CLEAR: Covariant Least-Square Re-fitting

Charles DELEDALLE, Nicolas PAPADAKIS, Joseph SALMON & **Samuel VAITER**

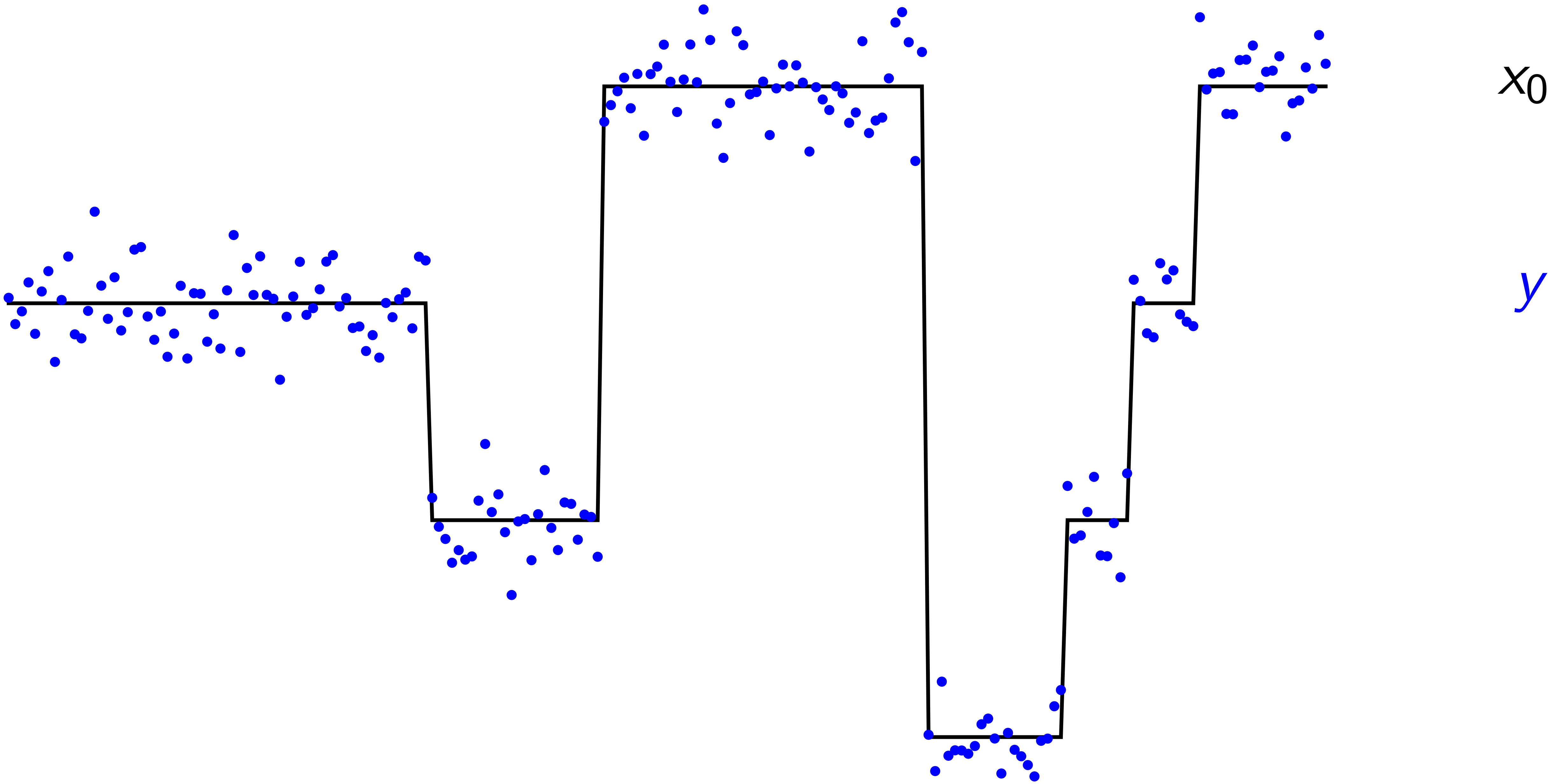
CNRS & IMB, Dijon, France

02/10/2016
SIGMA'16

A Starting Point

Noisy observations

$$y = x_0 + w$$



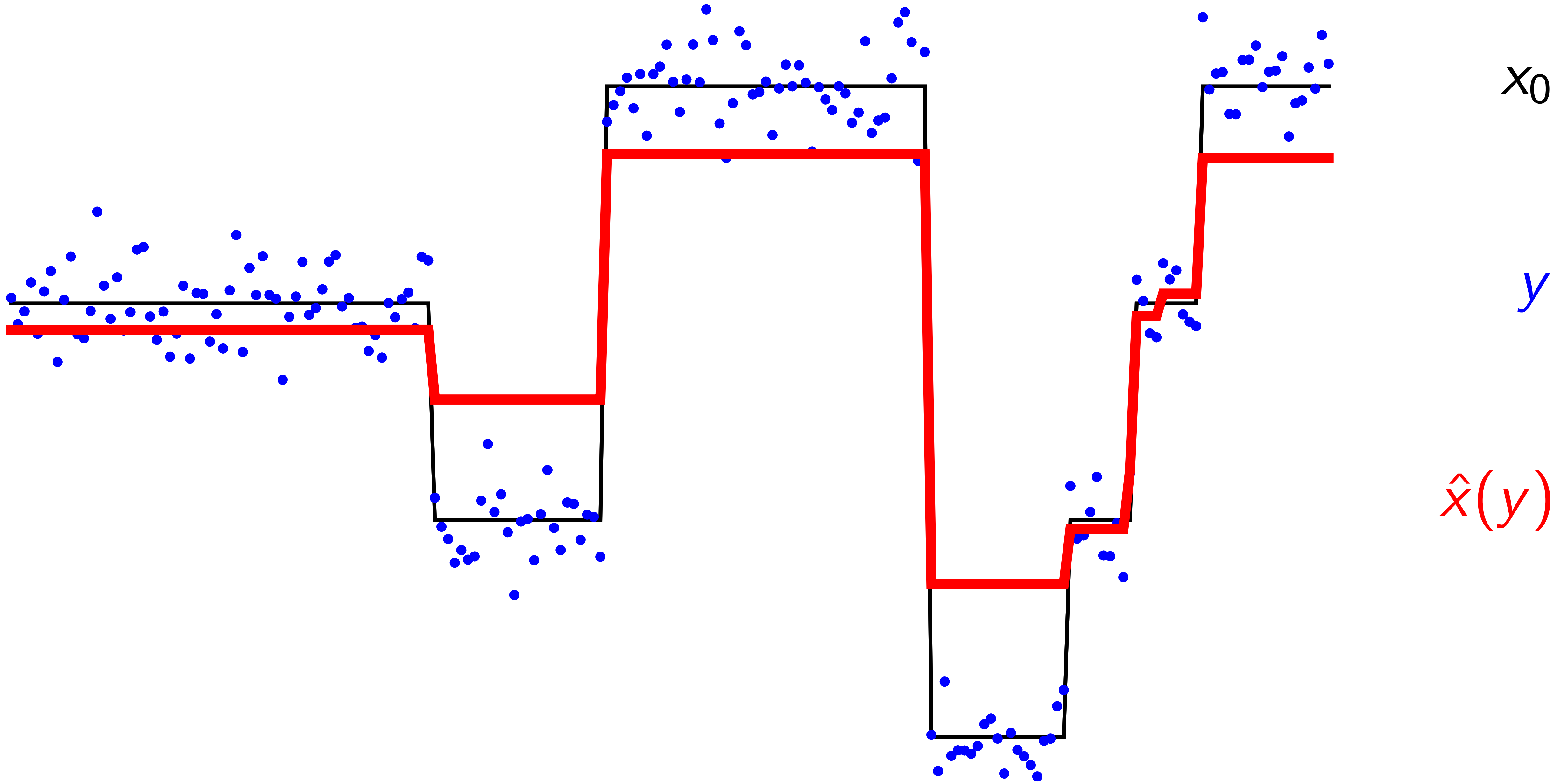
A Starting Point

Noisy observations

$$y = x_0 + w$$

Total Variation regularization [Rudin et al. 92]

$$\hat{x}(y) = \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \ \frac{1}{2}\|x - y\|_2^2 + \lambda \|\nabla x\|_1$$



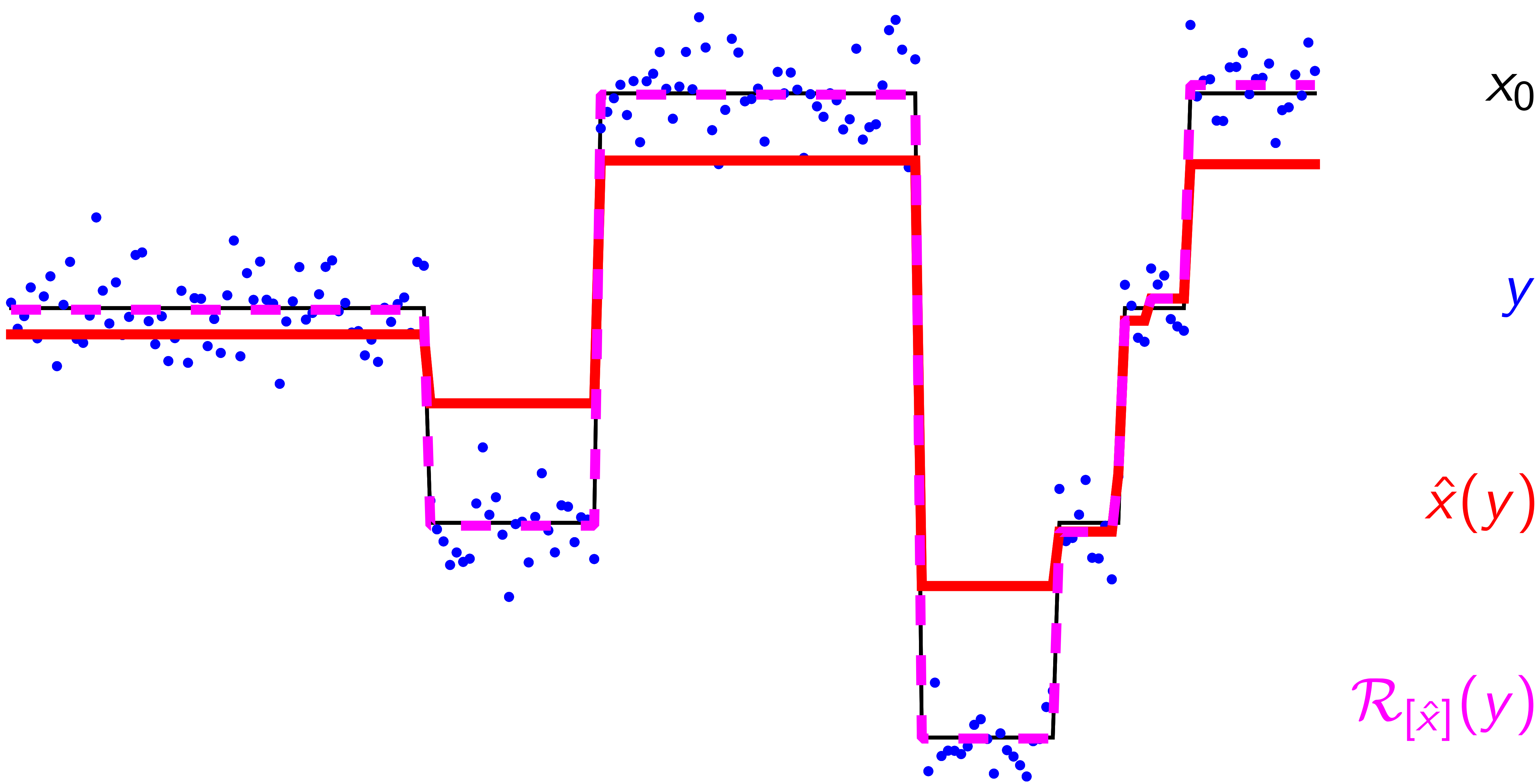
A Starting Point

Noisy observations

$$y = x_0 + w$$

Total Variation regularization [Rudin et al. 92]

$$\hat{x}(y) = \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{2}\|x - y\|_2^2 + \lambda \|\nabla x\|_1$$



Generalization of the Mean-Jump

Preserving the subspace $\{x \ : \ \mathrm{supp}(\nabla x) = \mathrm{supp}(\nabla \hat{x}(y))\}$

$\longrightarrow$ refitting [Efron et al. 2004], [Lederer 2013]

Generalization of the Mean-Jump

Preserving the subspace $\{x \ : \ \mathrm{supp}(\nabla x) = \mathrm{supp}(\nabla \hat{x}(y))\}$

$\longrightarrow$ refitting [Efron et al. 2004], [Lederer 2013]

General point of view ?

Generalization of the Mean-Jump

Preserving the subspace $\{x \ : \ \text{supp}(\nabla x) = \text{supp}(\nabla \hat{x}(y))\}$

$\longrightarrow$ refitting [Efron et al. 2004], [Lederer 2013]

General point of view ?

TV

Model space (for a weakly differentiable estimator)

$$\mathcal{M}_{[\hat{x}]}(y) = \hat{x}(y) + \text{Im}[J_{\hat{x}}(y)]$$

Jacobian of $\hat{x}$ at y

Generalization of the Mean-Jump

Preserving the subspace $\{x \ : \ \mathrm{supp}(\nabla x) = \mathrm{supp}(\nabla \hat{x}(y))\}$

$\longrightarrow$ refitting [Efron et al. 2004], [Lederer 2013]

General point of view ?

TV

Model space (for a weakly differentiable estimator)

$$\mathcal{M}_{[\hat{x}]}(y) = \hat{x}(y) + \mathrm{Im}[J_{\hat{x}}(y)]$$

Jacobian of $\hat{x}$ at y

Invariant Re-enhancement

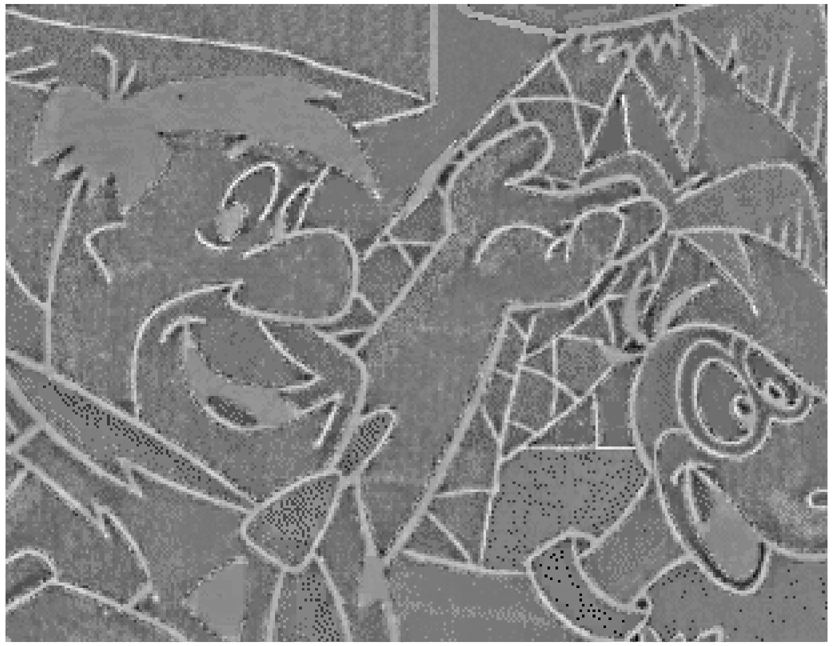
$$\mathcal{R}^{\mathrm{inv}}_{[\hat{x}]}(y) = \underset{x \in \mathcal{M}_{[\hat{x}]}(y)}{\mathrm{argmin}} \ \|y - x\|_2^2$$

$$\mathcal{R}^{\mathrm{inv}}_{[\hat{x}]}(y) = \hat{x}(y) + JJ^{+}(y - \hat{x}(y))$$

Performance on Anisotropic TV

y

$\hat{x}(y)$

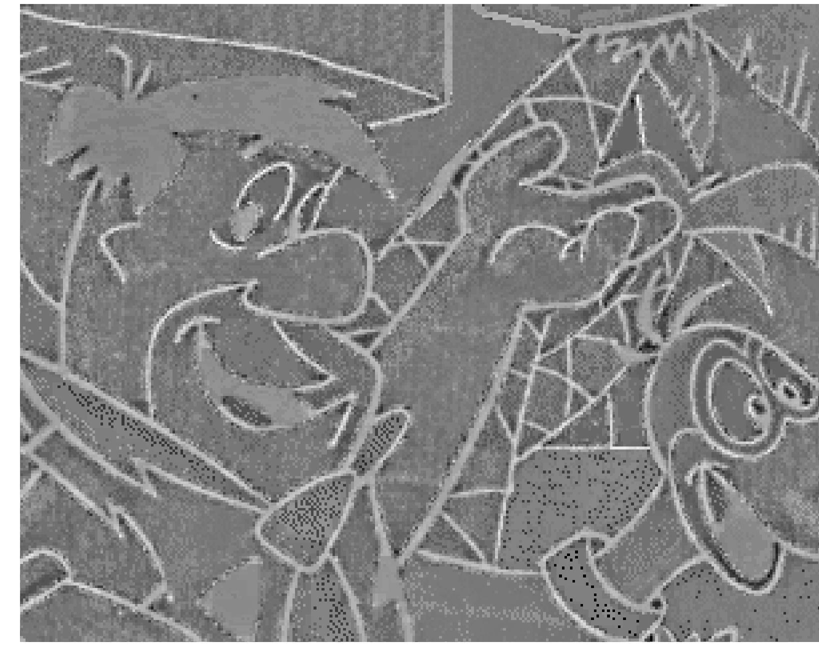


$\hat{x}(y) - x_0$

Performance on Anisotropic TV

y

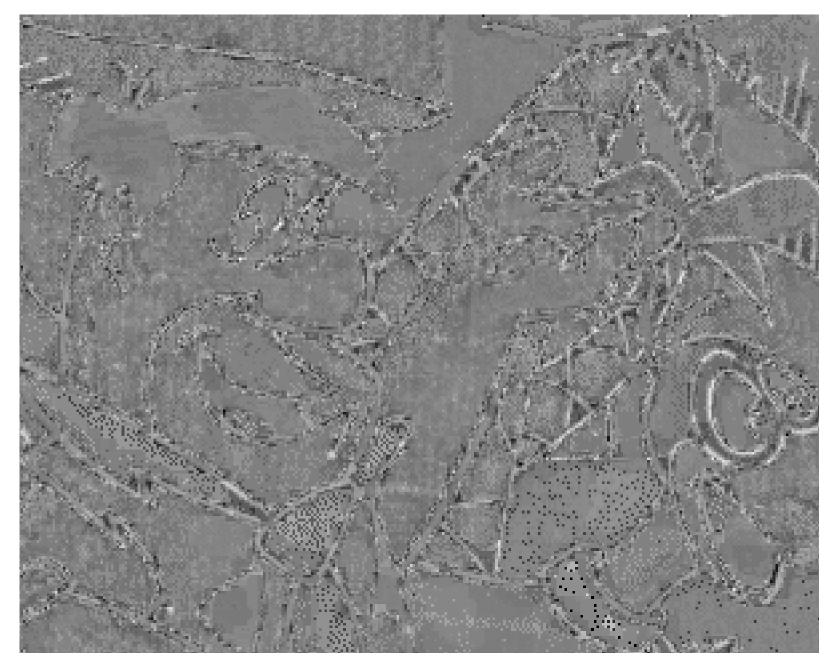
$\hat{x}(y)$



$\hat{x}(y) - x_0$

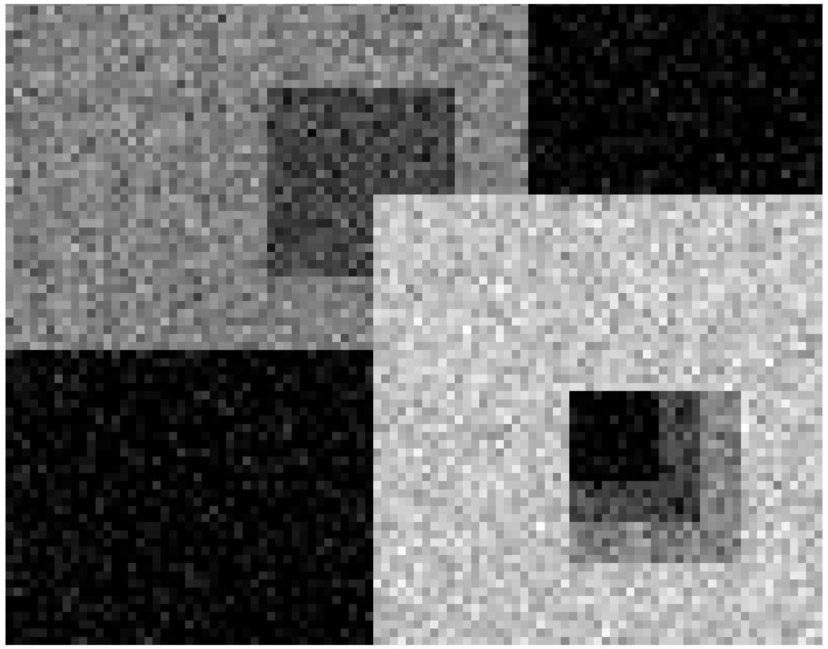
$$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y) = J_{\hat{x}}(y)$$

$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y)$



$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y) - x_0$

Performance on Isotropic TV

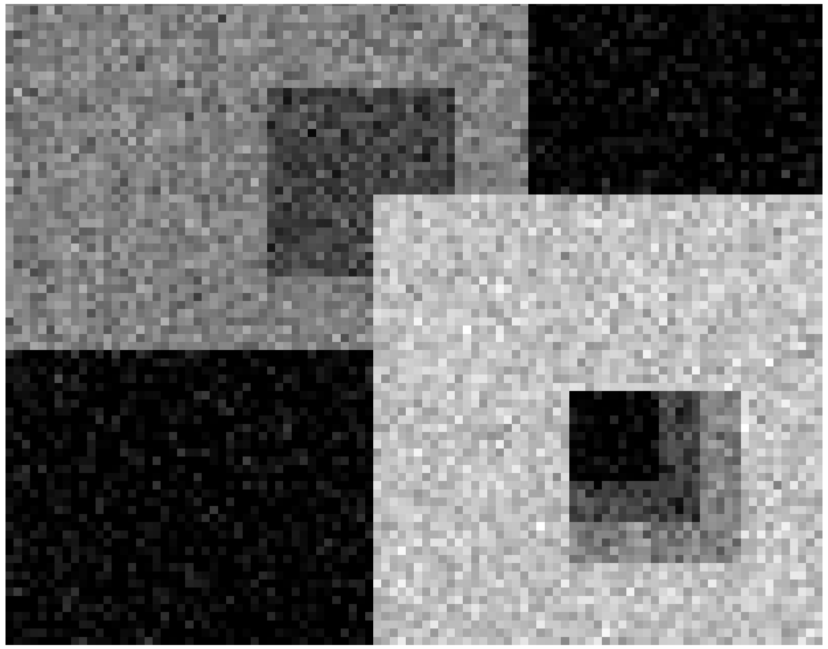


y

$\hat{x}(y)$

$\hat{x}(y) - x_0$

Performance on Isotropic TV



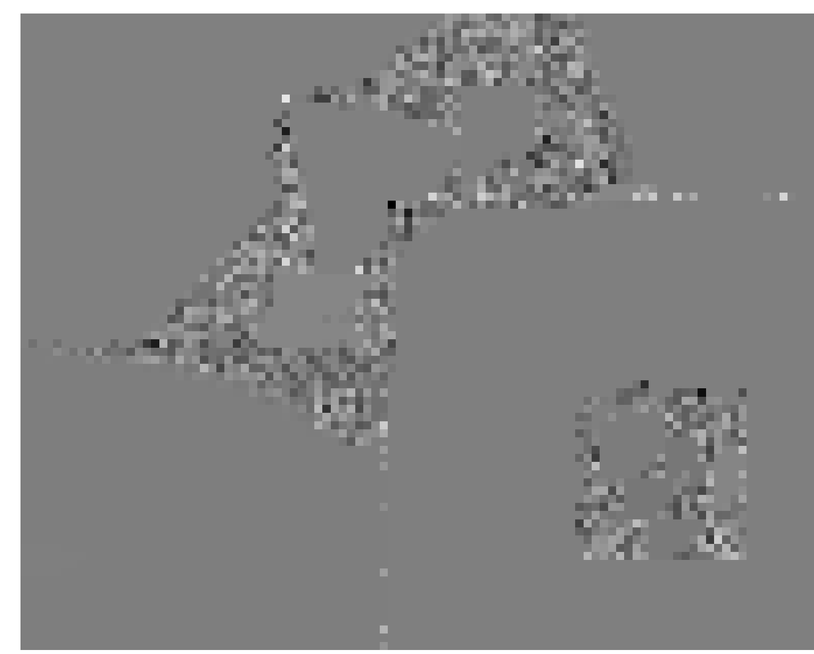
y

$\hat{x}(y)$

$\hat{x}(y) - x_0$

$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y) \neq J_{\hat{x}}(y)$

$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y)$



$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y) - x_0$

From Invariant to Covariant Renhancement

the model space captures only linear invariance

$$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y) = \underset{x \in \mathcal{M}_{[\hat{x}]}(y)}{\text{argmin}} \; \|y - x\|_2^2$$

need more constraints

From Invariant to Covariant Renhancement

the model space captures only linear invariance

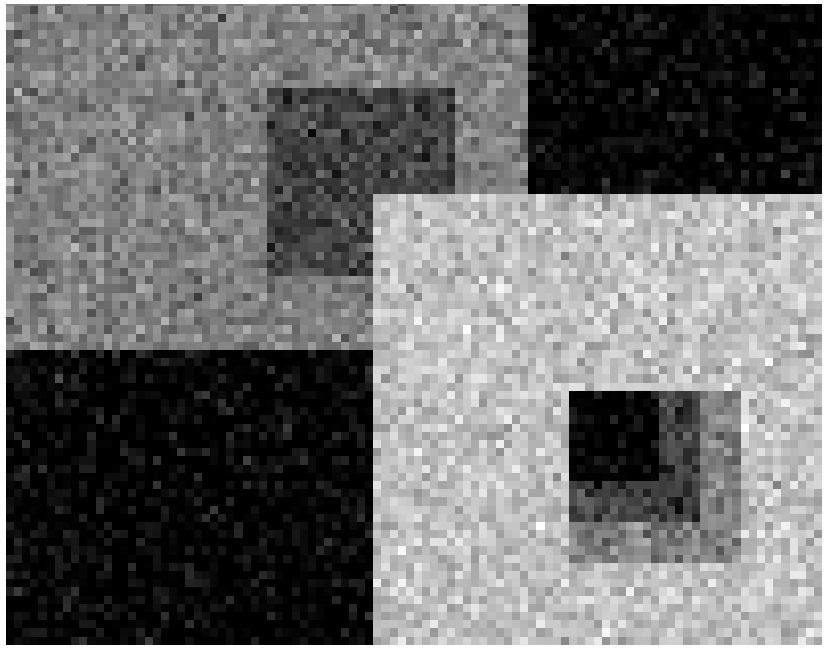
$$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y) = \underset{x \in \mathcal{M}_{[\hat{x}]}(y)}{\text{argmin}} \; \|y - x\|_2^2$$

need more constraints

Empirical observation

The Jacobian captures more invariances

From Invariant to Covariant Renhancement



y

$\hat{x}(y)$

$\hat{x}(y) - x_0$

$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y)$

$J_{\hat{x}}(y)$

Local Approach

$$\begin{aligned}\hat{x} : \mathbb{R}^p &\to \mathbb{R}^p \\ y &\mapsto \hat{x}(y)\end{aligned}$$

Local Approach

$$\begin{aligned} \hat{x} : \mathbb{R}^p &\to \mathbb{R}^p \\ y &\mapsto \hat{x}(y) \end{aligned}$$

enhance locally

$$\begin{aligned} \mathcal{D} : ((\mathbb{R}^p \to \mathbb{R}^p) \times \mathbb{R}^p) &\to (\mathbb{R}^p \to \mathbb{R}^p) \\ (\hat{x}, y) &\mapsto \mathcal{D}_{\hat{x}(y)} \end{aligned}$$

Local Approach

$$\begin{aligned} \hat{x} : \mathbb{R}^p &\to \mathbb{R}^p \\ y &\mapsto \hat{x}(y) \end{aligned}$$

enhance locally

$$\begin{aligned} \mathcal{D} : ((\mathbb{R}^p \to \mathbb{R}^p) \times \mathbb{R}^p) &\to (\mathbb{R}^p \to \mathbb{R}^p) \\ (\hat{x}, y) &\mapsto \mathcal{D}_{\hat{x}(y)} \end{aligned}$$

apply point-wise

$$\begin{aligned} \mathcal{R}_{[\hat{x}]} : \mathbb{R}^p &\to \mathbb{R}^p \\ y &\mapsto \mathcal{R}_{[\hat{x}]}(y) = \mathcal{D}_{\hat{x}(y)}(y) \end{aligned}$$

Covariant Re-enhancement

Local constraints

Affine map	$\mathcal{D}_{\hat{x}(y)}(z) = Az + b$
Covariant preserving	$J_{\mathcal{D}_{\hat{x}(y)}}(y) = J_{\hat{x}}(y)$
Coherent map	$\mathcal{D}_{\hat{x}(y)}(\hat{x}(y)) = \hat{x}(y)$

Covariant Re-enhancement

Local constraints

Affine map	$\mathcal{D}_{\hat{x}(y)}(z) = Az + b$
Covariant preserving	$J_{\mathcal{D}_{\hat{x}(y)}}(y) = J_{\hat{x}}(y)$
Coherent map	$\mathcal{D}_{\hat{x}(y)}(\hat{x}(y)) = \hat{x}(y)$

Characterization

$$\mathcal{D}_{\hat{x}(y)}(z) = \hat{x}(y) + J(z - \hat{x}(y))$$

Covariant Re-enhancement

Local constraints

Affine map $\mathcal{D}_{\hat{x}(y)}(z) = Az + b$

Covariant preserving $J_{\mathcal{D}_{\hat{x}(y)}}(y) = J_{\hat{x}}(y)$

Coherent map $\mathcal{D}_{\hat{x}(y)}(\hat{x}(y)) = \hat{x}(y)$

Characterization

$$\mathcal{D}_{\hat{x}(y)}(z) = \hat{x}(y) + J(z - \hat{x}(y))$$

Definition

$$\mathcal{R}_{[\hat{x}]}(y) = \mathcal{D}_{\hat{x}(y)}(y)$$

$$\mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

Covariant Re-enhancement

Local constraints

Affine map $\mathcal{D}_{\hat{x}(y)}(z) = Az + b$

Covariant preserving $J_{\mathcal{D}_{\hat{x}(y)}}(y) = J_{\hat{x}}(y)$

Coherent map $\mathcal{D}_{\hat{x}(y)}(\hat{x}(y)) = \hat{x}(y)$

Characterization

$$\mathcal{D}_{\hat{x}(y)}(z) = \hat{x}(y) + J(z - \hat{x}(y))$$

Definition

$$\mathcal{R}_{[\hat{x}]}(y) = \mathcal{D}_{\hat{x}(y)}(y)$$

$$\mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

Invariant vs Covariant

$$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y) = \hat{x}(y) + JJ^{+}(y - \hat{x}(y)) \qquad \mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

Covariant Re-enhancement

Local constraints

Affine map $\mathcal{D}_{\hat{x}(y)}(z) = Az + b$

Covariant preserving $J_{\mathcal{D}_{\hat{x}(y)}}(y) = J_{\hat{x}}(y)$

Coherent map $\mathcal{D}_{\hat{x}(y)}(\hat{x}(y)) = \hat{x}(y)$

Characterization

$$\mathcal{D}_{\hat{x}(y)}(z) = \hat{x}(y) + J(z - \hat{x}(y))$$

Definition

$$\mathcal{R}_{[\hat{x}]}(y) = \mathcal{D}_{\hat{x}(y)}(y)$$

$$\mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

Invariant vs Covariant

$$\mathcal{R}^{\text{inv}}_{[\hat{x}]}(y) = \hat{x}(y) + JJ^{+}(y - \hat{x}(y)) \qquad \mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

In practice ?

Numerical Stability

$$\hat{x}(y) = \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \ \frac{1}{2}\|x - y\|_2^2 + \lambda \|\nabla x\|_1$$

Computing $\mathcal{R}_{[\hat{x}]}(y)$ requires the knowledge of $\operatorname{supp}(\nabla \hat{x}(y))$

But in practice, $\hat{x}(y)$ is approximated through a sequence $\hat{x}^k(y)$

Unfortunately, $\hat{x}^k(y) \approx \hat{x}(y) \not\Rightarrow \operatorname{supp}(\hat{x}^k(y)) \approx \operatorname{supp}(\hat{x}(y))$

Numerical Stability

$$\hat{x}(y) = \underset{x \in \mathbb{R}^p}{\operatorname{argmin}} \ \frac{1}{2}\|x - y\|_2^2 + \lambda \|\nabla x\|_1$$

Computing $\mathcal{R}_{[\hat{x}]}(y)$ requires the knowledge of $\operatorname{supp}(\nabla \hat{x}(y))$

But in practice, $\hat{x}(y)$ is approximated through a sequence $\hat{x}^k(y)$

Unfortunately, $\hat{x}^k(y) \approx \hat{x}(y) \not\Rightarrow \operatorname{supp}(\hat{x}^k(y)) \approx \operatorname{supp}(\hat{x}(y))$

y

$\hat{x}(y)$

$\mathcal{R}_{[\hat{x}^k]}(y)$

$\mathcal{R}_{[\hat{x}]}(y)$

Joint Estimation—Re-enhancement

$$\mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

Iterative algorithm

$$\hat{x}^{k+1}(y) = \mathcal{A}_k(\hat{x}^k(y), y)$$

Joint Estimation — Re-enhancement

$$\mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

Iterative algorithm

$$\hat{x}^{k+1}(y) = \mathcal{A}_k(\hat{x}^k(y), y)$$

Chained expression

$$\hat{x}^k(y) = \mathcal{A}_k \circ \mathcal{A}_{k-1} \circ \cdots \circ \mathcal{A}_0(\hat{x}^0(y), y)$$

Joint Estimation—Re-enhancement

$$\mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

Iterative algorithm

$$\hat{x}^{k+1}(y) = \mathcal{A}_k(\hat{x}^k(y), y)$$

Chained expression

$$\hat{x}^k(y) = \mathcal{A}_k \circ \mathcal{A}_{k-1} \circ \cdots \circ \mathcal{A}_0(\hat{x}^0(y), y)$$

Chain rule

$$\frac{\partial}{\partial y}\mathcal{A}_k \circ \frac{\partial}{\partial y}\mathcal{A}_{k-1} \circ \cdots \circ \frac{\partial}{\partial y}\mathcal{A}_0(y) \to Jy$$

One-step vs Two-step Evaluation

Two step

$$\frac{\partial}{\partial y}\mathcal{A}_k \circ \frac{\partial}{\partial y}\mathcal{A}_{k-1} \circ \cdots \circ \frac{\partial}{\partial y}\mathcal{A}_0(y) \to Jy$$

$$\frac{\partial}{\partial y}\mathcal{A}_k \circ \frac{\partial}{\partial y}\mathcal{A}_{k-1} \circ \cdots \circ \frac{\partial}{\partial y}\mathcal{A}_0(\hat{x}(y)) \to J\hat{x}(y)$$

$$\mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

One-step vs Two-step Evaluation

Two step

$$\frac{\partial}{\partial y}\mathcal{A}_k \circ \frac{\partial}{\partial y}\mathcal{A}_{k-1} \circ \cdots \circ \frac{\partial}{\partial y}\mathcal{A}_0(y) \to Jy$$

$$\frac{\partial}{\partial y}\mathcal{A}_k \circ \frac{\partial}{\partial y}\mathcal{A}_{k-1} \circ \cdots \circ \frac{\partial}{\partial y}\mathcal{A}_0(\hat{x}(y)) \to J\hat{x}(y)$$

$$\mathcal{R}_{[\hat{x}]}(y) = \hat{x}(y) + J(y - \hat{x}(y))$$

One step

$$\text{If } J\hat{x}(y) = \hat{x}(y) \implies \mathcal{R}_{[\hat{x}]}(y) = Jy$$

$\to$ theoretical general results

$\to$ true for aniso-TV, iso-TV, Lasso, ...

Example for Anisotropic TV

$$\hat{x}(y) = \operatorname*{argmin}_{x \in \mathbb{R}^p} \frac{1}{2}\|x - y\|_2^2 + \lambda \|\nabla x\|_1$$

$$\begin{aligned} z^{k+1} &= \Pi_{B_\lambda}(z^k + \sigma \nabla v^k) \\ x^{k+1} &= (1+\tau)^{-1}\left(x^k + \tau(y + \operatorname{div} z^{k+1})\right) \\ v^{k+1} &= x^{k+1} + \theta(x^{k+1} - x^k) \end{aligned}$$

Chambolle-Pock

Π_{B_λ} projection on the λ-ball

Example for Anisotropic TV

$$\hat{x}(y) = \operatorname*{argmin}_{x \in \mathbb{R}^p} \frac{1}{2}\|x - y\|_2^2 + \lambda \|\nabla x\|_1$$

$$\begin{aligned}
z^{k+1} &= \Pi_{B_\lambda}(z^k + \sigma \nabla v^k) \\
x^{k+1} &= (1+\tau)^{-1}\left(x^k + \tau(y + \operatorname{div} z^{k+1})\right) \\
v^{k+1} &= x^{k+1} + \theta(x^{k+1} - x^k)
\end{aligned}$$

Chambolle-Pock

$$\begin{aligned}
\tilde{z}^{k+1} &= \Psi_{z^k + \sigma \nabla v^k}(\tilde{z}^k + \sigma \nabla \tilde{v}^k) \\
\tilde{x}^{k+1} &= (1+\tau)^{-1}\left(\tilde{x}^k + \tau(y + \operatorname{div} \tilde{z}^{k+1})\right) \\
\tilde{v}^{k+1} &= \tilde{x}^{k+1} + \theta(\tilde{x}^{k+1} - \tilde{x}^k)
\end{aligned}$$

Differentiation of CP

Π_{B_λ} projection on the λ-ball

Ψ_z hard-thresholding

Example for Anisotropic TV

$$\hat{x}(y) = \operatorname*{argmin}_{x \in \mathbb{R}^p} \frac{1}{2}\|x - y\|_2^2 + \lambda \|\nabla x\|_1$$

$$z^{k+1} = \Pi_{B_\lambda}(z^k + \sigma \nabla v^k)$$
$$x^{k+1} = (1+\tau)^{-1}\left(x^k + \tau(y + \operatorname{div} z^{k+1})\right)$$
$$v^{k+1} = x^{k+1} + \theta(x^{k+1} - x^k)$$

Chambolle-Pock

$$\tilde{z}^{k+1} = \Psi_{z^k + \sigma \nabla v^k}(\tilde{z}^k + \sigma \nabla \tilde{v}^k)$$
$$\tilde{x}^{k+1} = (1+\tau)^{-1}\left(\tilde{x}^k + \tau(y + \operatorname{div} \tilde{z}^{k+1})\right)$$
$$\tilde{v}^{k+1} = \tilde{x}^{k+1} + \theta(\tilde{x}^{k+1} - \tilde{x}^k)$$

Differentiation of CP

Π_{B_λ} projection on the λ-ball

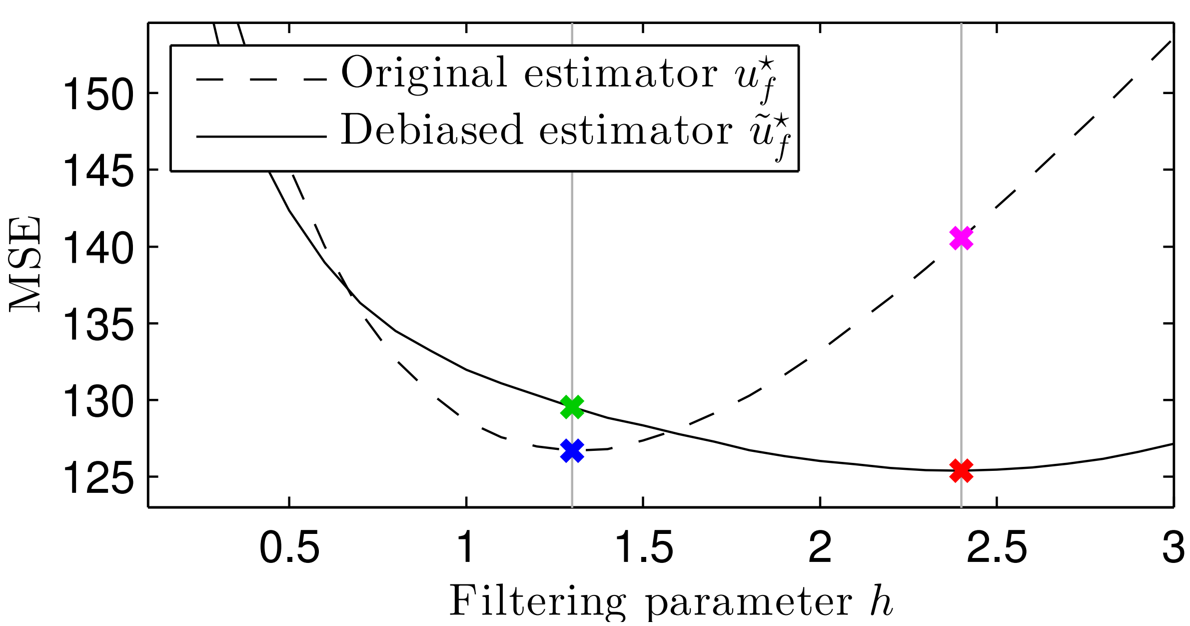
Ψ_z hard-thresholding

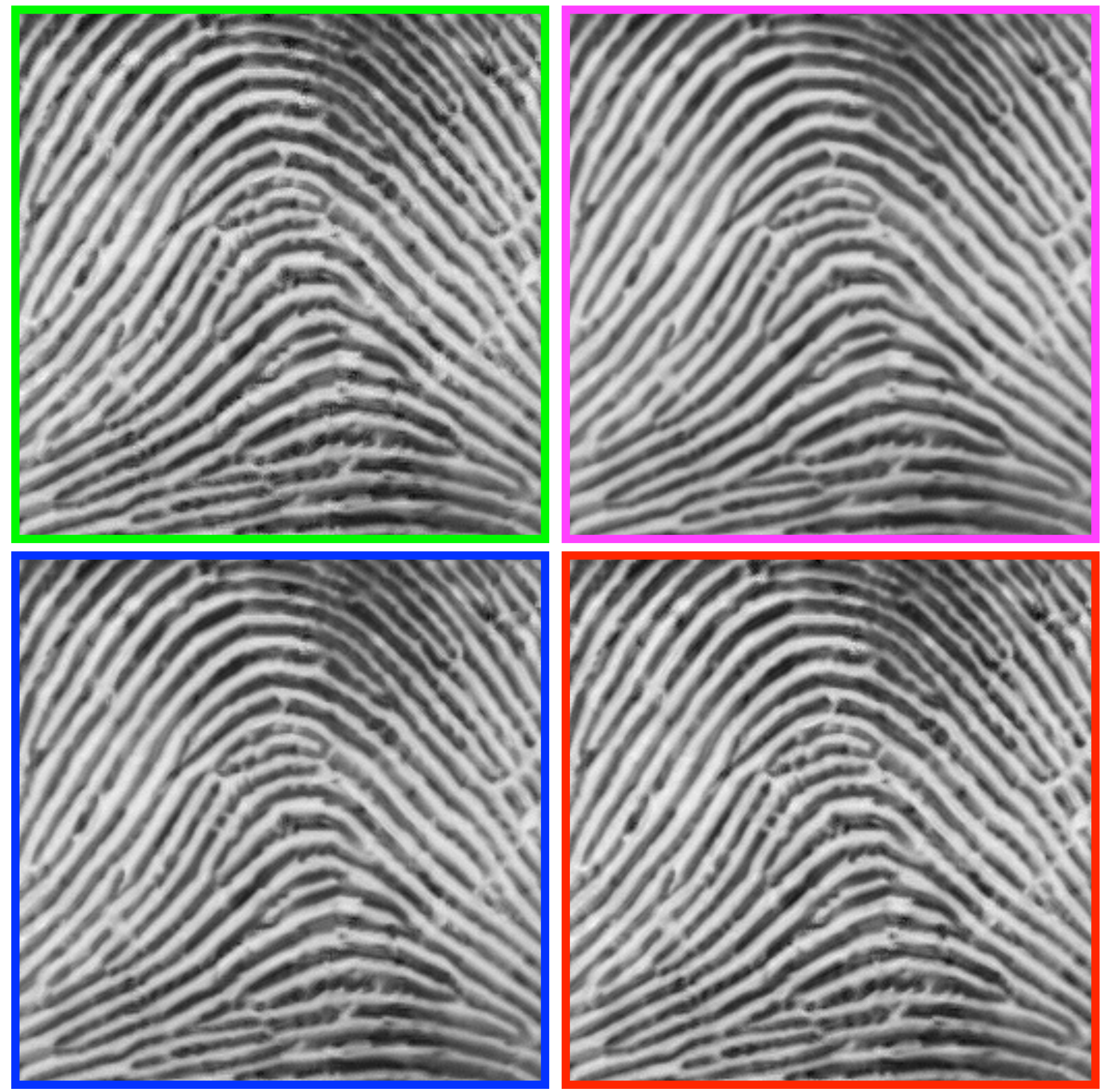
Complexity: 2x CP

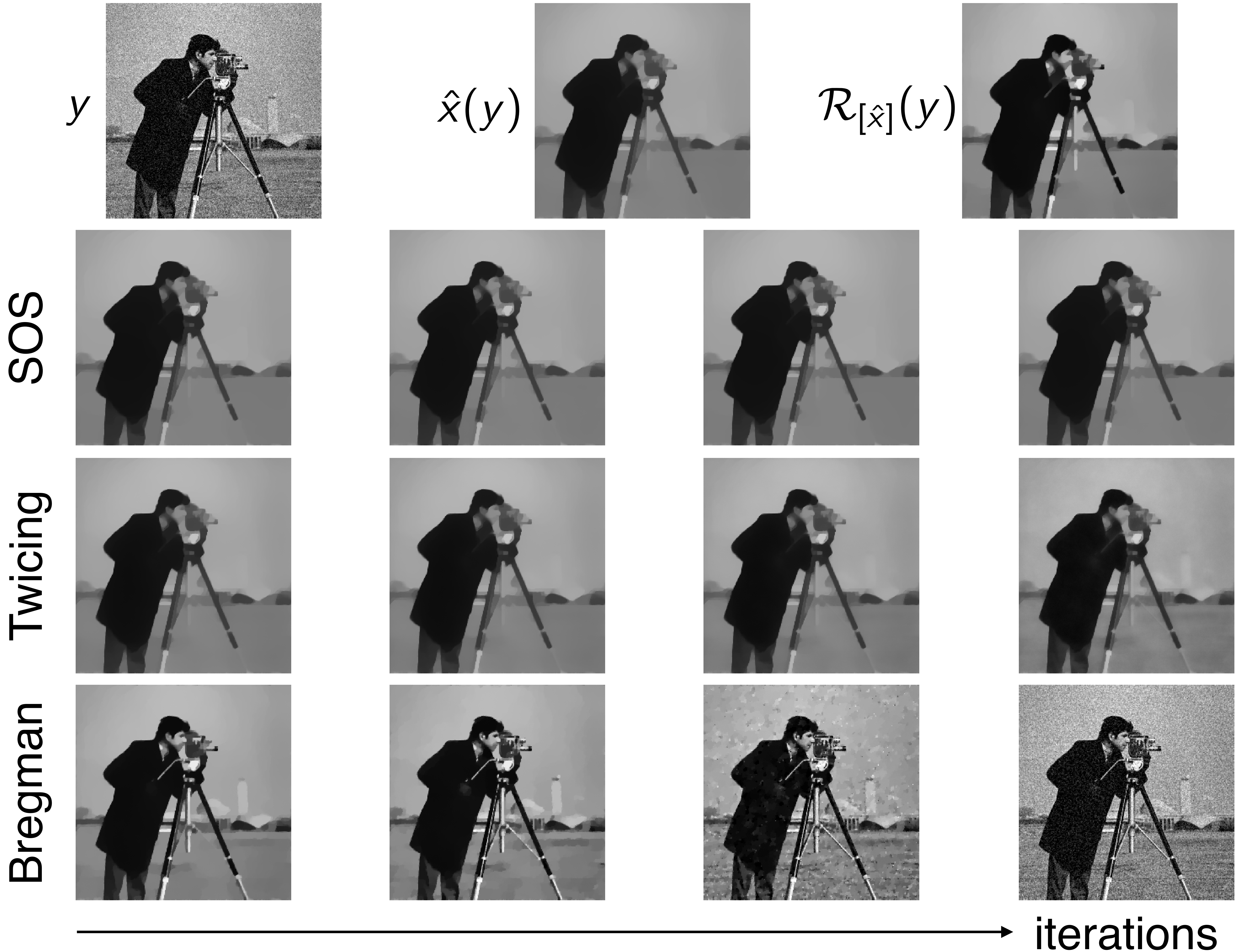
Numerical Evaluation for NLM

y

$\hat{x}(y)$

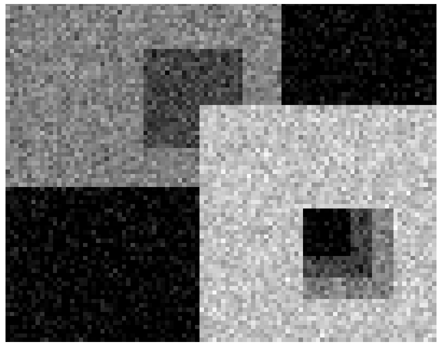


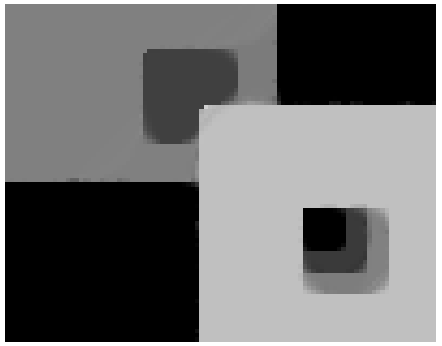




y
$\hat{x}(y)$
$\mathcal{R}_{[\hat{x}]}(y)$
SOS
Twicing
Bregman
iterations

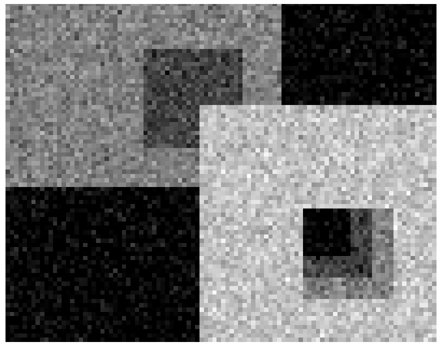
Conclusion

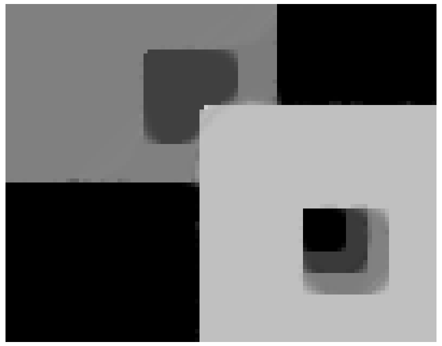




Fast and accurate denoising re-enhancement

Conclusion



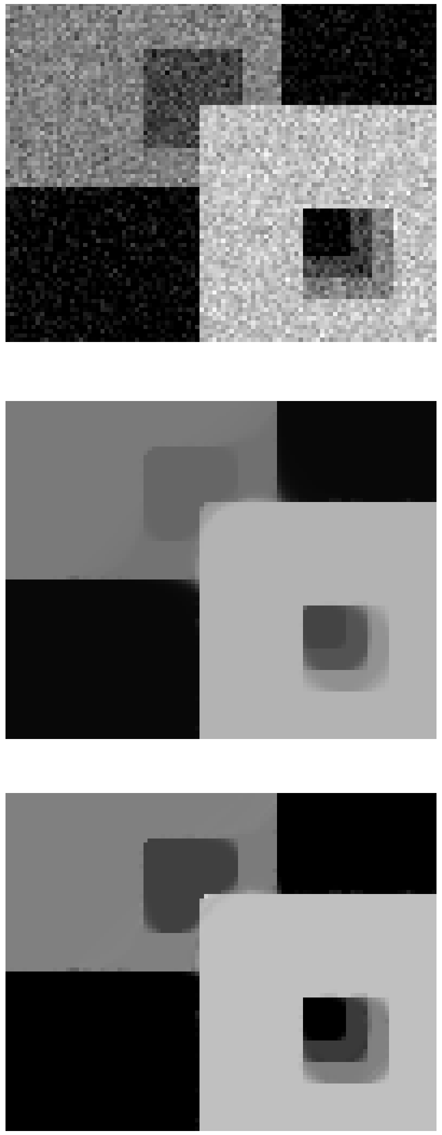


Fast and accurate denoising re-enhancement

Not covered today:

- Bias reduction (theoretical) results
- Approximate Jacobian preserving

Conclusion



Fast and accurate denoising re-enhancement

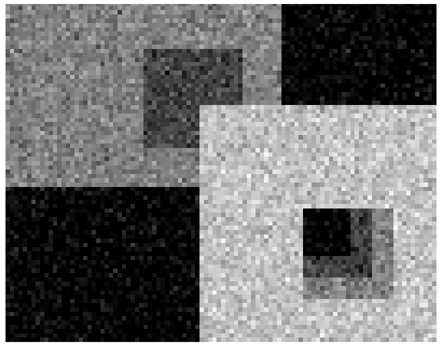
Not covered today:

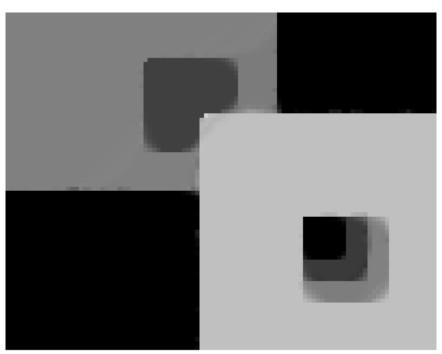
- Bias reduction (theoretical) results
- Approximate Jacobian preserving

This talk: denoising $y = x_0 + w$

→ works for inverse problems too $y = \Phi x_0 + w$

Conclusion





Fast and accurate denoising re-enhancement

Not covered today:

- Bias reduction (theoretical) results
- Approximate Jacobian preserving

This talk: denoising $y = x_0 + w$

works for inverse problems too $y = \Phi x_0 + w$

Thanks for your attention

Part of this work was presented at SSVM’15 (invariant)

Preprint online (covariant)