Multiresolution Adaptive Block-Coordinate Methods for Image Reconstruction

E. Desainte-Maréville, M. Foare, P. Gonçalves, N. Pustelnik, E. Riccietti

LIP, ENS Lyon

SIAM Conference on Optimization - 04/06/2026

Contents

1. Introduction to inverse problems and classical algorithms
2. Dealing with large images: multilevel methods
3. Equivalence with block-coordinate descent methods and block selection strategies

Inverse problems in imaging

Inverse problems

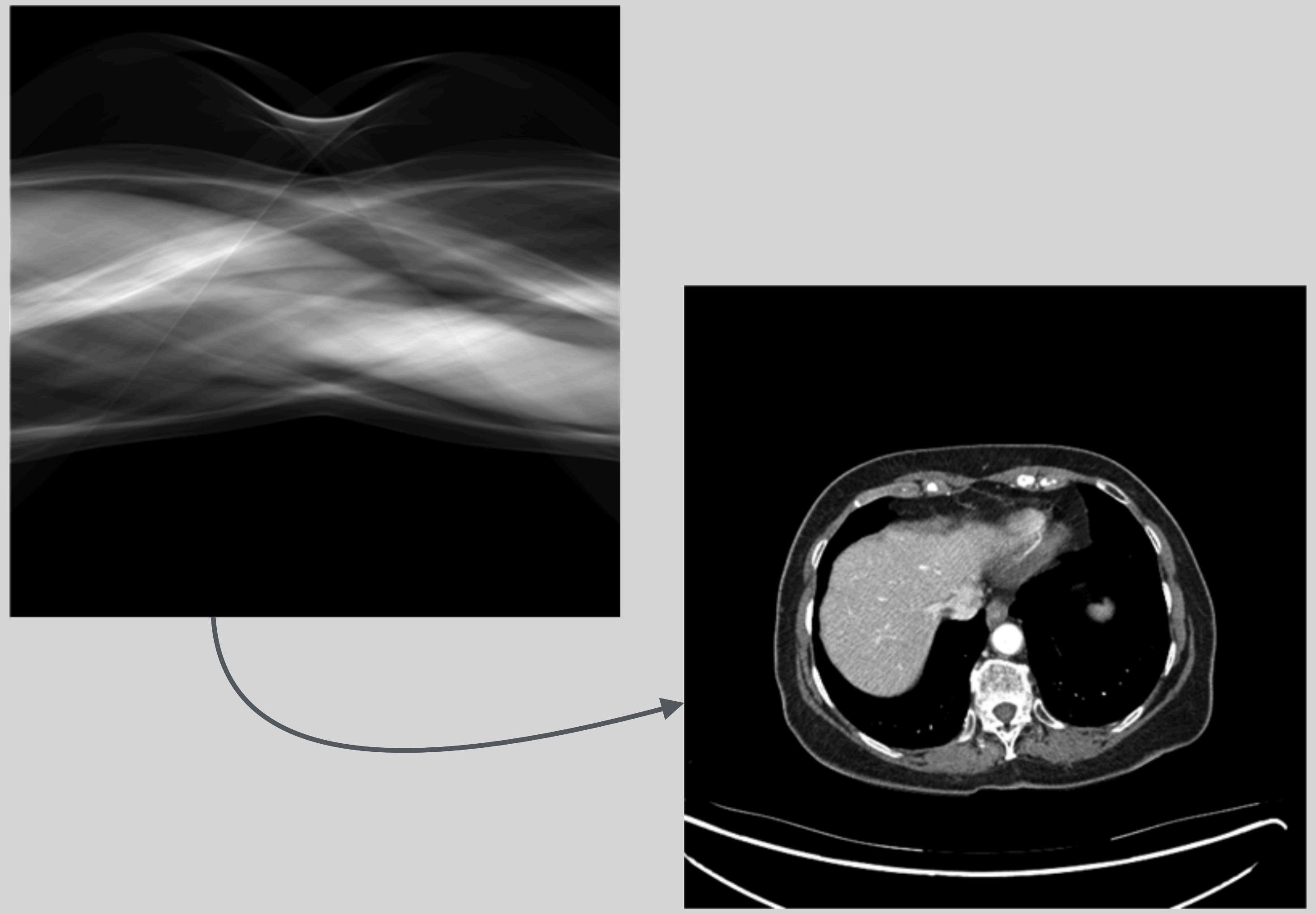
Introduction

Goal: recover a signal from incomplete / degraded measurements

Some examples

Computational photography[1] - Image denoising, deblurring, inpainting, super-resolution…

Astronomical Imaging - Image of a black hole reconstructed from data collected by the Event Horizon telescope[2]



Medical imaging - CT Scan: recover an object from energy dissipation measurements[3]

[1][*Beck & Teboulle.*, 2009]

[2][*Akiyama et al.*, 2019]

[3][Vo *et al.*, 2024]

Inverse problems

Physical model

$$y = A\bar{x} + \eta$$

$y \in \mathbb{R}^m$: observation

$\bar{x} \in \mathbb{R}^n$: ground truth

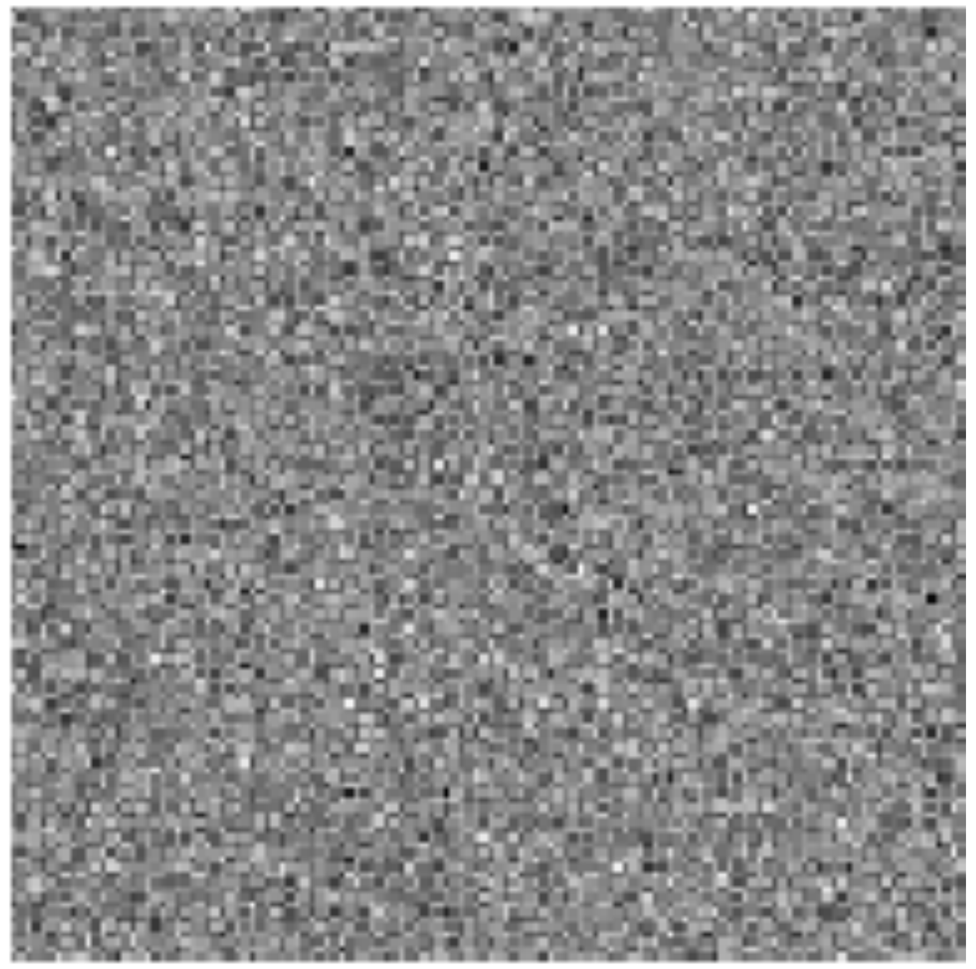
$A \in \mathbb{R}^{m \times n}$: linear degradation

$\eta \sim \mathcal{N}(0, \sigma^2 I_m)$: Gaussian noise

Inverse problems

Examples

Denoising

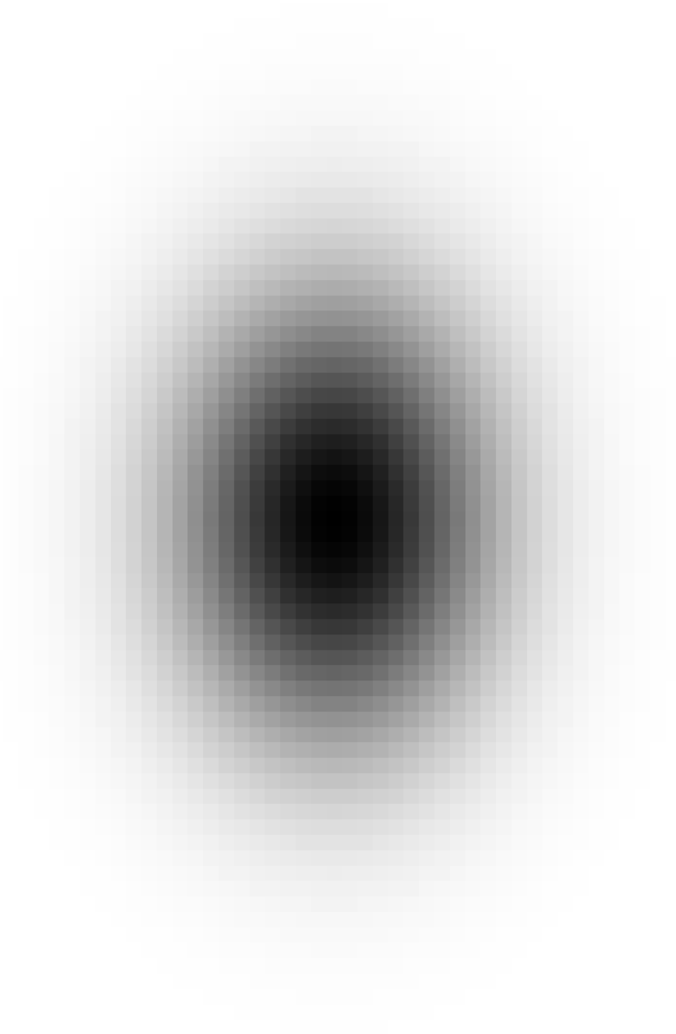
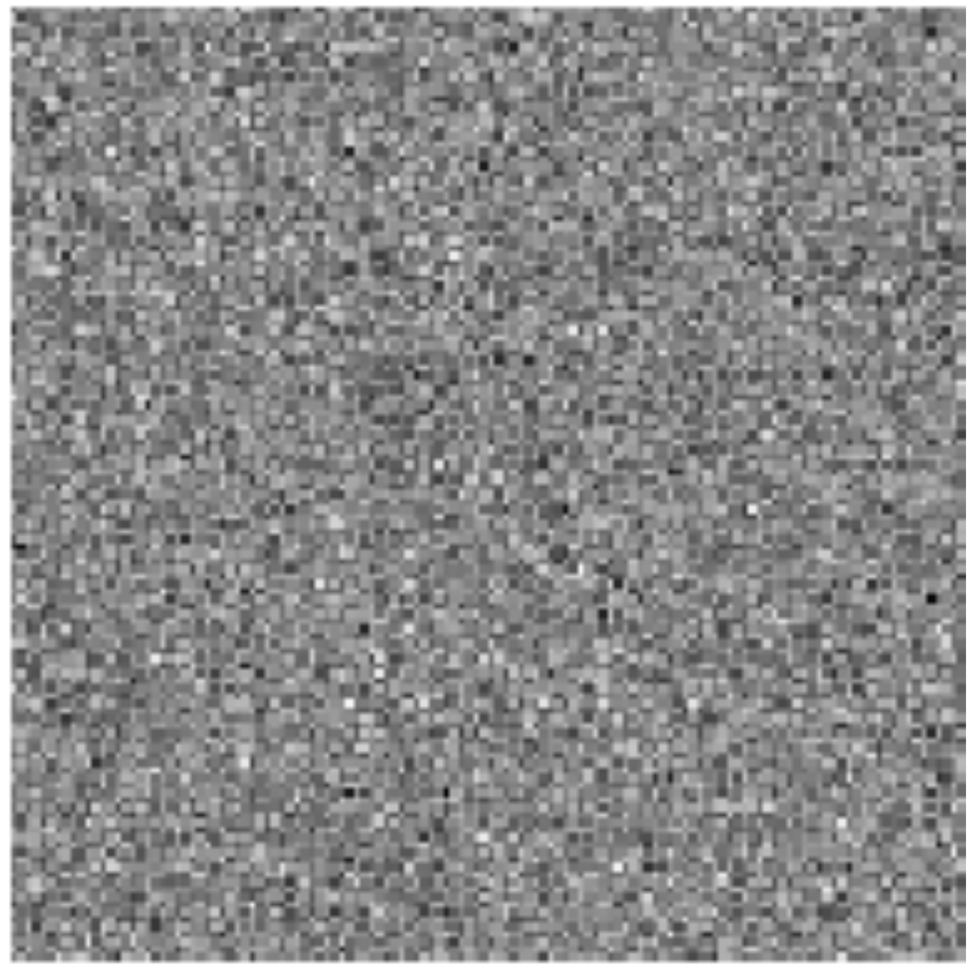
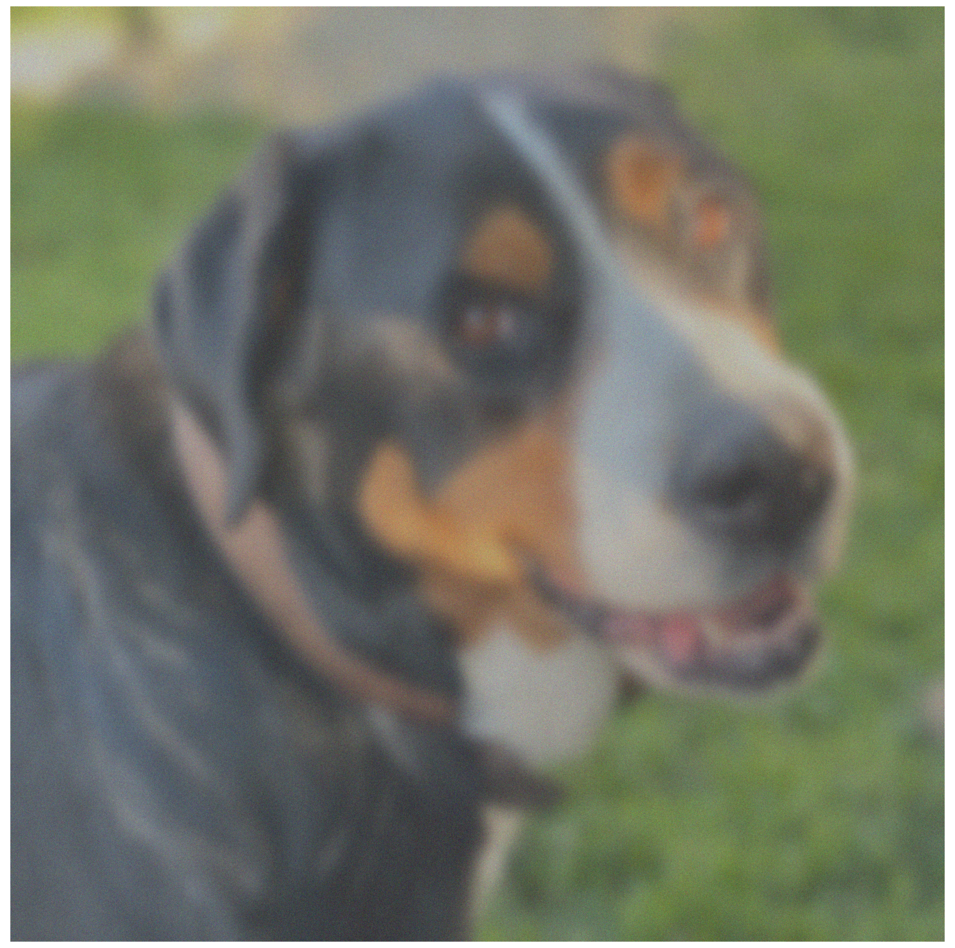
 $\times$ Id $+$  $=$ 

$\bar{x}$ A η y

Inverse problems

Examples

Deblurring

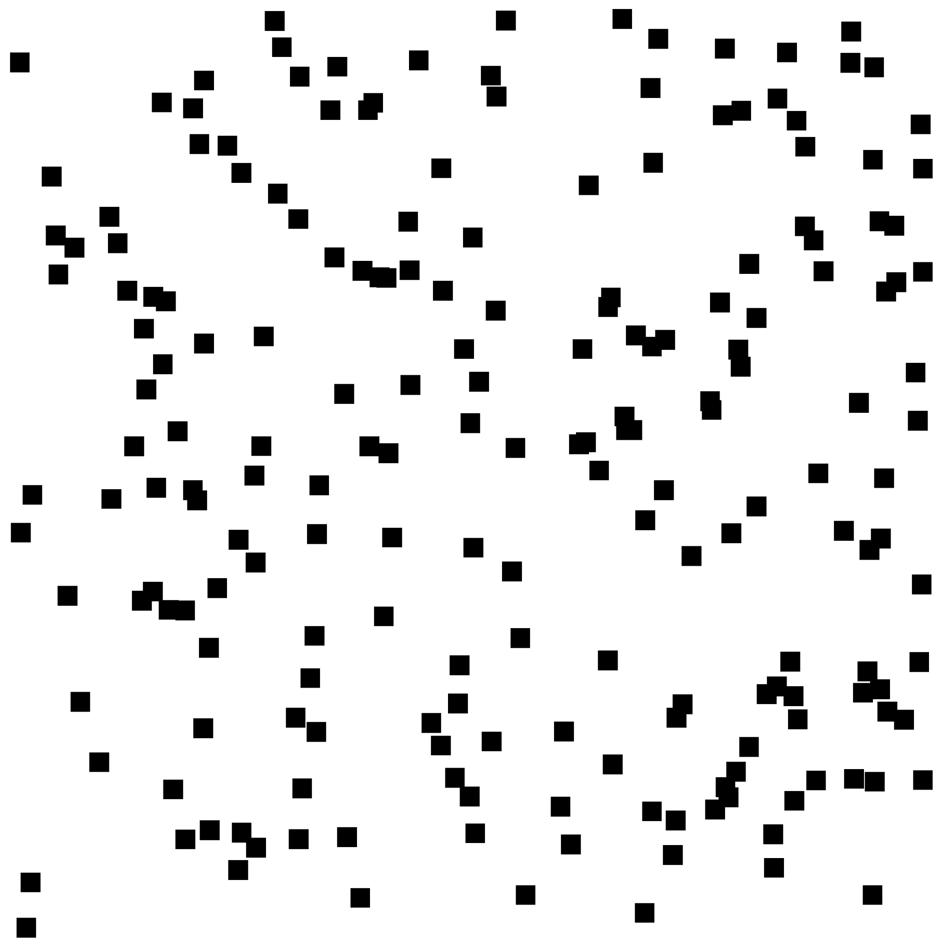
 *  +  = 

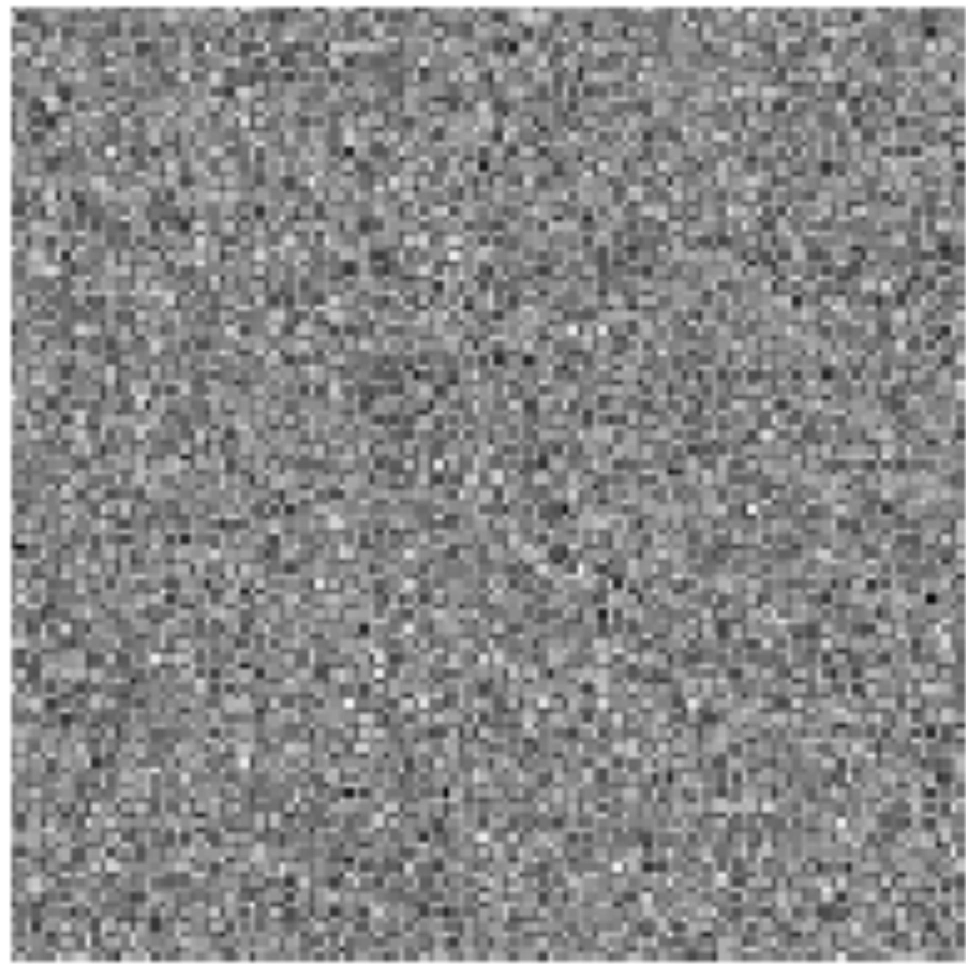
$$\bar{x} \quad * \quad A \quad + \quad \eta \quad = \quad y$$

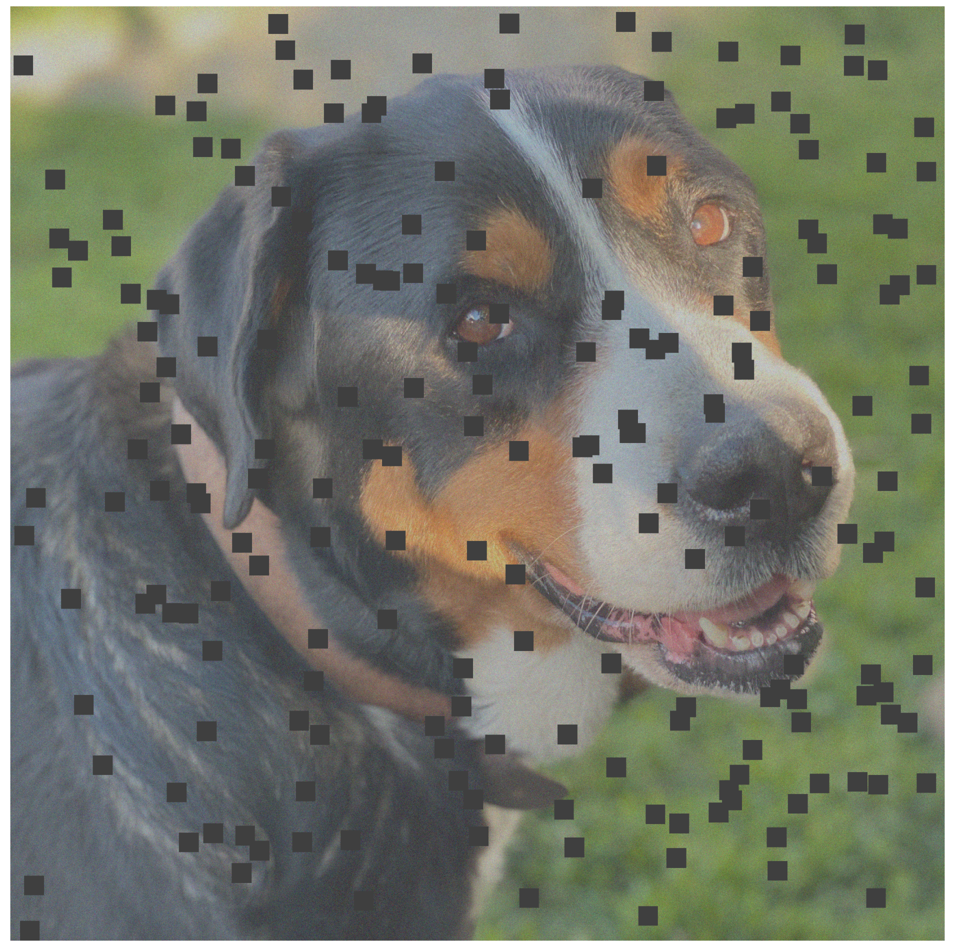
Inverse problems

Examples

Inpainting







$$\bar{x} \odot A + \eta = y$$

Inverse problems

Variational formulation

Goal: estimate $\bar{x}$ given $y = A\bar{x} + \varepsilon$

$$\text{Find} \quad \hat{x} \in \operatorname*{Argmin}_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - y\|^2 + \lambda R(x)$$

$x \mapsto \frac{1}{2}\|Ax - y\|^2$: data-fidelity term

Ensures that $A\hat{x} \approx y$

λR: regularization term

Stabilizes the optimization problem

Inverse problems

The Forward-Backward algorithm

To minimize $x \mapsto \frac{1}{2}\|Ax - y\|^2 + \lambda R(x)$:

Initialize $x_0 \in \mathbb{R}^n$, set $\tau > 0$,

Iterate:

$$x_{k+\frac{1}{2}} = x_k - \tau A^\top (Ax_k - y)$$ ⟵ Gradient step on smooth term

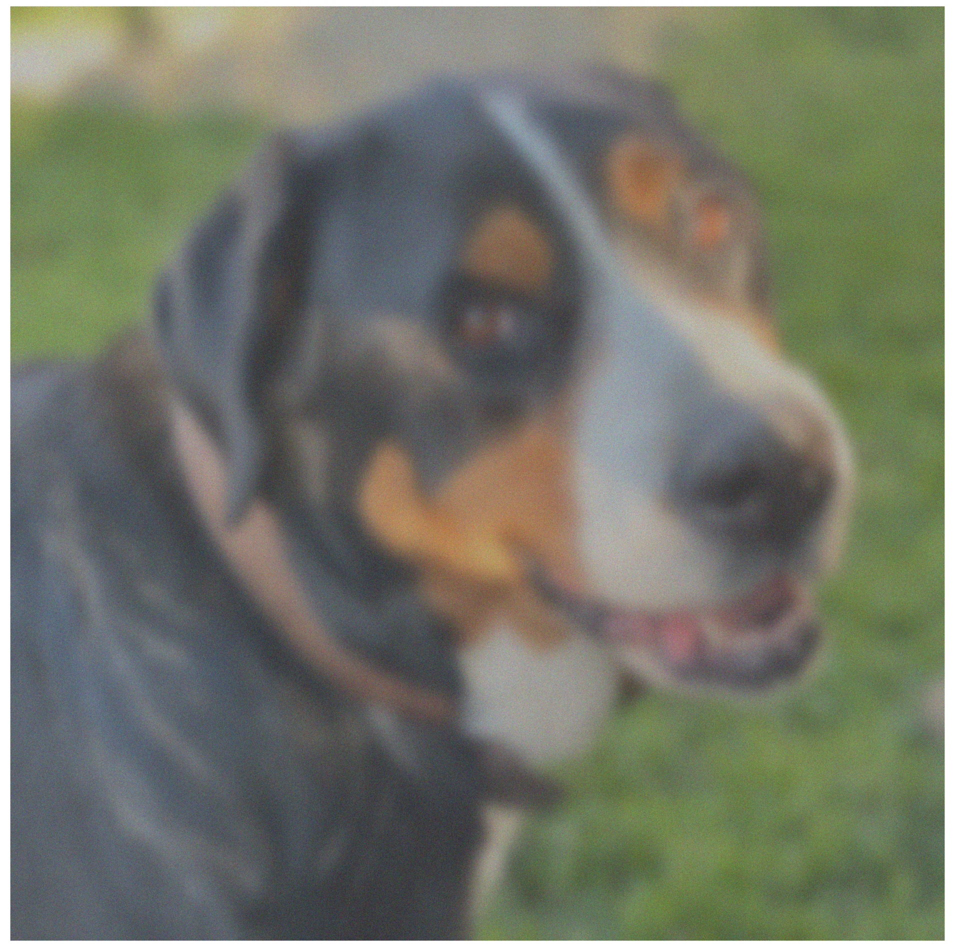
$$x_{k+1} = \mathrm{prox}_{\tau\lambda R}\left(x_{k+\frac{1}{2}}\right)$$ ⟵ Proximal step on nonsmooth term

Where $\mathrm{prox}_f(y) = \underset{x \in \mathbb{R}^n}{\mathrm{argmin}} \frac{1}{2}\|x - y\|^2 + f(y)$

Inverse problems

Examples of reconstructions (deblurring)

y



Observation

$\hat{x}_1$

Reconstruction with
$R(x) = \|Wx\|_1$

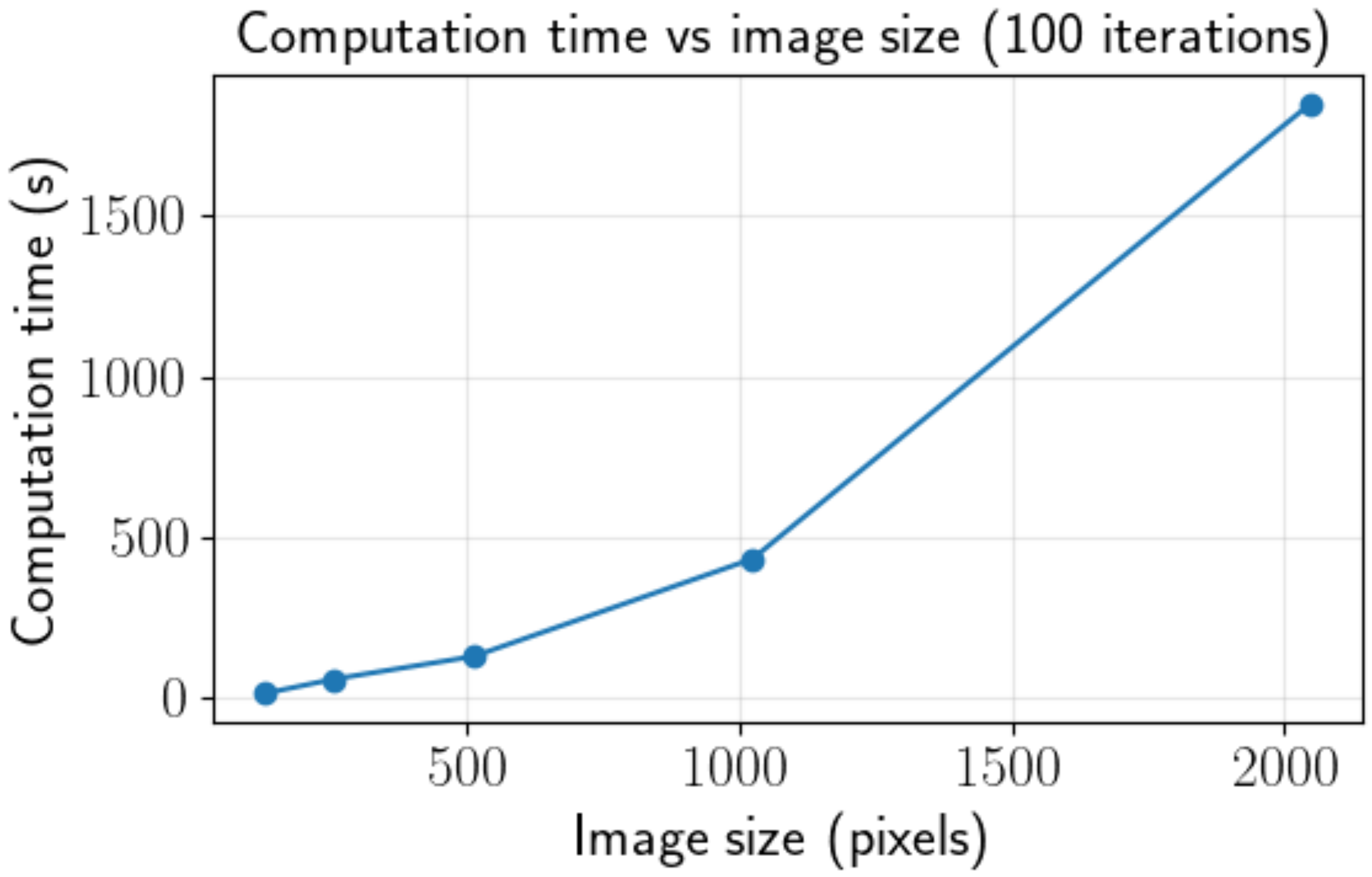
$\hat{x}_2$

Reconstruction with
$R(x) = \mathrm{TV}(x) = \|Dx\|_{1,2}$

Inverse problems

Limitations of standard methods

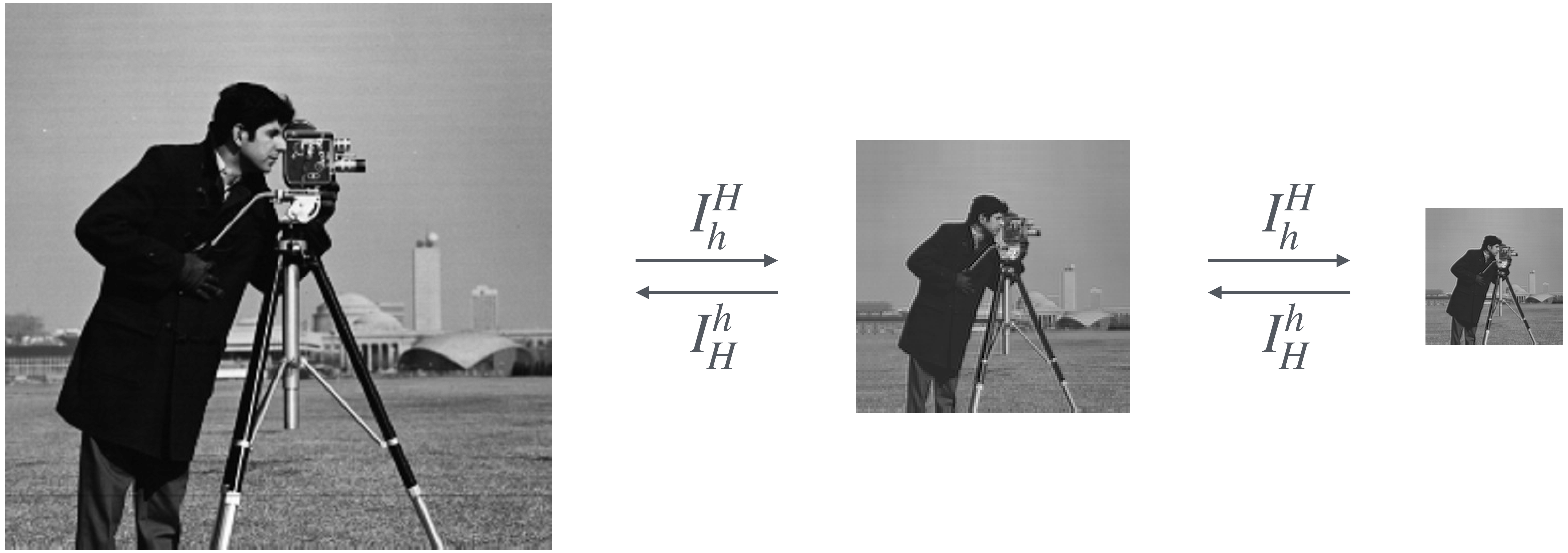
The complexity of the FB algorithm does not scale well with the dimension of the image



Multilevel methods

Multilevel methods

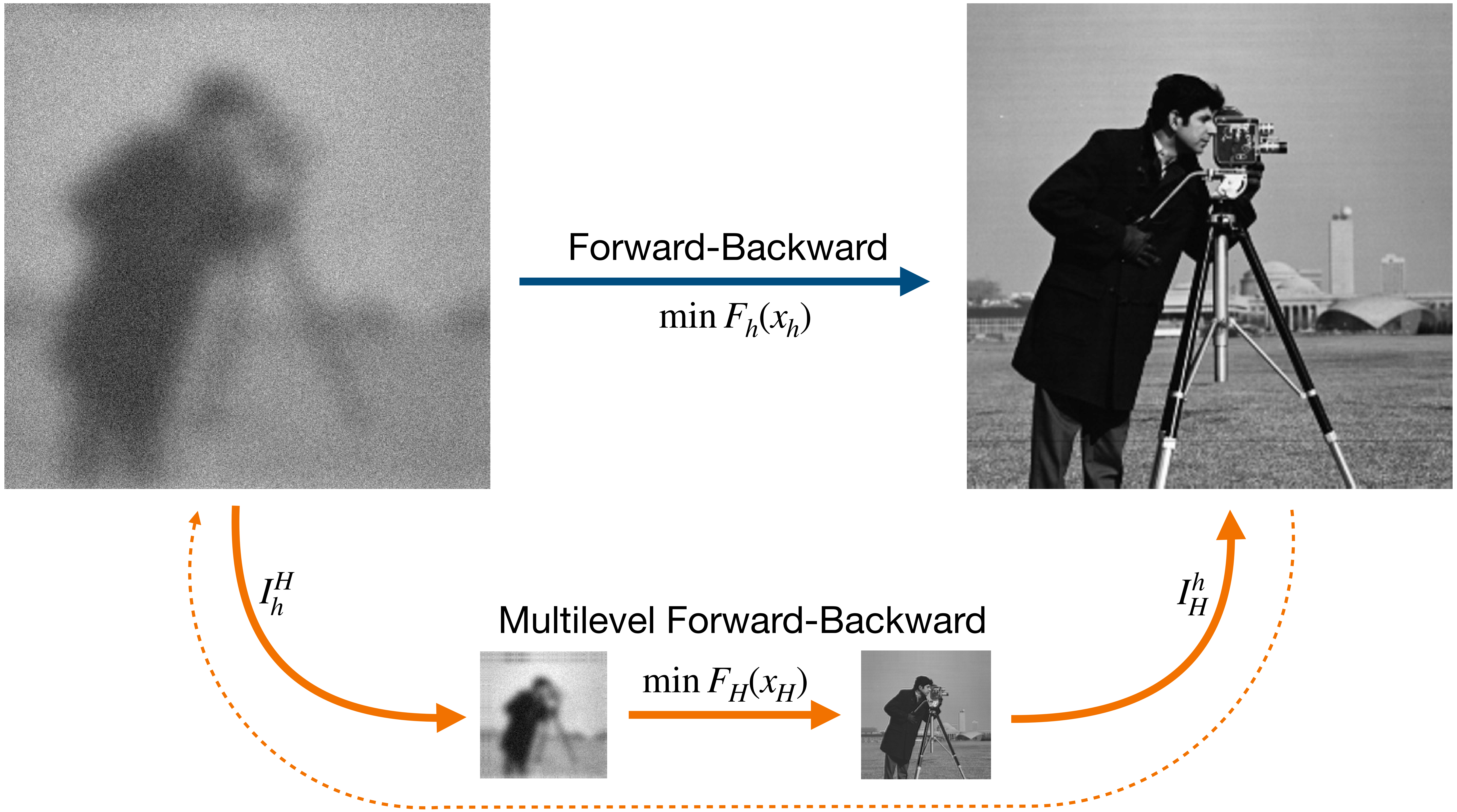
Goal: Solve large-scale problems by exploiting representations of an image at different resolutions.



I_h^H: filtering + downsampling

Multilevel methods

Principle



[*Lauga et al.*, 2023]

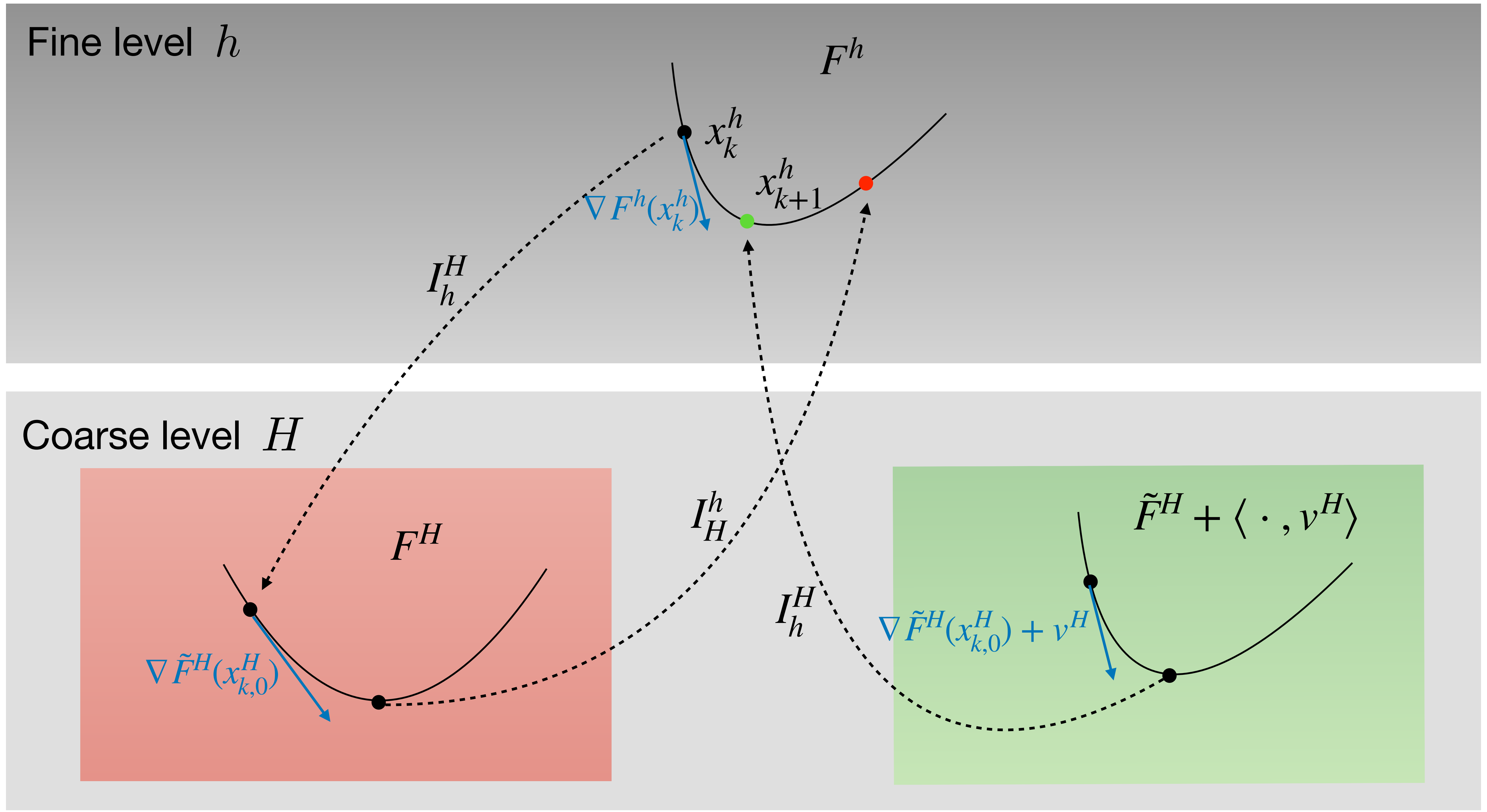
Multilevel methods

Principle

F_H is defined as

$$F_H(x_H) = \underbrace{\frac{1}{2}\|A_H x_H - y_H\|^2 + \lambda R(x_H)}_{\tilde{F}_H} + \langle v_H, x_H \rangle$$

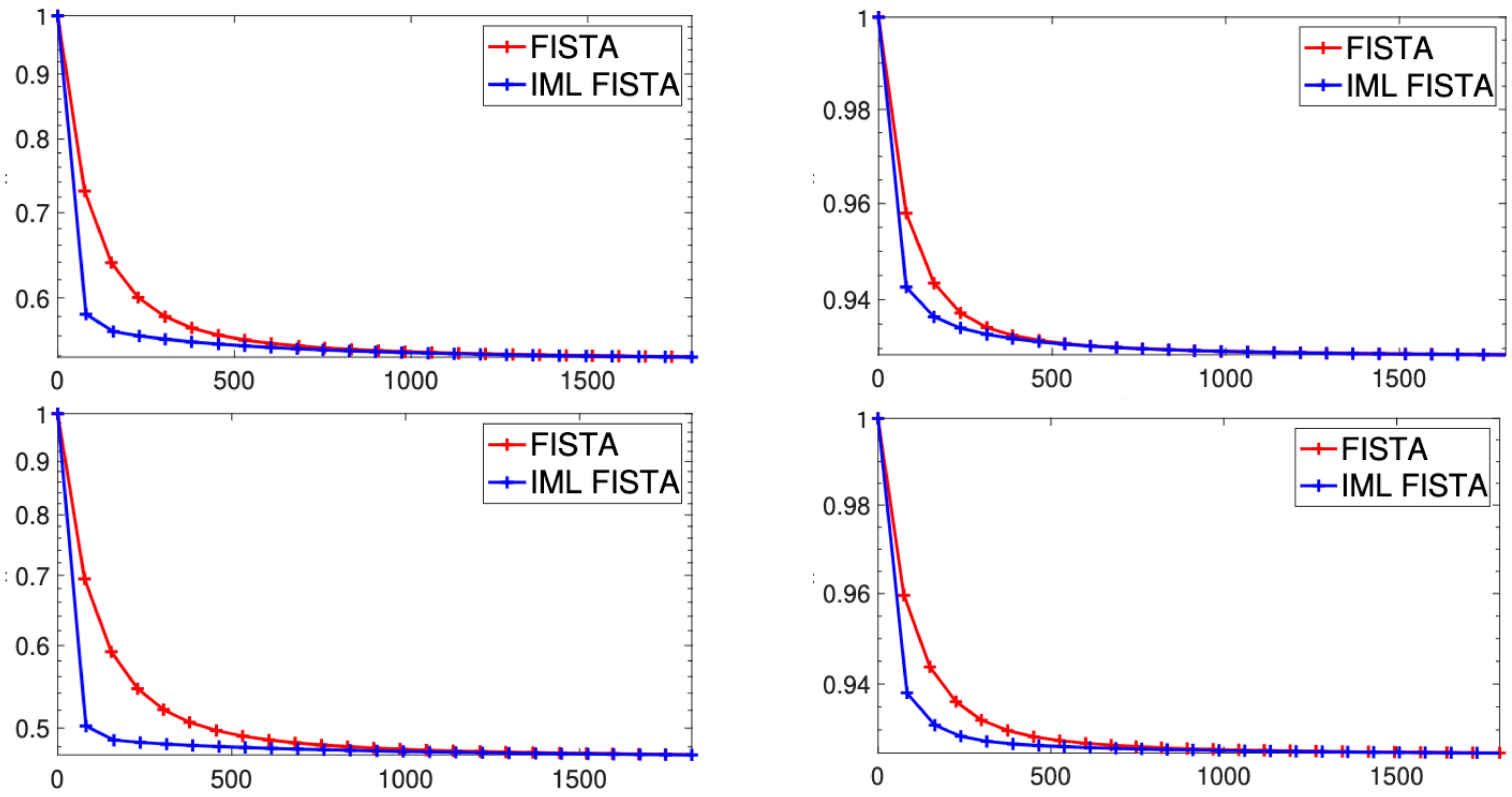
where v_H is called the **first-order coherence** term, that incorporates the influence of the high frequencies on the coarse gradient



[*Lauga et al.*, 2023]

Multilevel methods

Numerical experiments on deblurring problems



Red: single level

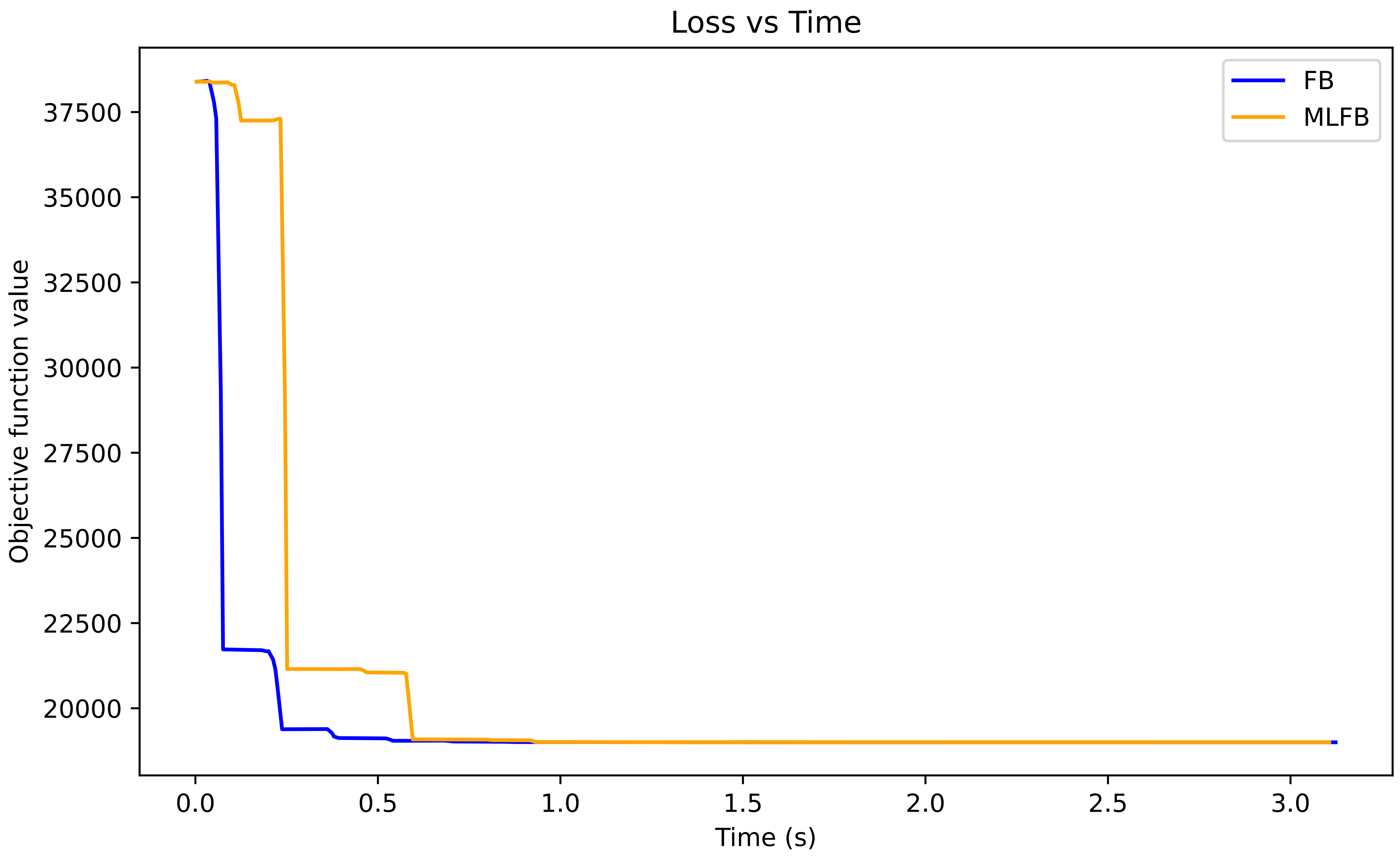
Blue: multilevel

Figure from [*Lauga et al.*, 2023]

Multilevel methods

Room for improvement

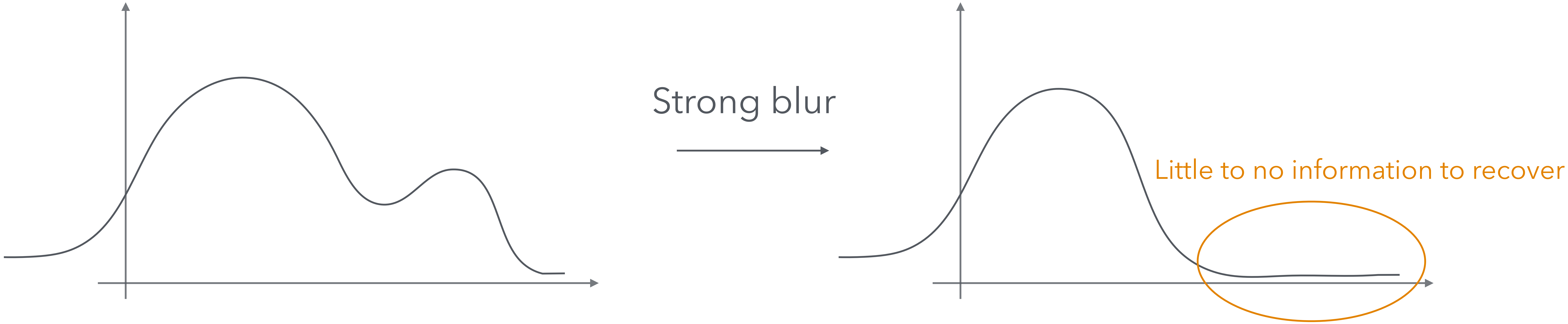
In the case of a lower blur / stronger noise, the classical FB strategy remains the best



Multilevel methods

Room for improvement

Possible reason: a strong blur cuts the high frequencies.



Original frequency spectrum

New frequency spectrum

The information to recover is contained in the low frequencies, *i.e.* in the coarse representations.

Multilevel methods

Room for improvement

Solution: incorporate more high frequency iterations based on the context.

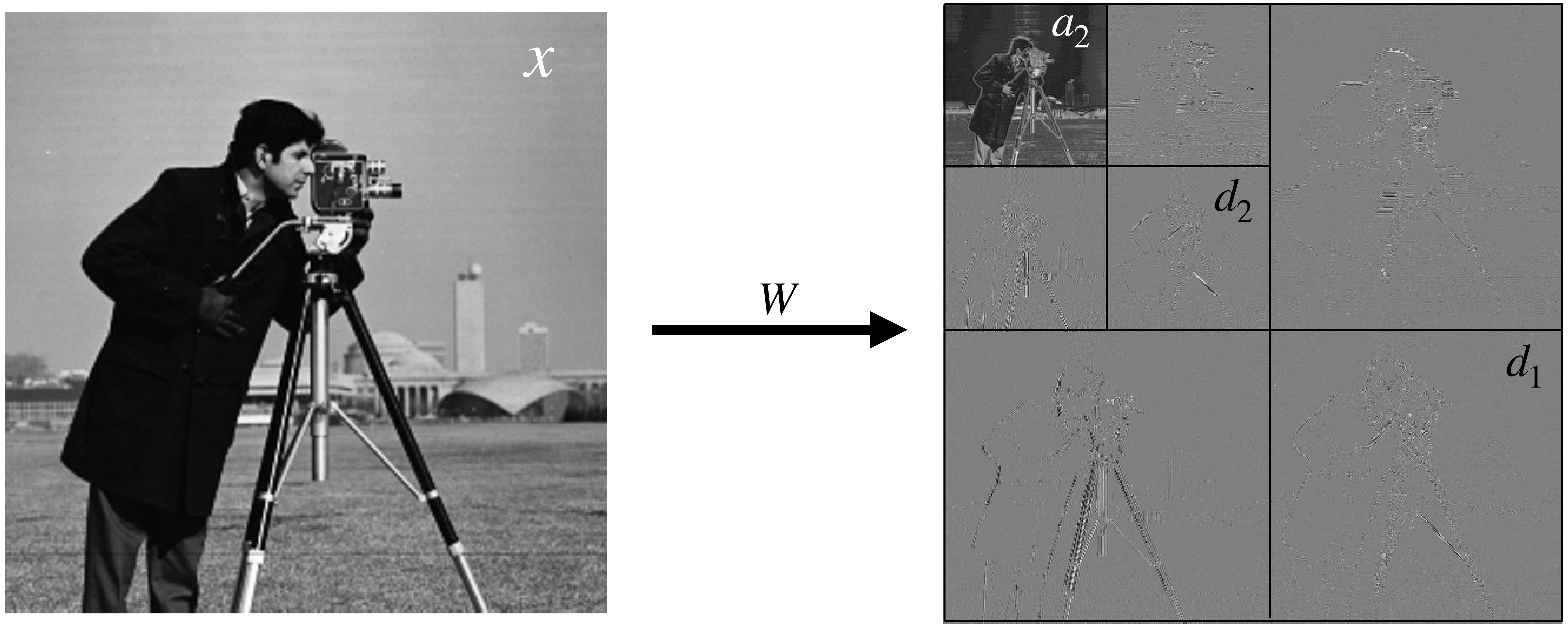
Block-coordinate descent methods

Block-coordinate descent methods

The Wavelet transform

An image can be decomposed into wavelet coefficient up to a scale J as $Wx = (a_J, d_J, \ldots, d_1)$, where:

- a_J is the approximation coefficient (lower frequencies)
- $d_J, \ldots, d_1$ are the detail coefficients (low to high frequencies)



Block-coordinate descent methods

The Wavelet transform

To exploit this multiresolution representation, the optimization problem can be defined in the wavelet domain:

$$\text{Find} \quad \hat{w} \in \underset{w \in \mathbb{R}^n}{\text{Argmin}} \underbrace{\frac{1}{2}\|AW^\top w - y\|^2}_{=:f(w)} + \sum_{i=0}^{J} g_i(w_i)$$

We assume that the regularization term is separable amongst wavelet blocks.

In practice, we will use $g_i = \lambda\| \cdot \|_1$

Block-coordinate descent methods

Block-coordinate forward-backward algorithm

The Forward-Backward algorithm can be decomposed onto wavelet blocks:

At iteration k:

Iterate for each $0 \leq i \leq J$:

$w_i^{k+1} = \text{prox}_{\tau g_i}\left(w_i^k - \tau \nabla_i f(w^k)\right)$ if $\varepsilon_i^k = 1$, ← FB update on block i

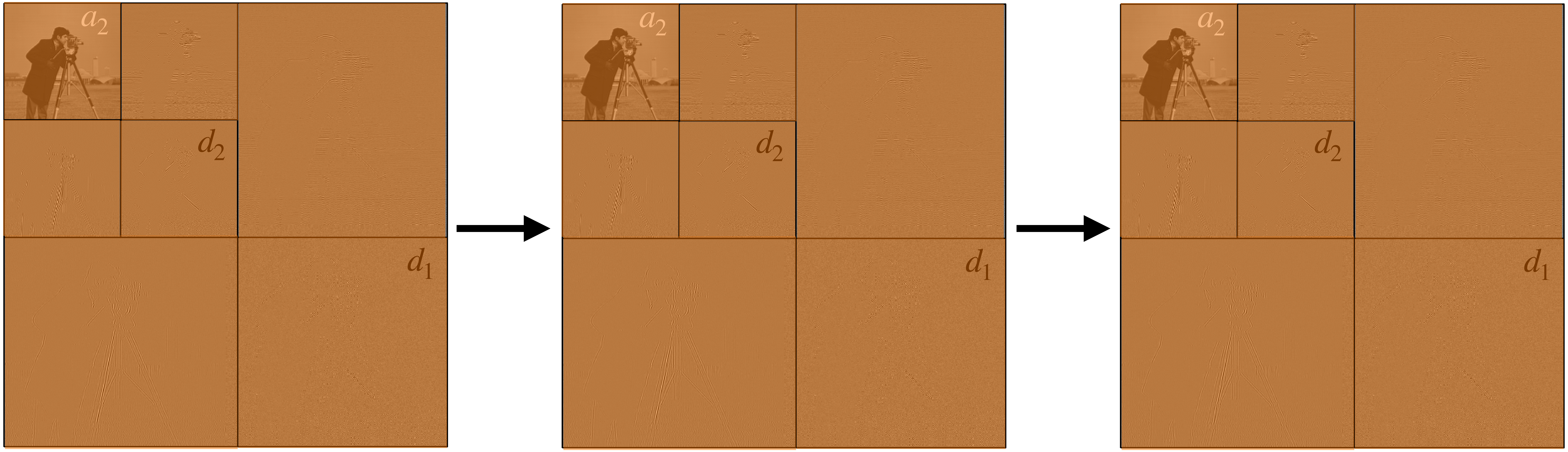
$w_i^{k+1} = w_i^k$ if $\varepsilon_i^k = 0$. ← no update on block i

The variables $\varepsilon_i^k \in \{0,1\}$ allow to choose if block i gets updated at iteration k.

Block-coordinate descent methods

Forward-Backward as a block-coordinate method

If $\forall k, i$, $\varepsilon_k^i = 1$, we recover the classical Forward-Backward method.

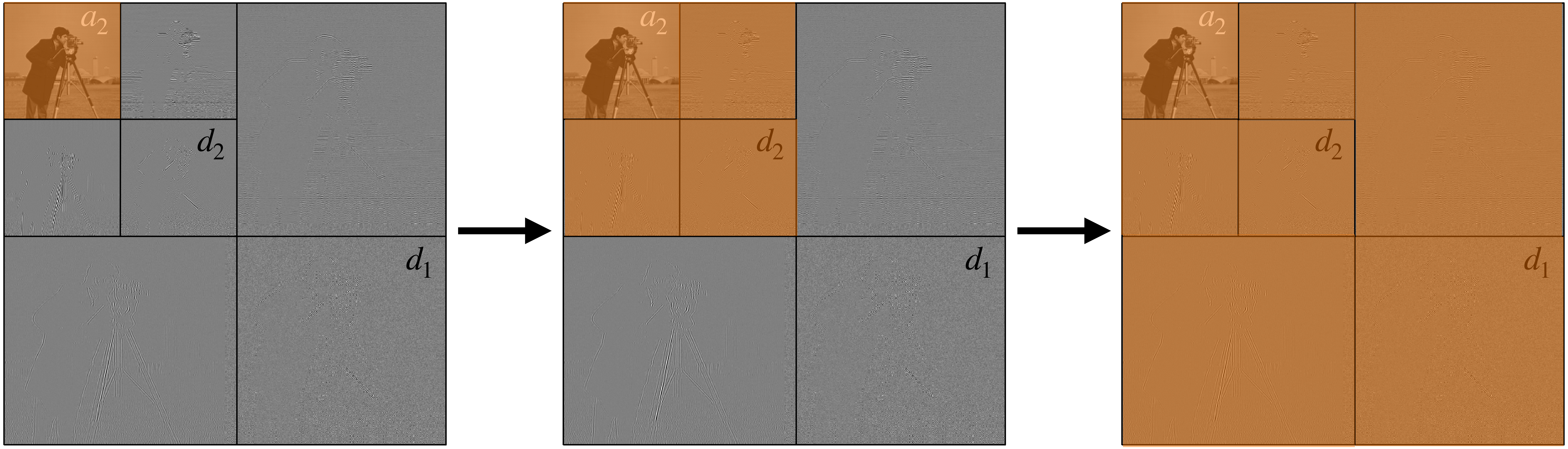


$\varepsilon^1 = (1,1,1)$ $\varepsilon^2 = (1,1,1)$ $\varepsilon^3 = (1,1,1)$

Block-coordinate descent methods

MLFB as a block-coordinate method

The Multilevel Forward-Backward algorithm can be seen as a block-coordinate descent algorithm:



$\varepsilon^1 = (1,0,0)$ $\varepsilon^2 = (1,1,0)$ $\varepsilon^3 = (1,1,1)$

Block-coordinate descent methods

MLFB as a block-coordinate method

The partial gradient with respect to block i naturally incorporates the coherence term:

$$\nabla_{w_i} f(w) = \sum_{j=0}^{J} \Pi_i A^\top A \Pi_j^\top w_j - \Pi_i A^\top y$$

where $\Pi_i x = w_i$

$$= \boxed{\Pi_i A^\top A \Pi_i^\top w_i - \Pi_i A^\top y} + \boxed{\sum_{j \neq i}^{J} \Pi_i A^\top A \Pi_j^\top w_j}$$

Gradient of the coarse data-fidelity function

$$w_i \mapsto \frac{1}{2}\|A\Pi_i^\top w_i - y\|^2$$

Contribution of block j on the i^{th} gradient: coherence term

Recall that

$$F_H(x_H) = \underbrace{\frac{1}{2}\|A_H x_H - y_H\|^2}_{\tilde{F}_H} + \langle v_H, x_H \rangle$$

Block-coordinate descent methods

Incorporating high frequency iterations when needed

Goal: accelerate convergence in every degradation regime.
Idea: adaptively select the blocks to update based on the current state of the algorithm.

Block-coordinate descent methods

Stochastic Gauss-Southwell selection rule

Define for every k, i :

$$\Delta_i^k = w_i^k - \text{prox}_{\tau g_i}(w_i^k - \tau \nabla_i f(w^k))$$

The "residual" of the FB update

Deterministic Gauss-Southwell: choose block $i^k \in \operatorname{argmax}_i \|\Delta_i^k\|$

Stochastic variant (*MAGIC-FB*)[1]: Define probabilities $p_i^k = \frac{\|\Delta_i^k\|}{\|\Delta^k\|}$, and activate block i according to

$$\varepsilon_i^k \sim \mathscr{B}(p_i^k)$$

- Allows parallel updates
- Variable number of blocks

[1][*D-M. et al.*, 2026]

Block-coordinate descent methods

Stochastic Gauss-Southwell selection rule

Define for every k, i :

$$\Delta_i^k = w_i^k - \mathrm{prox}_{\tau g_i}(w_i^k - \tau \nabla_i f(w^k))$$

The "residual" of the FB update

Problem: computing every $(\Delta_i^k)_i$ seems costly.

Solution: exploit the linearity of the problem:

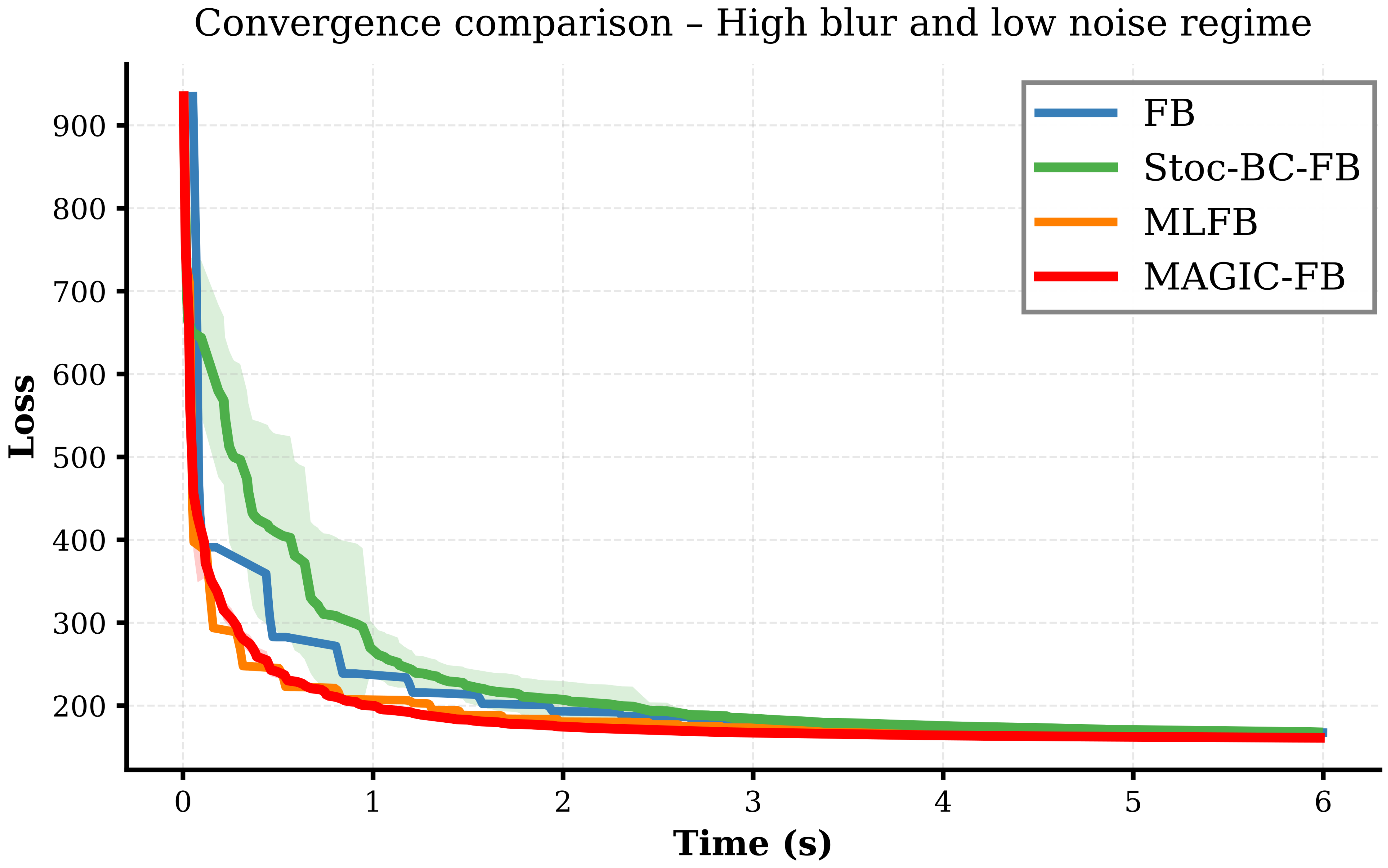
At each iteration, only the $J+1$ products $(\Pi_i A^\top A \Pi_{i^k}^\top w_{i^k})_{0 \leq i \leq J}$

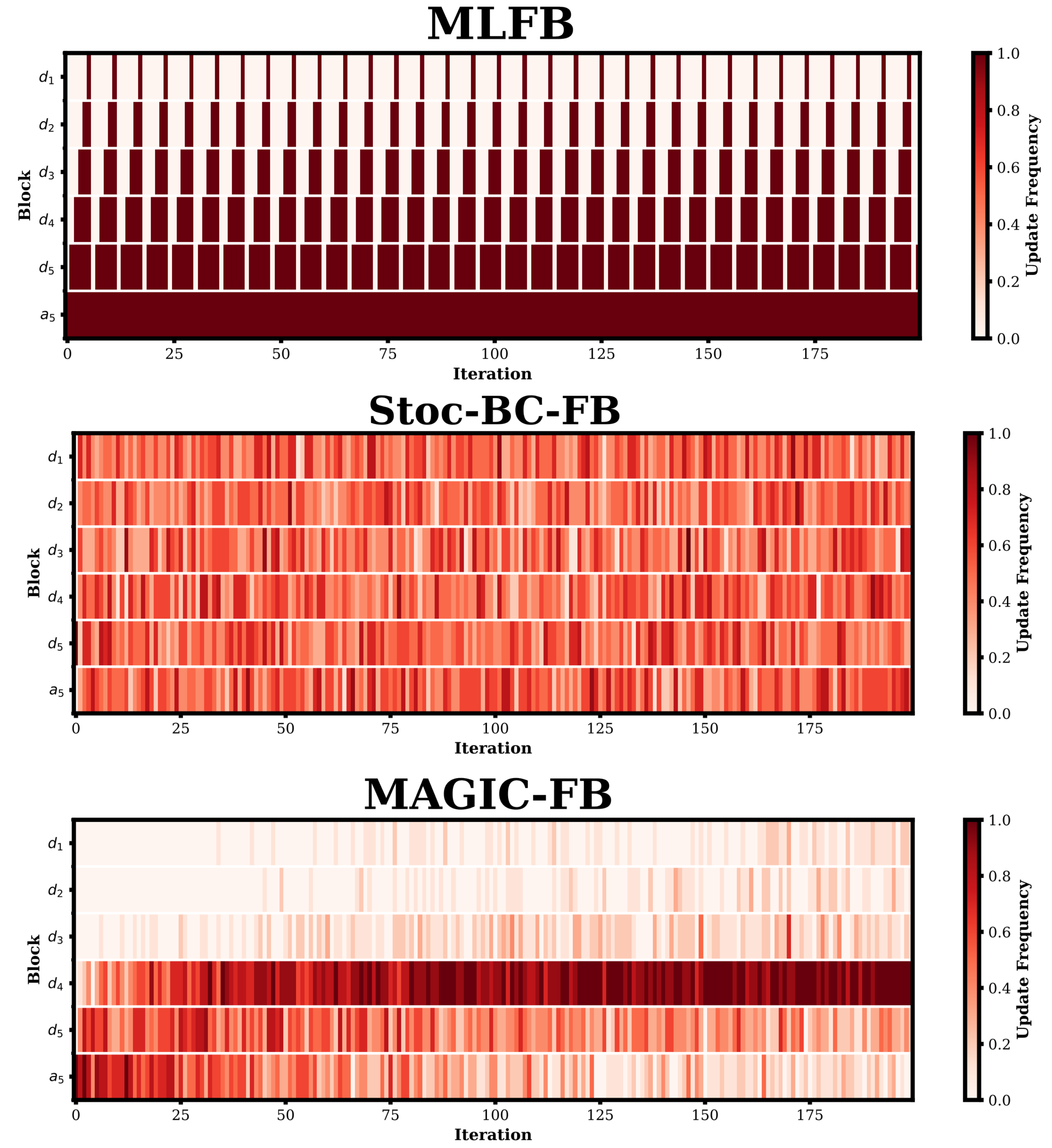
implying the updated block i^k are recomputed.

Block-coordinate descent methods

Stochastic Gauss-Soutwell rule: numerical results

Regime 1: $\sigma_{\text{blur}} = 7, \sigma_{\text{noise}} = 10^{-2}$

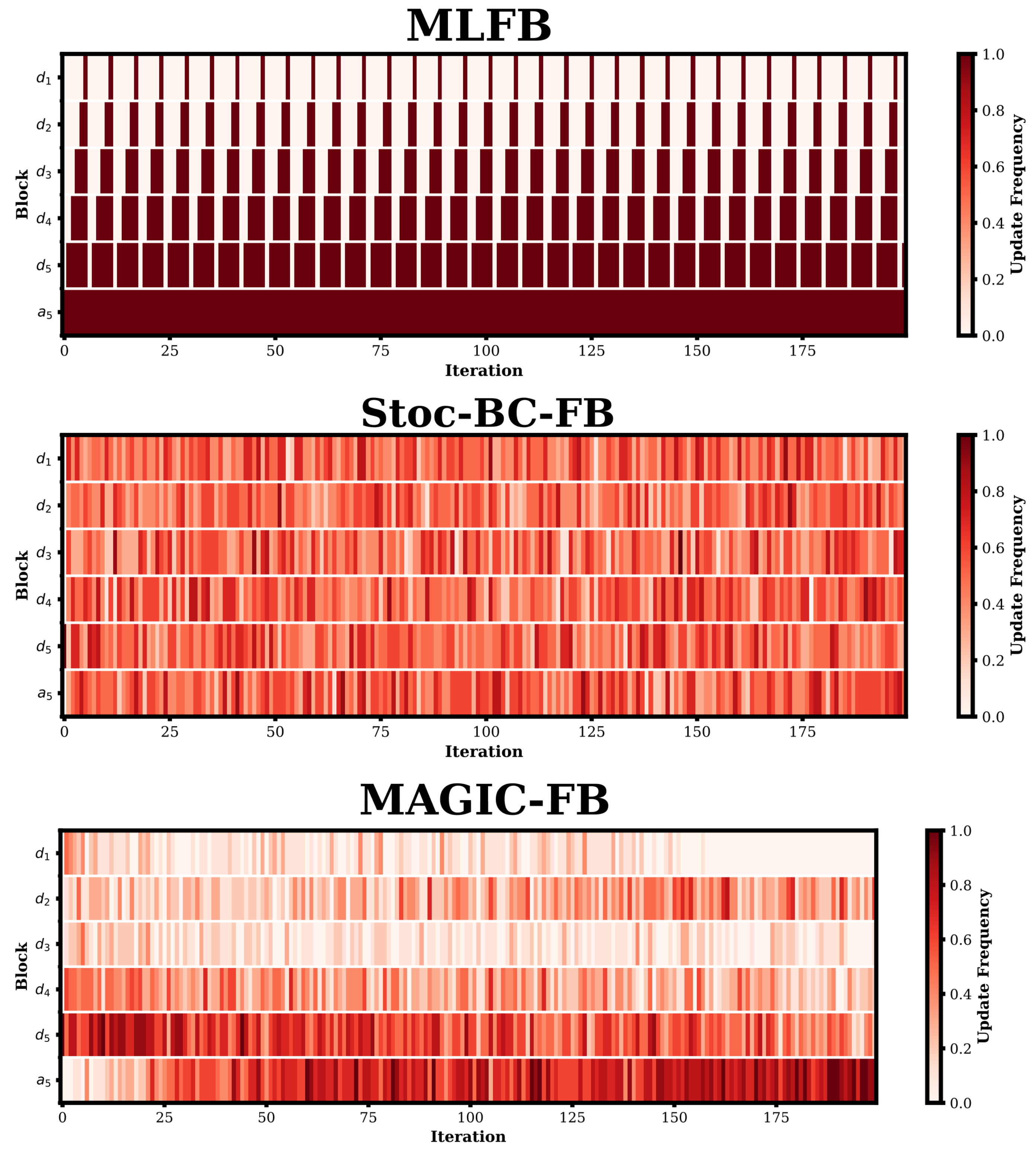


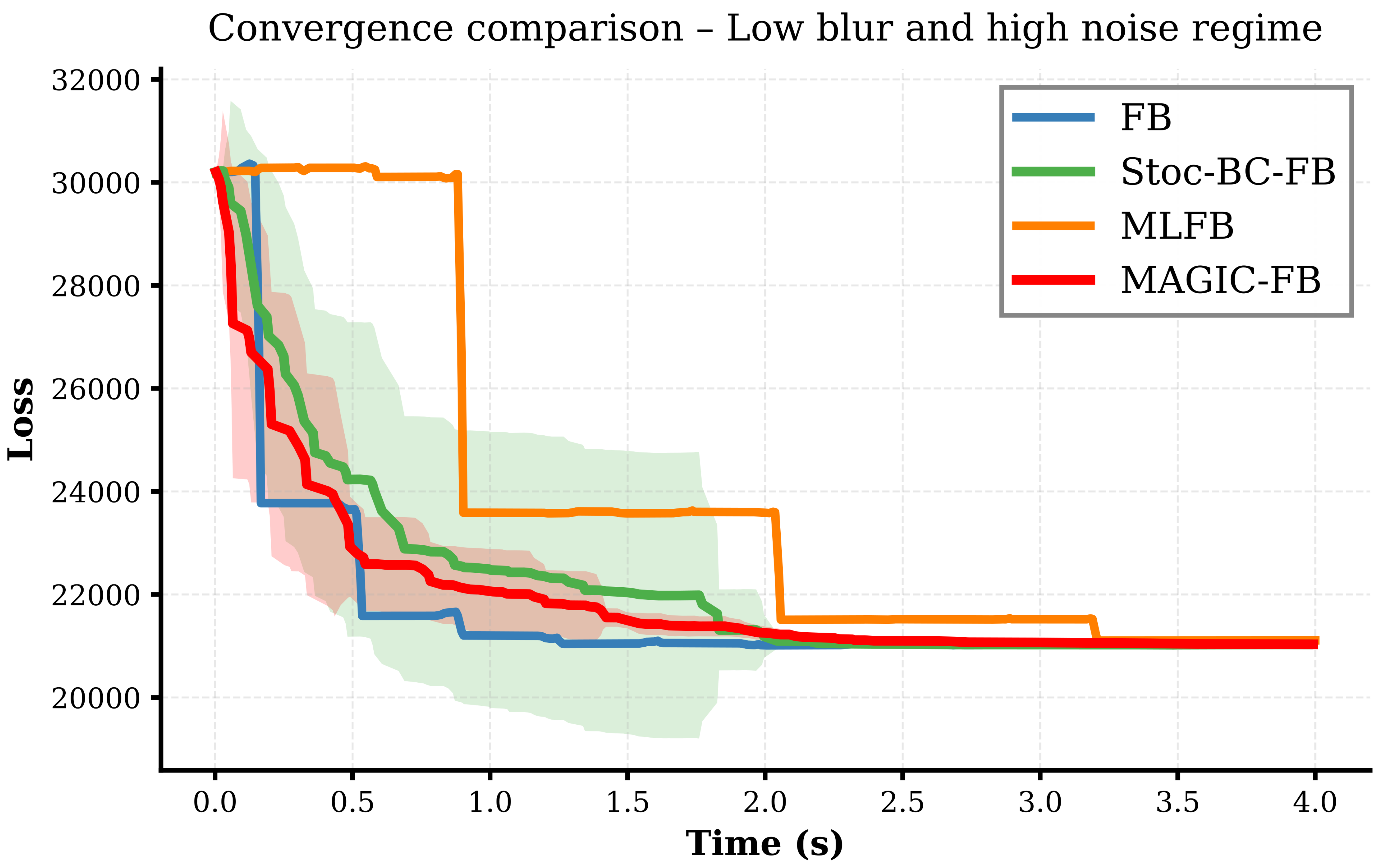


Block-coordinate descent methods

Stochastic Gauss-Soutwell rule: numerical results

Regime 2: $\sigma_{\text{blur}} = 1, \sigma_{\text{noise}} = 10^{-1}$

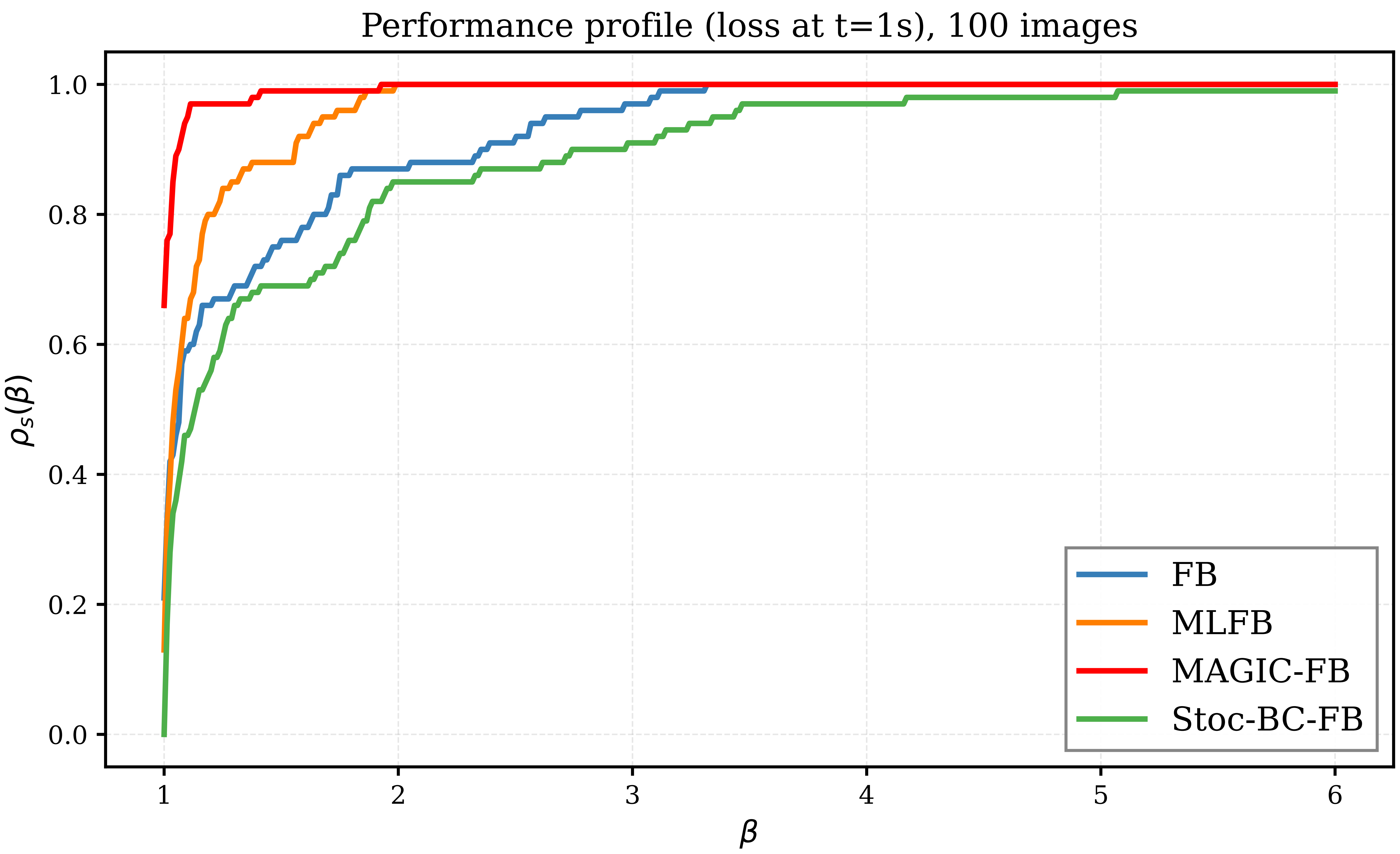




Block-coordinate descent methods

Stochastic Gauss-Soutwell rule: numerical results

Performance profile on 100 DIV2K images



Block-coordinate descent methods

Stochastic Gauss-Soutwell rule: numerical results

Conclusion of the experiments: MAGIC-FB matches the performance of the best method in different degradation regimes.

Multiresolution Adaptive Block-Coordinate Forward-Backward for Image Reconstruction, E. Desainte-Maréville, M. Foare, P. Gonçalves, N. Pustelnik, E. Riccietti, EUSIPCO 2026

Conclusion

Conclusion

Summary:

- Standard algorithms for inverse problems **do not scale well** with the dimension.
- **Multilevel methods** accelerate the convergence of standard algorithms by exploiting representations of images at different resolutions.
- The Multilevel Forward-Backward algorithm is **equivalent to a block-coordinate descent** algorithm on wavelet blocks, for a specific choice of update rule.
- In some degradation regimes, MLFB fails to exploit high frequency information: we define an **adaptive block selection rule** that matches the best method in every regime.

Conclusion

Perspectives:

1. Theoretical analysis of this adaptive algorithm:
    - Proof of convergence
    - Study on other inverse problems, and study which problems are well suited for an adaptive block selection

Conclusion

Perspectives:

2. Apply this strategy to learning-based methods that provide reconstructions of better quality

Forward-Backward

Iterate:

$$x_{k+\frac{1}{2}} = x_k - \tau A^\top (A x_k - y)$$

$$x_{k+1} = \text{prox}_{\tau\lambda R}\left(x_{k+\frac{1}{2}}\right)$$

Plug-and-Play

Iterate:

$$x_{k+\frac{1}{2}} = x_k - \tau A^\top (A x_k - y)$$

$$x_{k+1} = D_\sigma\left(x_{k+\frac{1}{2}}\right)$$

where D_σ is a learned denoiser

Thank you !