Robust line fitting with IRLS

D.A. Forsyth,

University of Illinois at Urbana Champaign

Robust estimators

- General approach: find model parameters θ that minimize

$$\sum_i \rho_\sigma(r(x_i;\theta))$$

$r(x_i;\theta)$: residual of x_i w.r.t. model parameters θ

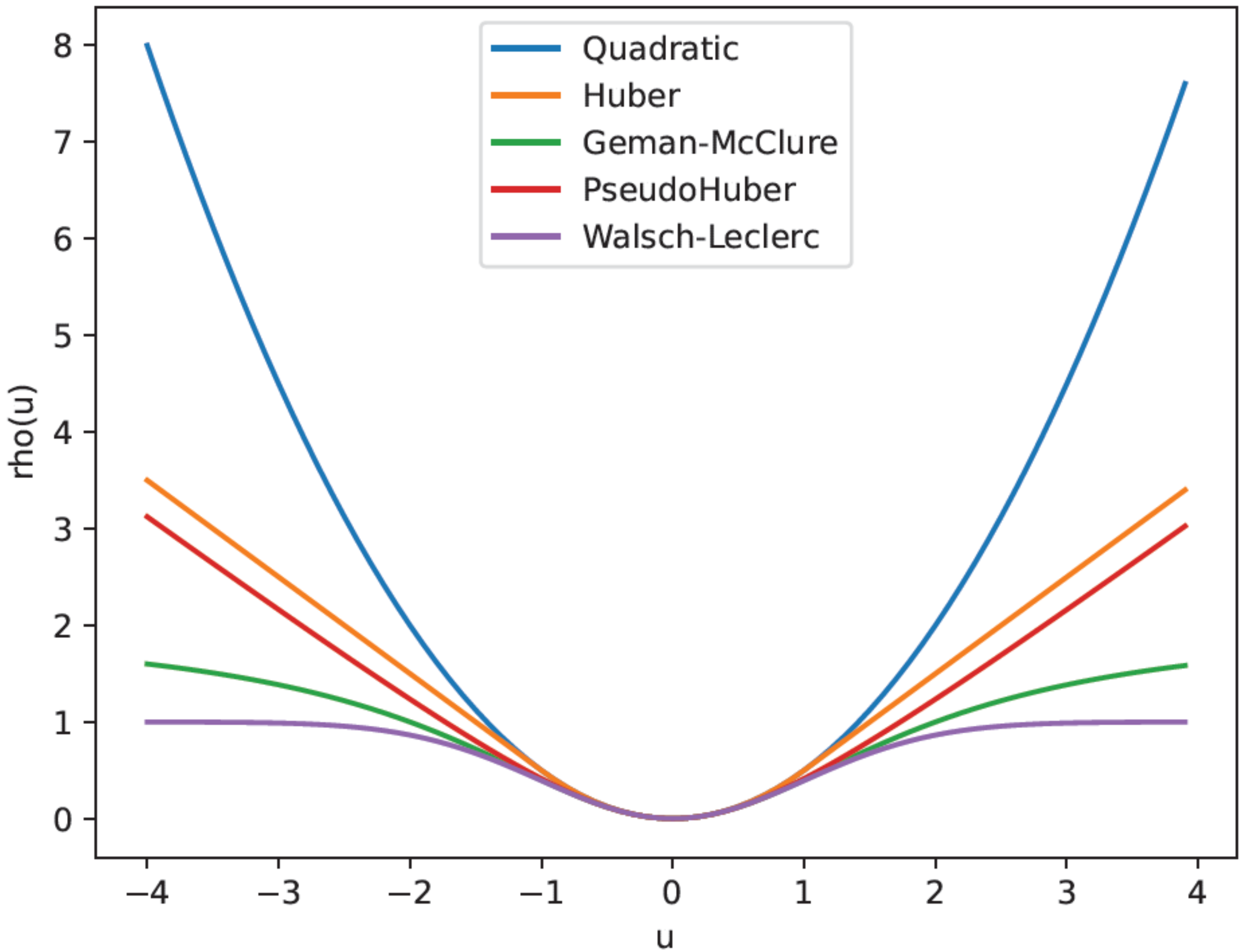
eg for line, $\theta = (a, b, d)$

residual $r(x_i;\theta) = (ax_i + by_i - d)$

ρ_σ: robust function with scale parameter σ

Notice that $\rho_\sigma(u) = u$^2 would give the original least squares loss

Robust estimators



The Huber loss

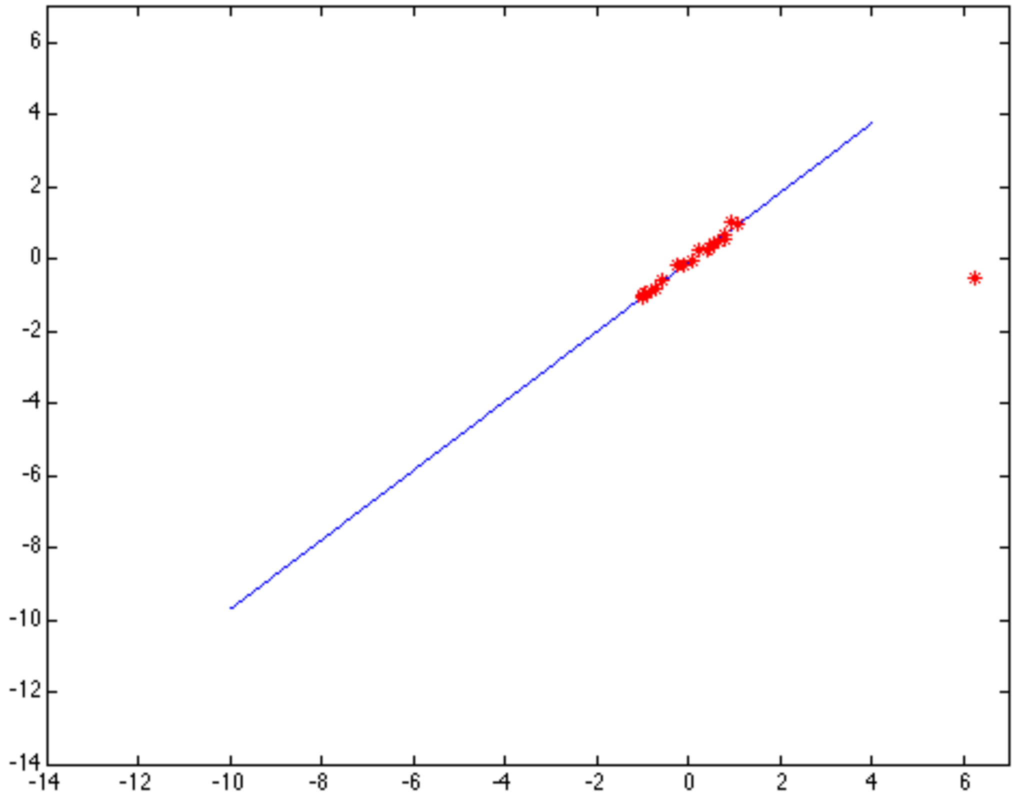
The *Huber loss* uses

Scale

$$\rho(u;\sigma) = \begin{cases} \frac{u^2}{2} & |u| < \sigma \\ \sigma|u| - \frac{\sigma^2}{2} & \end{cases}$$

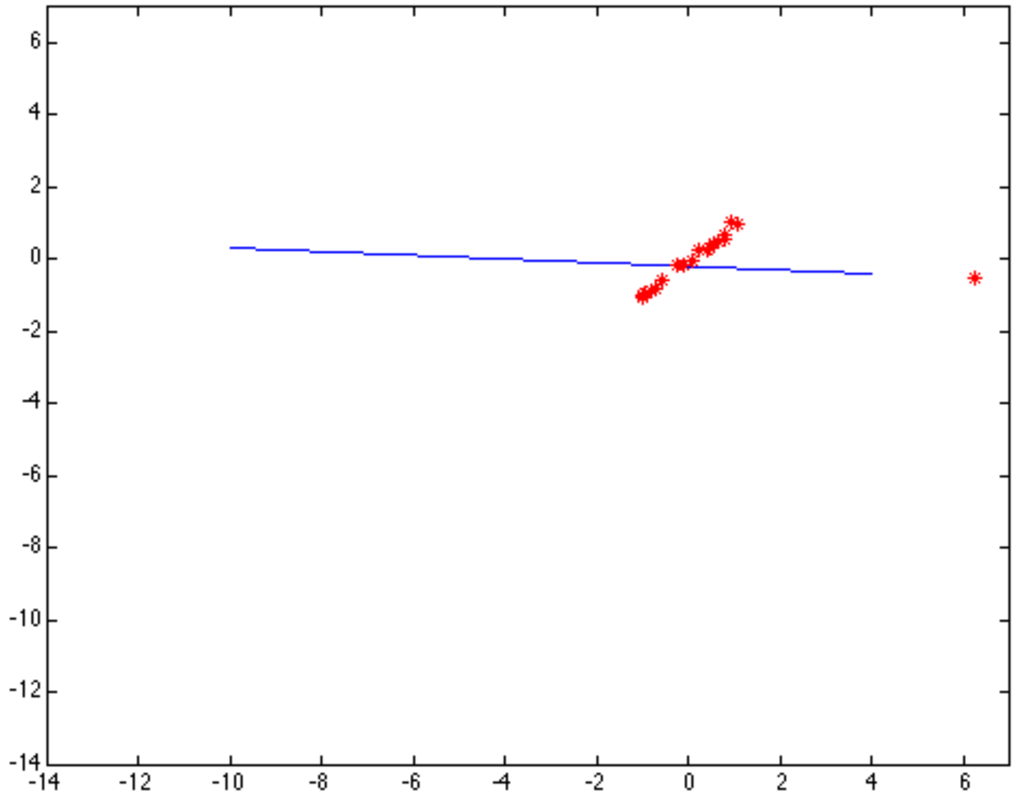
which is the same as $u^2/2$ for $-\sigma \leq u \leq \sigma$, switches to $|u|$ for larger (or smaller) σ, and has continuous derivative at the switch. The Huber loss is convex (meaning that there will be a unique minimum for our models) and differentiable, but is not smooth. The choice of the parameter σ (which is known as *scale*) has an effect on the estimate. You should interpret this parameter as the distance that a point can

Choosing the scale: Just right



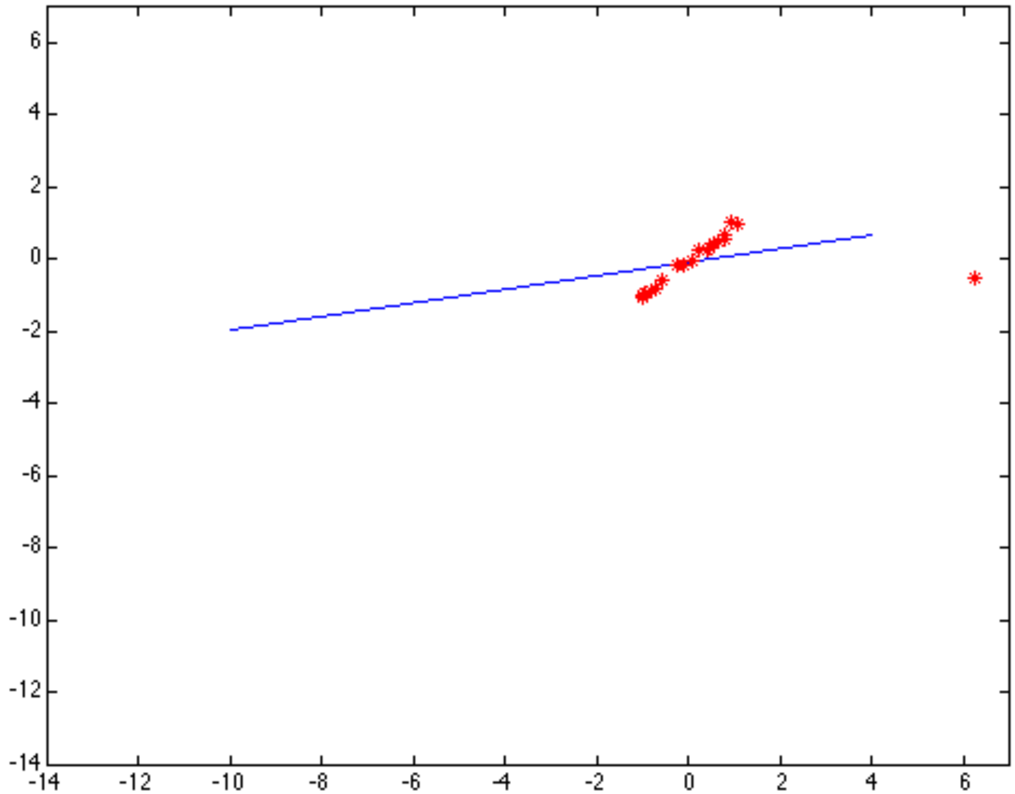
The effect of the outlier is minimized

Choosing the scale: Too small



The error value is almost the same for every point and the fit is very poor

Choosing the scale: Too large



Behaves much the same as least squares

Finding the line

Now the line is chosen by minimizing

$$(1/2)\sum_i \rho(r(\mathbf{x},\theta);\sigma)$$

with respect to $\theta = (a_1, a_2, c)$, subject to $a_1^2 + a_2^2 = 1$. The minimum occurs when

$$\begin{aligned}
\nabla_\theta \left(\sum_i \rho(r(\mathbf{x}_i,\theta);\sigma) \right) &= \sum_i \left[\frac{\partial \rho}{\partial u} \right] \nabla_\theta r(\mathbf{x}_i,\theta) \\
&= \lambda \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix}.
\end{aligned}$$

Finding the line - II

Here λ is a Lagrange multiplier and the derivative $\frac{\partial \rho}{\partial u}$ is evaluated at $r(\mathbf{x}_i, \theta)$, so it is a function of θ. Now notice that

$$\begin{aligned}\sum_i \left[\frac{\partial \rho}{\partial u}\right] \nabla_\theta r(\mathbf{x}_i, \theta) &= \sum_i \left[\left(\frac{\frac{\partial \rho}{\partial u}}{r(\mathbf{x}_i, \theta)}\right)\right] r(\mathbf{x}_i, \theta) \nabla_\theta r(\mathbf{x}_i, \theta) \\ &= \sum_i \left[\left(\frac{\frac{\partial \rho}{\partial u}}{r(\mathbf{x}_i, \theta)}\right)\right] \nabla_\theta \left[(1/2) r(\mathbf{x}_i, \theta)\right]^2\end{aligned}$$

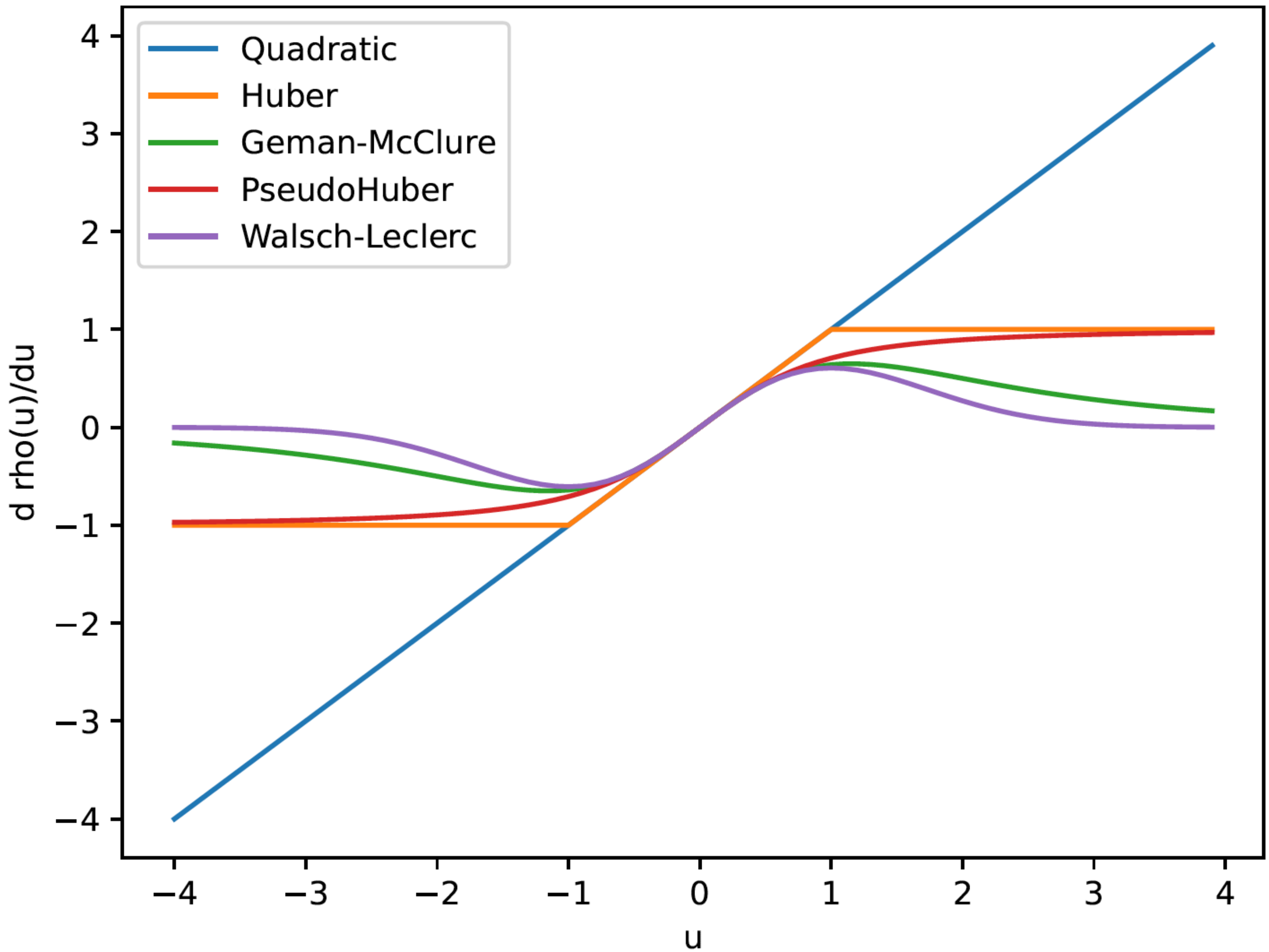
Now $[r(\mathbf{x}_i, \theta)]^2$ is the squared error. At the true minimum $\hat{\theta}$, writing

$$w_i = \left(\frac{\frac{\partial \rho}{\partial u}}{r(\mathbf{x}_i, \hat{\theta})}\right)$$

(where the derivatives are evaluated at that $\hat{\theta}$), then

$$\sum_i w_i \nabla_\theta \left[r(\mathbf{x}_i, \theta)\right]^2 = \lambda \begin{pmatrix} 2a_1 \\ 2a_2 \\ 0 \end{pmatrix}.$$

Influence functions



Idea – iteratively reweighted least squares

- Start with initial line
  - get weights, scale from line

- Iterate:
  - estimate line using weights, scale
  - estimate scale using line
  - estimate weights using scale, line

- We *know* that one stationary point is the true minimum
- No other guarantees I'm aware of, but quite well behaved

Starting IRLS

- Iterate:
  - Initial line:
    - Draw two points at random from dataset
    - Pass line through them
  - Fit line with IRLS
- Use the best you encounter
-
-

IRLS

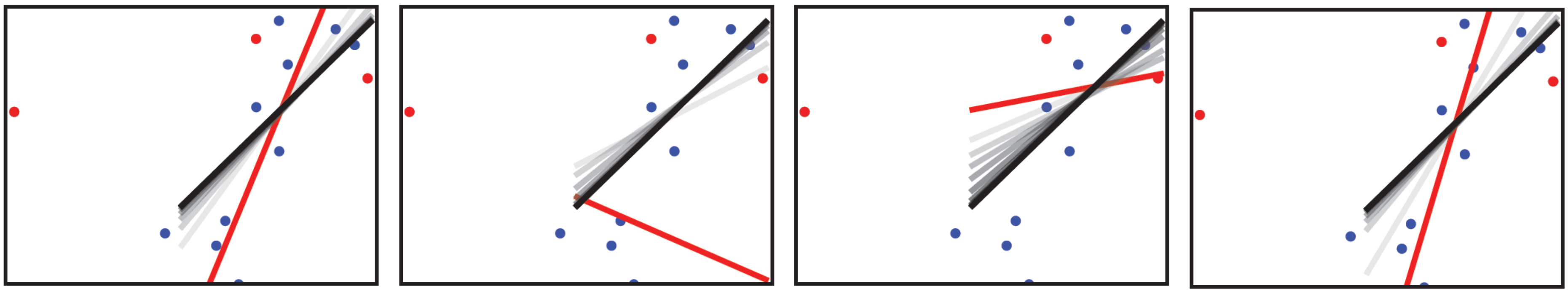


FIGURE 12.7: *Robust losses can control the influence of outliers.* **Blue** *points lie on a line, and have been perturbed by noise;* **red** *points are outliers. The* **red** *line shows a starting line, obtained by drawing a small random sample from the dataset, then fitting a line; the* **gray** *lines show iterates of IRLS applied to a Huber loss (later iterates are more opaque; scales are estimated as in the text). The procedure converges from a range of start points, some quite far from the "true" line. Notice how each start point results in the same line.*

IRLS isn't perfect...

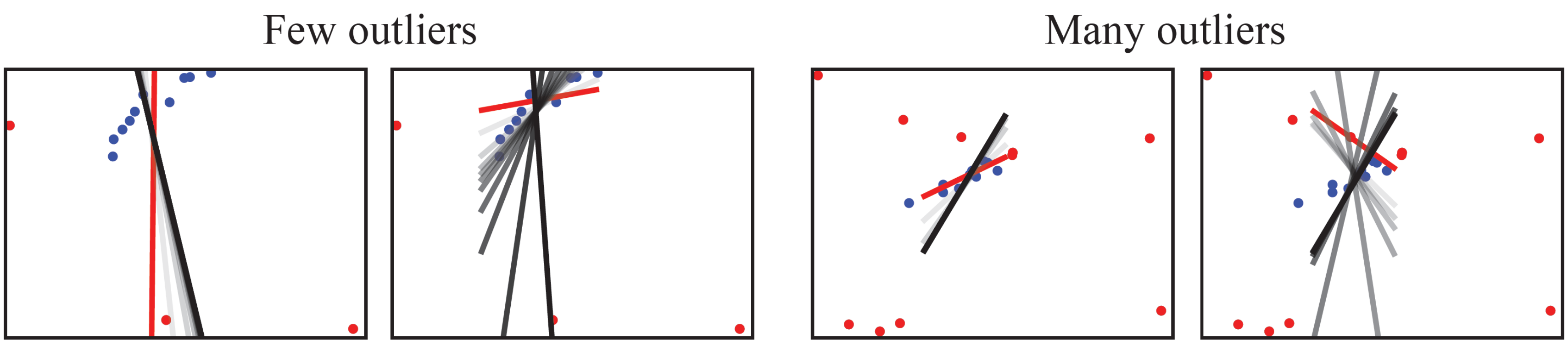


FIGURE 12.8: *Robust losses can fail, particularly when distant points still have some weight or if there are many outliers.* **Left***: a bad start point leads to a bad line;* **center left***: on the same data set, quite a good start point still converges to a bad line. Here there are few outliers, but they are far from the data and they contribute a significant weight to the loss. When there are many outliers, this effect worsens. Because each outlier still contributes a signficant weight to the loss, even a good start fails (***center right***). A poor start (***right***) also fails, and produces the same line as the good start – in fact, most starts end up close to this line. Again,* **blue** *points lie on a line, and have been perturbed by noise;* **red** *points are outliers; the* **red** *line shows a starting line, obtained by drawing a small random sample from the dataset, then fitting a line; the* **gray** *lines show iterates of IRLS (later iterates are more opaque).*

Things to think about…

14.3. Why is it fairly obvious that there should be local minima for a line fit using a robust loss?