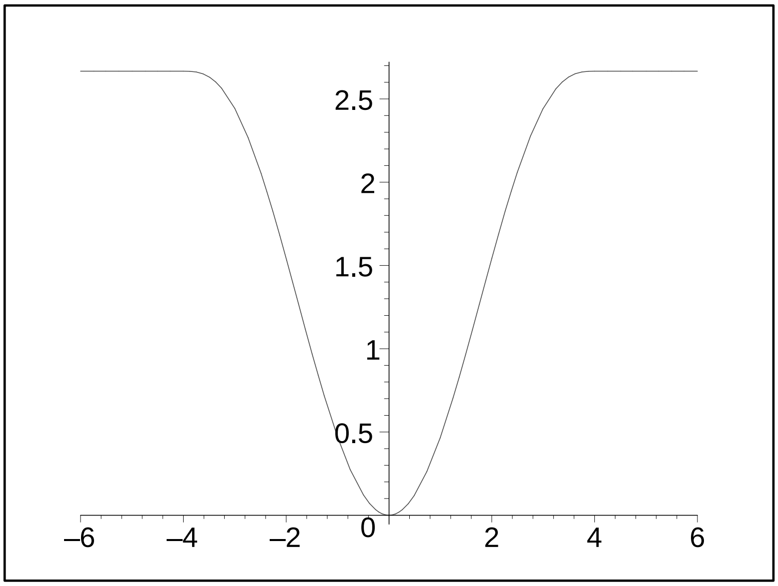
- Here's a plot of the Cauchy ρ function:



- Another, with a truncated ρ function, is the Beaton-Tukey biweight:

$$\rho_T(u) = \begin{cases} \frac{a^2}{6}\bigl[1 - (1 - \bigl(\frac{u}{a}\bigr)^2)^3\bigr] & |u| \le a \\ \frac{a^2}{6} & |u| > a \end{cases} \tag{5}$$

- Here's a plot of the Beaton-Tukey ρ function:



- Notice that each of the two ρ functions depends on σ and depends on a constant term. The constant is fairly easy to set, but σ is much more of a problem. It must either be provided in advance or estimated during minimization of the objective function.

Iteratively Reweighted Least-Squares — Derivation

Several algorithms to minimize (2) are in use. Here's the derivation of one of them, commonly known as "iteratively reweighted least-squares":

- To minimize $\sum_i \rho(u_i)$, take the derivative with respect to $\mathbf{a}$ and set the result equal to $\mathbf{0}$

$$\frac{\partial E}{\partial \mathbf{a}} = \sum_i \psi(u_i) \frac{du_i}{d\mathbf{a}} = \sum_i \frac{-\psi(u_i)\mathbf{x}_i}{\sigma} = \mathbf{0}. \tag{6}$$

  where $\psi(u) = \rho'(u)$. These are called the *estimating equations*.

- Define the function $w(u_i) = \psi(u_i)/u_i$, and make the substitution, $\psi(i) = w(u_i) \cdot u_i$. This yields

$$\sum_i w(u_i) \big[y_i \mathbf{x}_i - (\mathbf{x}_i \mathbf{x}_i^\top) \mathbf{a} \big] = \mathbf{0}, \tag{7}$$

  where we have factored out the σ term.

- Now, temporarily fix $w(u_i)$ for each i. Define $w_i = w(u_i)$.

- This gives the "solution" as a simple least-squares problem:

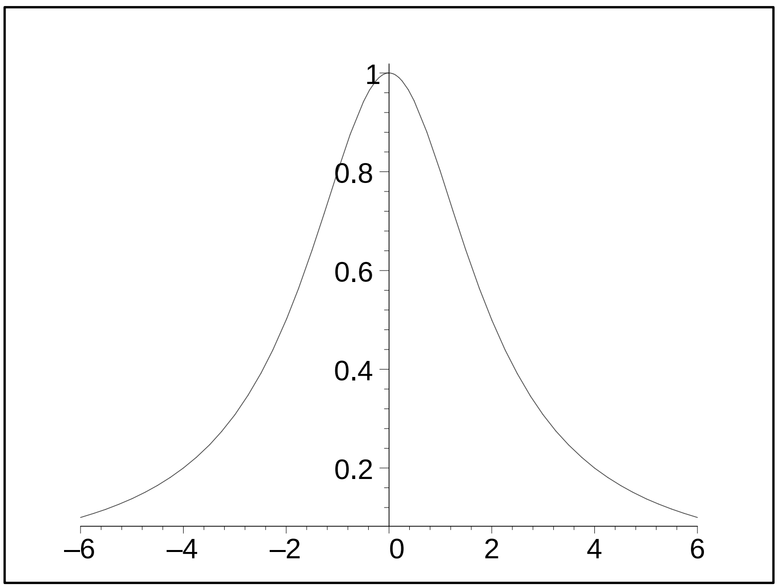
$$\hat{\mathbf{a}} = \big(\sum_i w_i \mathbf{x}_i \mathbf{x}_i^\top\big)^{-1} \sum_i w_i y_i \mathbf{x}_i. \tag{8}$$

  Note that this solution is depends on the w_i values which in turn depend on $\hat{\mathbf{a}}$.

- The idea is to alternate calculating $\hat{\mathbf{a}}$ and recalculating $w_i = w((y_i - \hat{\mathbf{a}}^\top \mathbf{x}_i)/\sigma_i)$.

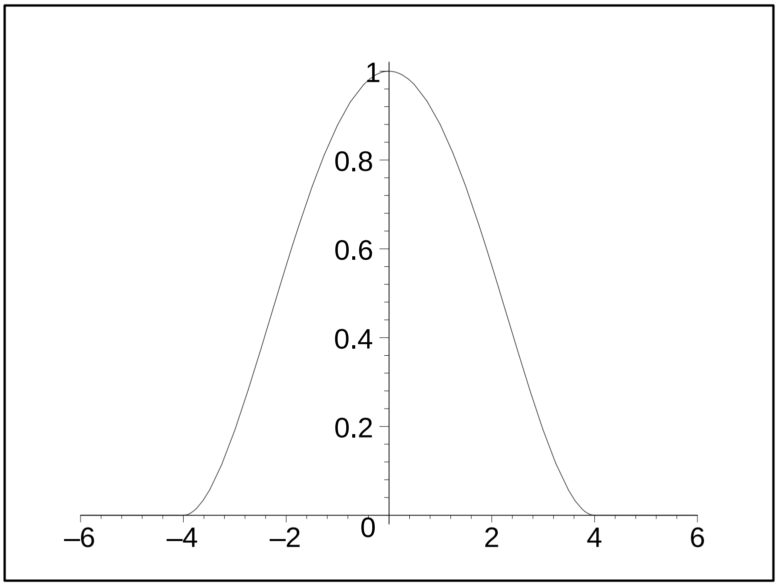
- Here are the weight functions associated with the two estimates. For the Cauchy ρ function,

$$w_C(u) = \frac{u}{1 + (u/c)^2} \tag{9}$$



  and, for the Beaton-Tukey ρ function,

$$w_T(u) = \begin{cases} \left[1 - \left(\frac{u}{a}\right)^2\right]^2 & |u| \le a \\ 0 & |u| > a \end{cases}. \tag{10}$$



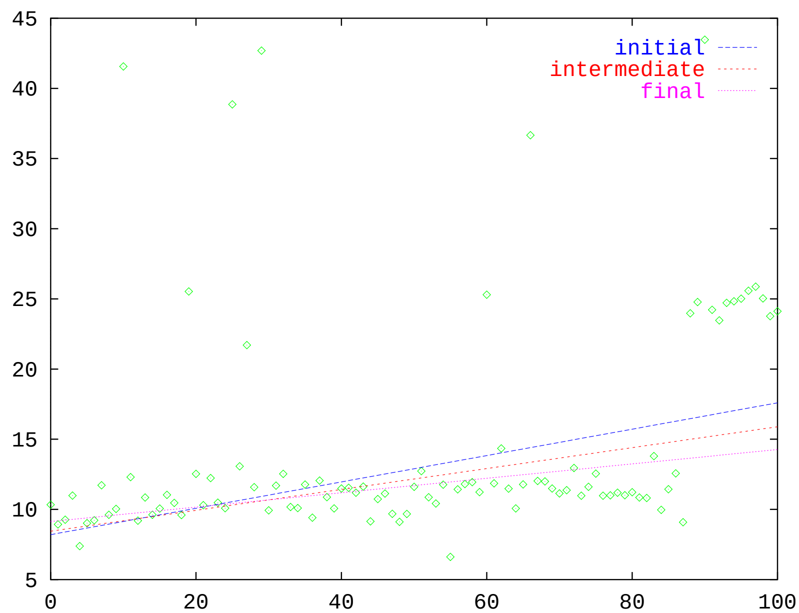
Iteratively Reweighted Least-Squares — Algorithm

Here's a summary of the resulting iteratively reweighted least-squares algorithm (IRLS). It assumes that the σ values are given.

1. Form an initial parameter estimate $\hat{\mathbf{a}}_0$. The simplest method is to use regular least-squares.
2. $t = 1$
3. Repeat for n iterations or until the change in the parameter estimate is small enough:
4. Calculate $w_i = w((y_i - \hat{\mathbf{a}}_{t-1}^\top \mathbf{x}_i)/\sigma)$.
5. Solve the weighted least-squares equation (8) to yield the new estimate, $\hat{\mathbf{a}}_t$.
6. $t++$;

Example and Discussion

The following figure illustrates several estimated lines during the stages of IRLS. The initial estimate was obtained from least-squares.



- Notice that the procedure doesn't quite converge all the way to the "correct" line — the one corresponding to the largest fraction of the data.
- The least-squares initialization does not allow IRLS to recover from all bad initializations. The earlier, "leverage" example, is one such case.
- A different robust estimation technique is needed to handle these more difficult examples.

Least-Median of Squares (LMS)

Instead of replacing the quadratic function, as is done in an M-Estimator, the summation is replaced by the median. This yields the objective function

$$E(\mathbf{a}) = \operatorname*{median}_i (y_i - \mathbf{a}^\top \mathbf{x}_i)^2 \tag{11}$$

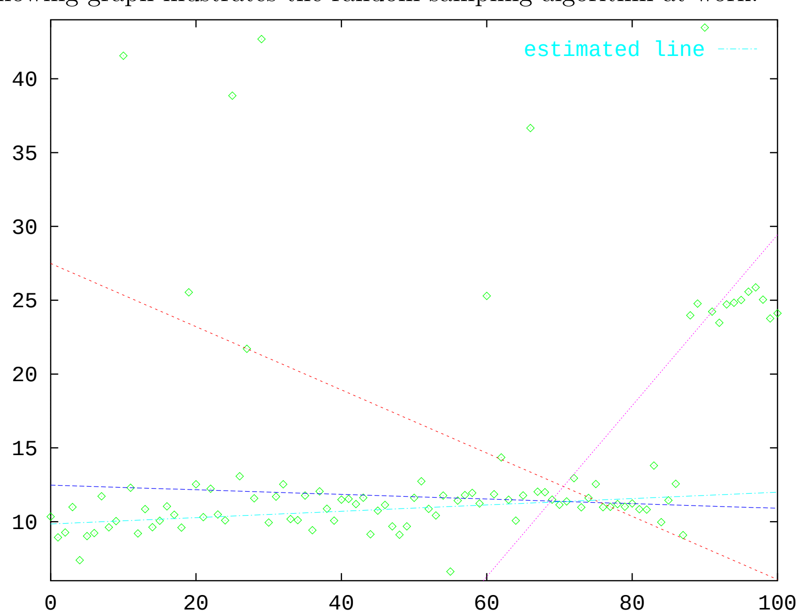
- Evaluating $E(\mathbf{a})$ for hypothesized parameter set $\mathbf{a}$ requires calculating the residual of each data point and then finding the median of the square residuals.
- For any $\mathbf{a}$, up to half the points can be arbitrarily far away without from the line (plane, hyperplane) without affecting $E(\mathbf{a})$.
- The parameter vector minimizing $E(\mathbf{a})$ corresponds to the line (plane, hyperplane) with the smallest surrounding interval (in the y direction) containing half the points.

LMS Minimization via Random-Sampling Search

- The objective function is not differentiable, so there is no hope of a closed-form solution (as in least-squares) and no hope of an iterative solution (as in IRLS).
- We need to be able to efficiently search the space of possible parameter vectors $\mathbf{a}$ to find the one that minimizes $E(\mathbf{a})$.
- The problem is doing this efficiently.
- The idea of a *random-sampling search* technique is to use the data points themselves to generate the $\mathbf{a}$ to test.
- For parameter vectors with k degrees of freedom, k data points are chosen randomly to generate a fit, and the objective function is evaluated with respect to the remaining $n-k$ points.
- This is repeated S times with different samples, and the generated $\mathbf{a}$ corresponding to the smallest median is taken as the estimate.
- The number of samples required is set to ensure a high probability of obtaining at least one that contains all inliers. The derivation will be given in class.

Example and Discussion

The following graph illustrates the random-sampling algorithm at work.



- The resulting estimate is better than the M-estimator result, but still not perfect.
- The instability in the estimate is due in part to the random-sampling search technique and in part to the use of the median as the objective function.
- Thus, even though LMS can tolerate up to 50% bad data, there are still drawbacks.
- Some of these may be corrected by combining LMS with an M-Estimator, as we will see below.
- First, however, we will look at a different random-sampling based estimator.

Random-Sample Consensus — RANSAC

LMS can not be applied when the fraction of inliers could be less than 50%. In vision applications, however, we can often have a rough idea of the standard deviation (σ) of the noise in the data.

- The RANSAC (Random Sample Consensus) objective function is as follows
$$E(\mathbf{a}) = \underset{i}{\text{card}}(|y_i - \mathbf{a}^\top \mathbf{x}_i| < \theta), \tag{12}$$
where θ is a constant, typically 2σ to 3σ.
- The same random sampling algorithm as used in LMS can be applied here. In fact, the RANSAC algorithm predates LMS by three years.

Combining Robust Estimators and Non-Robust Estimation

- The power of the random sampling algorithms is their ability to tolerate a large number of outliers. The weakness is the instability of the estimate.
- The power of M-estimators is the ability to produce stable estimates, but only when initialized in the vicinity of the correct estimate.
- The obvious solution, used in practice, is to start with a random sampling algorithm and then switch to an M-estimator.
- To run the M-Estimator, a scale estimate must either be given or estimated following LMS.

Practice Problems

1. This question explores the relationship between the RANSAC objective function (12) which is maximized and M-estimator objective functions, which are minimized.

   (a) Determine a function f such that

   $$\operatorname*{card}_i(|y_i - \mathbf{a}^\top \mathbf{x}_i| < \theta) = \sum_i f(y_i - \mathbf{a}^\top \mathbf{x}_i).$$

   Define

   $$F(\mathbf{a}) = \sum_i f(y_i - \mathbf{a}^\top \mathbf{x}_i).$$

   (b) Find a new function $G(\mathbf{a}) = 1 - F(\mathbf{a})$ and write the resulting function g such that

   $$G(\mathbf{a}) = \sum_i g(y_i - \mathbf{a}^\top \mathbf{x}_i).$$

   (c) What is the relationship between g and the ρ function of an M-estimator?

   (d) Based on this, summarize the primary differences between M-estimators and RANSAC.

2. Outline a RANSAC procedure of estimating a planar homography given a set of correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, some of which may be incorrect (outliers). Include a summary of the sampling procedure, the distance measure, and the revision to the estimate after the best transformation has been estimated.

3. Derive the complete expression for the Sampson error for estimating the 2D homography. Use

   $$\mathbf{A}_i = \begin{pmatrix} \mathbf{0}^T & -\mathbf{x}_i^T & v'_i \mathbf{x}_i^T \\ \mathbf{x}_i^T & \mathbf{0}^T & -u'_i \mathbf{x}_i^T \end{pmatrix},$$

   and $\mathbf{x}_i = (u_i, v_i, 1)^T$.

4. Derive the DLT constraint matrix for the estimation of the parameters of the camera matrix $\mathbf{P}$ given corresponding sets of points $\mathbf{X}_i \leftrightarrow \mathbf{x}_i$, where

   $$\mathbf{X}_i = \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

   is a known point in $\mathbb{R}^3$ (e.g. on a calibration object) and

   $$\mathbf{x}_i = \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix}$$

   is a corresponding image point. Derive the normalization equations.

Homework Problem

1. An estimator is "affine equivariant" if an affine transformation of the data produces an affine transformation to the resulting estimate. In the case of regression, the affine transformation would be applied to the $\mathbf{x}_i$ values, i.e.

$$\mathbf{x}_i' = \mathbf{B}\mathbf{x} + \mathbf{t}$$

and a separate linear scaling would be applied to the y_i values,

$$y_i' = by_i + c.$$

For each robust estimator of linear regression we have discussed, determine whether or not it is affine equivariant, and when the answer is "no", explain the cause of the problem.