Lecture 3

Computing Sparse (and Robust?) Solutions

M D Sacchi
University of Alberta

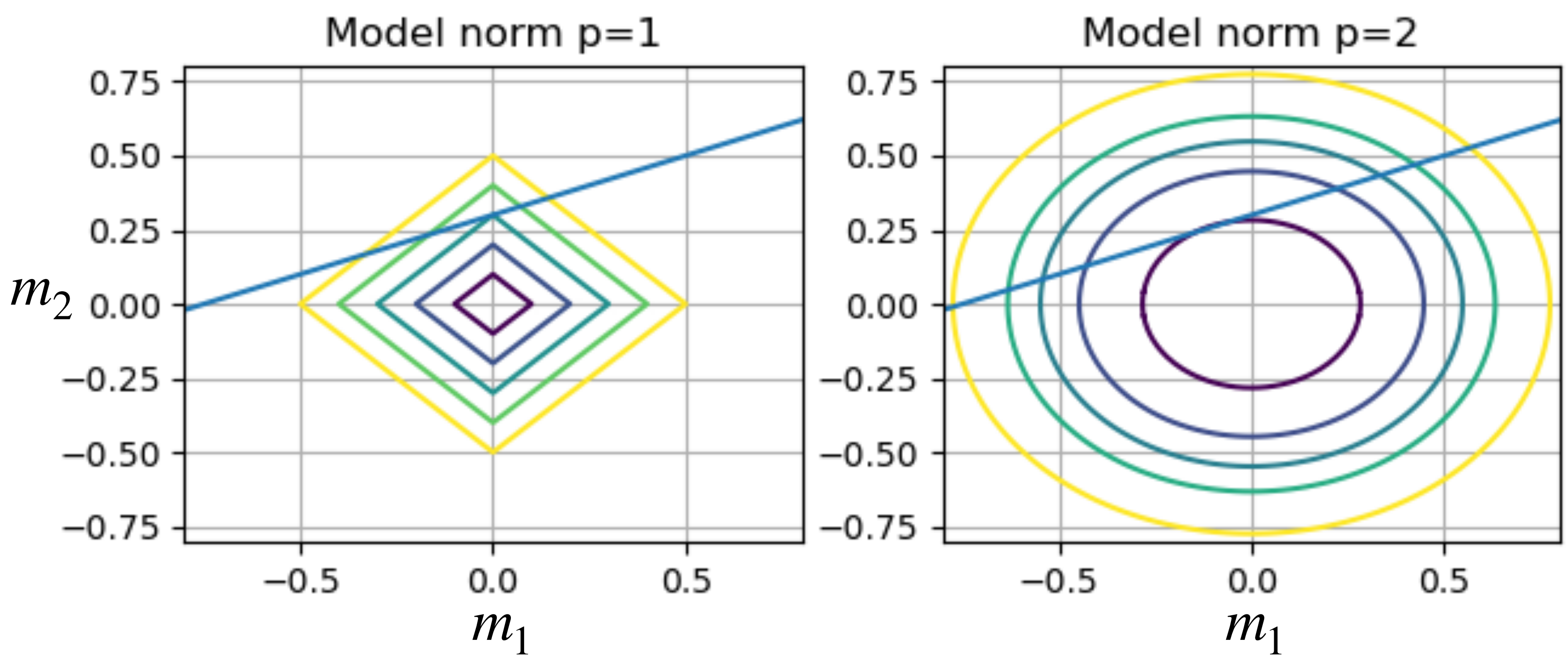
Course information

- Email: msacchi@ualberta.ca

- https://github.com/msacchi/SEP_lectures

Changing the norm to obtain sparse solutions (lp-norm)

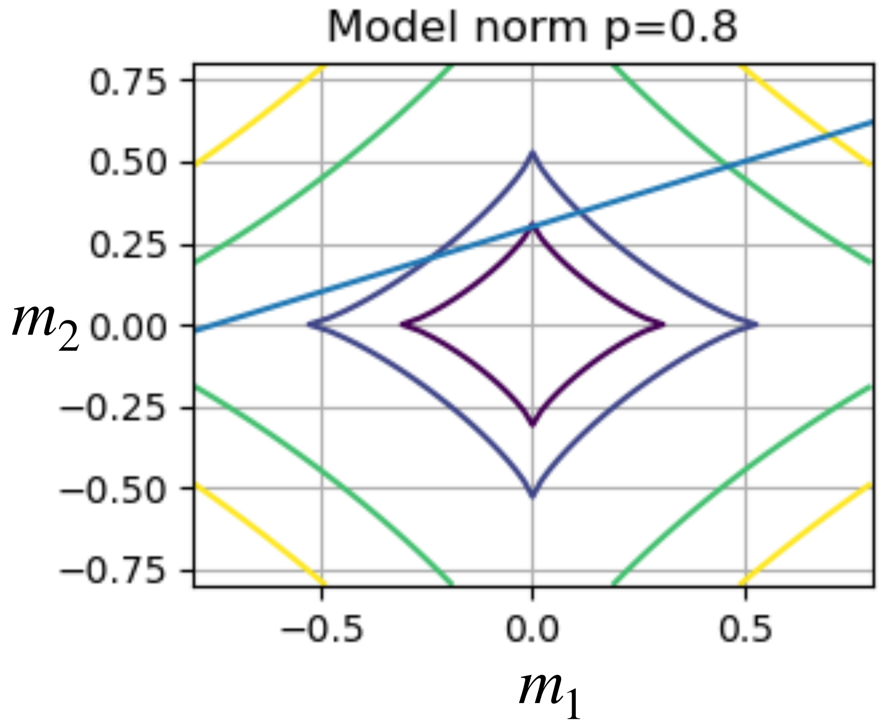
minimize $\|\mathbf{m}\|_p^p = |m_1|^p + |m_2|^p$ **suject to** $a_1 m_1 + a_2 m_2 = 1$



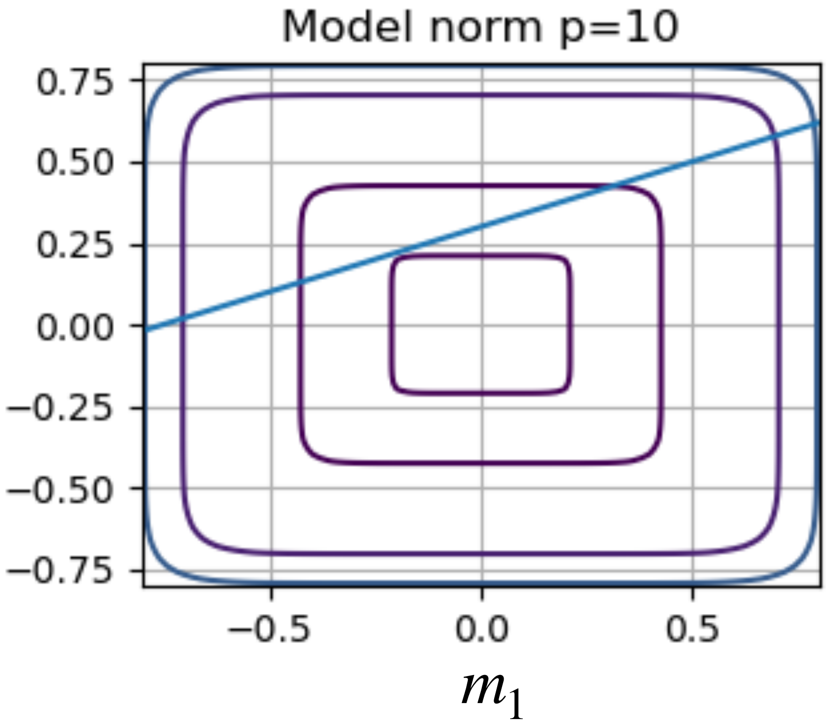
Sparse solution **non-sparse solution**

Changing the norm to obtain sparse solutions (lp-norm)

minimize $\|\mathbf{m}\|_p^p = |m_1|^p + |m_2|^p$ **suject to** $a_1 m_1 + a_2 m_2 = 1$



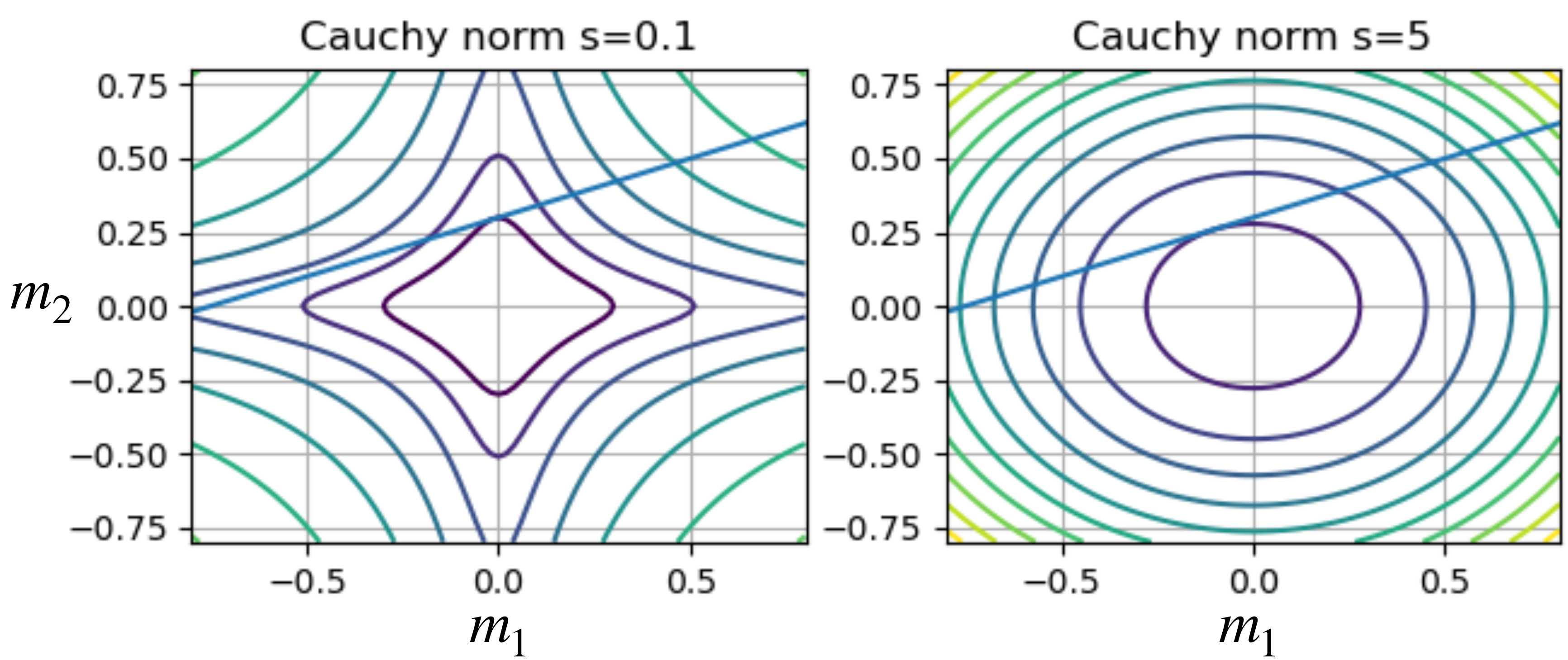
Sparse solution



non-sparse solution

Changing the norm to obtain sparse solutions: Cauchy Criterion (not really a norm)

minimize $\mathscr{R}_c(\mathbf{m}) = \log(m_1^2 + s^2) + \log(m_2^2 + s^2)$ **suject to** $a_1 m_1 + a_2 m_2 = 1$



Sparse solution **non-sparse solution**

Lp-norm, Cauchy and Hyperbolic regulatization

- We can produce sparse solution by using regularization terms of the form

$$\|\mathbf{m}\|_p^p = \sum_i |m_i|^p = \sum_i f(m_i) \rightarrow f(u) = |u|^p$$

$$\mathcal{R}_c(\mathbf{m}) = \sum_i \log(m_i^2 + s^2) = \sum_i f(m_i) \rightarrow f(u) = \log(u^2 + s^2)$$

$$\mathcal{R}_h(\mathbf{m}) = \sum_i \sqrt{m_i^2 + s^2} = \sum_i f(m_i) \rightarrow f(u) = \sqrt{u^2 + s^2}$$

- For Lp-norm, p close to 1 produce sparse solutions. The other two criteria resemble the L2-norm when the scale parameter s is large.

- Other functions/norms: Huber, Geman-Geman,…….

Linear inverse problem with sparsity constraint

- We use sparsity to regularize the problem and produce sparse solutions. We find the solution $\mathbf{m}$ that minimizes

$$J = \frac{1}{2}\|\mathbf{L}\mathbf{m} - \mathbf{d}\|_2^2 + \mu\|\mathbf{m}\|_1$$

- There are many ways of minimizing the regularized cost function above: IRLS, ISTA are two methods that can be easily derived.

IRLS

- Iterative re-weighted least-squares method

- First proposed for robust regression problems but then adopted to find sparse solutions

$$\frac{\partial J}{\partial \mathbf{m}} = \mathbf{L}^T\mathbf{L}\mathbf{m} - \mathbf{L}^T\mathbf{d} + \mu\frac{\partial \|\mathbf{m}\|_1}{\partial \mathbf{m}} = 0$$

$$\frac{\partial J}{\partial \mathbf{m}} = \mathbf{L}^T\mathbf{L}\mathbf{m} - \mathbf{L}^T\mathbf{d} + \mu\mathbf{v} = 0, \textbf{ with } v_i = sign(m_i) \approx \frac{m_i}{|m_i| + \epsilon}$$

IRLS

- Iterative re-weighted least-squares method

$$\frac{\partial J}{\partial \mathbf{m}} = \mathbf{L}^T\mathbf{L}\mathbf{m} - \mathbf{L}^T\mathbf{d} + \mu\mathbf{v} = 0, \textbf{ with } v_i = sign(m_i) \approx \frac{m_i}{|m_i| + \epsilon}$$

$$\mathbf{v} = \underbrace{\mathbf{Q}}_{\text{Diagonal Matrix}}\mathbf{m} \textbf{ with } Q_{ii} = \frac{1}{|m_i| + \epsilon}$$

- Condition for minimum

$$\frac{\partial J}{\partial \mathbf{m}} = \mathbf{L}^T\mathbf{L}\mathbf{m} - \mathbf{L}^T\mathbf{d} + \mu\mathbf{Q}\mathbf{m} = 0$$

IRLS

- The algorithm (explicit-form solution)

$$\mathbf{Q} = \mathbf{I}$$

for $\nu = 1$ **to** $MaxIter$

$$\mathbf{m} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{Q})^{-1}\mathbf{L}^T\mathbf{d}$$

$$\mathbf{Q} = \textbf{diag}(\frac{1}{|\mathbf{m}| + \epsilon)})$$

Element-wise operation

end

IRLS

- Condition for minimum

$$\frac{\partial J}{\partial \mathbf{m}} = \mathbf{L}^T\mathbf{L}\mathbf{m} - \mathbf{L}^T\mathbf{d} + \mu \mathbf{Q}\mathbf{m} = 0$$

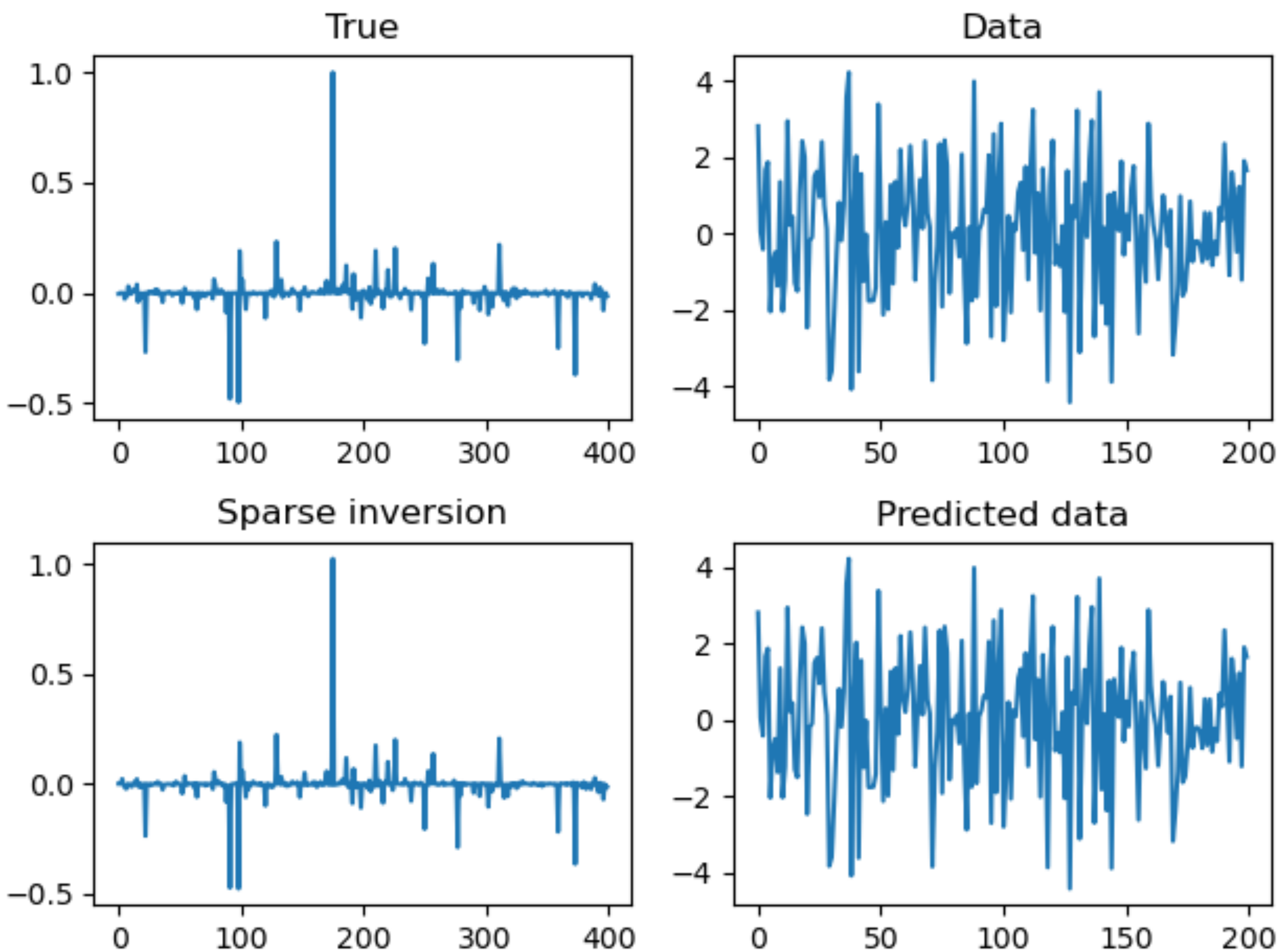
- But, the diagonal matrix depends on the solution $\mathbf{m}$, so we don't have a closed-form solution. We have an iterative solution:

$$\mathbf{L}^T\mathbf{L}\mathbf{m} + \mu \mathbf{Q}\mathbf{m} = \mathbf{L}^T\mathbf{d} \;\; \rightarrow \;\; \mathbf{m}^\nu = (\mathbf{L}^T\mathbf{L} + \mu \mathbf{Q}^{\nu-1})^{-1}\mathbf{L}^T\mathbf{d}$$

$$Q_{ii}^{\nu-1} = \frac{1}{|\, m_i^{\nu-1} \,| + \epsilon}$$

`IRLS_tests.ipynb`

$\mathbf{m} = sign(\mathbf{r}) \, . \, \mathbf{r}^{\beta} \quad \mathbf{r} \sim \mathcal{N}(0,\sigma) \qquad \beta = 4$



$$\mathbf{d} = \mathbf{L}\mathbf{m}$$

$$\mathbf{d}^{Pred} = \mathbf{L}\hat{\mathbf{m}}$$

$\hat{\mathbf{m}} = argmin_{\mathbf{m}} \; [\|\mathbf{L}\mathbf{m} - \mathbf{d}\|_2^2 + \mu \|\mathbf{m}\|_1]$

IRLS via preconditioning

- The solution $\mathbf{m}^{\nu} = (\mathbf{L}^T\mathbf{L} + \mathbf{Q}^{\nu-1})^{-1}\mathbf{L}^T\mathbf{d}$ can be interpreted as the minimum of the cost

$$J = \|\mathbf{Lm} - \mathbf{d}\|_2^2 + \mu\|\mathbf{Q}^{1/2}\mathbf{m}\|_2^2$$

- Where I considered $\mathbf{Q}$ independent of $\mathbf{m}$. The cost function can be written as follows:

$$J = \|\mathbf{LPu} - \mathbf{d}\|_2^2 + \mu\|\mathbf{u}\|_2^2, \qquad \mathbf{m} = \mathbf{Pu} \qquad \mathbf{u} = \mathbf{Q}^{1/2}\mathbf{m}$$

$$P_{ii} = \sqrt{|m_i| + \epsilon}$$

IRLS

- The algorithm (implicit-form solution)

$$\mathbf{P} = \mathbf{I}$$

$$\textbf{for } \nu = 1 \textbf{ to } MaxIter$$

$$\mathbf{u} = \textbf{argmin}(\|\mathbf{L}\mathbf{P}\,\mathbf{u} - \mathbf{d}\|_2^2 + \mu\|\mathbf{u}\|_2^2)$$

$$\mathbf{m} = \mathbf{P}\mathbf{u}$$

$$\mathbf{P} = \textbf{diag}(\sqrt{|\,\mathbf{m}\,| + \epsilon}) \longleftarrow \text{Element-wise operation}$$

$$\textbf{end}$$

IRLS

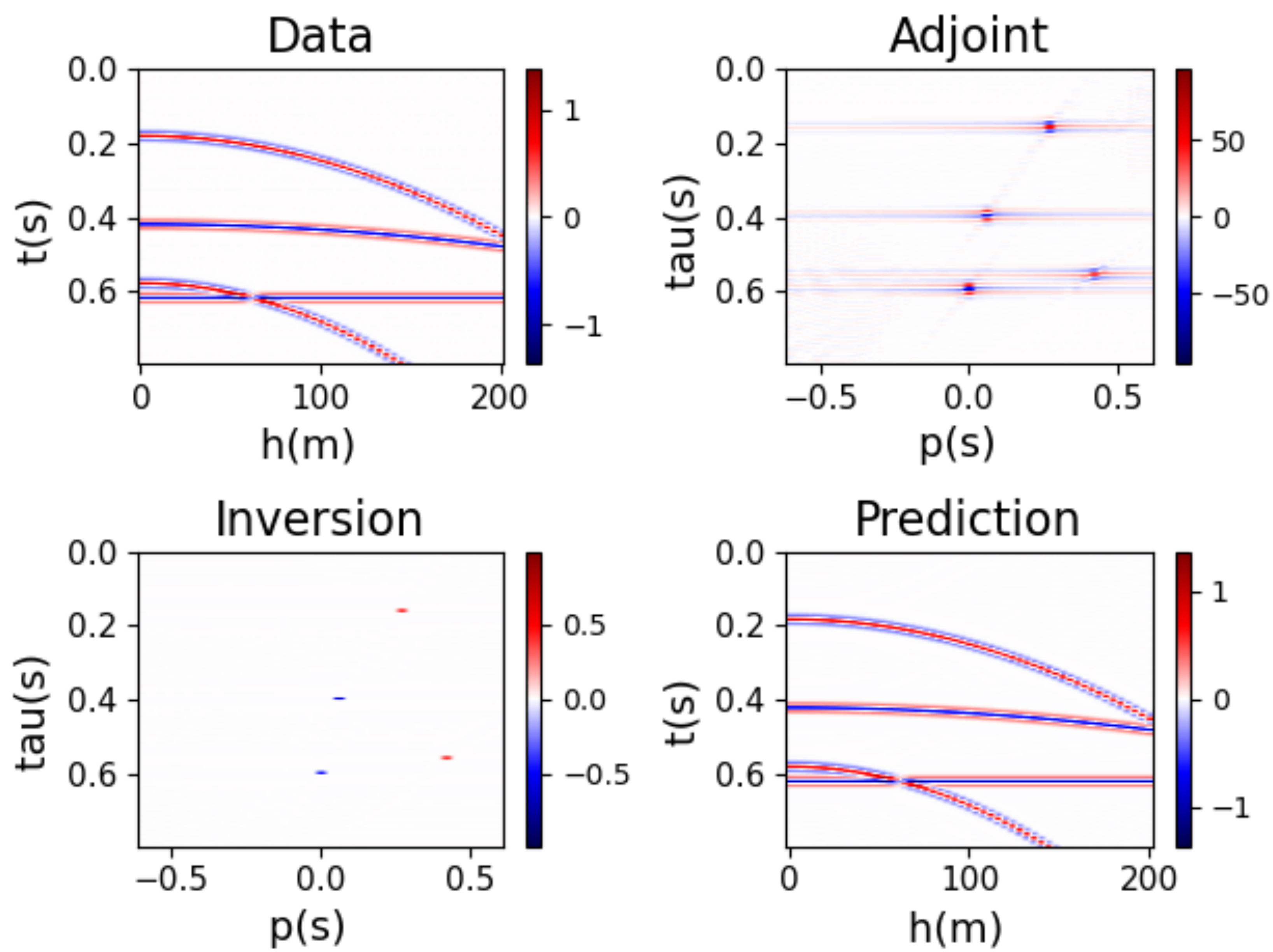
- The algorithm (implicit-form solution where you only need forward and adjoint operators:

$$
\begin{aligned}
&\mathbf{P} = \mathbf{I} \\
&\quad \textbf{for } \nu = 1 \textbf{ to } MaxIter \\
&\qquad \mathbf{u} = \textbf{CGLS}(\mathbf{d}, [\mathbf{L}, \mathbf{P}], \mu) \qquad \longleftarrow \text{ CGLS with lenient stopping criteria} \\
&\qquad \mathbf{m} = \mathbf{P}\mathbf{u} \\
&\qquad \mathbf{P} = \textbf{diag}(\sqrt{|\mathbf{m}| + \epsilon}) \\
&\quad \textbf{end}
\end{aligned}
$$

- We have two iteration loops. For practical problems, the main idea is to reduce as much as possible the internal iteration and do a small number of external updates. The latter requires some experimentation (example: high-res Radon transforms)

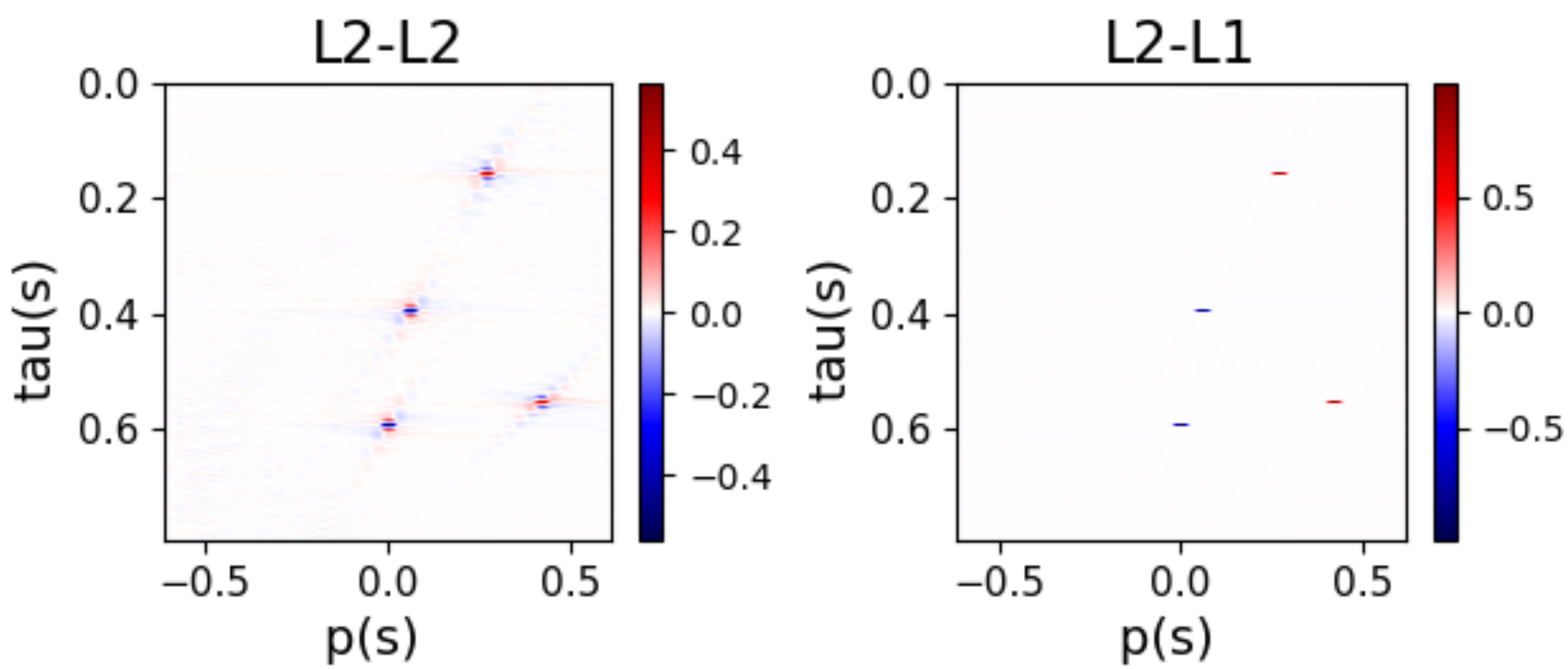
IRLS with implicit Radon operator

`IRLS_tests.ipynb`



Comparison of sparse vs non-sparse solution

CGLS and IRLS solutions



$\mathbf{P} = \mathbf{I}$

for $\nu = 1$ **to** $MaxIter$

$\mathbf{u} = \textbf{argmin}(\|\mathbf{LP} - \mathbf{d}\|_2^2 + \mu\|\mathbf{u}\|_2^2)$

$\mathbf{m} = \mathbf{Pu}$

$\mathbf{P} = \textbf{diag}(\sqrt{|\mathbf{m}| + \epsilon})$

end

L2-L2 (non-sparse) corresponds to $MaxIter$=1
L2-L1 (sparse) corresponds to $MaxIter$=4

`IRLS_tests.ipynb`

Robust Solutions

$$J = \frac{1}{2}\|\mathbf{Lm} - \mathbf{d}\|_1 + \mu\|\mathbf{m}\|_1$$

$\mathbf{P} = \mathbf{I}$

for $\nu = 1$ **to** $MaxIter$

$\quad \mathbf{u} = \textbf{argmin}\|\, \mathbf{Q}[\mathbf{LP}\,\mathbf{u} - \mathbf{d}\,]\,\|_2^2 + \mu\|\mathbf{u}\|_2^2$

$\quad \mathbf{m} = \mathbf{Pu}$

$\quad \mathbf{r} = \mathbf{d} - \mathbf{Lm}$

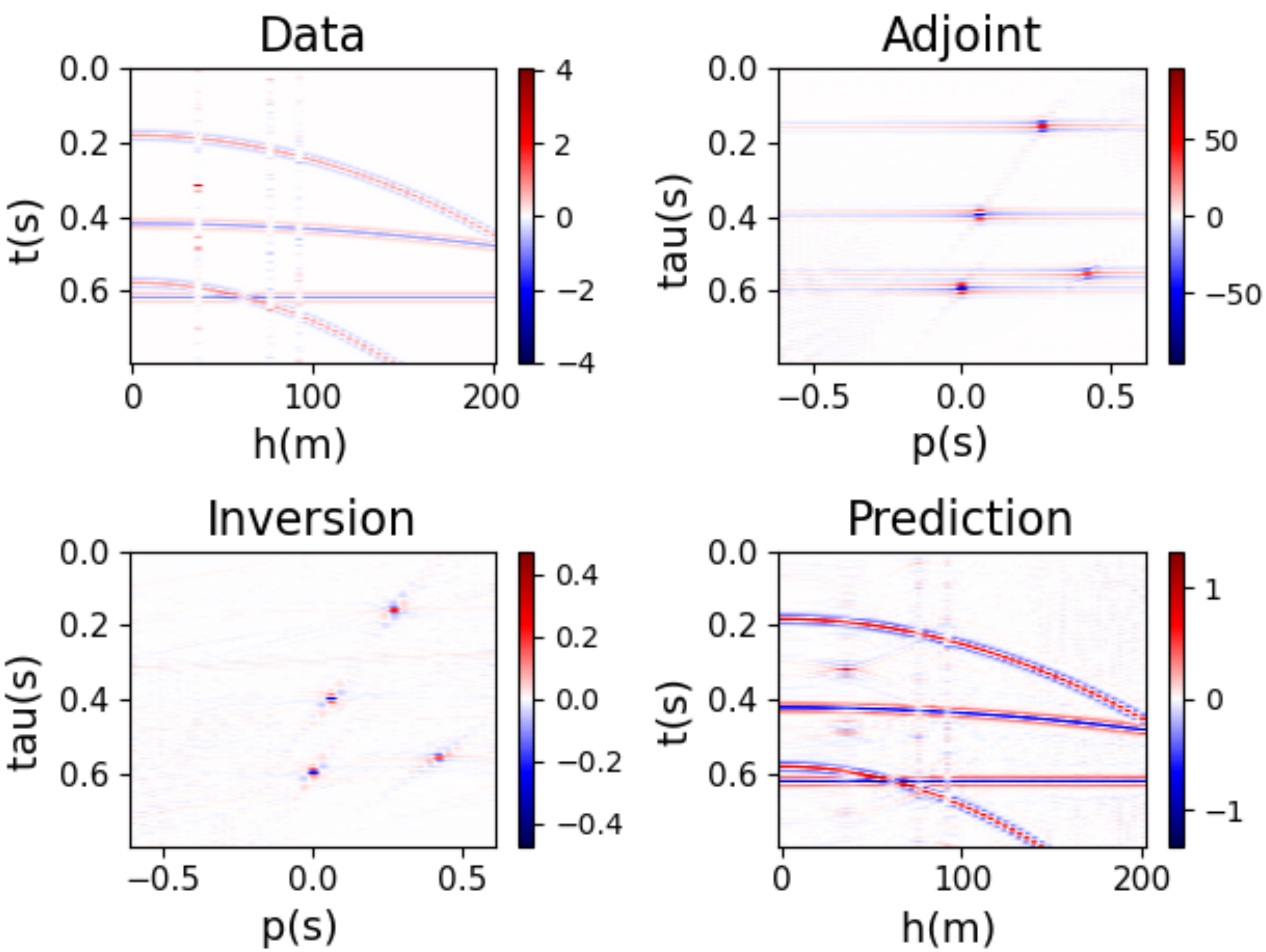
$\quad \mathbf{P} = \textbf{diag}(\sqrt{|\,\mathbf{m}\,| + \epsilon_1})$

$\quad \mathbf{Q} = \textbf{diag}(1/\sqrt{|\,\mathbf{r}\,| + \epsilon_2})$

end

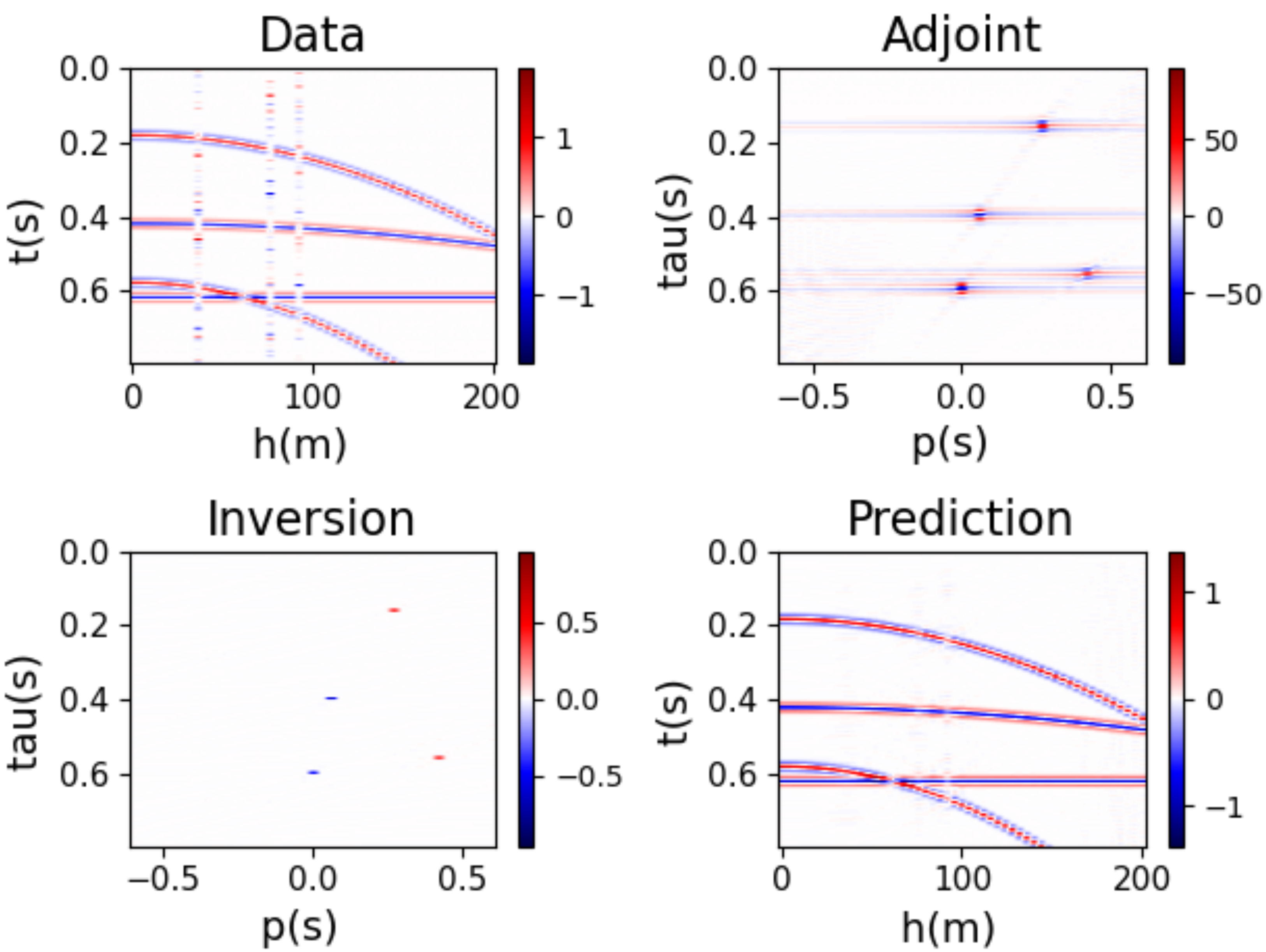
Non-robust inversion

$$J = \|\mathbf{Lm} - \mathbf{d}\|_2^2 + \mu\|\mathbf{m}\|_2^2$$



Robust and Sparse Inversion

$$J = \|\mathbf{Lm} - \mathbf{d}\|_1^1 + \mu\|\mathbf{m}\|_1$$



IRLS_tests.ipynb

```
Q=I
P=I
for k = 1:4

Par_P = Dict(:w=>P)
Par_Q = Dict(:w=>Q)
u,cost1 =  ConjugateGradients(Q.*d,[WeightingOp, SeisConv, SeisRadon_tx, WeightingOp],[Par_Q,Par_W,Par_R,Par_P])
           mirls = P.*u
           res = de -  SeisConv(SeisRadon_tx(mirls,false;Par_R...),false;Par_W...)
           Q = 1.0./sqrt.((abs.(res).+0.1))
           P = sqrt.(abs.(mirls).+0.0001)
           append!(cost,cost1)
end
```