Point sets, Maps and Navigation - II

D.A. Forsyth

Robustness is a serious problem

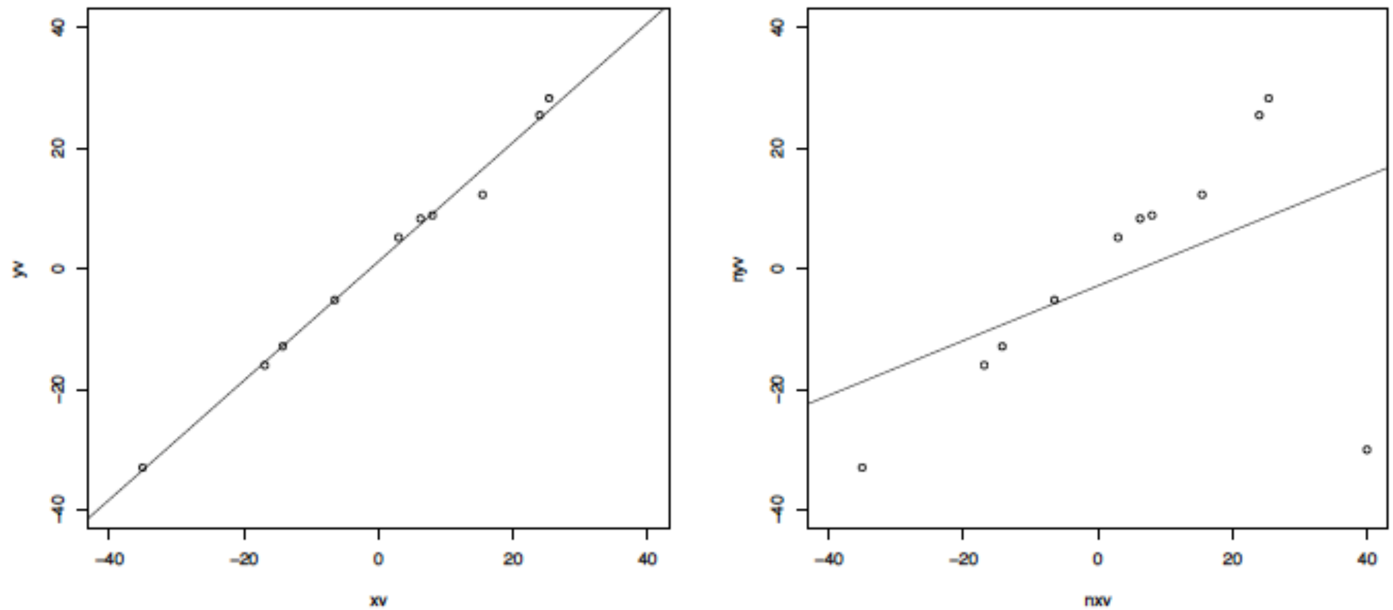


FIGURE 10.6: *On the* **left**, *a synthetic dataset with one independent and one explanatory variable, with the regression line plotted. Notice the line is close to the data points, and its predictions seem likely to be reliable. On the* **right**, *the result of adding a single outlying datapoint to that dataset. The regression line has changed significantly, because the regression line tries to minimize the sum of squared vertical distances between the data points and the line. Because the outlying datapoint is far from the line, the squared vertical distance to this point is enormous. The line has moved to reduce this distance, at the cost of making the other points further from the line.*

Robustness is a serious problem

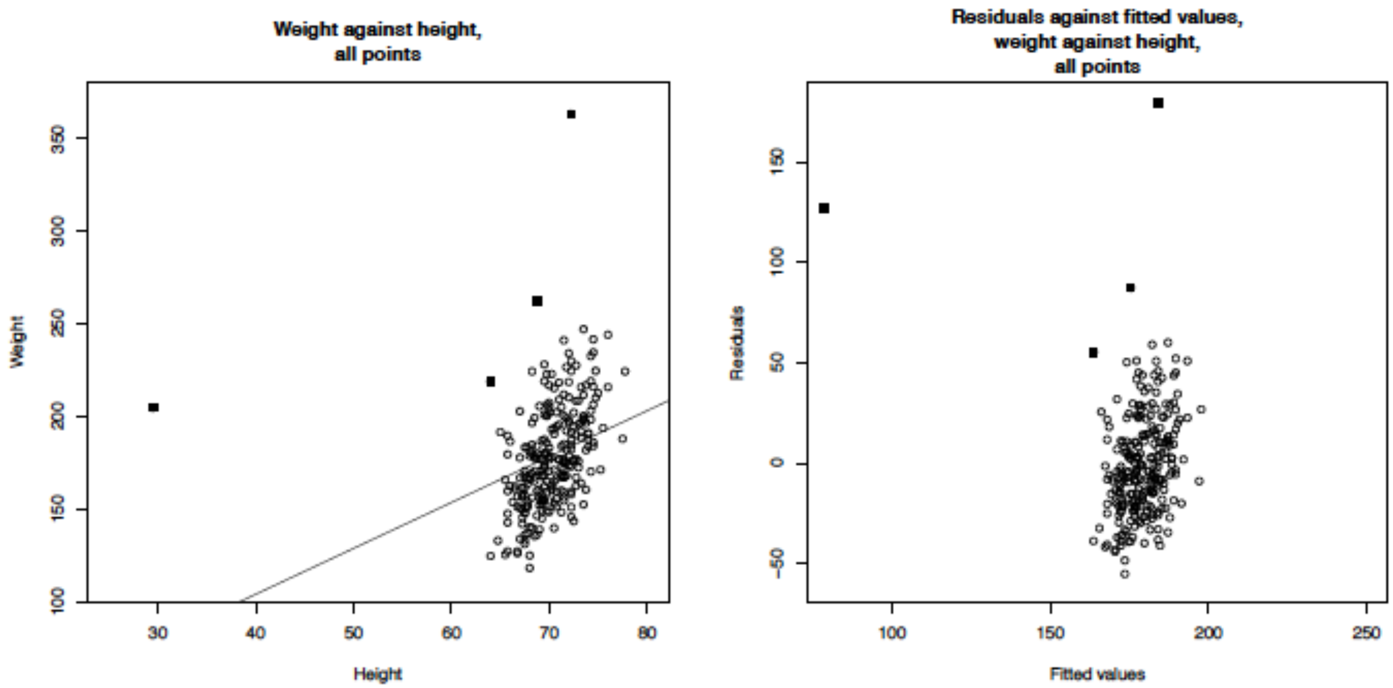


FIGURE 10.7: *On the* **left**, *weight regressed against height for the bodyfat dataset. The line doesn't describe the data particularly well, because it has been strongly affected by a few data points (filled-in markers). On the* **right**, *a scatter plot of the residual against the value predicted by the regression. This doesn't look like noise, which is a sign of trouble.*

Key issue:

- Squaring a large number produces a huge number
- A few wildly mismatch points can throw off R, t
- Fixes:
    - remove matches with “large” distances
        - actually, quite good
        - but what happens if new such pairs emerge?
    - apply an M-estimator
        - deals with new pairs

You should have watched the IRLS movie for regression by now

Robust estimation

- Iteratively reweighted least squares (IRLS)
  - Estimates weights in every iteration to down-weigh outliers

$$\min_x \|w(Ax - b)\|$$

- Non-linear estimation with robust loss function (M-estimators)
  - A robust loss function is used that down-weighs the influence of outliers

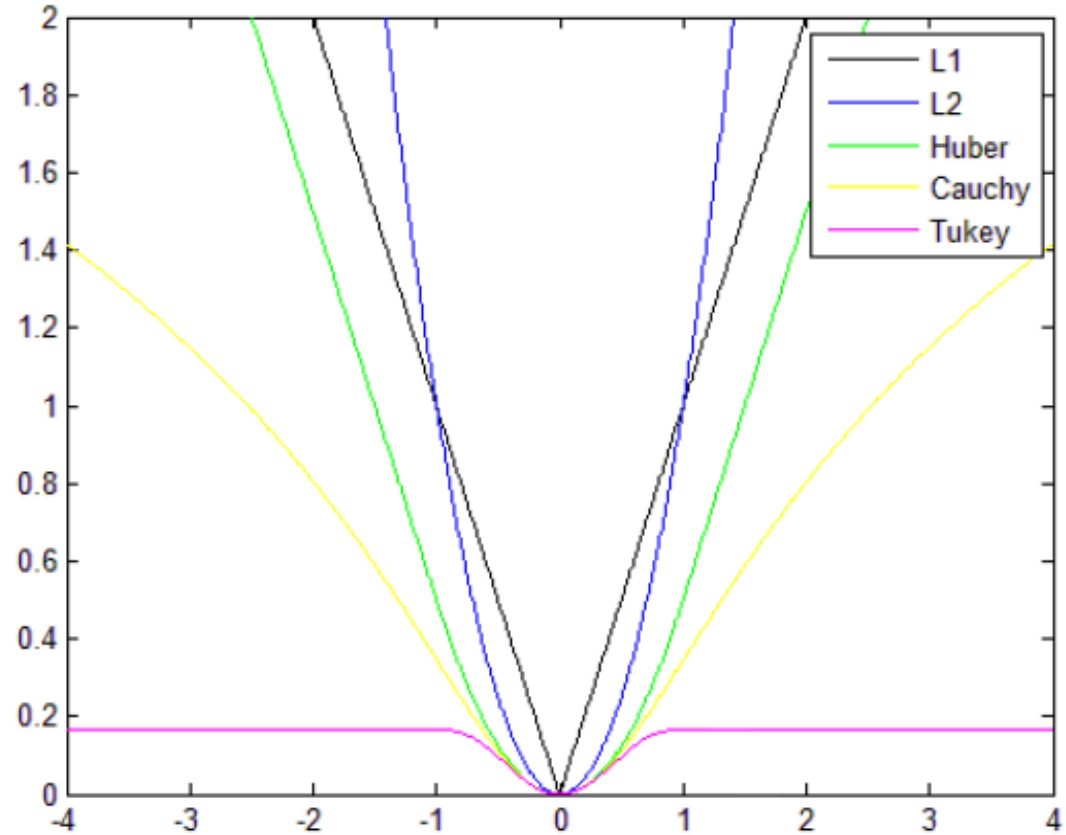
$$\min_x \sum_i \rho_i(\|f_i(x_i)\|^2)$$

$$\min_x \|A(x) - b\|$$

Non-linear estimation with robust loss function

$$\min_{x} \sum_{i} \rho_i(\|f_i(x_i)\|^2)$$

- Non-linear optimization (e.g. Levenberg-Marquard)
- Iteration necessary
- No explicit weight computation necessary
- Loss function should be differentiable
- Jacobian needs to be calculated



Weighted Least-Squares

- Not all residuals are equally important. Weights define the importance.
- Weights are collected in diagonal weight matrix

$$\min_x \|w(Ax-b)\|$$
$$x = (A^TWA)^{-1}A^TWb$$

- Still one-step closed form solution.
- But, how to find the weights?

- Online estimation of the weights from robust cost function -> IRLS

Iteratively reweighted least squares (IRLS)

1. Write the problem to solve as a weighted least squares optimization

$$C(x, w) = \sum_i w_i f_i(x)$$

cost function

2. Solve iteratively: At each step define weights

$$w_i^t = w_i(x^t)$$

and define update step

$$x^{t+1} = \operatorname*{argmin}_x C(x, w^t)$$

$$= \operatorname*{argmin}_x \sum_i w_i^t f_i(x)$$

$$x = (A^T W A)^{-1} A^T W b$$

3. Hope that it converges to what you want

How to choose the weights

- Weighted least squares minimizes the following cost

$$C(x, w) = \sum_i w_i f_i(x)$$

- We wish to minimize the cost with robustifier ρ

$$C_\rho(x) = \sum_i \rho(f_i(x))$$

- Minima of both cost functions need to be the same

$$\nabla C(x, w) = 0 \; if \; and \; only \; if \, \nabla C_\rho(x) = 0$$

$$\nabla w_i f_i(x) = \nabla \rho(f_i(x))$$
$$w_i \nabla f_i(x) = \rho'(f_i(x)) \nabla f_i(x)$$
$$w_i = \rho'(f_i(x))$$

required weights

Example L_1

- L_1 norm means absolute distance

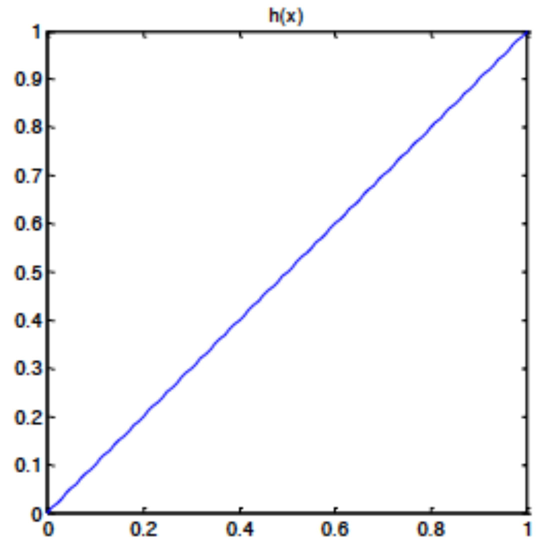
$$f_i(x) = d(x, y_i)^2$$
$$\rho(s) = \sqrt{s}$$

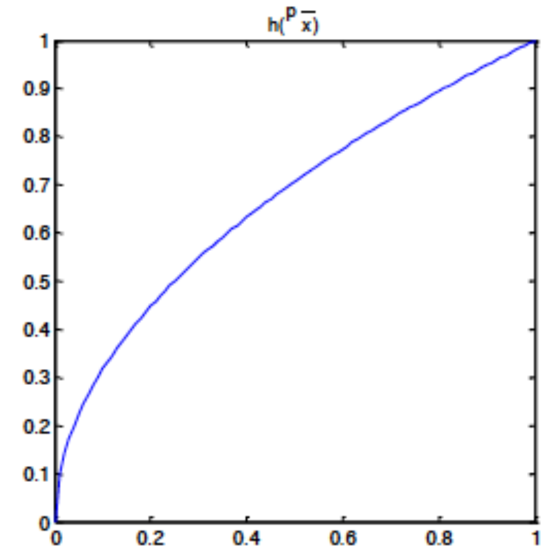
$$C_\rho(x) = \sum_i \rho(f_i(x)) = \sum_i |d(x, y_i)|$$ sum of absolute distances

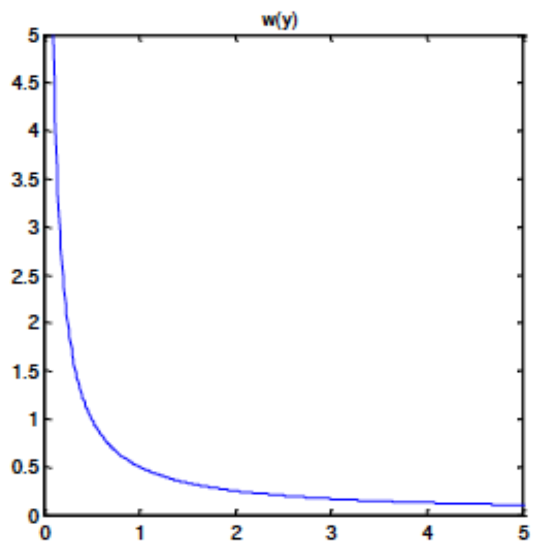
- Weight computation:

$$\begin{aligned} w_i &= \rho'\big(f_i(x)\big) \\ &= \tfrac{1}{2} f(x)^{-1/2} \\ &= \frac{1}{2} d(x, y_i)^{-1} \end{aligned}$$

L_1



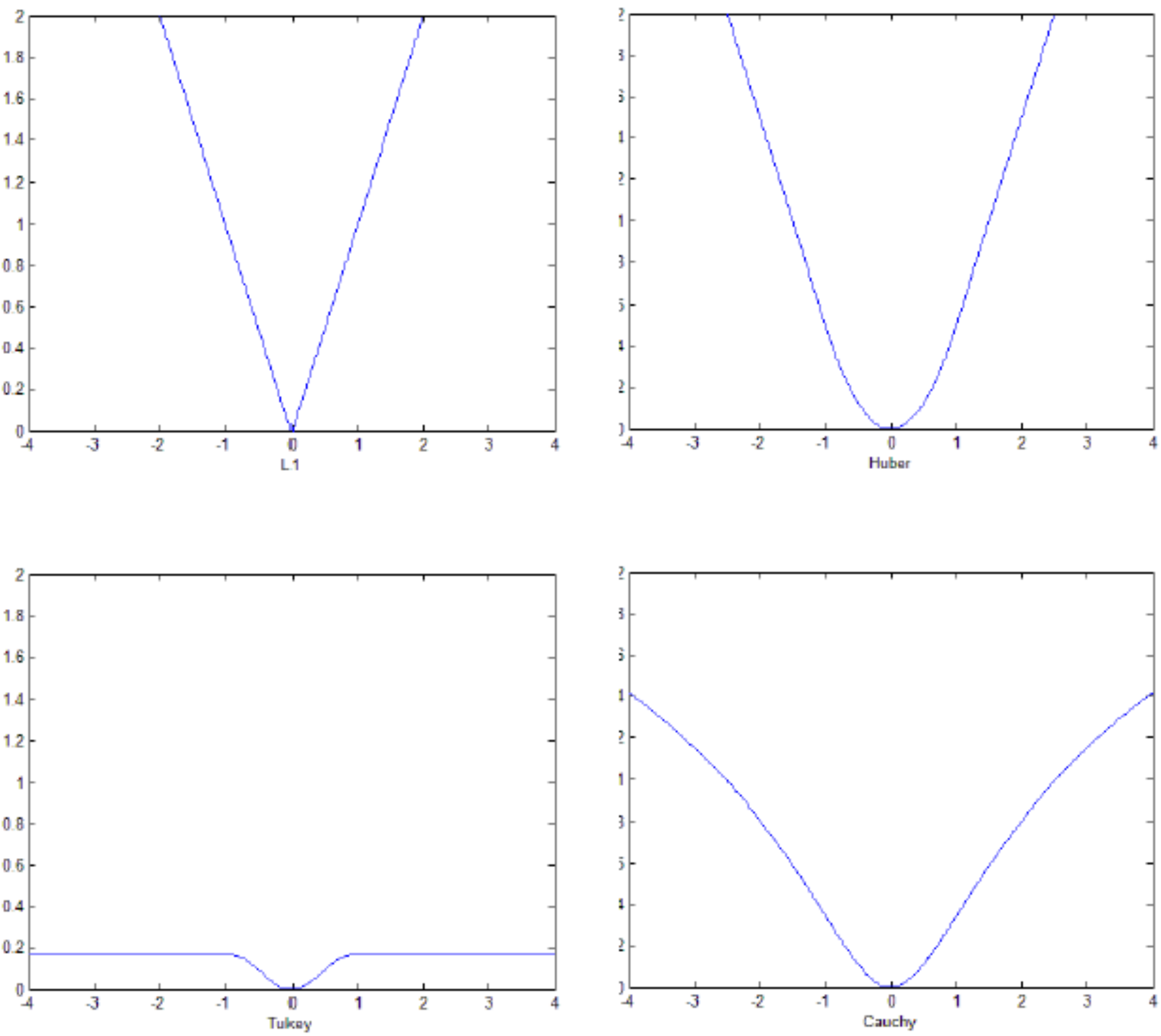




Advantage: Robust

Disadvantage:
- Weights not defined at 0
- Can stop at non-minimum

More robust loss functions



loss function	p(x)
L_1	$\lvert x\rvert$
$Huber \begin{cases} if\ \lvert x\rvert \le k \\ if\ \lvert x\rvert \ge k \end{cases}$	$\begin{cases} x^2/2 \\ k(\lvert x\rvert - \frac{k}{2}) \end{cases}$
$Tukey \begin{cases} if\ \lvert x\rvert \le k \\ if\ \lvert x\rvert > k \end{cases}$	$\begin{cases} k^2/6(1-\left(1-\left(\frac{x}{c}\right)\right. \\ k^2/6 \end{cases}$
$Cauchy$	$\frac{k^2}{2}\log(1+(x/k)^2)$