Texture segmentation based on fractal attributes using convex functional minimization with generalized Stein formalism for automated regularization parameter selection.

Barbara Pascal

October 12$^{\text{th}}$ 2021

Under the supervision of **Patrice Abry** and **Nelly Pustelnik**

Collaboration with **Valérie Vidal** and **Samuel Vaiter**

Image segmentation

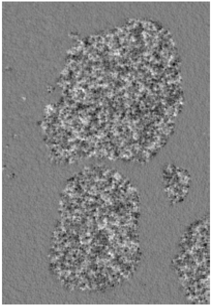
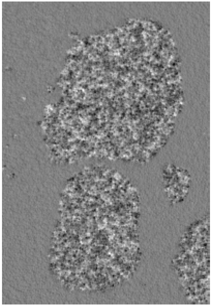
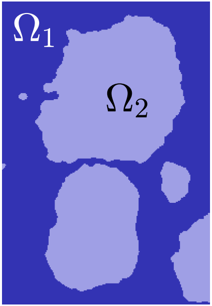


Image segmentation

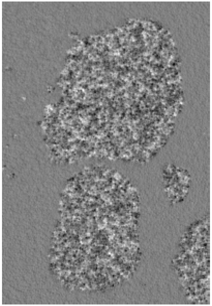


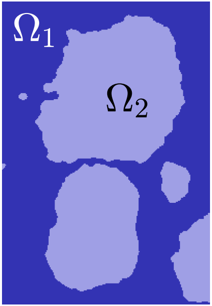


Goal: obtain a partition of the image into K homogeneous regions

$$\Omega = \Omega_1 \bigsqcup \ldots \bigsqcup \Omega_K$$

Image segmentation





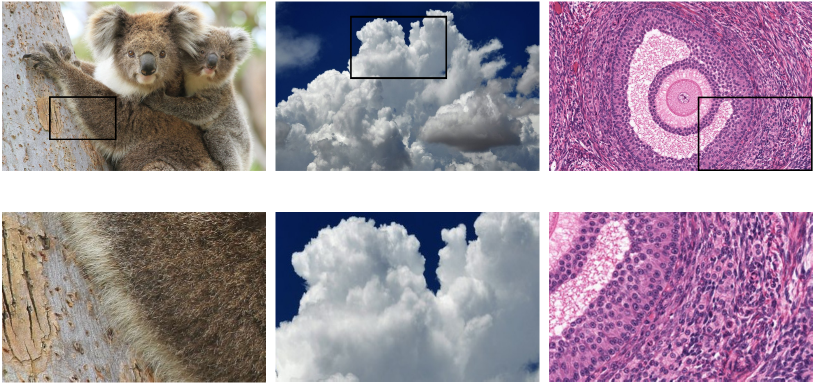
Goal: obtain a partition of the image into K **homogeneous** regions

$$\Omega = \Omega_1 \bigsqcup \ldots \bigsqcup \Omega_K$$

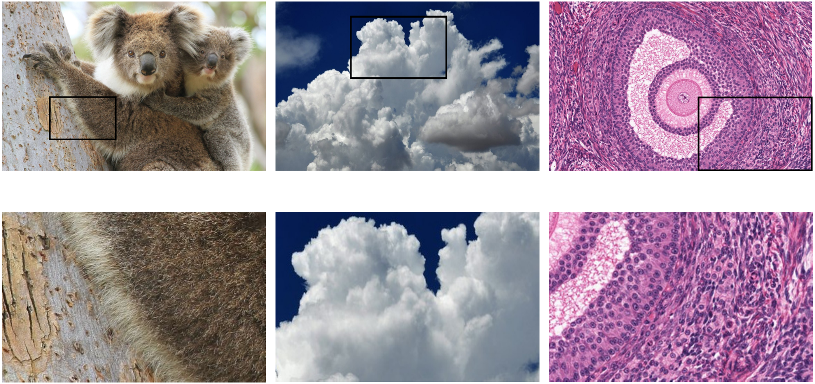
Textures

Textures

Textures



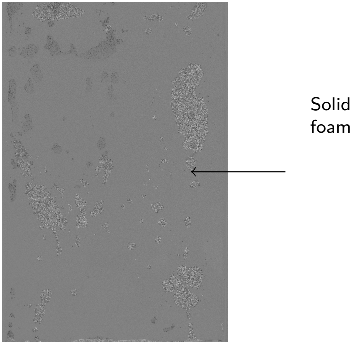
Textures



Crucial to describe real-world images

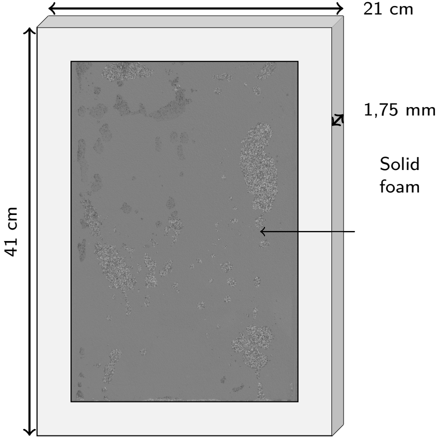
Multiphase flow through porous media

Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)



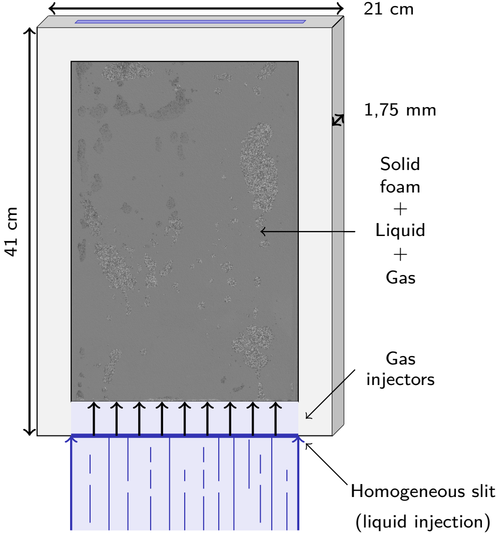
Multiphase flow through porous media

Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)



Multiphase flow through porous media

Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)



→ 1600 × 1100 pixels
→ video: ∼ 1000 images
→ phase diagram: ∼ 10 flow rates

Presentation outline

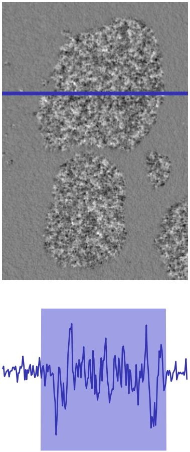
1. Texture characterization: *fractals attributes*
   - local variance σ^2
   - local regularity h

2. Design of functionals: *penalized least-squares*
   - free contours
   - co-localized contours

3. Accelerated minimization algorithms: *splitting proximal algorithms*
   - computation of proximal operators
   - acceleration based on strong-convexity

4. Hyperparameters tuning: *SURE under Gaussian correlated noise*
   - projected estimation error
   - quasi-Newton minimization

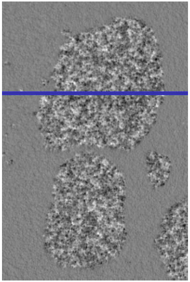
Piecewise monofractal model

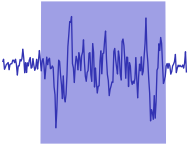


Piecewise monofractal model

Fractals attributes

- variance σ^2 *amplitude of variations*

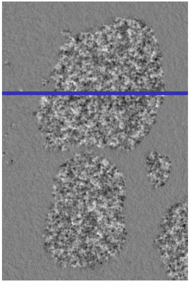


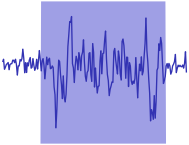


Piecewise monofractal model

Fractals attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*



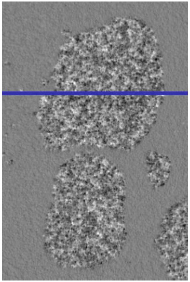


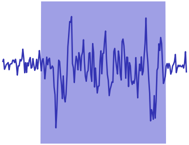
Piecewise monofractal model

Fractals attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

$$|f(x) - f(y)| \leq \sigma(x)|x - y|^{h(x)}$$





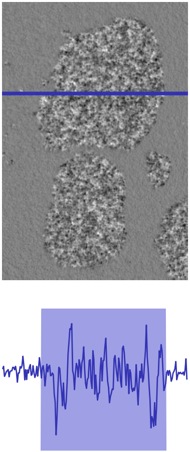
Piecewise monofractal model

Fractals attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

$$|f(x) - f(y)| \leq \sigma(x)|x-y|^{h(x)}$$

$h(x) \equiv h_1 = 0.9$ $\qquad$ $h(x) \equiv h_2 = 0.3$



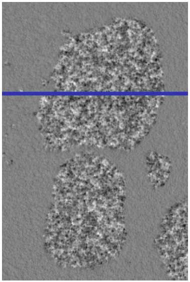
Piecewise monofractal model

Fractals attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

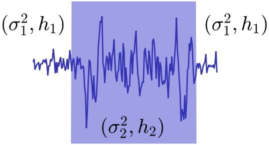
$$|f(x) - f(y)| \leq \sigma(x)|x - y|^{h(x)}$$

$h(x) \equiv h_1 = 0.9$ $\quad$ $h(x) \equiv h_2 = 0.3$



Segmentation

- σ^2 and h piecewise constant

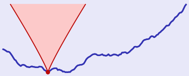


Piecewise monofractal model

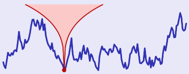
Fractals attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

$$|f(x) - f(y)| \leq \sigma(x)|x - y|^{h(x)}$$



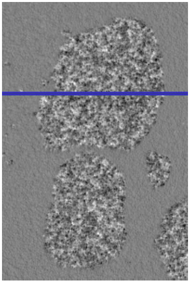
$h(x) \equiv h_1 = 0.9$

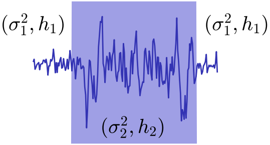


$h(x) \equiv h_2 = 0.3$

Segmentation

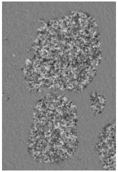
- σ^2 and h piecewise constant
- region Ω_k characterized by (σ_k^2, h_k)





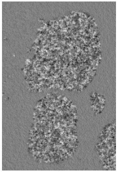
Multiscale analysis

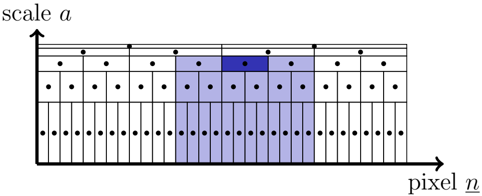
Textured image



Multiscale analysis

Textured image

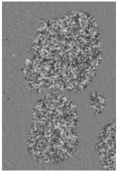




Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

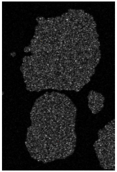
Multiscale analysis

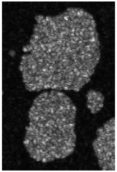
Textured image



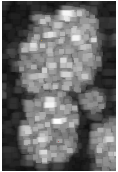
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

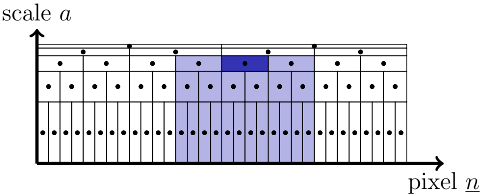
Scale $a = 2^1$ $a = 2^2$ $a = 2^5$





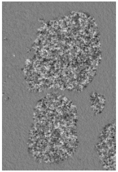
$\cdots$





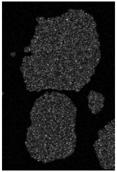
Multiscale analysis

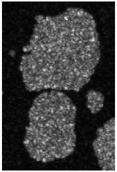
Textured image

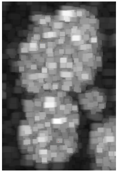


Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$





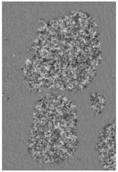


Proposition *(Jaffard, 2004), (Wendt, 2008)*

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$

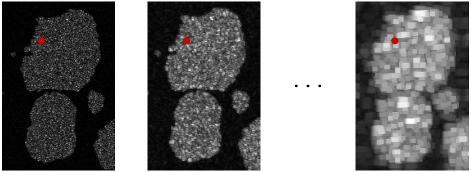
Multiscale analysis

Textured image



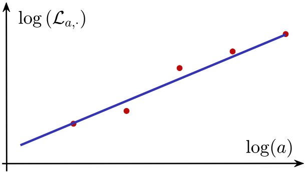
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$



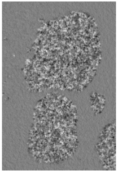
Proposition *(Jaffard, 2004), (Wendt, 2008)*

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



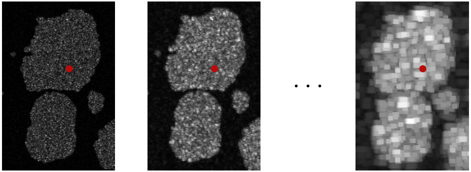
Multiscale analysis

Textured image



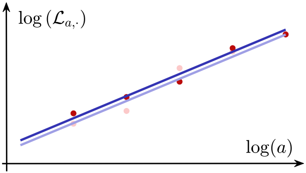
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$



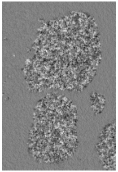
Proposition *(Jaffard, 2004), (Wendt, 2008)*

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



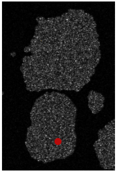
Multiscale analysis

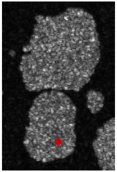
Textured image

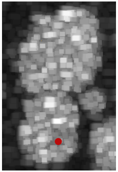


Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ $\dots$ $a = 2^5$

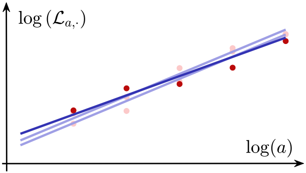






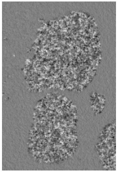
Proposition *(Jaffard, 2004), (Wendt, 2008)*

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



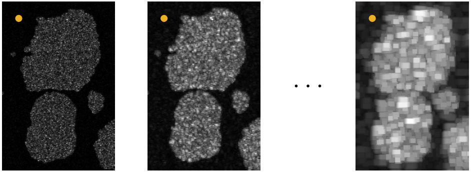
Multiscale analysis

Textured image



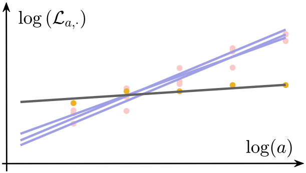
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ ... $a = 2^5$



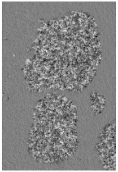
Proposition *(Jaffard, 2004), (Wendt, 2008)*

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



Multiscale analysis

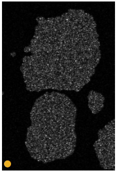
Textured image



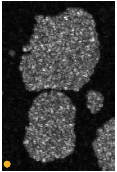
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale

$a = 2^1$

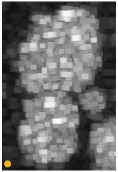


$a = 2^2$



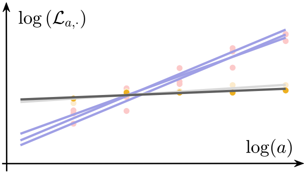
...

$a = 2^5$



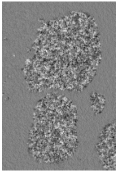
Proposition *(Jaffard, 2004), (Wendt, 2008)*

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



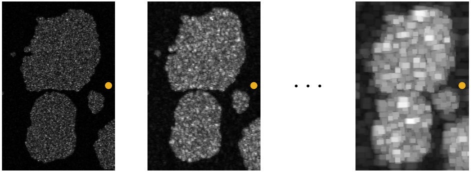
Multiscale analysis

Textured image



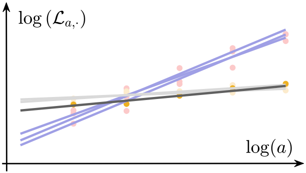
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ $\dots$ $a = 2^5$



Proposition *(Jaffard, 2004), (Wendt, 2008)*

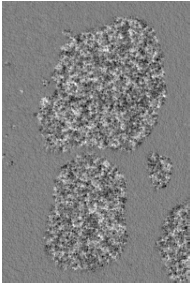
$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a \to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



Direct punctual estimation

Linear regression $\log\left(\mathcal{L}_{a,\cdot}\right) \simeq \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\propto \log(\boldsymbol{\sigma}^2)}$

Textured image

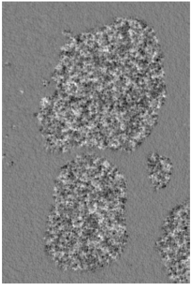


Direct punctual estimation

Linear regression $\log\left(\mathcal{L}_{a,\cdot}\right) \simeq \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\propto \log(\boldsymbol{\sigma}^2)}$

$$\left(\widehat{\boldsymbol{h}}^{\mathrm{LR}}, \widehat{\boldsymbol{v}}^{\mathrm{LR}}\right) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a=a_{\min}}^{a_{\max}} \left\|\log\left(\mathcal{L}_{a,\cdot}\right) - \log(a)\boldsymbol{h} - \boldsymbol{v}\right\|^2$$

Textured image

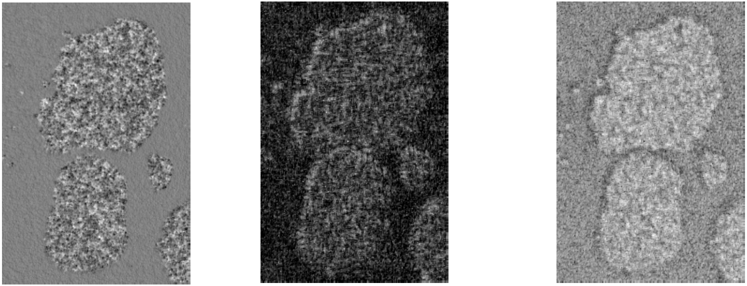


Direct punctual estimation

Linear regression $\log\left(\mathcal{L}_{a,\cdot}\right) \simeq \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\propto \log(\boldsymbol{\sigma}^2)}$

$$\left(\widehat{\boldsymbol{h}}^{\mathrm{LR}}, \widehat{\boldsymbol{v}}^{\mathrm{LR}}\right) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a=a_{\min}}^{a_{\max}} \left\|\log\left(\mathcal{L}_{a,\cdot}\right) - \log(a)\boldsymbol{h} - \boldsymbol{v}\right\|^2$$

Textured image | Local regularity $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ | Local power $\widehat{\boldsymbol{v}}^{\mathrm{LR}}$

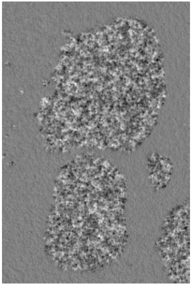


Direct punctual estimation

Linear regression $\underbrace{\mathbb{E}\log\left(\mathcal{L}_{a,\cdot}\right)}_{\text{expected value}} = \log(a)\underbrace{\bar{\boldsymbol{h}}}_{\text{regularity}} + \underbrace{\bar{\boldsymbol{v}}}_{\propto\log(\boldsymbol{\sigma}^2)}$

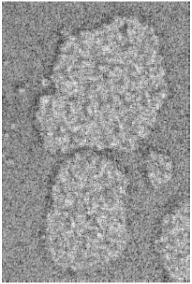
$$\left(\widehat{\boldsymbol{h}}^{\mathrm{LR}}, \widehat{\boldsymbol{v}}^{\mathrm{LR}}\right) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a=a_{\min}}^{a_{\max}} \left\|\log\left(\mathcal{L}_{a,\cdot}\right) - \log(a)\boldsymbol{h} - \boldsymbol{v}\right\|^2$$

Textured image



Local regularity $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

Local power $\widehat{\boldsymbol{v}}^{\mathrm{LR}}$



$\longrightarrow$ large estimation variance

A posteriori regularization

Linear regression $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

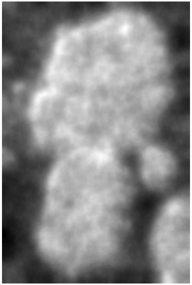
A posteriori regularization

Filter smoothing (linear)

$$\left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \widehat{\boldsymbol{h}}^{\mathrm{LR}}$$

Linear regression $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

Lissage



A posteriori regularization

Filter smoothing (linear)

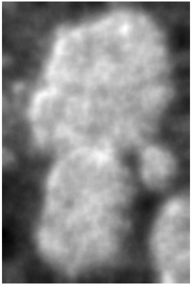
$$\left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \widehat{\boldsymbol{h}}^{\mathrm{LR}}$$

ROF denoising (nonlinear)

$$\underset{\boldsymbol{h}}{\operatorname{argmin}} \; \|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathrm{LR}}\|^2 + \lambda \|\mathbf{D}\boldsymbol{h}\|_{2,1}$$

Linear regression $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

Lissage



ROF

A posteriori regularization

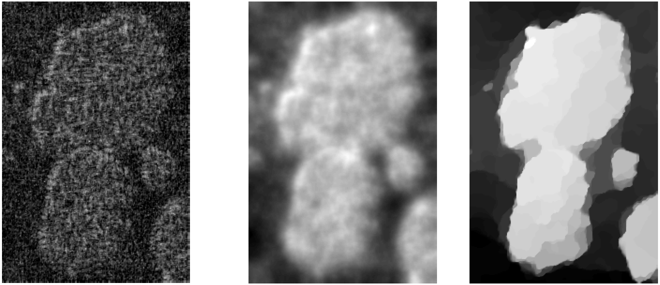
Filter smoothing (linear)

$$\left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \widehat{\boldsymbol{h}}^{\mathrm{LR}}$$

ROF denoising (nonlinear)

$$\underset{\boldsymbol{h}}{\operatorname{argmin}} \ \|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathrm{LR}}\|^2 + \lambda \|\mathbf{D}\boldsymbol{h}\|_{2,1}$$

Linear regression $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ | Lissage | ROF

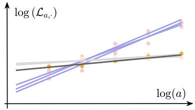


$\longrightarrow$ cumulative estimation variance and regularization bias

Functionals with either free or co-localized contours

$$\sum_a \underbrace{\|\log \mathcal{L}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}}$$

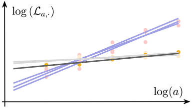
$\rightarrow$ fidelity to the log-linear model



Functionals with either free or co-localized contours

$$\underbrace{\sum_a \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

→ fidelity to the log-linear model

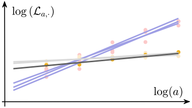


→ favors piecewise constancy

Functionals with either free or co-localized contours

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

→ fidelity to the log-linear model



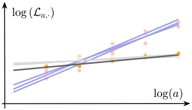
→ favors piecewise constancy

Functionals with either free or co-localized contours

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

→ fidelity to the log-linear model

→ favors piecewise constancy



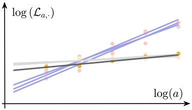
Finite differences $\mathbf{D}_1\boldsymbol{x}$ (horizontal), $\mathbf{D}_2\boldsymbol{x}$ (vertical) in each pixel

Functionals with either free or co-localized contours

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \underbrace{\sum_a \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

→ fidelity to the log-linear model

→ favors piecewise constancy



Finite differences $\mathbf{D}\boldsymbol{x} = [\mathrm{D}_1\boldsymbol{x}, \mathrm{D}_2\boldsymbol{x}]$

Free: $\boldsymbol{h}$, $\boldsymbol{v}$ are **independently** piecewise constant

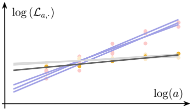
$$\mathcal{Q}_{\mathrm{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha) = \alpha\|\mathbf{D}\boldsymbol{h}\|_{2,1} + \|\mathbf{D}\boldsymbol{v}\|_{2,1}$$

Functionals with either free or co-localized contours

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \quad \underbrace{\sum_a \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

→ fidelity to the log-linear model

→ favors piecewise constancy



Finite differences $\mathbf{D}\boldsymbol{x} = [\mathbf{D}_1\boldsymbol{x}, \mathbf{D}_2\boldsymbol{x}]$

Free: $\boldsymbol{h}$, $\boldsymbol{v}$ are **independently** piecewise constant

$$\mathcal{Q}_{\mathrm{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha) = \alpha\|\mathbf{D}\boldsymbol{h}\|_{2,1} + \|\mathbf{D}\boldsymbol{v}\|_{2,1}$$

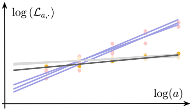
Co-localized: $\boldsymbol{h}$, $\boldsymbol{v}$ are **concomitantly** piecewise constant

$$\mathcal{Q}_{\mathrm{C}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha) = \|[\alpha\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}]\|_{2,1}$$

Functionals with either free or co-localized contours

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \quad \underbrace{\sum_a \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \quad + \quad \lambda \; \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

Least-Squares → fidelity to the log-linear model



Total Variation → favors piecewise constancy

Disjoint contours

$\boldsymbol{h} \in \mathbb{R}^{2\times 2}$ $\quad$ $\boldsymbol{v} \in \mathbb{R}^{2\times 2}$

Common contours

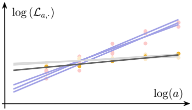
$\boldsymbol{h} \in \mathbb{R}^{2\times 2}$ $\quad$ $\boldsymbol{v} \in \mathbb{R}^{2\times 2}$

Functionals with either free or co-localized contours

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

→ fidelity to the log-linear model

→ favors piecewise constancy



Disjoint contours

$\boldsymbol{h} \in \mathbb{R}^{2\times 2}$ $\boldsymbol{v} \in \mathbb{R}^{2\times 2}$

$\mathcal{Q}_{\mathsf{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; 1) = 4$

Common contours

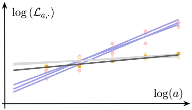
$\boldsymbol{h} \in \mathbb{R}^{2\times 2}$ $\boldsymbol{v} \in \mathbb{R}^{2\times 2}$

$\mathcal{Q}_{\mathsf{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; 1) = 4$

Functionals with either free or co-localized contours

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \quad \underbrace{\sum_a \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \quad + \quad \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

Least-Squares → fidelity to the log-linear model



Total Variation → favors piecewise constancy

Disjoint contours

$\boldsymbol{h} \in \mathbb{R}^{2\times 2}$ $\quad \boldsymbol{v} \in \mathbb{R}^{2\times 2}$

$\mathcal{Q}_{\mathrm{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; 1) = 4$

$\mathcal{Q}_{\mathrm{C}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; 1) = 2 + \sqrt{2} \simeq 3.4$

Common contours

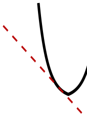
$\boldsymbol{h} \in \mathbb{R}^{2\times 2}$ $\quad \boldsymbol{v} \in \mathbb{R}^{2\times 2}$

$\mathcal{Q}_{\mathrm{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; 1) = 4$

$\mathcal{Q}_{\mathrm{C}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; 1) = 2\sqrt{2} \simeq 2.8$

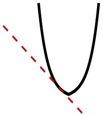
Functionals minimization

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



Functionals minimization

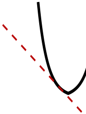
$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \ \sum_a \underbrace{\|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \ + \ \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



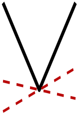
- gradient descent $\boldsymbol{x}^{n+1} = \boldsymbol{x}^n - \tau \nabla \varphi(\boldsymbol{x}^n)$

Functionals minimization

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \ \sum_a \underbrace{\|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



nonsmooth

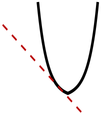


- gradient descent $\boldsymbol{x}^{n+1} = \boldsymbol{x}^n - \tau\nabla\varphi(\boldsymbol{x}^n)$
- implicit subgradient descent: proximal point algorithm

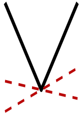
$$\boldsymbol{x}^{n+1} = \boldsymbol{x}^n - \tau\boldsymbol{u}^n, \ \boldsymbol{u}^n \in \partial\varphi(\boldsymbol{x}^{n+1}) \ \Leftrightarrow \ \boldsymbol{x}^{n+1} = \operatorname{prox}_{\tau\varphi}(\boldsymbol{x}^n)$$

Functionals minimization

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \ \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \ + \ \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



nonsmooth



- gradient descent $\boldsymbol{x}^{n+1} = \boldsymbol{x}^n - \tau \nabla \varphi(\boldsymbol{x}^n)$
- implicit subgradient descent: proximal point algorithm

$$\boldsymbol{x}^{n+1} = \boldsymbol{x}^n - \tau \boldsymbol{u}^n, \ \boldsymbol{u}^n \in \partial\varphi(\boldsymbol{x}^{n+1}) \ \Leftrightarrow \ \boldsymbol{x}^{n+1} = \mathrm{prox}_{\tau\varphi}(\boldsymbol{x}^n)$$

- splitting proximal algorithm

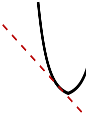
$$\boldsymbol{y}^{n+1} = \mathrm{prox}_{\sigma(\lambda\mathcal{Q})^*}\left(\boldsymbol{y}^n + \sigma \mathbf{D}\bar{\boldsymbol{x}}^n\right)$$

$$\boldsymbol{x}^{n+1} = \mathrm{prox}_{\tau\|\boldsymbol{\mathcal{L}} - \boldsymbol{\Phi}\cdot\|_2^2}\left(\boldsymbol{x}^n - \tau \mathbf{D}^\top \boldsymbol{y}^{n+1}\right), \qquad \boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$$

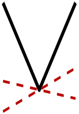
$$\bar{\boldsymbol{x}}^{n+1} = 2\boldsymbol{x}^{n+1} - \boldsymbol{x}^n$$

Functionals minimization

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



nonsmooth



- gradient descent $\boldsymbol{x}^{n+1} = \boldsymbol{x}^n - \tau \nabla \varphi(\boldsymbol{x}^n)$
- implicit subgradient descent: proximal point algorithm

$$\boldsymbol{x}^{n+1} = \boldsymbol{x}^n - \tau \boldsymbol{u}^n, \;\; \boldsymbol{u}^n \in \partial\varphi(\boldsymbol{x}^{n+1}) \;\Leftrightarrow\; \boldsymbol{x}^{n+1} = \mathrm{prox}_{\tau\varphi}(\boldsymbol{x}^n)$$

- splitting proximal algorithm $\quad \mathrm{prox}_{\tau\varphi}(\boldsymbol{x}) = \underset{\boldsymbol{u}}{\mathrm{argmin}} \, \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{u}\|^2 + \tau\varphi(\boldsymbol{u})$

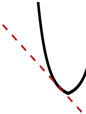
$$\boldsymbol{y}^{n+1} = \mathrm{prox}_{\sigma(\lambda\mathcal{Q})^*}\left(\boldsymbol{y}^n + \sigma\mathbf{D}\bar{\boldsymbol{x}}^n\right)$$

$$\boldsymbol{x}^{n+1} = \mathrm{prox}_{\tau\|\boldsymbol{\mathcal{L}}-\boldsymbol{\Phi}\cdot\|_2^2}\left(\boldsymbol{x}^n - \tau\mathbf{D}^\top\boldsymbol{y}^{n+1}\right), \qquad \boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$$

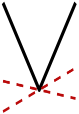
$$\bar{\boldsymbol{x}}^{n+1} = 2\boldsymbol{x}^{n+1} - \boldsymbol{x}^n$$

Computation of proximal operators

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

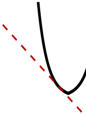


nonsmooth

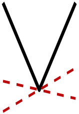


Computation of proximal operators

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \, \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



nonsmooth

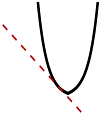


Ex. **Mixed norm:** for $\boldsymbol{z} = [\boldsymbol{z}_1; \ldots, ; \boldsymbol{z}_I]$

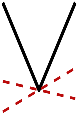
$$\mathcal{Q}(\boldsymbol{z}) = \|\boldsymbol{z}\|_{2,1} = \sum_{\underline{n}\in\Omega} \sqrt{\sum_{i=1}^{I} z_i^2(\underline{n})} = \sum_{\underline{n}\in\Omega} \|\boldsymbol{z}(\underline{n})\|_2$$

Computation of proximal operators

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



nonsmooth



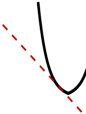
Ex. Mixed norm: for $\boldsymbol{z} = [\boldsymbol{z}_1; \dots; \boldsymbol{z}_I]$

$$\mathcal{Q}(\boldsymbol{z}) = \|\boldsymbol{z}\|_{2,1} = \sum_{\underline{n}\in\Omega} \sqrt{\sum_{i=1}^{I} z_i^2(\underline{n})} = \sum_{\underline{n}\in\Omega} \|\boldsymbol{z}(\underline{n})\|_2$$

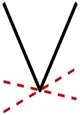
$$\boldsymbol{p} = \mathrm{prox}_{\lambda\|\cdot\|_{2,1}}(\boldsymbol{z}) \quad \Leftrightarrow \quad p_i(\underline{n}) = \max\left(0, 1 - \frac{\lambda}{\|\boldsymbol{z}(\underline{n})\|_2}\right) z_i(\underline{n})$$

Computation of proximal operators

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



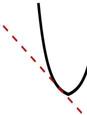
nonsmooth



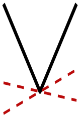
Least-Squares: $\|\log \boldsymbol{\mathcal{L}} - \boldsymbol{\Phi}(\boldsymbol{h}, \boldsymbol{v})\|^2, \quad \boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$

Computation of proximal operators

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \sum_a \underbrace{\|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \, \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



nonsmooth



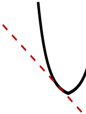
Least-Squares: $\|\log \mathcal{L} - \Phi(\boldsymbol{h}, \boldsymbol{v})\|^2, \quad \Phi : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$

Proposition *(Pascal, 2019)*

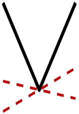
$$(\widetilde{\boldsymbol{h}}, \widetilde{\boldsymbol{v}}) = \mathrm{prox}_{\tau\|\mathcal{L} - \Phi\cdot\|^2}(\boldsymbol{h}, \boldsymbol{v}) \Longleftrightarrow (\widetilde{\boldsymbol{h}}, \widetilde{\boldsymbol{v}}) = \left(\mathbf{I} + \tau\Phi^\top\Phi\right)^{-1}\left((\boldsymbol{h}, \boldsymbol{v}) + \tau\Phi^\top \log \mathcal{L}\right)$$

Computation of proximal operators

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



nonsmooth



Least-Squares: $\|\log \boldsymbol{\mathcal{L}} - \boldsymbol{\Phi}(\boldsymbol{h}, \boldsymbol{v})\|^2, \quad \boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$

Proposition *(Pascal, 2019)*

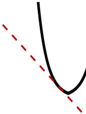
Let $S_m = \sum_a \log^m(a)$, $\mathcal{D} = (1 + \tau S_2)(1 + \tau S_0) - \tau^2 S_1^2$,
$\boldsymbol{\mathcal{T}} = \sum_a \log \boldsymbol{\mathcal{L}}_a$ and $\boldsymbol{\mathcal{G}} = \sum_a \log(a) \log \boldsymbol{\mathcal{L}}_a$, alors

$$(\widetilde{\boldsymbol{h}}, \widetilde{\boldsymbol{v}}) = \text{prox}_{\tau\|\boldsymbol{\mathcal{L}} - \boldsymbol{\Phi}\cdot\|^2}(\boldsymbol{h}, \boldsymbol{v}) \iff (\widetilde{\boldsymbol{h}}, \widetilde{\boldsymbol{v}}) = \left(\mathbf{I} + \tau \boldsymbol{\Phi}^\top \boldsymbol{\Phi}\right)^{-1} \left((\boldsymbol{h}, \boldsymbol{v}) + \tau \boldsymbol{\Phi}^\top \log \boldsymbol{\mathcal{L}}\right)$$

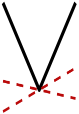
$$\iff \begin{cases} \widetilde{\boldsymbol{h}} = \mathcal{D}^{-1}\left((1 + \tau S_0)(\tau \boldsymbol{\mathcal{G}} + \boldsymbol{h}) - \tau S_1(\tau \boldsymbol{\mathcal{T}} + \boldsymbol{v})\right) \\ \widetilde{\boldsymbol{v}} = \mathcal{D}^{-1}\left((1 + \tau S_2)(\tau \boldsymbol{\mathcal{T}} + \boldsymbol{v}) - \tau S_1(\tau \boldsymbol{\mathcal{G}} + \boldsymbol{h})\right) \end{cases}$$

Accelerated algorithm based on strong-convexity

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \;\; \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \;+\; \lambda \, \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



nonsmooth



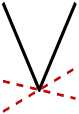
Primal-dual algorithm *(Chambolle, 2011)*

$$\delta\text{: duality gap, } \delta(\boldsymbol{x}^n, \boldsymbol{y}^n) \underset{n\to\infty}{\longrightarrow} 0$$

Accelerated algorithm based on strong-convexity

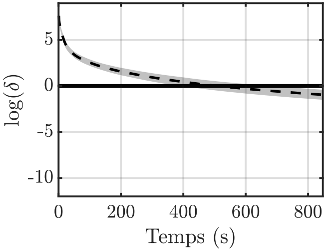
$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

nonsmooth



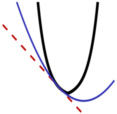
Primal-dual algorithm *(Chambolle, 2011)*

δ: duality gap, $\delta(\boldsymbol{x}^n, \boldsymbol{y}^n) \underset{n\to\infty}{\longrightarrow} 0$



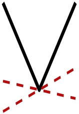
Convexity properties

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \, \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



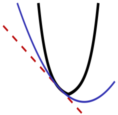
μ-strongly convex

nonsmooth



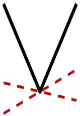
Convexity properties

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



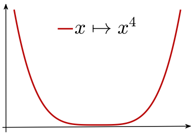
$\boldsymbol{\mu}$-strongly convex

nonsmooth

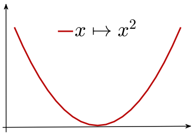


Strong-convexity

- φ $\boldsymbol{\mu}$-strongly convex iff $\varphi - \dfrac{\boldsymbol{\mu}}{2}\|\cdot\|^2$ convex



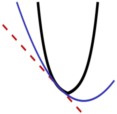
✓ strictly convex
✗ non strongly convex



✓ strictly convex
✓ 1-strongly convex

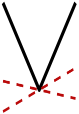
Convexity properties

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



μ-strongly convex

nonsmooth

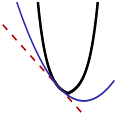


Strong-convexity

- φ μ-strongly convex iff $\varphi - \frac{\mu}{2}\|\cdot\|^2$ convex
- φ $\mathcal{C}^2$ with Hessian matrix $\boldsymbol{H}\varphi \succeq 0 \implies \mu = \min \mathrm{Sp}(\boldsymbol{H}\varphi)$

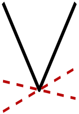
Convexity properties

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



μ-strongly convex

nonsmooth



Strong-convexity

- φ μ-strongly convex iff $\varphi - \dfrac{\mu}{2}\|\cdot\|^2$ convex
- φ $\mathcal{C}^2$ with Hessian matrix $\boldsymbol{H}\varphi \succeq 0 \implies \mu = \min \mathrm{Sp}(\boldsymbol{H}\varphi)$

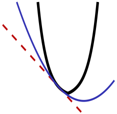
Proposition *(Pascal, 2019)*

$\sum_a \|\log \boldsymbol{\mathcal{L}} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2$ est μ-strongly convex.

$a_{\min} = 2^1$, $a_{\max}$	2^2	2^3	2^4	2^5	2^6
$\mu = \min \mathrm{Sp}\left(2\boldsymbol{\Phi}^\top \boldsymbol{\Phi}\right)$	0.29	**0.72**	1.20	1.69	2.20

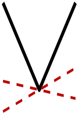
Accelerated algorithm based on strong-convexity

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



μ-strongly convex

nonsmooth



Accelerated Primal-dual algorithm *(Chambolle, 2011)*

for $n = 0, 1, \dots$ $\qquad\qquad \boldsymbol{x} = (\boldsymbol{h}, \boldsymbol{v})$

$$\boldsymbol{y}^{n+1} = \operatorname{prox}_{\sigma_n(\lambda\mathcal{Q})^*}\left(\boldsymbol{y}^n + \sigma_n \mathbf{D}\bar{\boldsymbol{x}}^n\right)$$

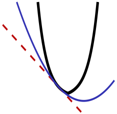
$$\boldsymbol{x}^{n+1} = \operatorname{prox}_{\tau_n\|\boldsymbol{\mathcal{L}} - \boldsymbol{\Phi}\cdot\|_2^2}\left(\boldsymbol{x}^n - \tau_n \mathbf{D}^\top \boldsymbol{y}^{n+1}\right)$$

$$\theta_n = \sqrt{1 + 2\mu\tau_n}, \quad \tau_{n+1} = \tau_n/\theta_n, \quad \sigma_{n+1} = \theta_n\sigma_n$$

$$\bar{\boldsymbol{x}}^{n+1} = \boldsymbol{x}^{n+1} + \theta_n^{-1}\left(\boldsymbol{x}^{n+1} - \boldsymbol{x}^n\right)$$

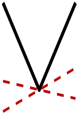
Algorithme accéléré par forte-convexité

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$



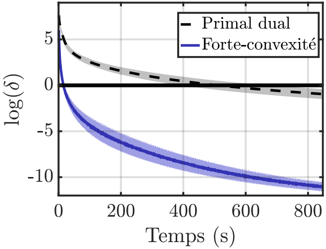
μ-strongly convex

nonsmooth



Accelerated Primal-dual algorithm *(Chambolle, 2011)*

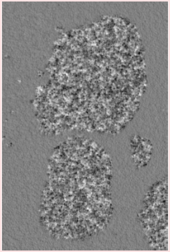
δ: duality gap, $\delta(\boldsymbol{x}^n, \boldsymbol{y}^n) \underset{n\to\infty}{\longrightarrow} 0$



Segmentation *via* iterated thresholding

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \mathcal{L}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \;+\; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

Textured image

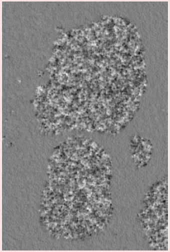


Segmentation *via* iterated thresholding

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \sum_a \underbrace{\|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

Textured image

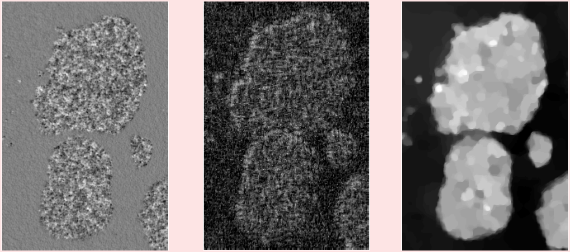
Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$



Segmentation *via* iterated thresholding

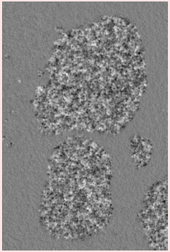
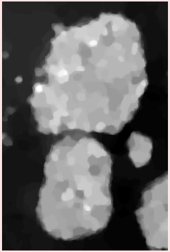
$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \ \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

Textured image | Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ | Co-localized contours $\widehat{\boldsymbol{h}}^{\mathrm{C}}$



Segmentation *via* iterated thresholding

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

Textured image	Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$	Co-localized contours $\widehat{\boldsymbol{h}}^{\mathrm{C}}$	Threshold estimate† $T\widehat{\boldsymbol{h}}^{\mathrm{C}}$
			

† (Cai, 2013)

State-of-the-art methods for texture segmentation

Threshold-ROF on $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

(Nafornita, 2014), (Pustelnik, 2016)

$$\underset{\boldsymbol{h}}{\operatorname{argmin}} \|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathrm{LR}}\|^2 + \lambda \|\mathbf{D}\boldsymbol{h}\|_{2,1}$$

Lin. reg.

ROF

Threshold

Only based on regularity $\boldsymbol{h}$.

State-of-the-art methods for texture segmentation

Threshold-ROF on $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

(Nafornita, 2014), (Pustelnik, 2016)

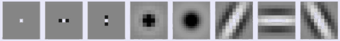
$$\underset{\boldsymbol{h}}{\operatorname{argmin}} \; \|\boldsymbol{h} - \widehat{\boldsymbol{h}}^{\mathrm{LR}}\|^2 + \lambda \|\mathbf{D}\boldsymbol{h}\|_{2,1}$$

Lin. reg. | ROF | Threshold

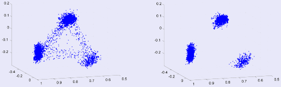
Only based on regularity $\boldsymbol{h}$.

Factorization based segmentation[†] *(Yuan, 2015)*

(i) local histograms



(ii) matrix factorization



(a) (b)

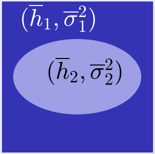
Fig. 2. Scatterplot of features in subspace. (a) Scatterplot of features projected onto the 3-d subspace. (b) Scatterplot after removing features with high edgeness.

[†] `https://sites.google.com/site/factorizationsegmentation/`

Compared segmentation performance on synthetic textures

Piecewise monofractal texture synthesis *(Pascal, 2019)*

- mask: $\Omega = \Omega_1 \sqcup \Omega_2$,
- attributes: $(\overline{h}_k, \overline{\sigma}_k^2)_{k=1,2}$

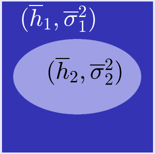


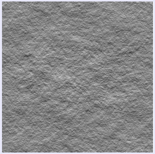
Compared segmentation performance on synthetic textures

Piecewise monofractal texture synthesis *(Pascal, 2019)*

- mask: $\Omega = \Omega_1 \sqcup \Omega_2$,
- attributes: $(\overline{h}_k, \overline{\sigma}_k^2)_{k=1,2}$

Ex. $\overline{h}_1 = 0.5$, $\overline{\sigma}_1^2 = 0.6$
$\overline{h}_2 = 0.6$, $\overline{\sigma}_2^2 = 0.7$



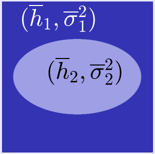


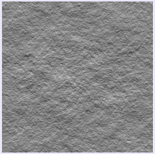
Compared segmentation performance on synthetic textures

Piecewise monofractal texture synthesis *(Pascal, 2019)*

- mask: $\Omega = \Omega_1 \sqcup \Omega_2$,
- attributes: $(\bar{h}_k, \bar{\sigma}_k^2)_{k=1,2}$

Ex. $\bar{h}_1 = 0.5$, $\bar{\sigma}_1^2 = 0.6$
$\bar{h}_2 = 0.6$, $\bar{\sigma}_2^2 = 0.7$





Averaged segmentation performances over 5 realizations

Yuan

$71.1 \pm 1.3\%$

T-ROF

$78.5 \pm 1.1\%$

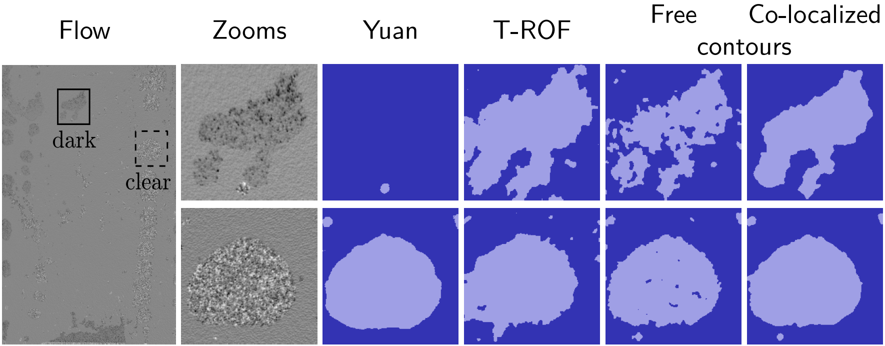
free contours

$90.2 \pm 1.9\%$

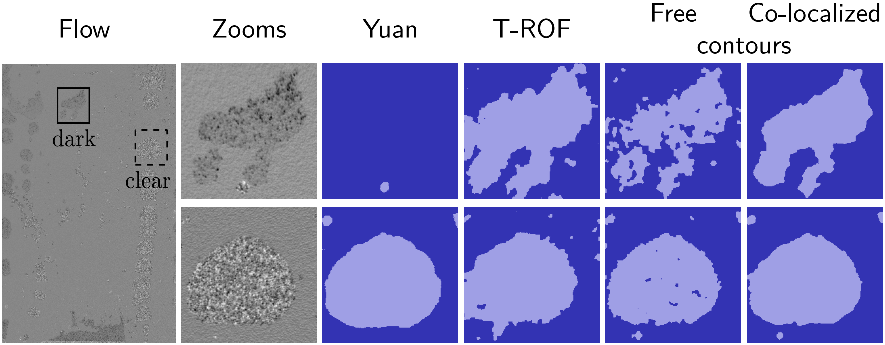
co-localized contours

$91.1 \pm 1.5\%$

Low activity: $Q_{\mathrm{G}} = 300\mathrm{mL}/\mathrm{min}$ - $Q_{\mathrm{L}} = 300\mathrm{mL}/\mathrm{min}$



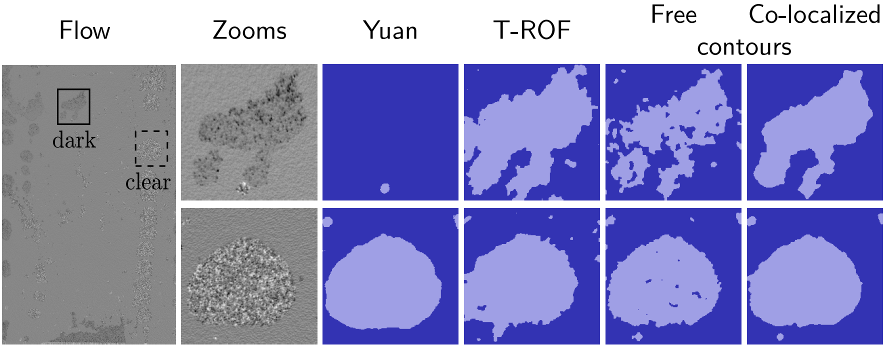
Low activity: $Q_{\mathrm{G}} = 300\mathrm{mL/min}$ - $Q_{\mathrm{L}} = 300\mathrm{mL/min}$



Liquid: $h_{\mathrm{L}} = 0.4$

Gas: $h_{\mathrm{G}} = 0.9$

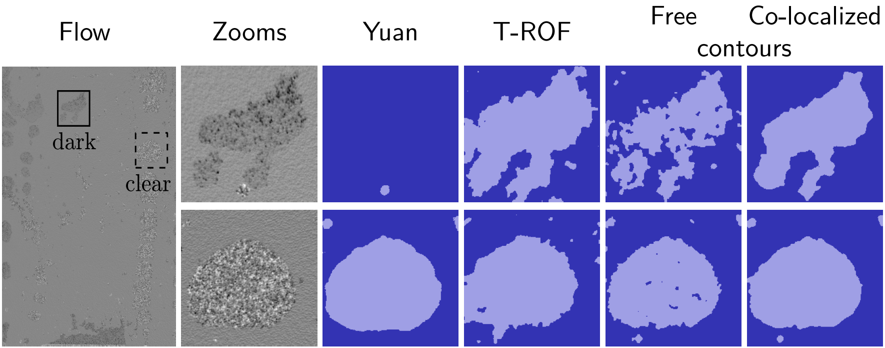
Low activity: $Q_{\mathrm{G}} = 300$mL/min - $Q_{\mathrm{L}} = 300$mL/min



Liquid: $h_{\mathrm{L}} = 0.4$ $\sigma^2_{\mathrm{dark}} = 10^{-2}$

Gas: $h_{\mathrm{G}} = 0.9$

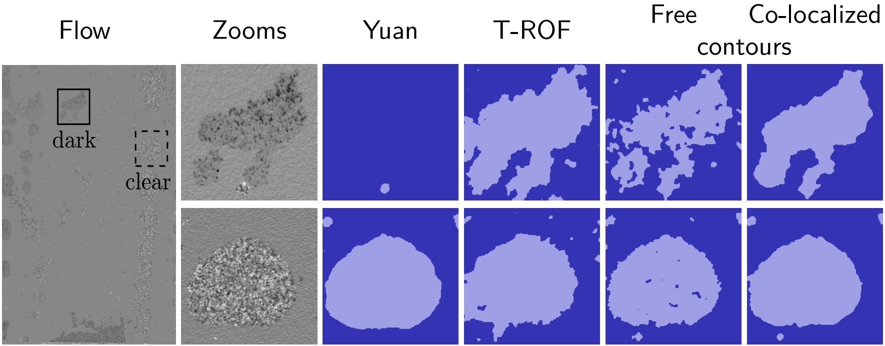
Low activity: $Q_{\mathrm{G}} = 300\text{mL}/\text{min}$ - $Q_{\mathrm{L}} = 300\text{mL}/\text{min}$



Liquid: $h_{\mathrm{L}} = 0.4$ $\sigma^2_{\mathrm{dark}} = 10^{-2}$

Gas: $h_{\mathrm{G}} = 0.9$

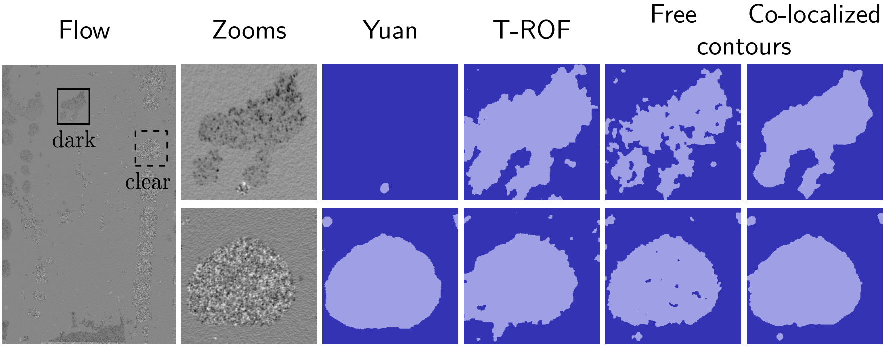
Low activity: $Q_{\mathrm{G}} = 300\mathrm{mL/min}$ - $Q_{\mathrm{L}} = 300\mathrm{mL/min}$



Liquid: $h_{\mathrm{L}} = 0.4$ $\quad \sigma^2_{\mathrm{dark}} = 10^{-2}$

Gas: $h_{\mathrm{G}} = 0.9$ $\quad \sigma^2_{\mathrm{dark}} = 10^{-2}$ (dark bubbles)

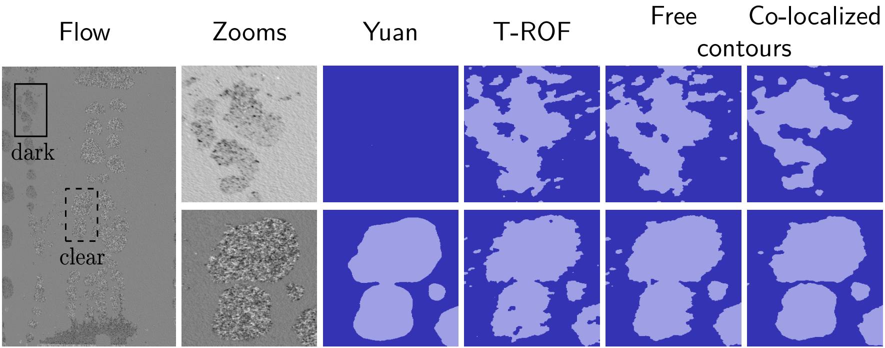
Low activity: $Q_{\mathrm{G}} = 300\mathrm{mL/min}$ - $Q_{\mathrm{L}} = 300\mathrm{mL/min}$



Liquid: $h_{\mathrm{L}} = 0.4$ $\sigma^2_{\mathrm{dark}} = 10^{-2}$

Gas: $h_{\mathrm{G}} = 0.9$ | $\sigma^2_{\mathrm{dark}} = 10^{-2}$ (dark bubbles)
$\sigma^2_{\mathrm{clear}} = 10^{-1}$ (clear bubbles)

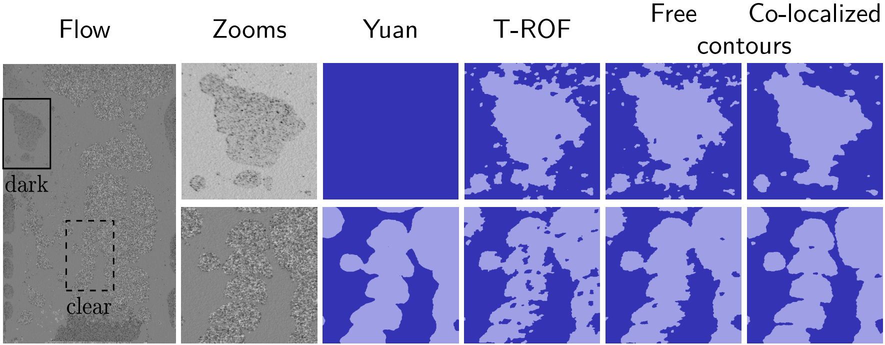
Transition: $Q_{\mathrm{G}} = 400\mathrm{mL}/\mathrm{min}$ - $Q_{\mathrm{L}} = 700\mathrm{mL}/\mathrm{min}$



Liquid: $h_{\mathrm{L}} = 0.4$ $\sigma^2_{\mathrm{dark}} = 10^{-2}$

Gas: $h_{\mathrm{G}} = 0.9$ $\left| \begin{array}{ll} \sigma^2_{\mathrm{dark}} = 10^{-2} & \text{(dark bubbles)} \\ \sigma^2_{\mathrm{clear}} = 10^{-1} & \text{(clear bubbles).} \end{array} \right.$

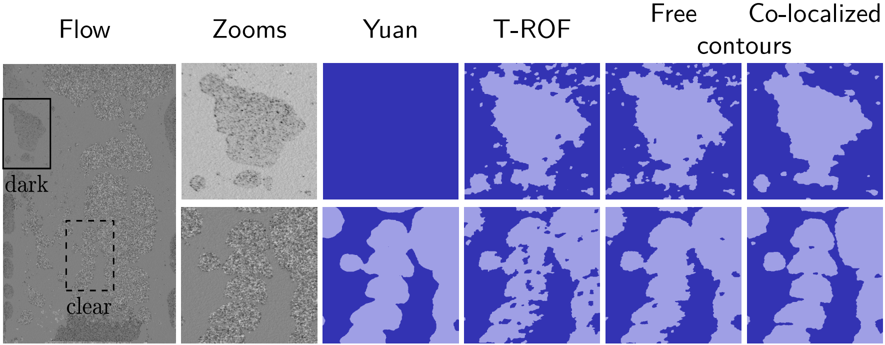
High activity: $Q_{\mathrm{G}} = 1200\mathrm{mL/min}$ - $Q_{\mathrm{L}} = 300\mathrm{mL/min}$



Liquid: $h_{\mathrm{L}} = 0.4 \qquad \sigma^2_{\mathrm{dark}} = 10^{-2}$

Gas: $h_{\mathrm{G}} = 0.9$ $\left| \begin{array}{ll} \sigma^2_{\mathrm{dark}} = 10^{-2} & \text{(dark bubbles)} \\ \sigma^2_{\mathrm{clear}} = 10^{-1} & \text{(clear bubbles).} \end{array} \right.$

High activity: $Q_{\mathrm{G}} = 1200\text{mL}/\text{min}$ - $Q_{\mathrm{L}} = 300\text{mL}/\text{min}$



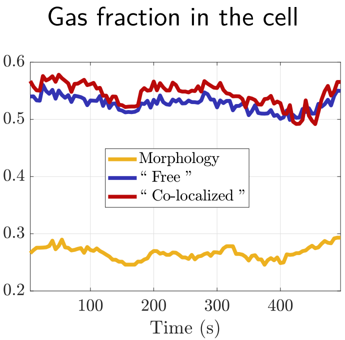
Computational load	1s	12s	700s	**2100s**

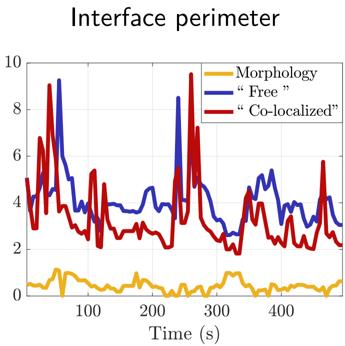
Liquid: $h_{\mathrm{L}} = 0.4$ $\quad \sigma^2_{\mathrm{dark}} = 10^{-2}$

Gas: $h_{\mathrm{G}} = 0.9$ $\quad \left| \begin{array}{ll} \sigma^2_{\mathrm{dark}} = 10^{-2} & \text{(dark bubbles)} \\ \sigma^2_{\mathrm{clear}} = 10^{-1} & \text{(clear bubbles).} \end{array} \right.$

Multiphase flow through porous media

Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)





Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

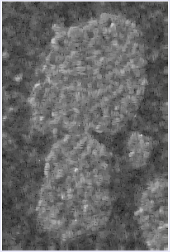
$(\lambda, \alpha) = (0, 0)$

Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ Co-localized contours estimate $\widehat{\boldsymbol{h}}^{\mathrm{C}}$

$(\lambda, \alpha) = (0, 0)$ $(\lambda, \alpha) = (0.5, 0.5)$



too small

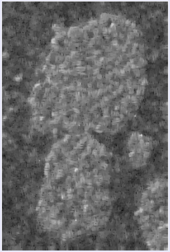
Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ Co-localized contours estimate $\widehat{\boldsymbol{h}}^{\mathrm{C}}$

$(\lambda, \alpha) = (0, 0)$

$(\lambda, \alpha) = (0.5, 0.5)$



too small

$(\lambda, \alpha) = (500, 500)$

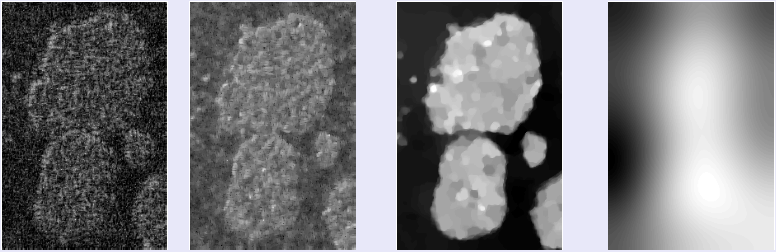
too large

Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ | Co-localized contours estimate $\widehat{\boldsymbol{h}}^{\mathrm{C}}$

$(\lambda, \alpha) = (0, 0)$ | $(\lambda, \alpha) = (0.5, 0.5)$ | $(\lambda^{\dagger}, \alpha^{\dagger}) = (11.5, 0.8)$ | $(\lambda, \alpha) = (500, 500)$



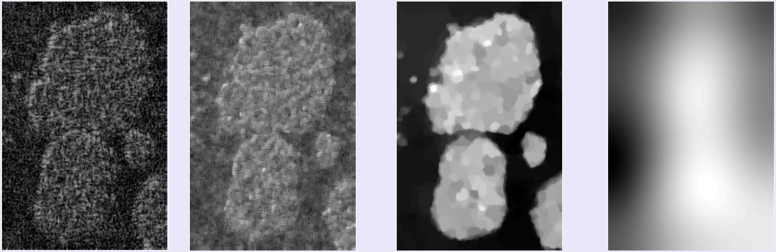
too small | optimal | too large

Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\mathcal{L}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ — Co-localized contours estimate $\widehat{\boldsymbol{h}}^{\mathrm{C}}$

$(\lambda, \alpha) = (0, 0)$ — $(\lambda, \alpha) = (0.5, 0.5)$ — $(\lambda^\dagger, \alpha^\dagger) = (11.5, 0.8)$ — $(\lambda, \alpha) = (500, 500)$



too small — optimal — too large

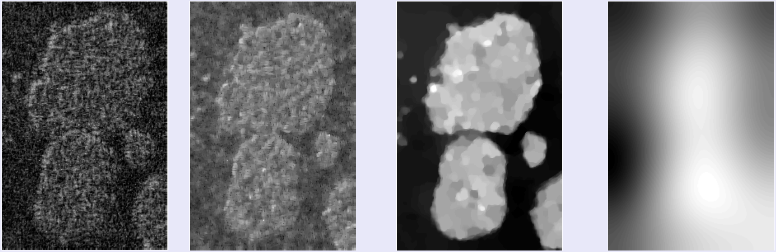
What *optimal* means?

Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\mathcal{L}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ | Co-localized contours estimate $\widehat{\boldsymbol{h}}^{\mathrm{C}}$

$(\lambda, \alpha) = (0, 0)$ | $(\lambda, \alpha) = (0.5, 0.5)$ | $(\lambda^\dagger, \alpha^\dagger) = (11.5, 0.8)$ | $(\lambda, \alpha) = (500, 500)$



too small | optimal | too large

What *optimal* means? How to determine $\lambda^\dagger$ and $\alpha^\dagger$?

Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$

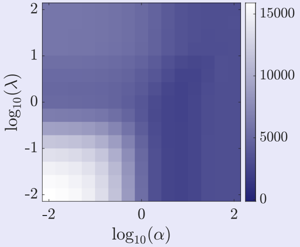
Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



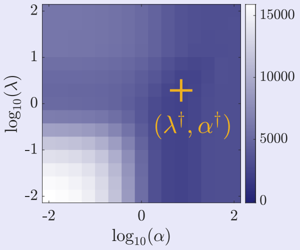
Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



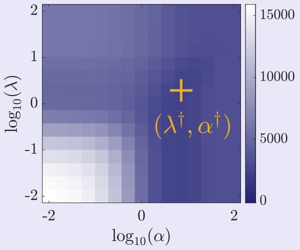
Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

?

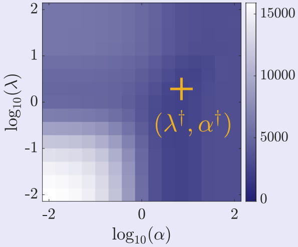
Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\| \widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}} \right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

?

Stein Unbiased Risk Estimate
(SURE)

Stein Unbiased Risk Estimate (Principe)

Observations $\boldsymbol{y} = \bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}}$: truth and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \rho^2 \mathbf{I})$

Stein Unbiased Risk Estimate (Principe)

Observations $\boldsymbol{y} = \bar{\boldsymbol{x}} + \zeta \in \mathbb{R}^P$, $\bar{\boldsymbol{x}}$: truth and $\zeta \sim \mathcal{N}(\mathbf{0}, \rho^2 \mathbf{I})$

Parametric estimator $(\boldsymbol{y}; \lambda) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)$

Ex.
$$\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) = \begin{cases} \left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \boldsymbol{y} & \text{(linear)} \\ \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{x}) & \text{(nonlinear)} \end{cases}$$

Stein Unbiased Risk Estimate (Principe)

Observations $\boldsymbol{y} = \bar{\boldsymbol{x}} + \zeta \in \mathbb{R}^P$, $\bar{\boldsymbol{x}}$: truth and $\zeta \sim \mathcal{N}(\mathbf{0}, \rho^2 \mathbf{I})$

Parametric estimator $(\boldsymbol{y}; \lambda) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)$

Ex. $$\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) = \begin{cases} \left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \boldsymbol{y} & \text{(linear)} \\ \operatorname*{argmin}_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{x}) & \text{(nonlinear)} \end{cases}$$

Quadratic error $R(\lambda) \triangleq \mathbb{E}_\zeta \|\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) - \bar{\boldsymbol{x}}\|^2 \overset{?}{=} \mathbb{E}_\zeta \widehat{R}(\boldsymbol{y}; \lambda)$ $\bar{\boldsymbol{x}}$ unknown

Stein Unbiased Risk Estimate (Principe)

Observations $\boldsymbol{y} = \bar{\boldsymbol{x}} + \zeta \in \mathbb{R}^P$, $\bar{\boldsymbol{x}}$: truth and $\zeta \sim \mathcal{N}(\mathbf{0}, \rho^2 \mathbf{I})$

Parametric estimator $(\boldsymbol{y}; \lambda) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)$

Ex. $$\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) = \begin{cases} \left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \boldsymbol{y} & \text{(linear)} \\ \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{x}) & \text{(nonlinear)} \end{cases}$$

Quadratic error $R(\lambda) \triangleq \mathbb{E}_\zeta \|\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) - \bar{\boldsymbol{x}}\|^2 \overset{?}{=} \mathbb{E}_\zeta \widehat{R}(\boldsymbol{y}; \lambda)$ $\bar{\boldsymbol{x}}$ unknown

Theorem *(Stein, 1981)*

Let $(\boldsymbol{y}; \lambda) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)$ an estimator of $\bar{\boldsymbol{x}}$

- weakly differentiable w.r.t. $\boldsymbol{y}$,
- such that $\zeta \mapsto \langle \widehat{\boldsymbol{x}}(\bar{\boldsymbol{x}} + \zeta; \lambda), \zeta \rangle$ is integrable w.r.t. $\mathcal{N}(\mathbf{0}, \rho^2 \mathbf{I})$.

$$\widehat{R}(\boldsymbol{y}; \lambda) \triangleq \|\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) - \boldsymbol{y}\|^2 + 2\rho^2 \operatorname{tr}\left(\partial_{\boldsymbol{y}} \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)\right) - \rho^2 P$$
$$\Longrightarrow R(\lambda) = \mathbb{E}_\zeta[\widehat{R}(\boldsymbol{y}; \lambda)].$$

Generalized *Stein Unbiased Risk Estimate*

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

Generalized *Stein Unbiased Risk Estimate*

Observations $\boldsymbol{y} = \Phi \bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\Phi : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \mathcal{S})$

E.g. the estimators $\widehat{\boldsymbol{h}}(\mathcal{L}; \lambda, \alpha)$ with free or co-localized contours

$\log \mathcal{L} = \Phi(\bar{\boldsymbol{h}}, \bar{\boldsymbol{v}}) + \zeta$

$\Phi : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$

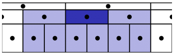
Generalized *Stein Unbiased Risk Estimate*

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \mathcal{S})$

E.g. the estimators $\widehat{\boldsymbol{h}}(\mathcal{L}; \lambda, \alpha)$ with free or co-localized contours

$\log \boldsymbol{\mathcal{L}} = \boldsymbol{\Phi}(\bar{\boldsymbol{h}}, \bar{\boldsymbol{v}}) + \boldsymbol{\zeta}$ $\quad\quad$ $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \mathcal{S})$

$\boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$



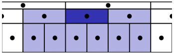
Generalized *Stein Unbiased Risk Estimate*

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

E.g. the estimators $\widehat{\boldsymbol{h}}(\mathcal{L}; \lambda, \alpha)$ with free or co-localized contours

$\log \boldsymbol{\mathcal{L}} = \boldsymbol{\Phi}(\bar{\boldsymbol{h}}, \bar{\boldsymbol{v}}) + \boldsymbol{\zeta}$ $\quad$ $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$ $\quad$ $\mathcal{R} = \|\widehat{\boldsymbol{h}} - \bar{\boldsymbol{h}}\|^2$

$\boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$



$\boldsymbol{\Pi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto (\boldsymbol{h}, \mathbf{0})$

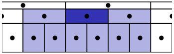
Generalized *Stein Unbiased Risk Estimate*

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

E.g. the estimators $\widehat{\boldsymbol{h}}(\mathcal{L}; \lambda, \alpha)$ with free or co-localized contours

$\log \boldsymbol{\mathcal{L}} = \boldsymbol{\Phi}(\bar{\boldsymbol{h}}, \bar{\boldsymbol{v}}) + \boldsymbol{\zeta}$ $\qquad$ $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$ $\qquad$ $\mathcal{R} = \|\widehat{\boldsymbol{h}} - \bar{\boldsymbol{h}}\|^2$

$\boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$



$\boldsymbol{\Pi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto (\boldsymbol{h}, \mathbf{0})$

Projected estimation error $\quad R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) \triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

$R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) \triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P, \quad \bar{\boldsymbol{x}} \in \mathbb{R}^N,\ \boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\mathcal{S}})$

$$
\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\| \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}}) \right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top
\end{aligned}
$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

$$\begin{aligned} R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\ &= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\ &= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \end{aligned}$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathcal{S})$

$$
\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2
\end{aligned}
$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathcal{S})$

$$\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_\zeta \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_\zeta \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_\zeta \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_\zeta \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_\zeta \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right]
\end{aligned}$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \mathcal{S})$

$$\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right]
\end{aligned}$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

$$
\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A}(\boldsymbol{\Phi}\bar{\boldsymbol{x}}+\boldsymbol{\zeta}), \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right]
\end{aligned}
$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathcal{S})$

$$
\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A} \qquad \boldsymbol{\zeta}\ , \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right]
\end{aligned}
$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P, \quad \bar{\boldsymbol{x}} \in \mathbb{R}^N, \ \boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\mathcal{S}})$

$$\mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top \boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top$$

$$\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A} \quad \boldsymbol{\zeta}, \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\mathbb{E}_{\boldsymbol{\zeta}}\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - \mathbb{E}_{\boldsymbol{\zeta}}\|\mathbf{A}\boldsymbol{\zeta}\|^2
\end{aligned}$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\mathcal{S}})$

$$
\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A}\qquad\boldsymbol{\zeta}\,, \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\mathbb{E}_{\boldsymbol{\zeta}}\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - \operatorname{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)
\end{aligned}
$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P, \quad \bar{\boldsymbol{x}} \in \mathbb{R}^N, \; \boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

$$\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - 2\left\langle \mathbf{A} \qquad \boldsymbol{\zeta}\;, \mathbf{A}\boldsymbol{\zeta}\right\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\mathbb{E}_{\boldsymbol{\zeta}} \left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\right\rangle - \underbrace{\operatorname{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)}_{\text{accessible}}
\end{aligned}$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \zeta \in \mathbb{R}^P, \quad \bar{\boldsymbol{x}} \in \mathbb{R}^N, \ \boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\zeta \sim \mathcal{N}(\mathbf{0}, \mathcal{S})$

$$\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_\zeta \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_\zeta \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_\zeta \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_\zeta \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \zeta)\|^2 \\
&= \mathbb{E}_\zeta \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\zeta\right\rangle + \|\mathbf{A}\zeta\|^2\right] \\
&= \mathbb{E}_\zeta \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\zeta\right\rangle - 2\left\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\zeta\right\rangle + \|\mathbf{A}\zeta\|^2\right] \\
&= \mathbb{E}_\zeta \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\zeta\right\rangle - 2\left\langle \mathbf{A} \qquad \zeta\ , \mathbf{A}\zeta\right\rangle + \|\mathbf{A}\zeta\|^2\right] \\
&= \mathbb{E}_\zeta \underbrace{\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2}_{\text{accessible}} + 2\mathbb{E}_\zeta \left\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\zeta\right\rangle - \underbrace{\operatorname{tr}\left(\mathbf{A}\mathcal{S}\mathbf{A}^\top\right)}_{\text{accessible}}
\end{aligned}$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

$$
\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - 2\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - 2\langle \mathbf{A} \qquad \boldsymbol{\zeta}\ , \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \underbrace{\mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2}_{\text{accessible}} + 2\mathbb{E}_{\boldsymbol{\zeta}} \langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - \underbrace{\operatorname{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)}_{\text{accessible}}
\end{aligned}
$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P, \quad \bar{\boldsymbol{x}} \in \mathbb{R}^N, \ \boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\mathcal{S}})$

$$
\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_\zeta \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_\zeta \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_\zeta \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_\zeta \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_\zeta \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_\zeta \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - 2\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_\zeta \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - 2\langle \mathbf{A} \qquad \boldsymbol{\zeta}\ , \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \underbrace{\mathbb{E}_\zeta \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2}_{\text{accessible}} + 2\mathbb{E}_\zeta \langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - \underbrace{\operatorname{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)}_{\text{accessible}}
\end{aligned}
$$

$$
\mathbb{E}_\zeta \langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle = \mathbb{E}_\zeta \langle \boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle
$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\mathcal{S}})$

$$
\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\zeta}\|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\zeta}\left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_{\zeta}\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\zeta}\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_{\zeta}\left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\zeta}\left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle\mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - 2\langle\mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\zeta}\left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle\mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - 2\langle\mathbf{A}\qquad\boldsymbol{\zeta}, \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \underbrace{\mathbb{E}_{\zeta}\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2}_{\text{accessible}} + 2\mathbb{E}_{\zeta}\langle\mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - \underbrace{\operatorname{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)}_{\text{accessible}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_{\zeta}\langle\mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle &= \mathbb{E}_{\zeta}\langle\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle \\
&= \int \langle\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta};\lambda), \mathbf{A}\boldsymbol{\zeta}\rangle \exp\left(-\frac{\boldsymbol{\zeta}^\top\boldsymbol{\mathcal{S}}^{-1}\boldsymbol{\zeta}}{2}\right)\mathrm{d}\boldsymbol{\zeta}
\end{aligned}
$$

Generalized *Stein Unbiased Risk Estimate* (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

$$\begin{aligned}
R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) &\triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left\|\boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{\Phi}(\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \bar{\boldsymbol{x}})\right\|^2 \qquad \mathbf{A} \triangleq \boldsymbol{\Pi}(\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{\Phi}\bar{\boldsymbol{x}})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y} + \boldsymbol{\zeta})\|^2 \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}), \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - 2\langle \mathbf{A}\boldsymbol{y}, \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \mathbb{E}_{\boldsymbol{\zeta}} \left[\|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - 2\langle \mathbf{A}\quad\boldsymbol{\zeta}, \mathbf{A}\boldsymbol{\zeta}\rangle + \|\mathbf{A}\boldsymbol{\zeta}\|^2\right] \\
&= \underbrace{\mathbb{E}_{\boldsymbol{\zeta}} \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y})\|^2}_{\text{accessible}} + 2\mathbb{E}_{\boldsymbol{\zeta}}\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle - \underbrace{\mathrm{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)}_{\text{accessible}}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\zeta}}\langle \mathbf{A}\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle &= \mathbb{E}_{\boldsymbol{\zeta}}\langle \boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}), \mathbf{A}\boldsymbol{\zeta}\rangle \\
&= \int \langle \boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta};\lambda), \mathbf{A}\boldsymbol{\zeta}\rangle \exp\left(-\frac{\boldsymbol{\zeta}^\top\boldsymbol{\mathcal{S}}^{-1}\boldsymbol{\zeta}}{2}\right) \mathrm{d}\boldsymbol{\zeta} \\
\text{(Gen. I.b.P.)} &= \mathbb{E}_{\boldsymbol{\zeta}}\mathrm{tr}\left(\boldsymbol{\mathcal{S}}\mathbf{A}^\top\boldsymbol{\Pi}\partial_{\boldsymbol{y}}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right)
\end{aligned}$$

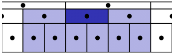
Generalized *Stein Unbiased Risk Estimate*

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

E.g. the estimators $\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha)$ with free or co-localized contours

$\log \boldsymbol{\mathcal{L}} = \boldsymbol{\Phi}(\bar{\boldsymbol{h}}, \bar{\boldsymbol{v}}) + \boldsymbol{\zeta}$ $\qquad \boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$ $\qquad \mathcal{R} = \|\widehat{\boldsymbol{h}} - \bar{\boldsymbol{h}}\|^2$

$\boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$



$\boldsymbol{\Pi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto (\boldsymbol{h}, \mathbf{0})$

Projected estimation error $\quad R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) \triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2$

Theorem *(Pascal, 2020)*

Let $(\boldsymbol{y}; \boldsymbol{\Lambda}) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda})$ an estimator of $\bar{\boldsymbol{x}}$

- weakly differentiable w.r.t. $\boldsymbol{y}$,
- such that $\boldsymbol{\zeta} \mapsto \langle \boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\bar{\boldsymbol{x}} + \boldsymbol{\zeta}; \lambda), \mathbf{A}\boldsymbol{\zeta}\rangle$ is integrable w.r.t. $\mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$.

$$\widehat{R}(\boldsymbol{\Lambda}) \triangleq \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\mathrm{tr}\left(\boldsymbol{\mathcal{S}}\mathbf{A}^\top \boldsymbol{\Pi}\partial_{\boldsymbol{y}}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda})\right) - \mathrm{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)$$

$$\Longrightarrow R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) = \mathbb{E}_{\boldsymbol{\zeta}}[\widehat{R}(\boldsymbol{\Lambda})].$$

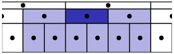
Generalized *Stein Unbiased Risk Estimate*

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \zeta \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\zeta \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

E.g. the estimators $\widehat{\boldsymbol{h}}(\mathcal{L}; \lambda, \alpha)$ with free or co-localized contours

$\log \boldsymbol{\mathcal{L}} = \boldsymbol{\Phi}(\bar{\boldsymbol{h}}, \bar{\boldsymbol{v}}) + \zeta$ $\qquad \zeta \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$ $\qquad \mathcal{R} = \|\widehat{\boldsymbol{h}} - \bar{\boldsymbol{h}}\|^2$

$\boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$



$\boldsymbol{\Pi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto (\boldsymbol{h}, \mathbf{0})$

Projected estimation error $\quad R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) \triangleq \mathbb{E}_\zeta \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2$

Theorem *(Pascal, 2020)*

Let $(\boldsymbol{y}; \boldsymbol{\Lambda}) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda})$ an estimator of $\bar{\boldsymbol{x}}$

- weakly differentiable w.r.t. $\boldsymbol{y}$,
- such that $\zeta \mapsto \langle \boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\bar{\boldsymbol{x}} + \zeta; \lambda), \mathbf{A}\zeta\rangle$ is integrable w.r.t. $\mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$.

$$\widehat{R}(\boldsymbol{\Lambda}) \triangleq \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\mathrm{tr}\left(\boldsymbol{\mathcal{S}}\mathbf{A}^\top \boldsymbol{\Pi}\partial_{\boldsymbol{y}}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda})\right) - \mathrm{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)$$

$$\Longrightarrow R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) = \mathbb{E}_\zeta[\widehat{R}(\boldsymbol{\Lambda})].$$

Computation of the degrees of freedom

Degrees of freedom $\mathrm{dof} \triangleq \mathrm{tr}\left(\boldsymbol{\mathcal{S}}\mathbf{A}^\top \boldsymbol{\Pi}\partial_{\boldsymbol{y}}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right)$

Computation of the degrees of freedom

Degrees of freedom $\quad \mathrm{dof} \triangleq \mathrm{tr}\left(\boldsymbol{\mathcal{S}}\mathbf{A}^\top \boldsymbol{\Pi} \partial_{\boldsymbol{y}} \widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda})\right)$

- **Monte Carlo strategy** (MC) **Large size matrix M** $\in \mathbb{R}^{P \times P}$

$$\mathrm{tr}(\mathbf{M}) = \mathbb{E}_{\boldsymbol{\varepsilon}} \langle \mathbf{M}\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \rangle, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_P)$$

Computation of the degrees of freedom

Degrees of freedom $\mathrm{dof} \triangleq \mathrm{tr}\left(\boldsymbol{\mathcal{S}}\mathbf{A}^\top \mathbf{\Pi} \partial_{\boldsymbol{y}} \widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda})\right)$

- **Monte Carlo strategy** (MC) **Large size matrix $\mathbf{M} \in \mathbb{R}^{P \times P}$**

$$\mathrm{tr}(\mathbf{M}) = \mathbb{E}_{\boldsymbol{\varepsilon}} \langle \mathbf{M}\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \rangle, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_P)$$

- **Finite Differences** (FD) **Inaccessible Jacobian matrix**

$$\partial_{\boldsymbol{y}} \widehat{\boldsymbol{x}}[\boldsymbol{\varepsilon}] \underset{\nu \to 0}{\simeq} \frac{1}{\nu} \left(\widehat{\boldsymbol{x}}(\boldsymbol{y} + \nu \boldsymbol{\varepsilon}; \boldsymbol{\Lambda}) - \widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) \right)$$

Computation of the degrees of freedom

Degrees of freedom $\mathrm{dof} \triangleq \mathrm{tr}\left(\boldsymbol{\mathcal{S}}\mathbf{A}^\top \boldsymbol{\Pi}\partial_{\boldsymbol{y}}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right)$

- **Monte Carlo strategy** (MC) — **Large size matrix $\mathbf{M} \in \mathbb{R}^{P\times P}$**

$$\mathrm{tr}(\mathbf{M}) = \mathbb{E}_{\boldsymbol{\varepsilon}}\left\langle \mathbf{M}\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}\right\rangle, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_P)$$

- **Finite Differences** (FD) — **Inaccessible Jacobian matrix**

$$\partial_{\boldsymbol{y}}\widehat{\boldsymbol{x}}\left[\boldsymbol{\varepsilon}\right] \underset{\nu\to 0}{\simeq} \frac{1}{\nu}\left(\widehat{\boldsymbol{x}}(\boldsymbol{y}+\nu\boldsymbol{\varepsilon};\boldsymbol{\Lambda}) - \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right)$$

Proposition *(Pascal, 2020)*

Let $(\boldsymbol{y};\boldsymbol{\Lambda}) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})$ an estimator of $\bar{\boldsymbol{x}}$

- uniformly Lipschitz continuous w.r.t. $\boldsymbol{y}$,
- such that $\forall \boldsymbol{\Lambda} \in \mathbb{R}^L$, $\widehat{\boldsymbol{x}}(\mathbf{0}_P;\boldsymbol{\Lambda}) = \mathbf{0}_N$. Then

$$\mathbb{E}_{\boldsymbol{\zeta}}\left[\mathrm{dof}\right] = \lim_{\nu\to 0}\mathbb{E}_{\boldsymbol{\zeta},\boldsymbol{\varepsilon}}\left[\frac{1}{\nu}\left\langle \boldsymbol{\mathcal{S}}\mathbf{A}^\top\boldsymbol{\Pi}\left(\widehat{\boldsymbol{x}}(\boldsymbol{y}+\nu\boldsymbol{\varepsilon};\boldsymbol{\Lambda}) - \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right), \boldsymbol{\varepsilon}\right\rangle\right]$$

Stein Unbiased Risk Estimate (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

Projected estimation error $R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) \triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2$

Generalized Finite Difference Monte Carlo SURE

$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y};\boldsymbol{\Lambda}\,|\,\boldsymbol{\mathcal{S}}) \triangleq \left\|\mathbf{A}\left(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}\right)\right\|^2 + \frac{2}{\nu}\left\langle \boldsymbol{\mathcal{S}}\mathbf{A}^\top\boldsymbol{\Pi}\left(\widehat{\boldsymbol{x}}(\boldsymbol{y}+\nu\boldsymbol{\varepsilon};\boldsymbol{\Lambda}) - \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right), \boldsymbol{\varepsilon}\right\rangle - \operatorname{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)$$

Stein Unbiased Risk Estimate (Computation)

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

Projected estimation error $R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) \triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2$

Generalized Finite Difference Monte Carlo SURE

$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y};\boldsymbol{\Lambda}\,|\,\boldsymbol{\mathcal{S}}) \triangleq \|\mathbf{A}\left(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}\right)\|^2 + \frac{2}{\nu}\left\langle \boldsymbol{\mathcal{S}}\mathbf{A}^\top\boldsymbol{\Pi}\left(\widehat{\boldsymbol{x}}(\boldsymbol{y}+\nu\boldsymbol{\varepsilon};\boldsymbol{\Lambda}) - \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right), \boldsymbol{\varepsilon}\right\rangle - \operatorname{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)$$

Theorem *(Pascal, 2020)*

Let $(\boldsymbol{y};\boldsymbol{\Lambda}) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})$ an estimator of $\bar{\boldsymbol{x}}$

- uniformly Lipschitz continuous w.r.t. $\boldsymbol{y}$,
- such that $\forall \boldsymbol{\Lambda} \in \mathbb{R}^L$, $\widehat{\boldsymbol{x}}(\mathbf{0}_P;\boldsymbol{\Lambda}) = \mathbf{0}_N$. Then

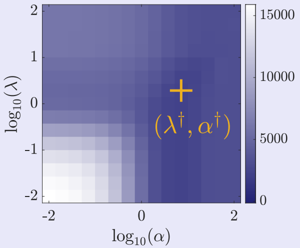
$$R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) = \lim_{\nu\to 0} \mathbb{E}_{\boldsymbol{\zeta},\boldsymbol{\varepsilon}}\left[\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y};\boldsymbol{\Lambda}\,|\,\boldsymbol{\mathcal{S}})\right]$$

Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

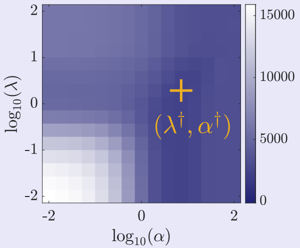
$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha | \boldsymbol{\mathcal{S}})$$

Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

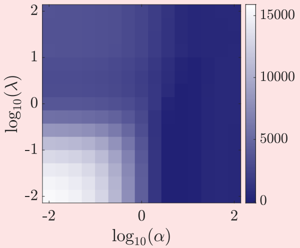
$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha | \boldsymbol{\mathcal{S}})$$

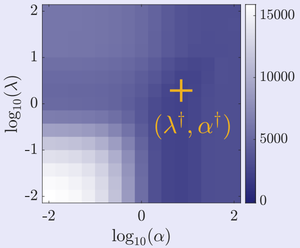


Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

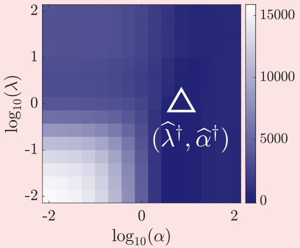
$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha | \boldsymbol{\mathcal{S}})$$

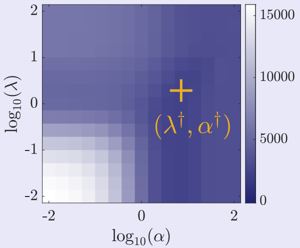


Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

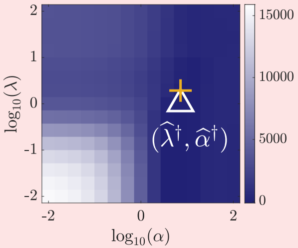
$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\| \widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}} \right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

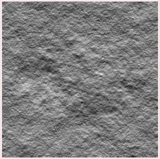
$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha | \boldsymbol{\mathcal{S}})$$

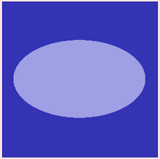


Grid search selection of regularization parameters

$$\left(\widehat{\boldsymbol{h}}^{\mathsf{F}}, \widehat{\boldsymbol{v}}^{\mathsf{F}}\right)(\boldsymbol{\mathcal{L}}; \boldsymbol{\Lambda}) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}_{\mathsf{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Example

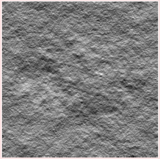


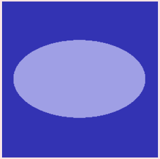


Grid search selection of regularization parameters

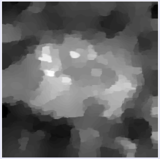
$$\left(\widehat{\boldsymbol{h}}^{\mathsf{F}}, \widehat{\boldsymbol{v}}^{\mathsf{F}}\right)(\boldsymbol{\mathcal{L}}; \boldsymbol{\Lambda}) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}_{\mathsf{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Example





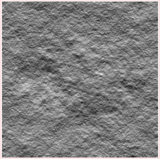
$\widehat{\boldsymbol{h}}^{\mathsf{F}}(\boldsymbol{\mathcal{L}}; \lambda^\dagger, \alpha^\dagger)$
(grid)

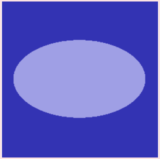


Grid search selection of regularization parameters

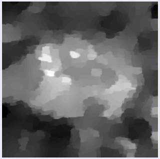
$$\left(\widehat{\boldsymbol{h}}^{\mathsf{F}}, \widehat{\boldsymbol{v}}^{\mathsf{F}}\right)(\boldsymbol{\mathcal{L}}; \boldsymbol{\Lambda}) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}_{\mathsf{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Example

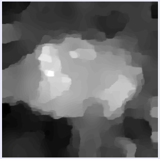




$\widehat{\boldsymbol{h}}^{\mathsf{F}}(\boldsymbol{\mathcal{L}}; \lambda^\dagger, \alpha^\dagger)$ (grid)



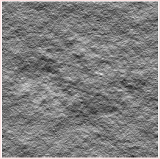
$\widehat{\boldsymbol{h}}^{\mathsf{F}}(\boldsymbol{\mathcal{L}}; \widehat{\lambda}^\dagger, \widehat{\alpha}^\dagger)$ (grid)

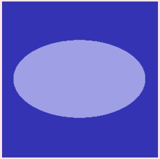


Grid search selection of regularization parameters

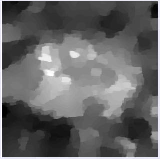
$$\left(\widehat{\boldsymbol{h}}^{\mathsf{F}}, \widehat{\boldsymbol{v}}^{\mathsf{F}}\right)(\boldsymbol{\mathcal{L}}; \boldsymbol{\Lambda}) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}_{\mathsf{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Example

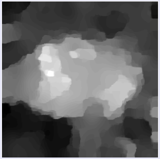




$\widehat{\boldsymbol{h}}^{\mathsf{F}}(\boldsymbol{\mathcal{L}}; \lambda^{\dagger}, \alpha^{\dagger})$ (grid)



$\widehat{\boldsymbol{h}}^{\mathsf{F}}(\boldsymbol{\mathcal{L}}; \widehat{\lambda}^{\dagger}, \widehat{\alpha}^{\dagger})$ (grid)



$15 \times 15 =$ **225** parameters $\longrightarrow$ grid search is very costly!

Automatic minimization of Generalized SURE

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P, \quad \bar{\boldsymbol{x}} \in \mathbb{R}^N, \boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

Generalized FDMC SURE $\quad \lim_{\nu \to 0} \mathbb{E}_{\boldsymbol{\zeta},\boldsymbol{\varepsilon}} \widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y}; \boldsymbol{\Lambda} \,|\, \boldsymbol{\mathcal{S}}) = R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda})$

Automatic minimization of Generalized SURE

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

Generalized FDMC SURE $\lim_{\nu\to 0} \mathbb{E}_{\boldsymbol{\zeta},\boldsymbol{\varepsilon}} \widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y}; \boldsymbol{\Lambda} \,|\, \boldsymbol{\mathcal{S}}) = R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda})$

But : $\underset{\boldsymbol{\Lambda}}{\text{minimize}}\ \widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y}; \boldsymbol{\Lambda} \,|\, \boldsymbol{\mathcal{S}})$ for given $\boldsymbol{y}$, $\boldsymbol{\mathcal{S}}$

Automatic minimization of Generalized SURE

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P\times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\mathcal{S}})$

Generalized FDMC SURE $\lim_{\nu\to 0} \mathbb{E}_{\boldsymbol{\zeta},\boldsymbol{\varepsilon}} \widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y}; \boldsymbol{\Lambda} \,|\, \boldsymbol{\mathcal{S}}) = R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda})$

But : $\underset{\boldsymbol{\Lambda}}{\text{minimize}}\ \widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y}; \boldsymbol{\Lambda} \,|\, \boldsymbol{\mathcal{S}}) \equiv \widehat{R}(\boldsymbol{\Lambda})$ for given $\boldsymbol{y}$, $\boldsymbol{\mathcal{S}}$

Automatic minimization of Generalized SURE

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P, \quad \bar{\boldsymbol{x}} \in \mathbb{R}^N, \ \boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\mathcal{S}})$

Generalized FDMC SURE $\lim_{\nu \to 0} \mathbb{E}_{\boldsymbol{\zeta},\boldsymbol{\varepsilon}} \widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y}; \boldsymbol{\Lambda} \,|\, \boldsymbol{\mathcal{S}}) = R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda})$

But : $\underset{\boldsymbol{\Lambda}}{\text{minimize}} \ \widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y}; \boldsymbol{\Lambda} \,|\, \boldsymbol{\mathcal{S}}) \equiv \widehat{R}(\boldsymbol{\Lambda})$ for given $\boldsymbol{y}$, $\boldsymbol{\mathcal{S}}$

Broyden-Fletcher-Goldfarb-Shanno Quasi-Newton *(Nocedal, 2006)*

for $t = 0, 1, \ldots$

$\boldsymbol{d}^{[t]} = -\boldsymbol{H}^{[t]} \partial_{\boldsymbol{\Lambda}} \widehat{R}(\boldsymbol{\Lambda}^{[t]})$ *descent direction*

$\alpha^{[t]} \in \underset{\alpha \in \mathbb{R}}{\text{Argmin}} \ \widehat{R}(\boldsymbol{\Lambda}^{[t]} + \alpha \boldsymbol{d}^{[t]})$ *line search*

$\boldsymbol{\Lambda}^{[t+1]} = \boldsymbol{\Lambda}^{[t]} + \alpha^{[t]} \boldsymbol{d}^{[t]}$

$\boldsymbol{u}^{[t]} = \partial_{\boldsymbol{\Lambda}} \widehat{R}(\boldsymbol{\Lambda}^{[t+1]}) - \partial_{\boldsymbol{\Lambda}} \widehat{R}(\boldsymbol{\Lambda}^{[t]})$ *gradient increment*

$\boldsymbol{H}^{[t+1]} = \text{BFGS}(\boldsymbol{H}^{[t]}, \boldsymbol{d}^{[t]}, \boldsymbol{u}^{[t]})$ *"inverse Hessian" update*

Automatic minimization of Generalized SURE

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

Generalized FDMC SURE $\lim_{\nu \to 0} \mathbb{E}_{\boldsymbol{\zeta},\boldsymbol{\varepsilon}} \widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y}; \boldsymbol{\Lambda} \,|\, \boldsymbol{\mathcal{S}}) = R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda})$

But : $\underset{\boldsymbol{\Lambda}}{\text{minimize}}\ \widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y}; \boldsymbol{\Lambda} \,|\, \boldsymbol{\mathcal{S}}) \equiv \widehat{R}(\boldsymbol{\Lambda})$ for given $\boldsymbol{y}$, $\boldsymbol{\mathcal{S}}$

Broyden-Fletcher-Goldfarb-Shanno Quasi-Newton *(Nocedal, 2006)*

for $t = 0, 1, \ldots$

$\boldsymbol{d}^{[t]} = -\boldsymbol{H}^{[t]} \partial_{\boldsymbol{\Lambda}} \widehat{R}(\boldsymbol{\Lambda}^{[t]})$ *descent direction*

$\alpha^{[t]} \in \underset{\alpha \in \mathbb{R}}{\text{Argmin}}\ \widehat{R}(\boldsymbol{\Lambda}^{[t]} + \alpha \boldsymbol{d}^{[t]})$ *line search*

$\boldsymbol{\Lambda}^{[t+1]} = \boldsymbol{\Lambda}^{[t]} + \alpha^{[t]} \boldsymbol{d}^{[t]}$

$\boldsymbol{u}^{[t]} = \partial_{\boldsymbol{\Lambda}} \widehat{R}(\boldsymbol{\Lambda}^{[t+1]}) - \partial_{\boldsymbol{\Lambda}} \widehat{R}(\boldsymbol{\Lambda}^{[t]})$ *gradient increment*

$\boldsymbol{H}^{[t+1]} = \text{BFGS}(\boldsymbol{H}^{[t]}, \boldsymbol{d}^{[t]}, \boldsymbol{u}^{[t]})$ *"inverse Hessian" update*

Stein Unbiased GrAdient Risk estimate

Generalized FDMC SURE

$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y};\boldsymbol{\Lambda}\,|\,\boldsymbol{\mathcal{S}}) = \left\|\mathbf{A}\left(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}\right)\right\|^2 +$$
$$\frac{2}{\nu}\left\langle \boldsymbol{\mathcal{S}}\mathbf{A}^\top\boldsymbol{\Pi}\left(\widehat{\boldsymbol{x}}(\boldsymbol{y}+\nu\boldsymbol{\varepsilon};\boldsymbol{\Lambda}) - \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right), \boldsymbol{\varepsilon}\right\rangle - \operatorname{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)$$

Stein Unbiased GrAdient Risk estimate

Generalized FDMC SURE

$$\widehat{R}_{\nu,\varepsilon}(\boldsymbol{y};\boldsymbol{\Lambda}\,|\,\boldsymbol{\mathcal{S}}) = \left\|\mathbf{A}\left(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}\right)\right\|^2 + \frac{2}{\nu}\left\langle \boldsymbol{\mathcal{S}}\mathbf{A}^\top\boldsymbol{\Pi}\left(\widehat{\boldsymbol{x}}(\boldsymbol{y}+\nu\boldsymbol{\varepsilon};\boldsymbol{\Lambda}) - \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right), \boldsymbol{\varepsilon}\right\rangle - \mathrm{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)$$

Generalized Finite Difference Monte Carlo SUGAR

$$\begin{aligned}\partial_{\boldsymbol{\Lambda}}\widehat{R}_{\nu,\varepsilon}(\boldsymbol{y};\boldsymbol{\Lambda}\,|\,\boldsymbol{\mathcal{S}}) = {} & 2\left(\mathbf{A}\boldsymbol{\Phi}\partial_{\boldsymbol{\Lambda}}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right)^\top\mathbf{A}\left(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}\right) \\ & + \frac{2}{\nu}\left\langle \boldsymbol{\mathcal{S}}\mathbf{A}^\top\boldsymbol{\Pi}\left(\partial_{\boldsymbol{\Lambda}}\widehat{\boldsymbol{x}}(\boldsymbol{y}+\nu\boldsymbol{\varepsilon};\boldsymbol{\Lambda}) - \partial_{\boldsymbol{\Lambda}}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right), \boldsymbol{\varepsilon}\right\rangle\end{aligned}$$

Stein Unbiased GrAdient Risk estimate

Generalized FDMC SURE

$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y};\boldsymbol{\Lambda}\,|\,\boldsymbol{\mathcal{S}}) = \|\mathbf{A}\left(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}\right)\|^2 + \frac{2}{\nu}\left\langle \boldsymbol{\mathcal{S}}\mathbf{A}^\top\boldsymbol{\Pi}\left(\widehat{\boldsymbol{x}}(\boldsymbol{y}+\nu\boldsymbol{\varepsilon};\boldsymbol{\Lambda}) - \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right), \boldsymbol{\varepsilon}\right\rangle - \mathrm{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)$$

Generalized Finite Difference Monte Carlo SUGAR

$$\partial_{\boldsymbol{\Lambda}}\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y};\boldsymbol{\Lambda}\,|\,\boldsymbol{\mathcal{S}}) = 2\left(\mathbf{A}\boldsymbol{\Phi}\partial_{\boldsymbol{\Lambda}}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right)^\top\mathbf{A}\left(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda}) - \boldsymbol{y}\right) + \frac{2}{\nu}\left\langle \boldsymbol{\mathcal{S}}\mathbf{A}^\top\boldsymbol{\Pi}\left(\partial_{\boldsymbol{\Lambda}}\widehat{\boldsymbol{x}}(\boldsymbol{y}+\nu\boldsymbol{\varepsilon};\boldsymbol{\Lambda}) - \partial_{\boldsymbol{\Lambda}}\widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})\right), \boldsymbol{\varepsilon}\right\rangle$$

Theorem *(Pascal, 2020)*

Let $(\boldsymbol{y};\boldsymbol{\Lambda}) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y};\boldsymbol{\Lambda})$ an estimator of $\bar{\boldsymbol{x}}$

- uniformly Lipschitz continuous w.r.t. $\boldsymbol{y}$
- such that $\forall \boldsymbol{\Lambda} \in \mathbb{R}^L$, $\widehat{\boldsymbol{x}}(\mathbf{0}_P;\boldsymbol{\Lambda}) = \mathbf{0}_N$,
- uniformly L-Lipschitz continuous w.r.t. $\boldsymbol{\Lambda}$, L ind. of $\boldsymbol{y}$. Then

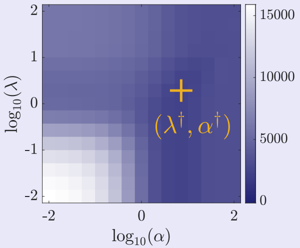
$$\partial_{\boldsymbol{\Lambda}} R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) = \lim_{\nu\to 0}\mathbb{E}_{\boldsymbol{\zeta},\boldsymbol{\varepsilon}}\left[\partial_{\boldsymbol{\Lambda}}\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{y};\boldsymbol{\Lambda}\,|\,\boldsymbol{\mathcal{S}})\right]$$

Parameter tuning (Automatic selection)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

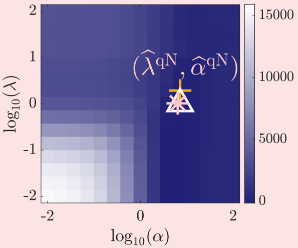
$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

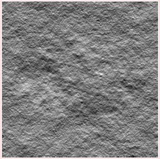
$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha | \boldsymbol{\mathcal{S}})$$

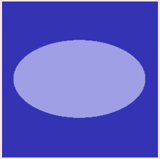


Automated selection of regularization parameters

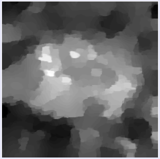
$$\left(\widehat{\boldsymbol{h}}^{\mathrm{F}}, \widehat{\boldsymbol{v}}^{\mathrm{F}}\right)(\boldsymbol{\mathcal{L}}; \boldsymbol{\Lambda}) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}_{\mathrm{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Example

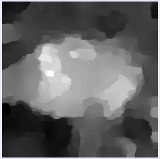




$\widehat{\boldsymbol{h}}^{\mathrm{F}}(\boldsymbol{\mathcal{L}}; \lambda^{\dagger}, \alpha^{\dagger})$
(grid)



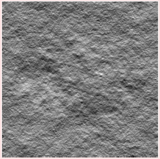
$\widehat{\boldsymbol{h}}^{\mathrm{F}}(\boldsymbol{\mathcal{L}}; \widehat{\lambda}^{\dagger}, \widehat{\alpha}^{\dagger})$
(grid)

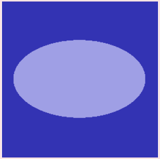


Automated selection of regularization parameters

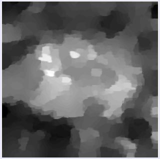
$$\left(\widehat{\boldsymbol{h}}^{\mathrm{F}}, \widehat{\boldsymbol{v}}^{\mathrm{F}}\right)(\boldsymbol{\mathcal{L}};\boldsymbol{\Lambda}) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}_{\mathrm{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

Example

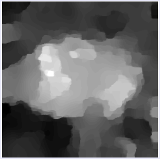




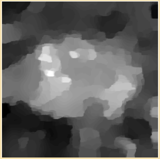
$\widehat{\boldsymbol{h}}^{\mathrm{F}}(\boldsymbol{\mathcal{L}}; \lambda^\dagger, \alpha^\dagger)$ (grid)



$\widehat{\boldsymbol{h}}^{\mathrm{F}}(\boldsymbol{\mathcal{L}}; \widehat{\lambda}^\dagger, \widehat{\alpha}^\dagger)$ (grid)



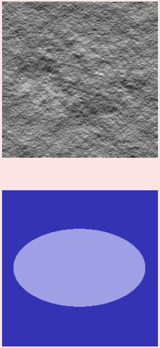
$\widehat{\boldsymbol{h}}^{\mathrm{F}}(\boldsymbol{\mathcal{L}}; \widehat{\lambda}^{\mathrm{qN}}, \widehat{\alpha}^{\mathrm{qN}})$ (quasi-Newton)



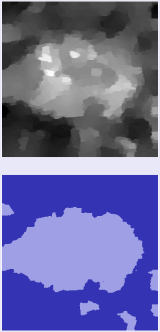
Automated selection of regularization parameters

$$\left(\widehat{\boldsymbol{h}}^{\mathsf{F}}, \widehat{\boldsymbol{v}}^{\mathsf{F}}\right)(\boldsymbol{\mathcal{L}}; \boldsymbol{\Lambda}) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}_{\mathsf{F}}(\mathbf{D}\boldsymbol{h}, \mathbf{D}\boldsymbol{v}; \alpha)$$

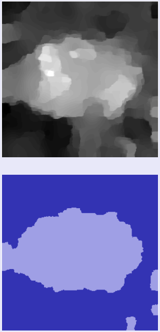
Example



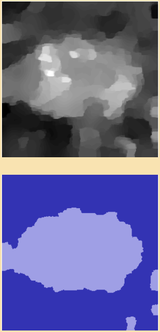
$\widehat{\boldsymbol{h}}^{\mathsf{F}}(\boldsymbol{\mathcal{L}}; \lambda^{\dagger}, \alpha^{\dagger})$ (grid)



$\widehat{\boldsymbol{h}}^{\mathsf{F}}(\boldsymbol{\mathcal{L}}; \widehat{\lambda}^{\dagger}, \widehat{\alpha}^{\dagger})$ (grid)



$\widehat{\boldsymbol{h}}^{\mathsf{F}}(\boldsymbol{\mathcal{L}}; \widehat{\lambda}^{\mathsf{qN}}, \widehat{\alpha}^{\mathsf{qN}})$ (quasi-Newton)



40 calls of the estimator v.s. **225** over a grid

Take home messages

- Fractal texture model based on local *regularity* and *variance*
  - ∗ appropriate for real-world texture characterization
  - ∗ complementary attributes → able to finely discriminate

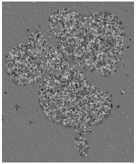
Take home messages

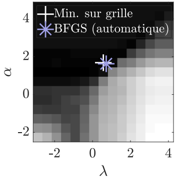
- Fractal texture model based on local *regularity* and *variance*
  - $*$ appropriate for real-world texture characterization
  - $*$ complementary attributes $\rightarrow$ able to finely discriminate
- Simultaneous estimation and regularization
  - $*$ significant decrease of the estimation error
  - $*$ accurate and regular contours thanks to co-localized penalization

Take home messages

- Fractal texture model based on local *regularity* and *variance*
  - appropriate for real-world texture characterization
  - complementary attributes → able to finely discriminate
- Simultaneous estimation and regularization
  - significant decrease of the estimation error
  - accurate and regular contours thanks to co-localized penalization
- Fast algorithms for automated tuning of hyperparameters
  - possibility to manage huge amount of data
  - amenable to process data corrupted by *correlated* Gaussian noise
  - ensured objectivity and reproducibility

Thank you for your attention.





Automated analysis of images from multiphase flow experiments