

First Order Geometric Multilevel Optimization For Discrete Tomography

Stefania Petra
Mathematical Imaging Group
Heidelberg University

**Workshop on Optimization and Operator Theory dedicated to
Professor Lev Bregman's 80th
TECHNION, 15-17 Nov. 2021**

joint work with Jan Plier¹, Fabrizio Savarino², Michal Kočvara³

¹HITS, Heidelberg

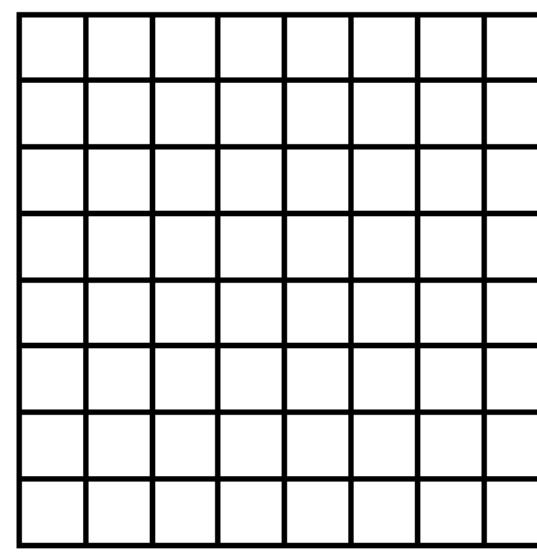
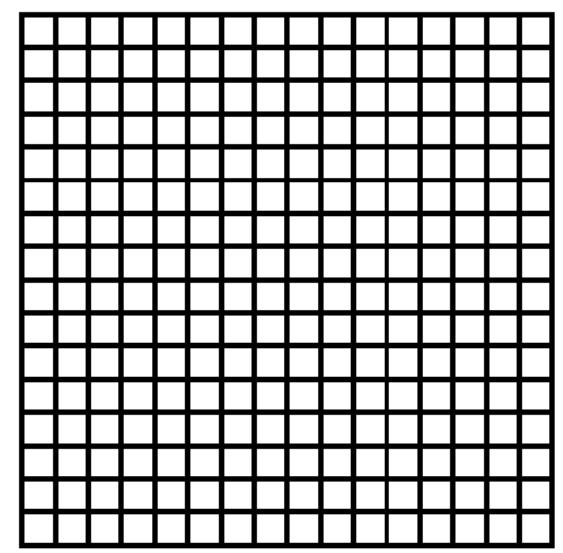
²IPA, Heidelberg University

³School of Mathematics - University of Birmingham

Motivation: Multigrid for PDEs

2D discrete Poisson equation

$$-\Delta u = f$$



finest grid



coarse grid

$$u \in \mathbf{R}^n$$

Poisson solver complexity

Gaussian elimination

$$O(n^2)$$

Jacobi iteration

$$O(n^2 \log \varepsilon)$$

Gauss–Seidel iteration

$$O(n^2 \log \varepsilon)$$

Conjugate gradient (CG)

$$O(n^{3/2} \log \varepsilon)$$

Fast Fourier transform (FFT)

$$O(n \log n)$$

Multigrid (iterative)

$$O(n \log \varepsilon)$$

Multigrid (FMG)

$$O(n)$$

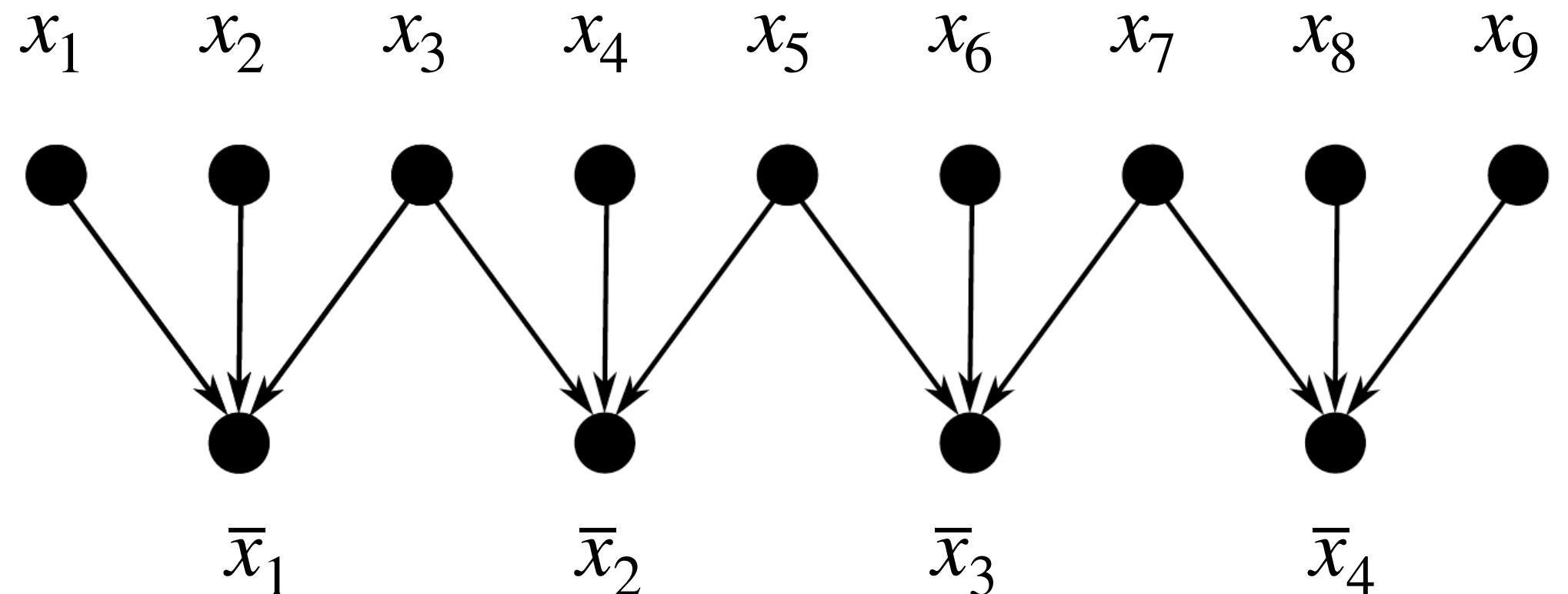
This Talk: Geometric Multilevel Optimization

problem discretization on a 3D grid graph

using the “natural” geometry
(positivity, box constraints)

Information transfer between levels

1. Geometric multigrid for PDEs



$$\min_{x \in C} f(x)$$

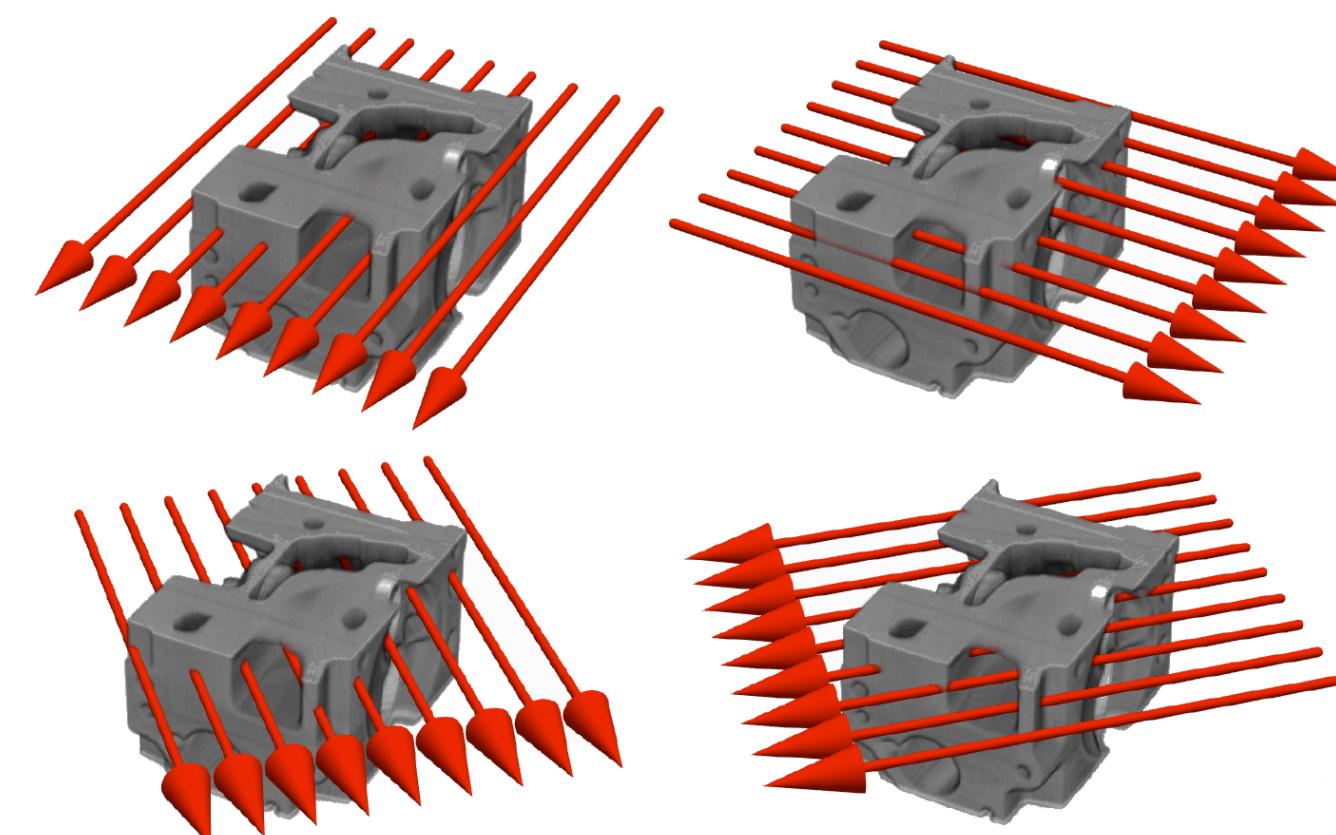
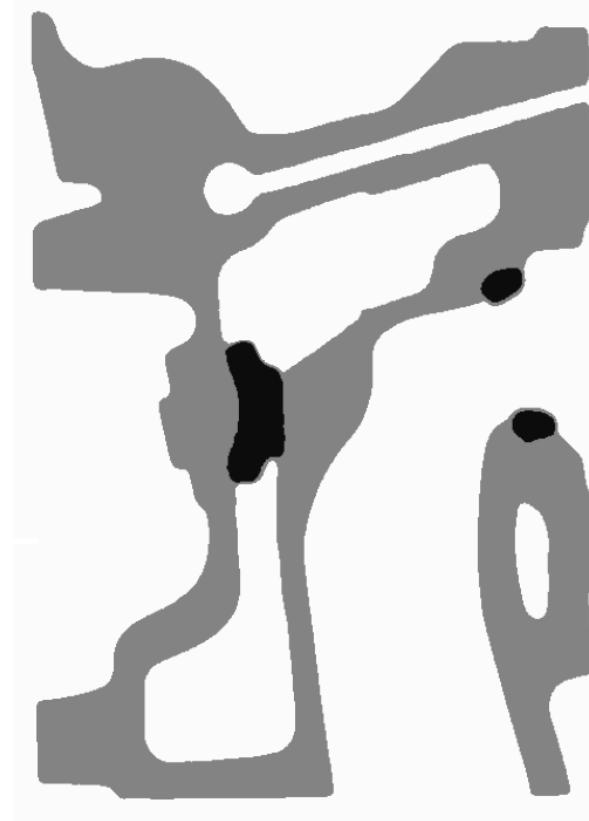
Challenges

- f convex but not quadratic
- non-smoothness due to constraints

2. Algebraic

Application Scenario: Discrete Tomography

3D image
with finite range



few projection angles

Challenges

- limited data scenarios: severely ill-posed problem
- intractable integrality constraint

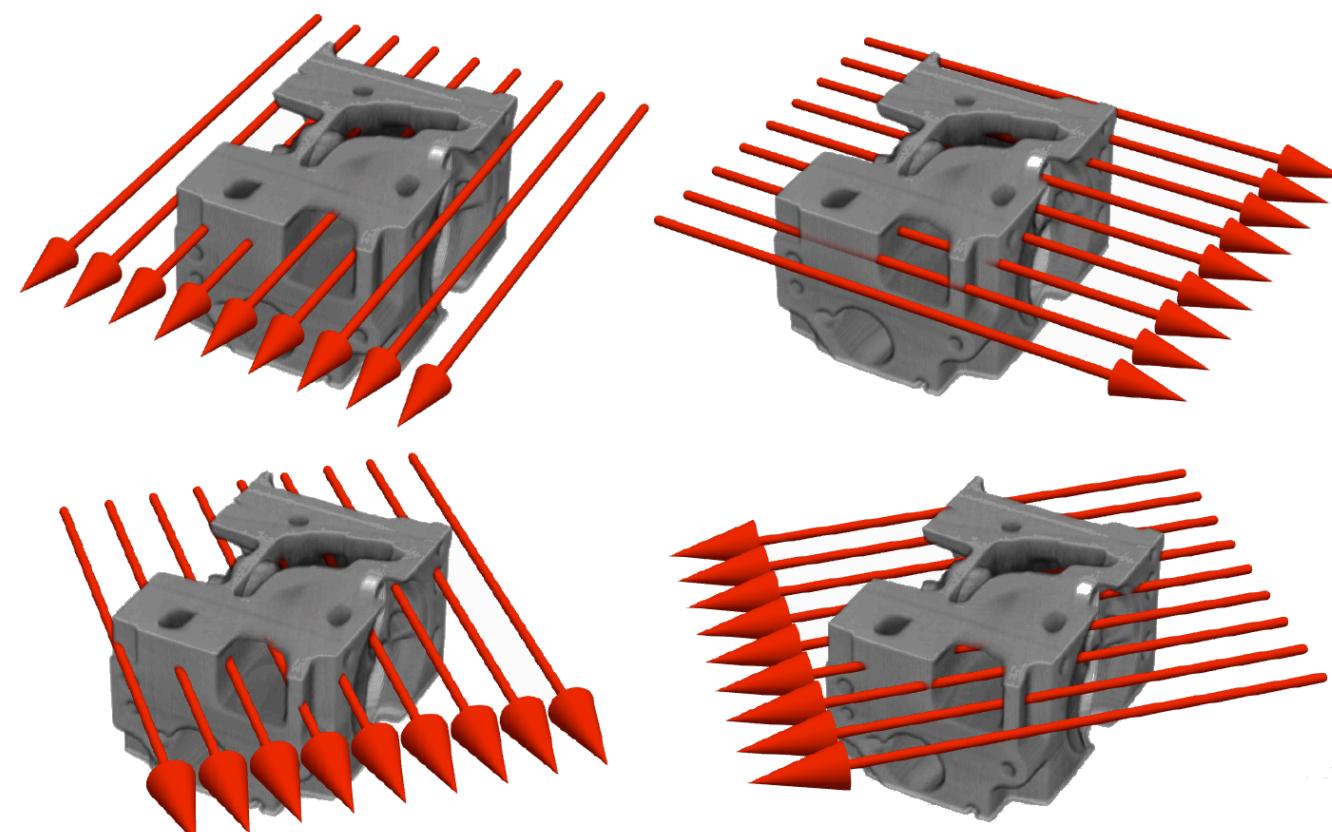
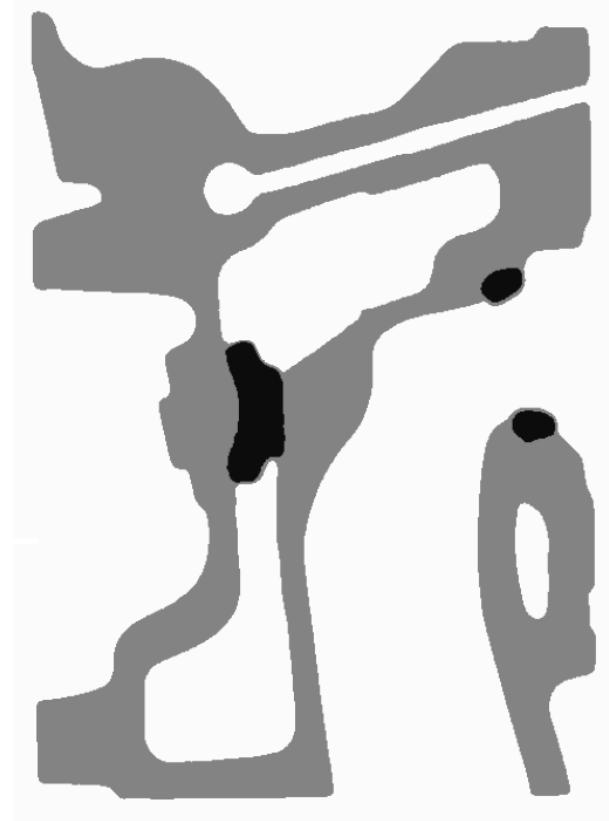
E.g. under suitable conditions (sparsity, under sampling)

$$\min_{x \in [0,1]} \|Ax - b\|^2 + \lambda \|\nabla x\|_1, \quad \lambda > 0$$

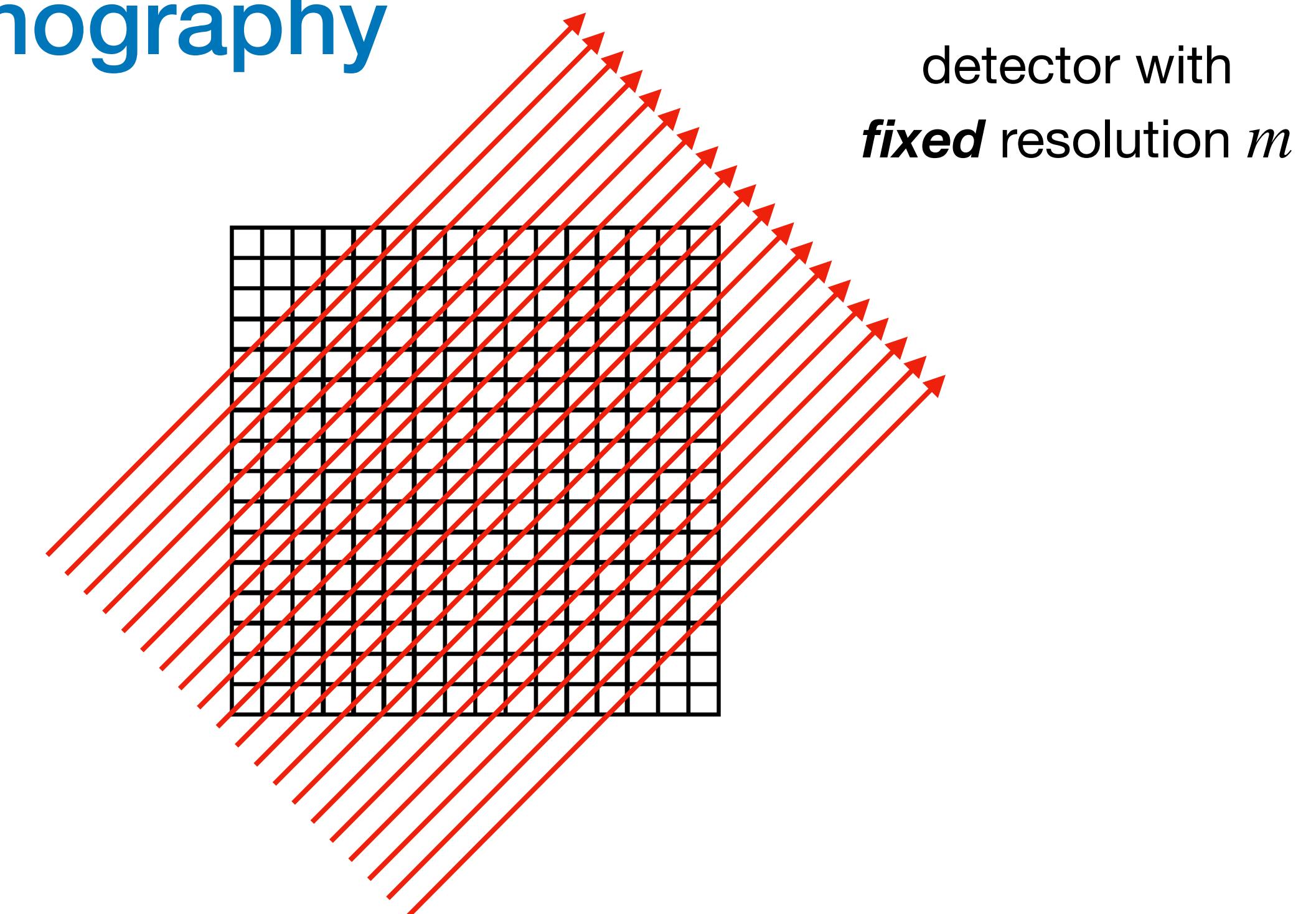
Kuske et al., **Performance Bounds For Co-/Sparse Box Constrained Signal Recovery**, 2019

Application Scenario: Discrete Tomography

3D image
with finite range



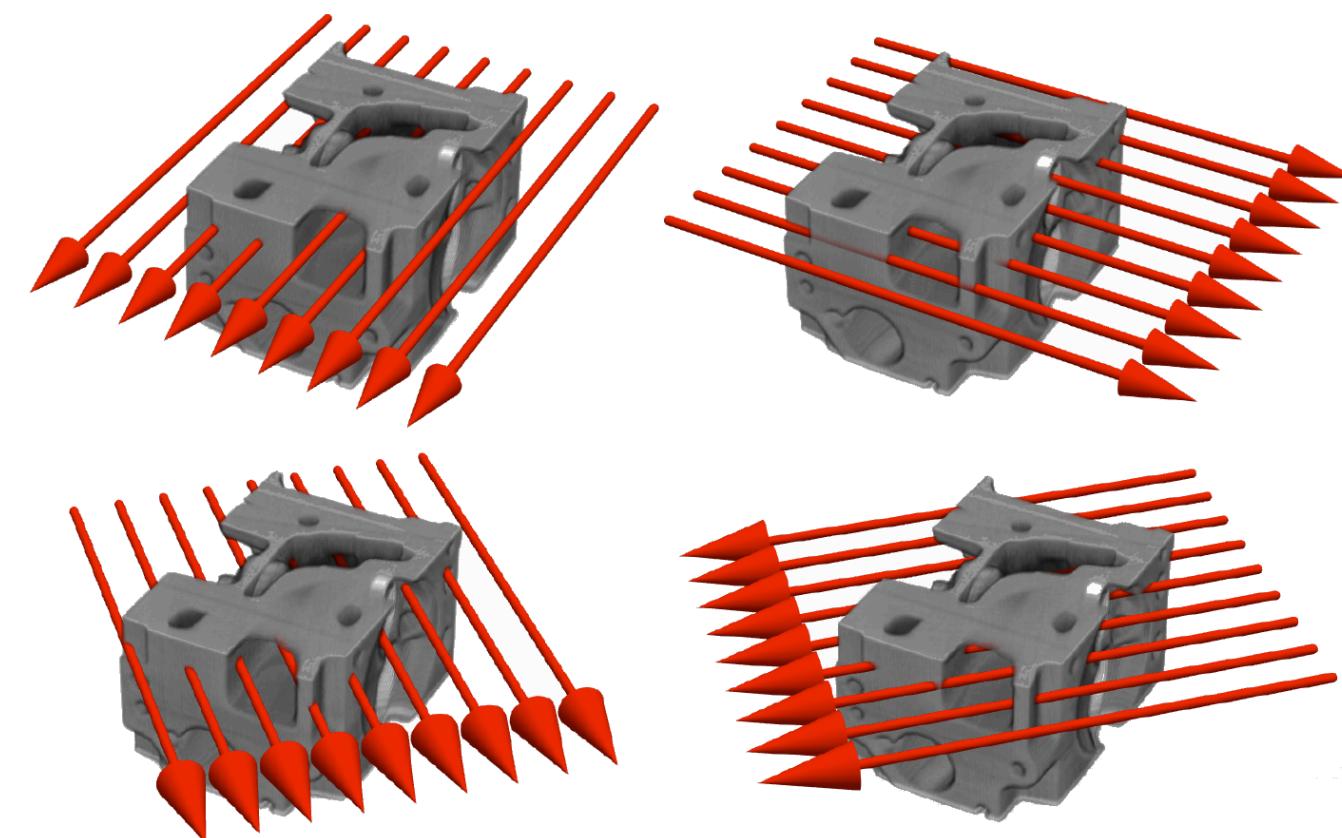
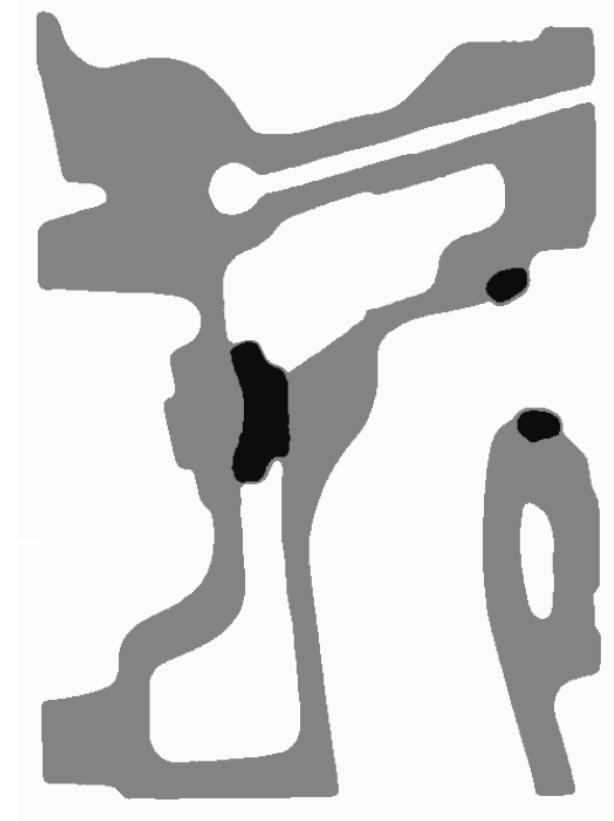
few projection angles



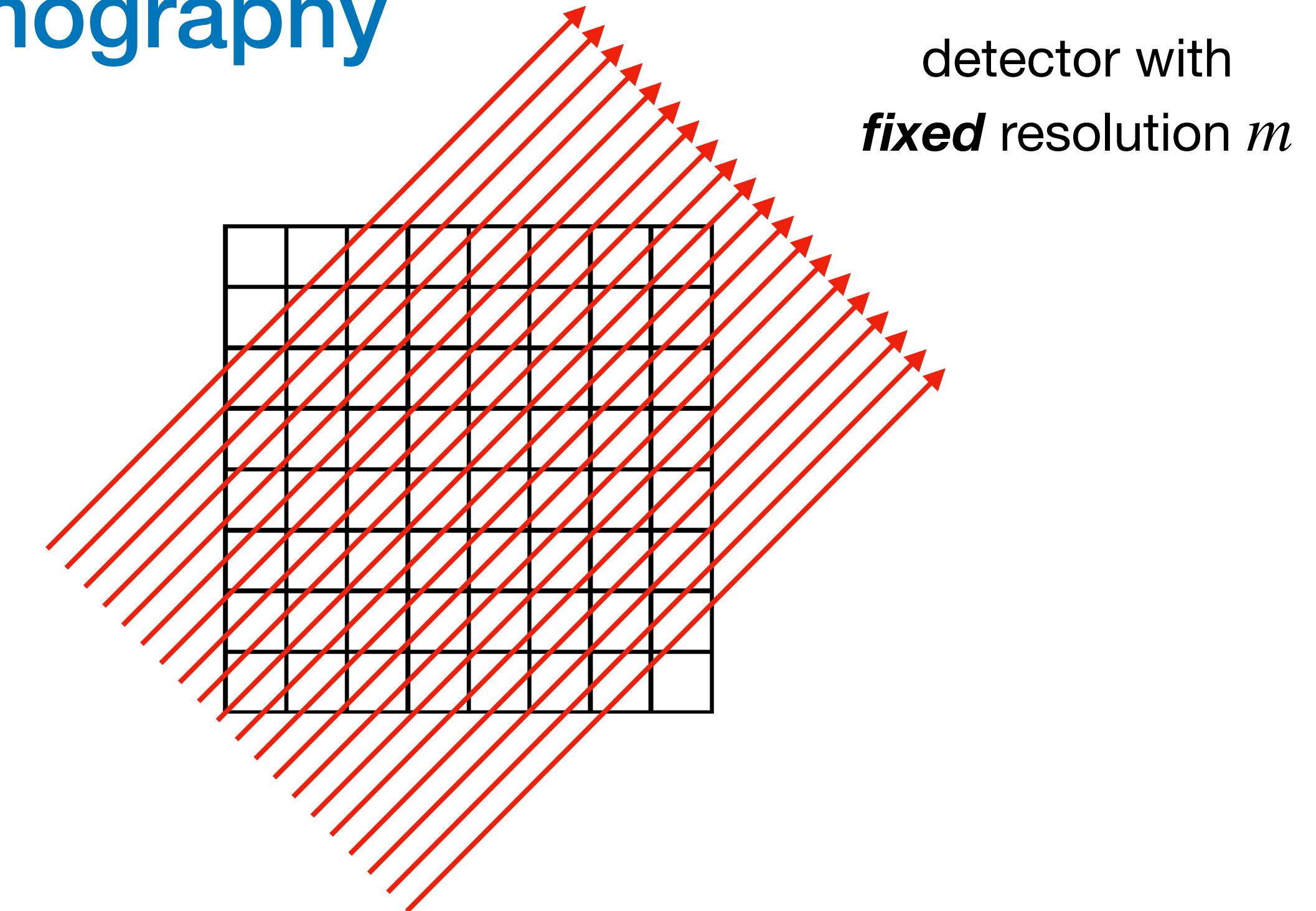
$$Ax = b, \quad A \in \mathbb{R}^{m \times n}, \quad m \ll n$$

Application Scenario: Discrete Tomography

3D image
with finite range



few projection angles

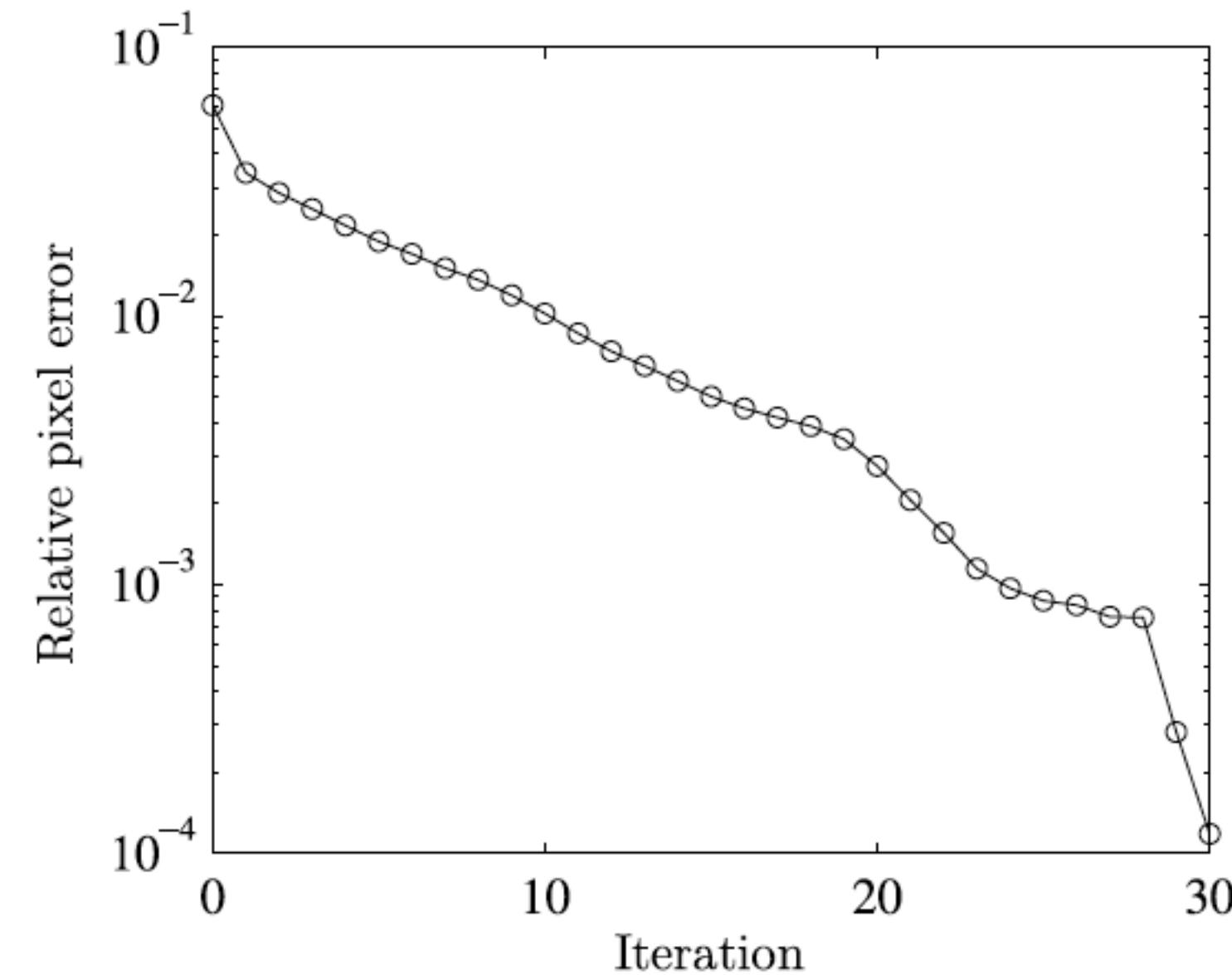


$$\bar{A}\bar{x} = b, \quad \bar{A} \in \mathbb{R}^{m \times \bar{n}}, \quad m/\bar{n} \quad \text{gets larger}$$

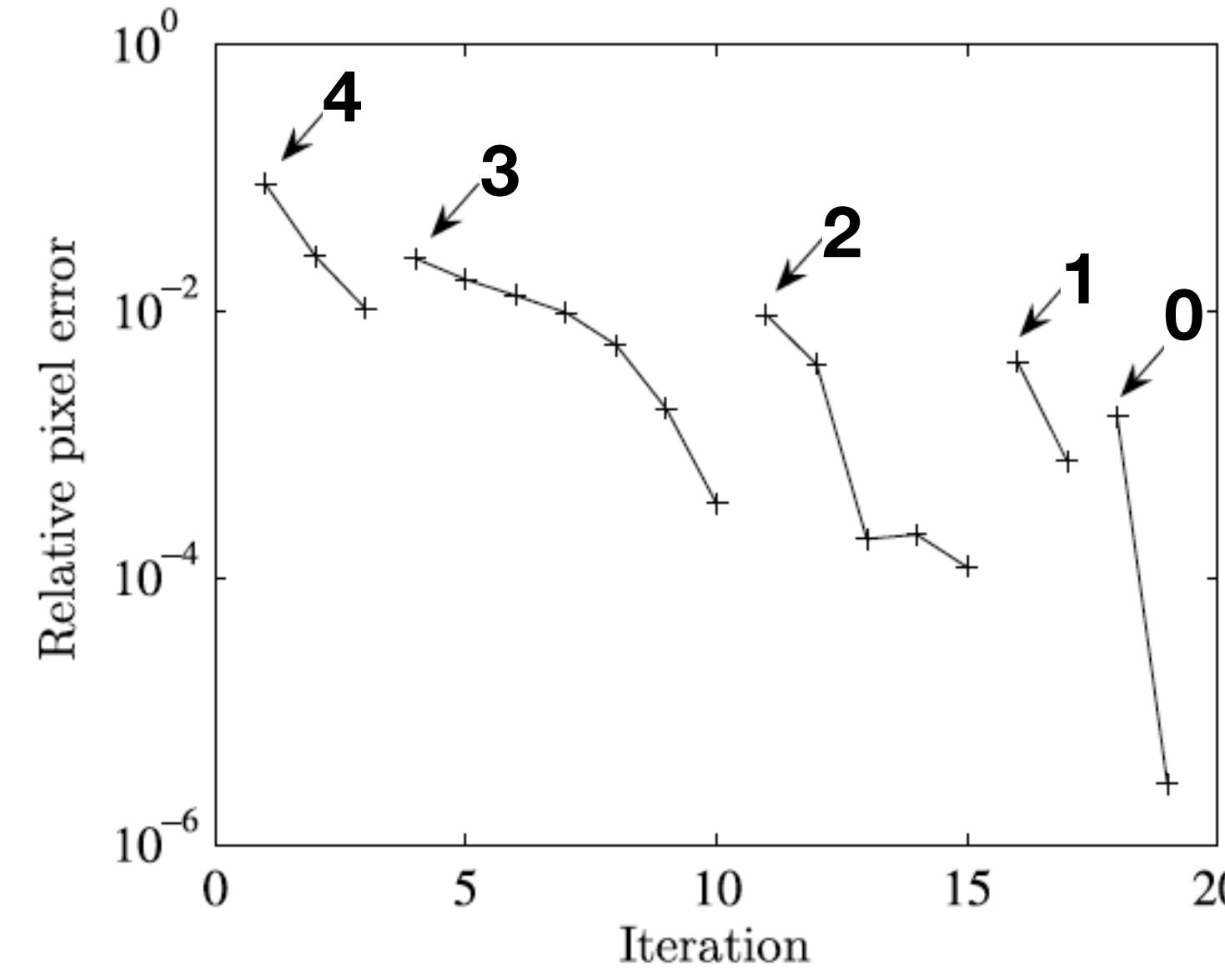
Related Work: Multiscale Acceleration for Discrete Tomography

S. Roux, H. Leclerc, F. Hild, **Efficient Binary Tomographic Reconstruction**, J Math Imaging Vis, 2014

A. Dabovolski, K. J. Batenburg, J. Sijbers, **A Multiresolution Approach to Discrete Tomography Using DART**, PLoS One. 2014



single scale



multiscale

- 0: fine scale 1024×1024
- 1: coarse scale 512×512
- 2: coarse scale 256×256
- 3: coarse scale 128×128
- 4: coarse scale 64×64

Challenges

- principled information transfer between levels

Related Work: Multilevel Optimization

smooth optimization

S. G. Nash, **A multigrid approach to discretized optimization problems.** Optimization Methods and Software, 2000

S. Gratton, A. Sartenaer, P. L. Toint, **Recursive trust-region methods for multiscale nonlinear optimization.** SIAM Journal on Optimization, 2008

W. Zaiwen, and D. Goldfarb, **A line search multigrid method for large-scale nonlinear optimization.** SIAM Journal on Optimization, 2009

C. P. Ho, M. Kocvara, P. Parpas, **Newton-type Multilevel Optimization Method.** Optimization Methods and Software, 2019

nonsmooth composite convex optimization

M. Kocvara and S. Mohammed, **A first-order multigrid method for bound-constrained convex optimization.** Optimization Methods and Software, 2016

P. Parpas, **A multilevel proximal gradient algorithm for a class of composite optimization problems.** SIAM J. Sci. Comput., 2017

First Order Geometric Multilevel Optimization For Discrete Tomography

Outline

Two-grid optimization

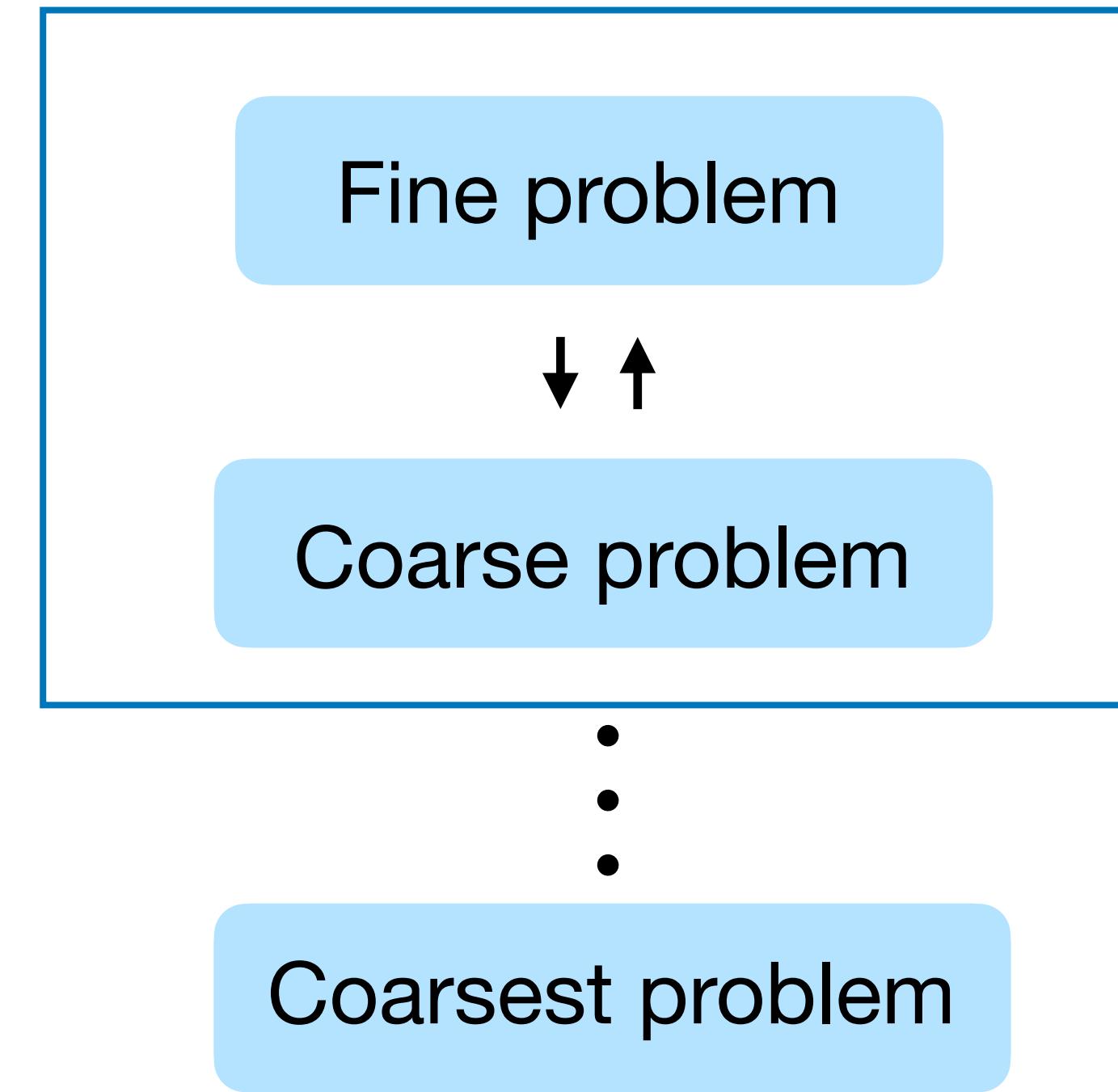
Connection to nonlinear multigrid (FAS)

Connection to standard local approximation

Geometric two-grid optimization and application to discrete tomography

Conclusion, outlook

Structure: Hierarchy of Grid Dependent Problems



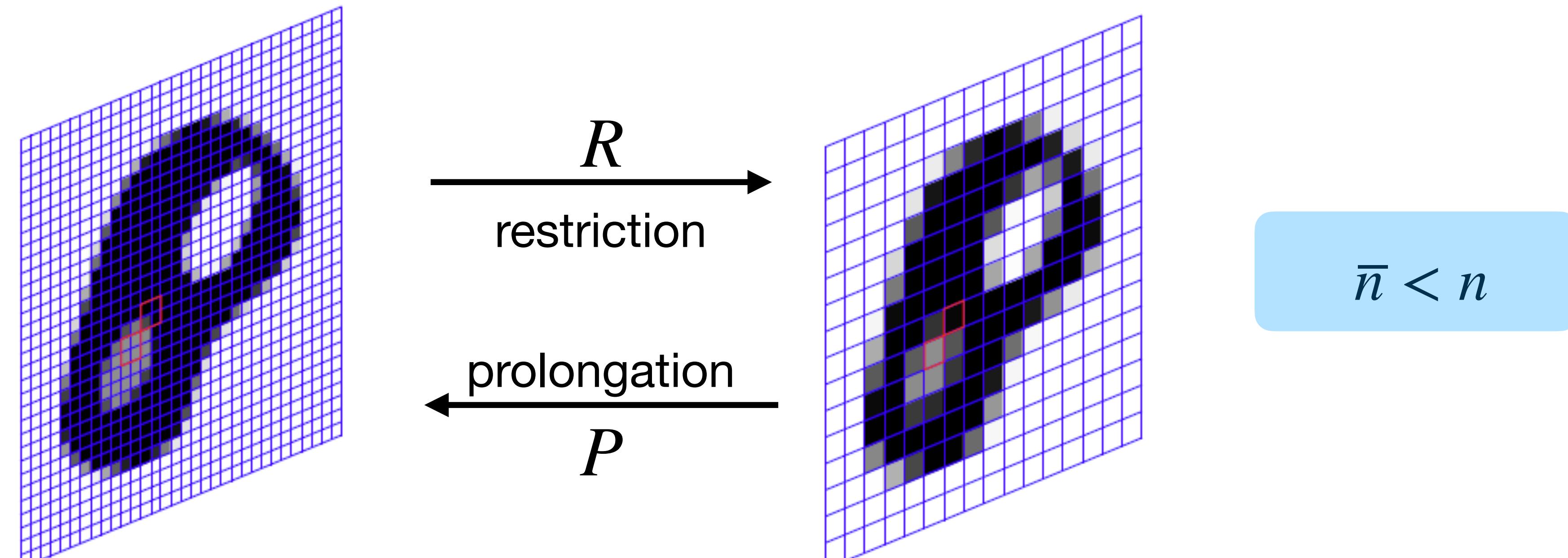
Two Grid Approach, Notation

fine grid variable $x \in \mathbf{R}^n$

fine objective $f \in C^1(\mathbf{R}^n, \mathbf{R})$

coarse grid variable $\bar{x} \in \mathbf{R}^{\bar{n}}$,

coarse discretization $\bar{f} \in C^1(\mathbf{R}^{\bar{n}}, \mathbf{R})$



intergrid transfer operators $R : \mathbf{R}^n \rightarrow \mathbf{R}^{\bar{n}}$ and $P : \mathbf{R}^{\bar{n}} \rightarrow \mathbf{R}^n$

Two Grid Approach, Coarse Model

for *current* fine grid variable $x \in \mathbf{R}^n$ define **coarse grid model**

$$\bar{\phi}(\bar{y}) = \bar{f}(\bar{y}) - \langle \bar{v}_x, \bar{y} \rangle$$

with $\bar{v}_x = \nabla \bar{f}(Rx) - R \nabla f(x)$

(Nash, 2000, Gratton et al., 2008, Wen and Goldfarb, 2009)

Two Grid Approach, Coarse Model

for *current* fine grid variable $x \in \mathbf{R}^n$ define **coarse grid model**

$$\bar{\phi}(\bar{y}) = \bar{f}(\bar{y}) - \langle \bar{v}_x, \bar{y} \rangle$$

with $\bar{v}_x = \nabla \bar{f}(Rx) - R \nabla f(x)$

first order coherence condition

$$\nabla \bar{\phi}(\bar{x}) = R \nabla f(x), \quad \boxed{\bar{x} = Rx}$$



starting iterate at coarse level¹³

(Nash, 2000, Gratton et al., 2008, Wen and Goldfarb, 2009)

Two Grid Approach, Coarse Model

for *current* fine grid variable $x \in \mathbf{R}^n$, set $\bar{x} = Rx$ and rewrite define **coarse grid model**

$$\begin{aligned}\bar{\phi}(\bar{x} + \bar{y}) &= \bar{f}(\bar{x} + \bar{y}) - \langle \nabla \bar{f}(\bar{x}) - R \nabla f(x), \bar{x} + \bar{y} \rangle, \\ &= \bar{f}(\bar{x} + \bar{y}) - \bar{f}(\bar{x}) - \langle \nabla \bar{f}(\bar{x}) - R \nabla f(x), \bar{x} + \bar{y} \rangle + \bar{f}(\bar{x}), \\ &= D_{\bar{f}}(\bar{x} + \bar{y}, \bar{x}) + \langle R \nabla f(x), \bar{y} \rangle + \cancel{\text{const}},\end{aligned}$$

with Bregman gap

$$D_{\bar{f}}(\bar{x} + \bar{y}, \bar{x}) = \bar{f}(\bar{x} + \bar{y}) - \bar{f}(\bar{x}) - \langle \nabla \bar{f}(\bar{x}), \bar{y} \rangle$$

Two Grid Approach, Coarse Model

for *current* fine grid variable $x \in \mathbf{R}^n$, set $\bar{x} = Rx$ and rewrite define **coarse grid model**

$$\begin{aligned}\bar{\phi}(\bar{x} + \bar{y}) &= D_{\bar{f}}(\bar{x} + \bar{y}, \bar{x}) + \langle R \nabla f(x), \bar{y} \rangle \\ &= D_{\bar{f}}(\bar{x} + \bar{y}, \bar{x}) + \langle \nabla f(x), P \bar{y} \rangle\end{aligned}$$

↑

$$R = P^\top$$

with Bregman gap

$$D_{\bar{f}}(\bar{x} + \bar{y}, \bar{x}) = \bar{f}(\bar{x} + \bar{y}) - \bar{f}(\bar{x}) - \langle \nabla \bar{f}(\bar{x}), \bar{y} \rangle$$

Two Grid Approach, Coarse Model

for *current* fine grid variable $x \in \mathbf{R}^n$, set $\bar{x} = Rx$

$$\bar{\phi}(\bar{x} + \bar{y}) = \underbrace{D_{\bar{f}}(\bar{x} + \bar{y}, \bar{x})}_{\geq 0} + \langle \nabla f(x), P\bar{y} \rangle \quad R = P^\top$$

if \bar{f} convex

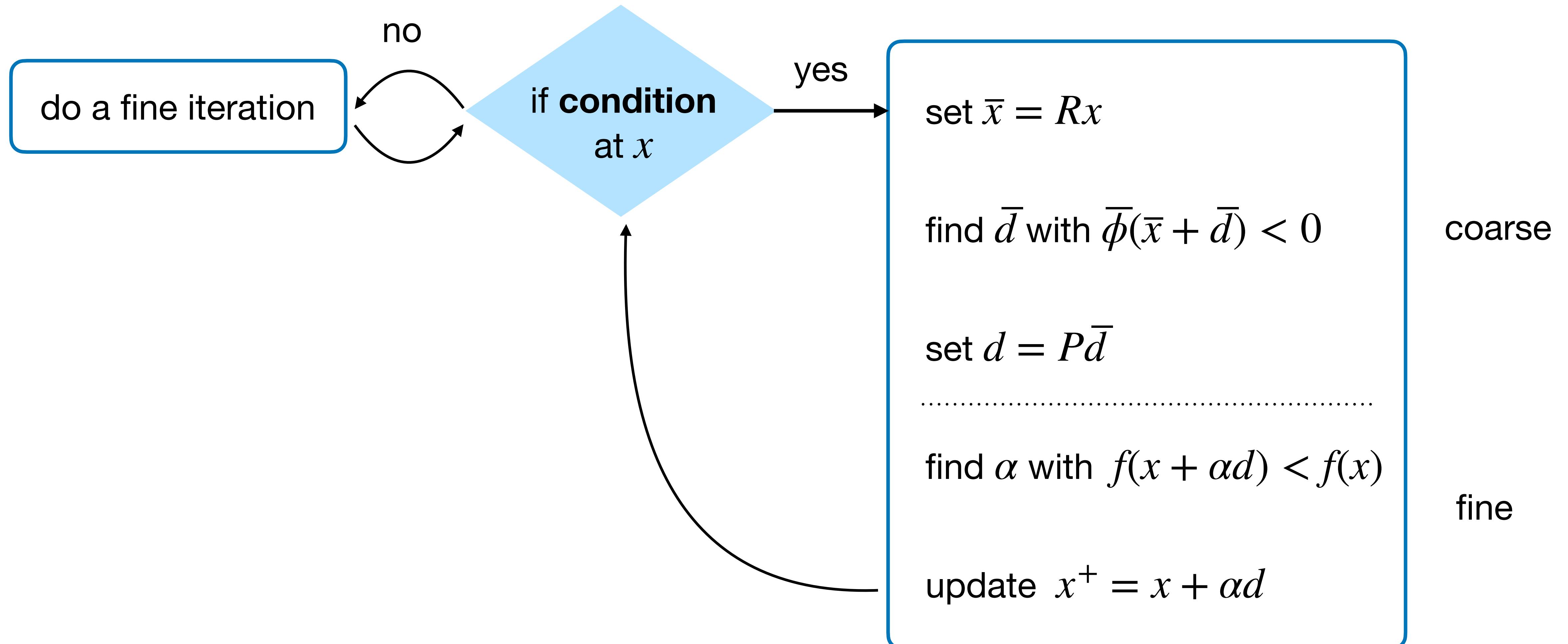
$$\bar{\phi}(\bar{x} + \bar{d}) < 0 \implies \langle \nabla f(x), d \rangle < 0, \quad d = P\bar{d}$$



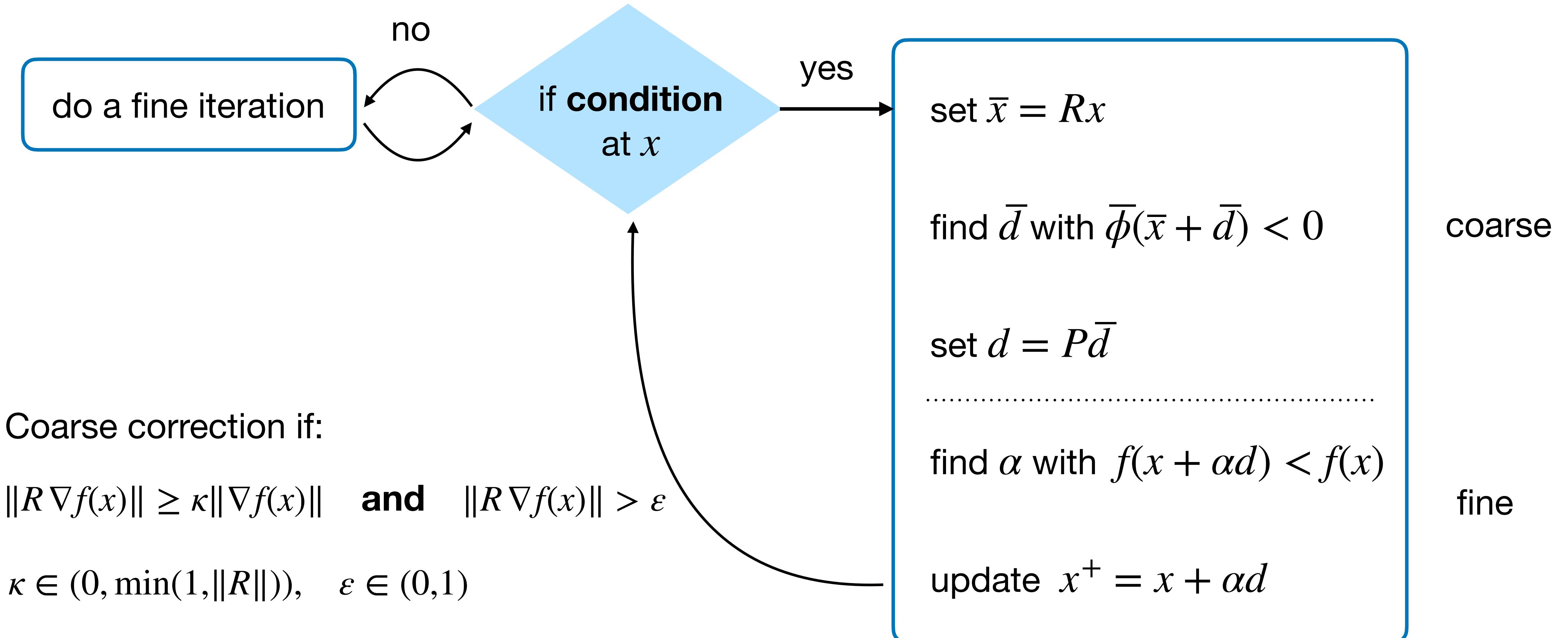
16

descent direction for f

Two Grid Approach, Coarse Model



Two Grid Approach, Coarse Model



First Order Geometric Multilevel Optimization For Discrete Tomography

Outline

Two-grid optimization

Connection to nonlinear multigrid (FAS)

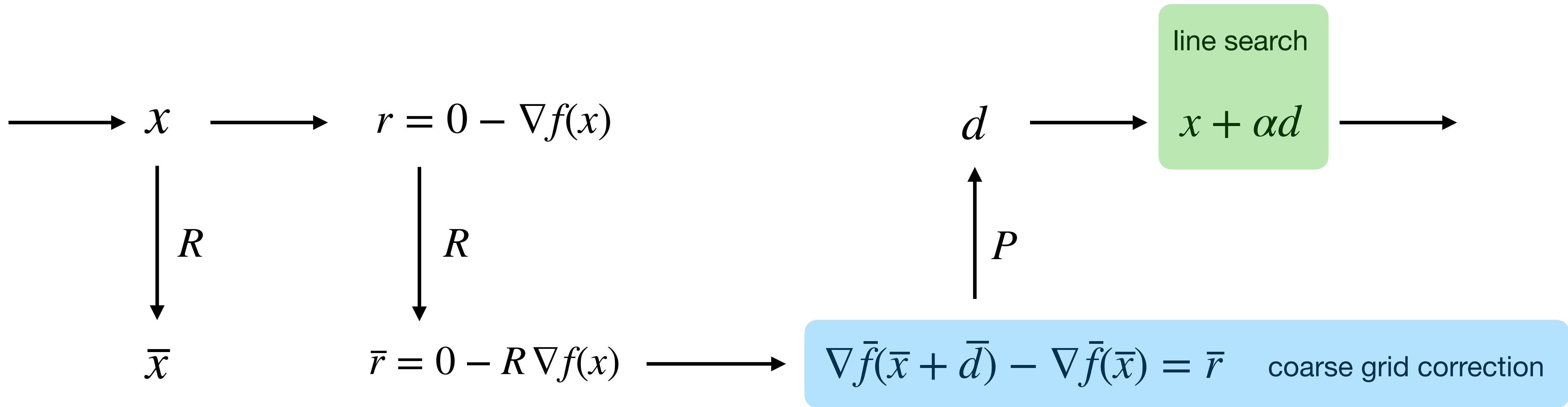
Connection to standard local approximation

Geometric two-grid optimization and application to discrete tomography

Conclusion, outlook

Full Approximation Scheme

$$\nabla f(x^*) = 0 \quad \text{nonlinear equation}$$



inexact optimization

$$\begin{aligned} 0 > \overset{!}{\bar{\phi}}(\bar{x} + \bar{d}) &= \bar{f}(\bar{x} + \bar{d}) - \langle \nabla \bar{f}(\bar{x}) - R \nabla f(x), \bar{d} \rangle + C(\bar{x}) \\ &= D_{\bar{f}}(\bar{x} + \bar{d}, \bar{x}) + \langle R \nabla f(x), \bar{d} \rangle + C'(\bar{x}) \end{aligned}$$

First Order Geometric Multilevel Optimization For Discrete Tomography

Outline

Two-grid optimization

Connection to nonlinear multigrid (FAS)

Connection to standard local approximation

Geometric two-grid optimization and application to discrete tomography

.....

Conclusion, outlook

Connection to Standard Local Approximation

“hierarchical approximation”, multilevel optimization

$$\bar{\phi}(\bar{x} + \bar{y}) = \langle \nabla f(x), P\bar{y} \rangle + D_{\bar{f}}(\bar{x} + \bar{y}, \bar{x})$$

1st order approximation



$$D_{\bar{f}}(\bar{x} + \bar{y}, \bar{x})$$

$$= \frac{1}{2} \langle \bar{y}, \nabla^2 \bar{f}(z), \bar{y} \rangle,$$

$$z \in \{(1-t)(\bar{x} + \bar{y}) + t\bar{y}\}_{t \in [0,1]}$$

2nd order approximation

$$q(x + y) = f(x) + \langle \nabla f(x), y \rangle + \frac{1}{2} \langle y, B_x y \rangle,$$

$$B_x \succ 0$$

quadratic approximation, quasi Newton

First Order Geometric Multilevel Optimization For Discrete Tomography

Outline

Two-grid optimization

Connection to nonlinear multigrid (FAS)

Connection to standard local approximation

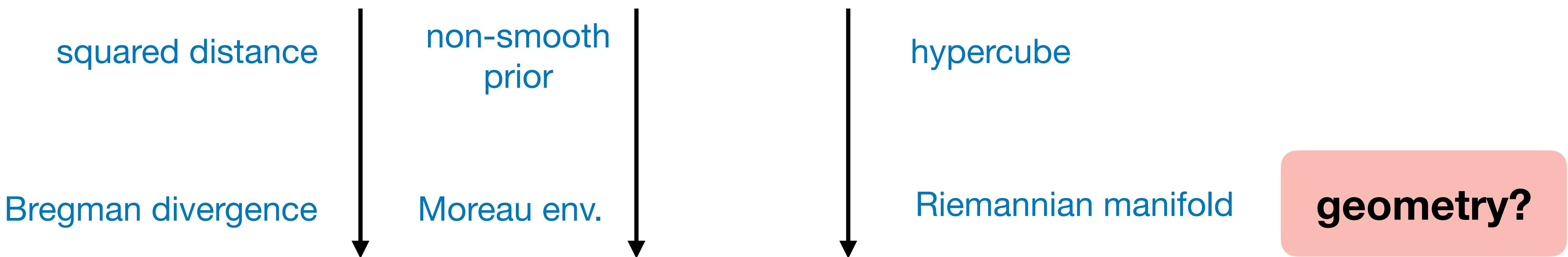
Geometric two-grid optimization and application to discrete tomography

Conclusion, outlook

Smooth Bound Constrained Convex Optimization

least-squares with sparsity prior and box constraints

$$\|Ax - b\|^2 + \lambda \|\nabla x\|_1, \quad x \geq \mathbf{0}, \quad x \leq \mathbf{1} \quad \lambda > 0$$



geometry!

$$f(x) = \text{KL}(Ax, b) + \lambda \|\nabla x\|_{1,\tau}, \quad x \in ((0,1)^n, h) = (\mathcal{M}, h), \quad \lambda > 0$$

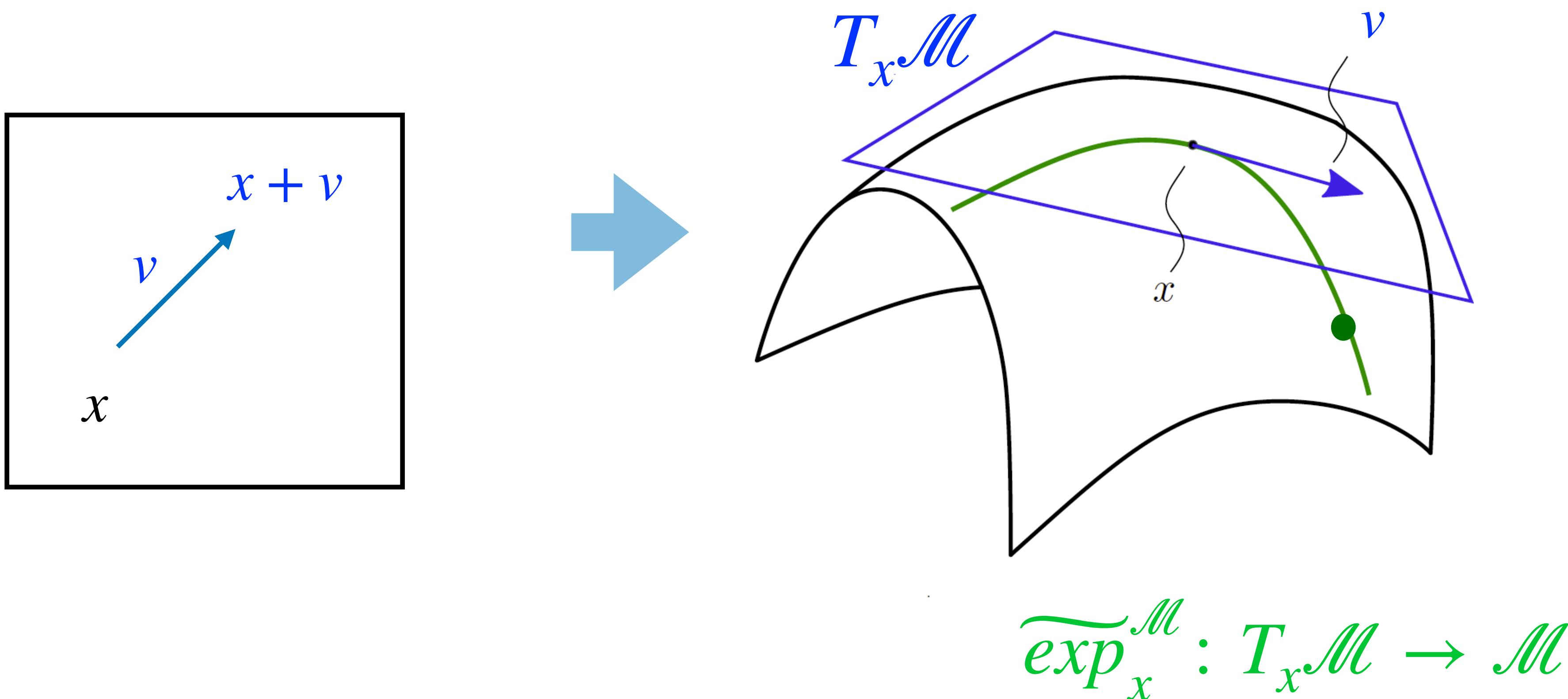
smooth non-quadratic data term with sparsity prior

$$\text{KL}(x, y) = \sum_{i \in [n]} \left(x_i \log \frac{x_i}{y_i} + y_i - x_i \right)$$

24

positive x : positive (non-normalised) discrete measure

Generalize Algorithmic Operations to Riemannian Manifolds



Geometry of the Box

Plier, Savarino, Kocvara, P., SSVM, 2021

taking into account **constraints**: $l \leq x \leq u$ (point-wise)

\implies change the Riemannian **metric**

\implies devise a **retraction** for first-order optimization

metric

convex Legendre-type function

Alvarez, Bolte, Brahic, SIAM J Control Optim, 2004

$$\varphi(x) = \langle u - l, (x - l)\log(x - l) + (u - x)\log(u - x) \rangle$$

$$h_x(v, w) = \langle v, \nabla^2 \varphi(x) w \rangle$$

retraction

$$\begin{array}{ccc} & \longleftrightarrow & \\ \mathcal{M} = ((l_i, u_i), h) & \xrightarrow{F_i} & \mathcal{N} = (\dot{\Delta}_2, g_{FR}) \\ & \xleftarrow{F_i^{-1}} & \end{array}$$

pullback

$$\widetilde{\exp}_x^{\mathcal{M}}: T_x \mathcal{M} \rightarrow \mathcal{M}, \quad \widetilde{\exp}_x^{\mathcal{M}}(v)_i = F_i^{-1} \left(\widetilde{\exp}_{F_i(x)}^{\mathcal{N}}(d_x F_i(v)) \right)$$

information geometry / e-connection

$$\text{Exp}_x = \frac{x \circ e^{\frac{v}{x}}}{\langle x, e^{\frac{v}{x}} \rangle}$$

Geometry of the Box

Plier, Savarino, Kocvara, P., SSVM, 2021

metric

$$h_x(v, w) = \langle v, \nabla^2 \varphi(x) w \rangle =: \langle v, H_x w \rangle$$

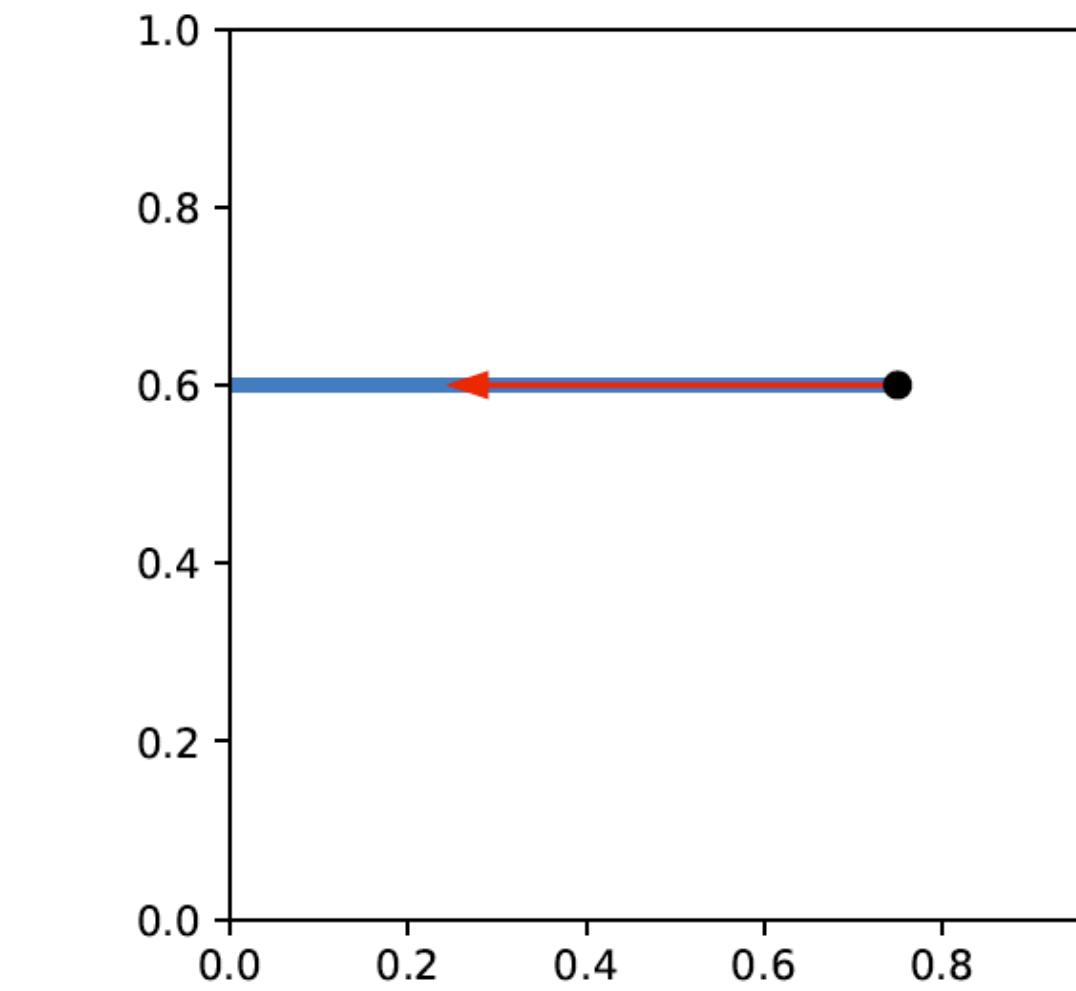
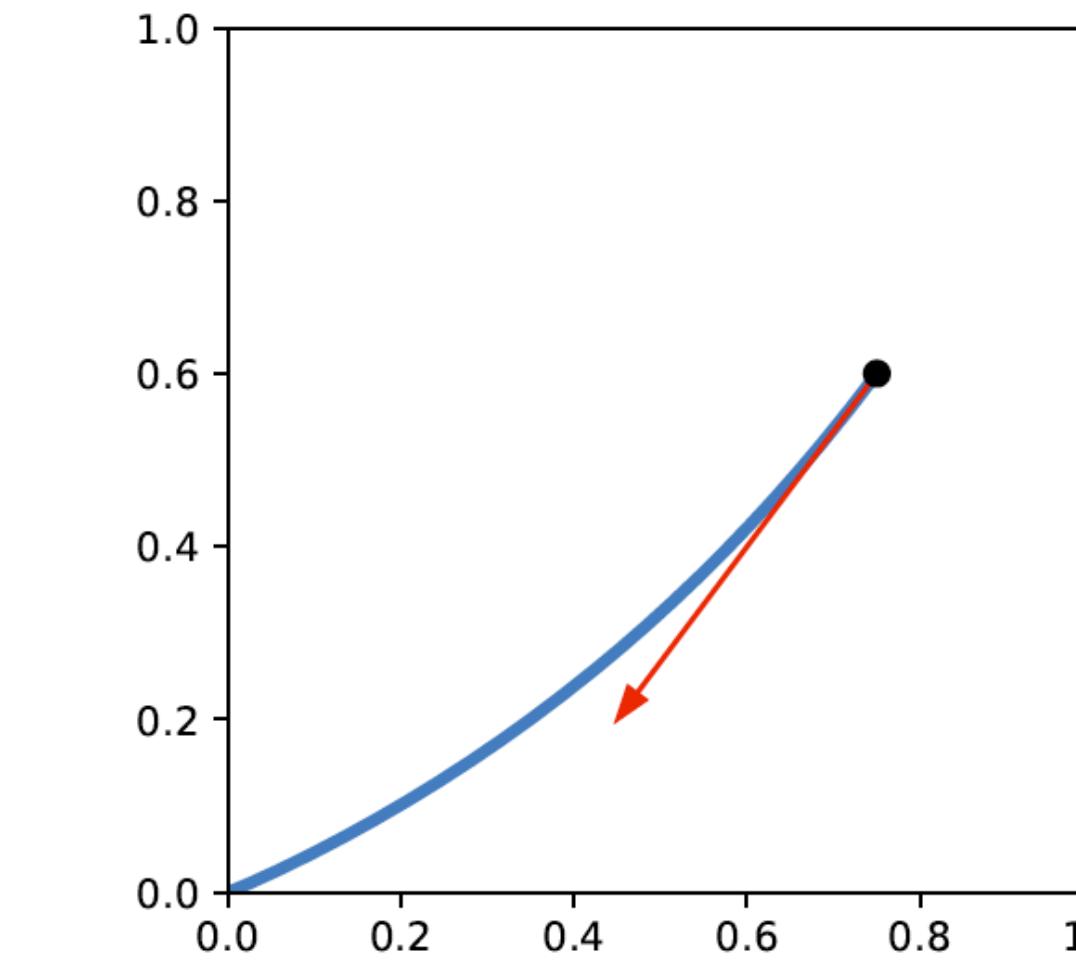
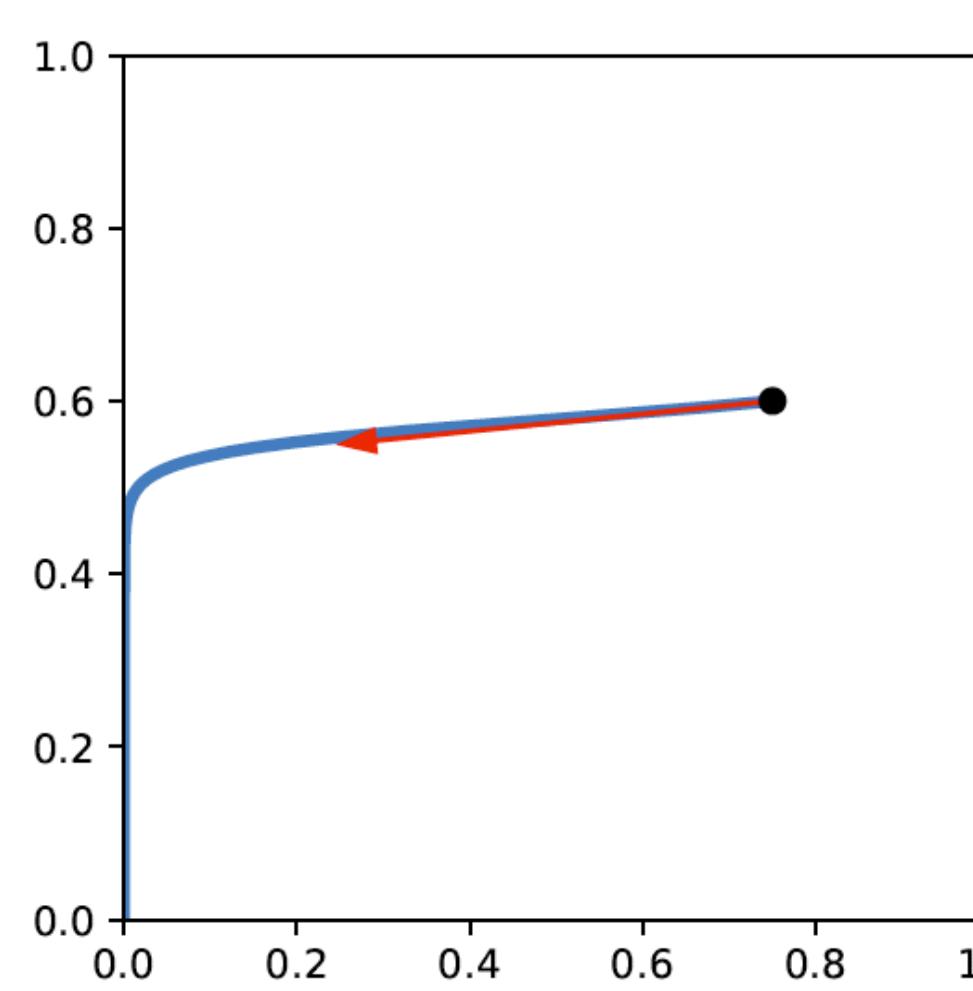
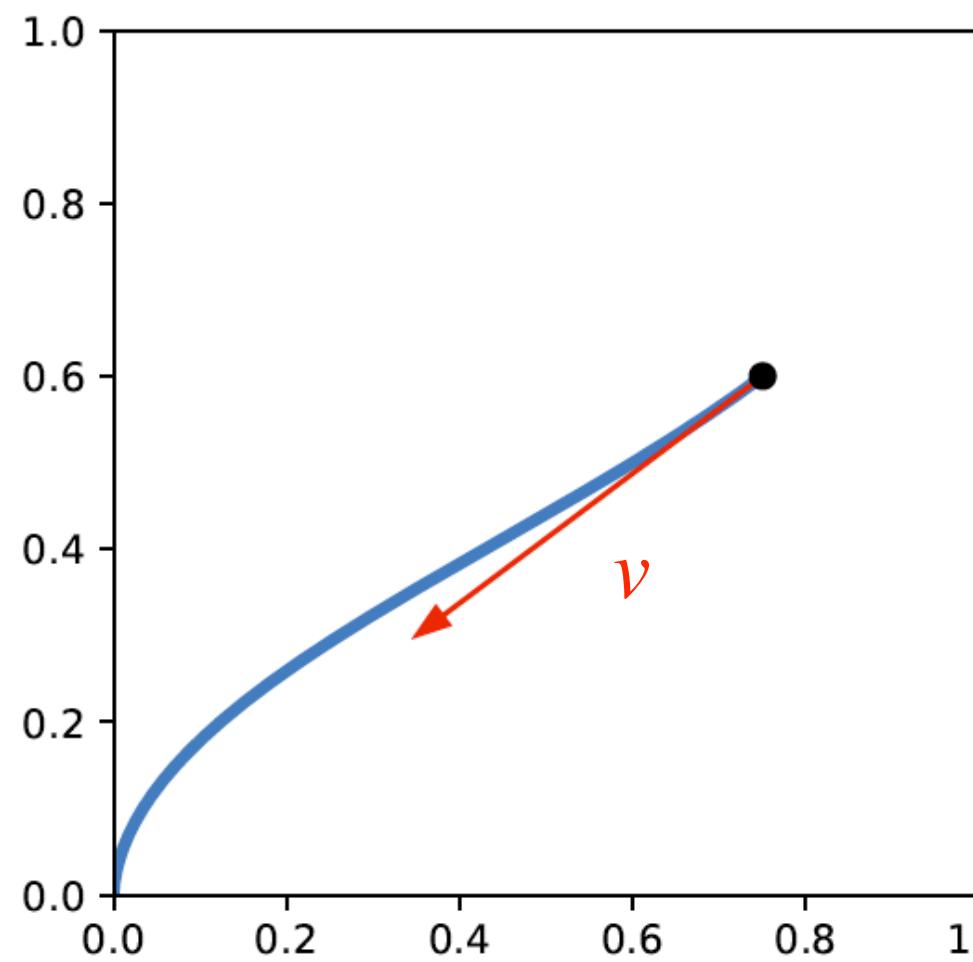
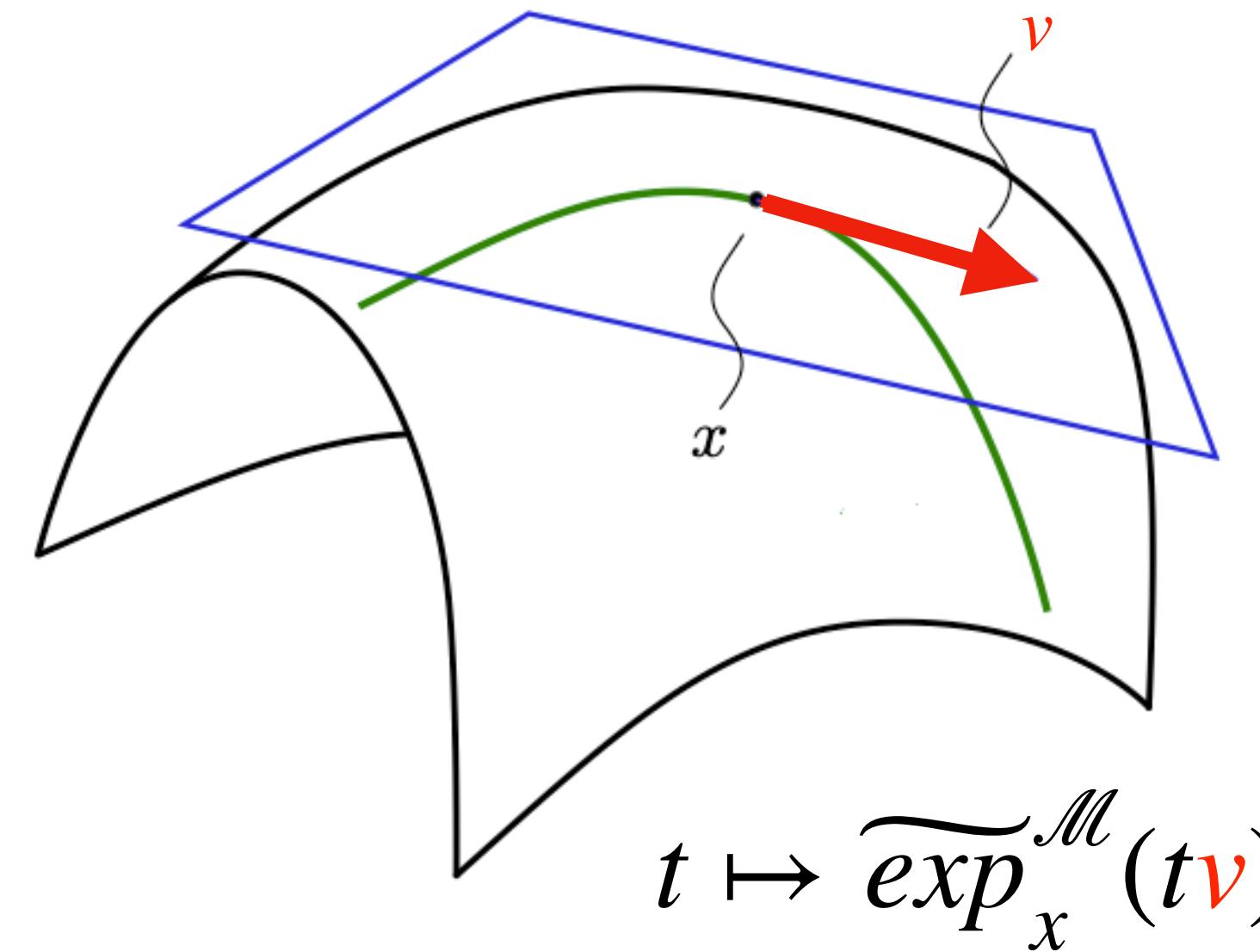
$$\nabla_{\mathcal{M}} f(x) = H_x^{-1} \nabla f(x) = \frac{(x - l)(u - x)}{(u - l)^2} \nabla f(x)$$

Riemannian gradient

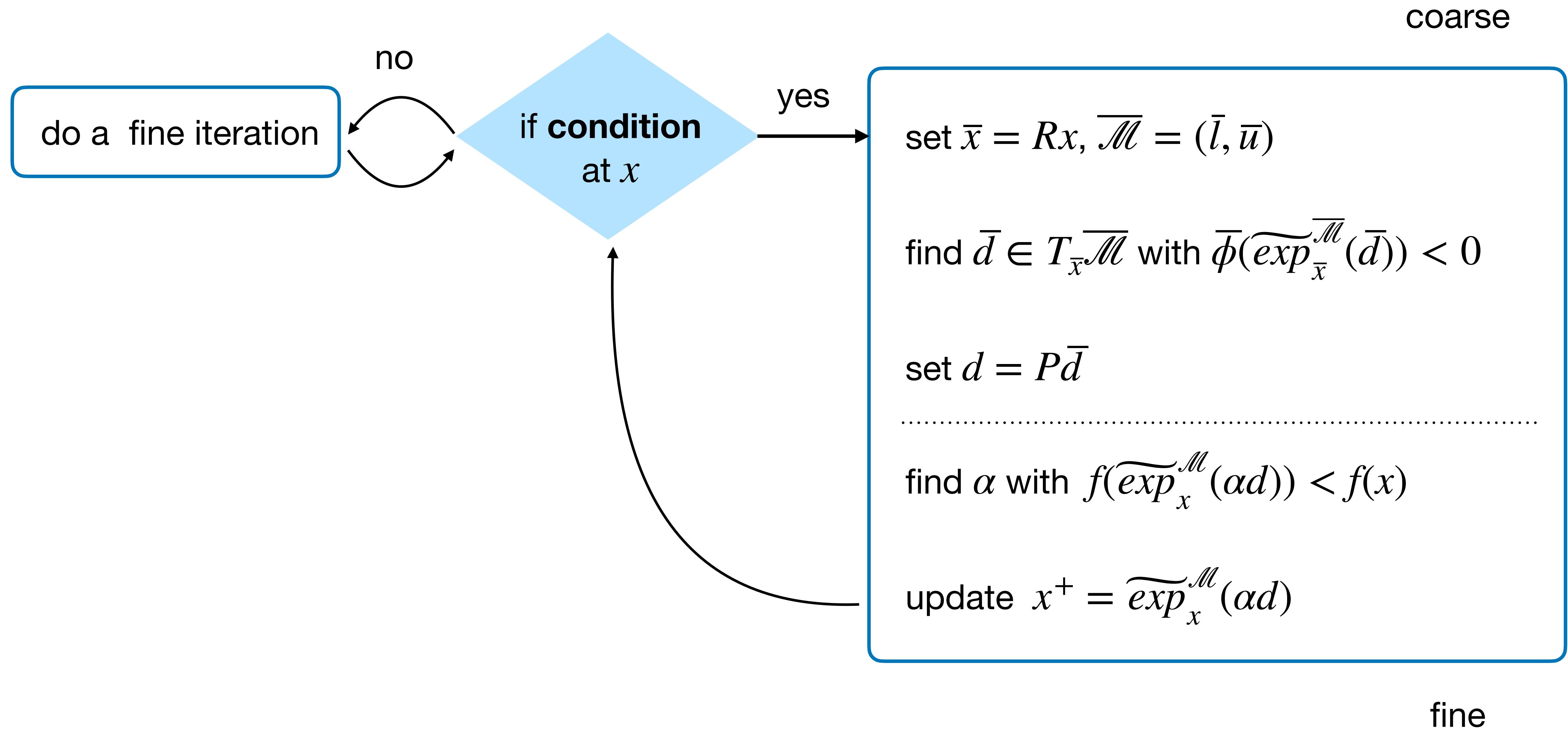
retraction

$$\widetilde{\exp}_x^{\mathcal{M}} : T_x \mathcal{M} \rightarrow \mathcal{M}, \quad \widetilde{\exp}_x^{\mathcal{M}}(v) = l + \frac{(u - l)(x - l)e^{\frac{u - l}{(x - l)(u - x)}v}}{(u - x) + (x - l)e^{\frac{u - l}{(x - l)(u - x)}v}}$$

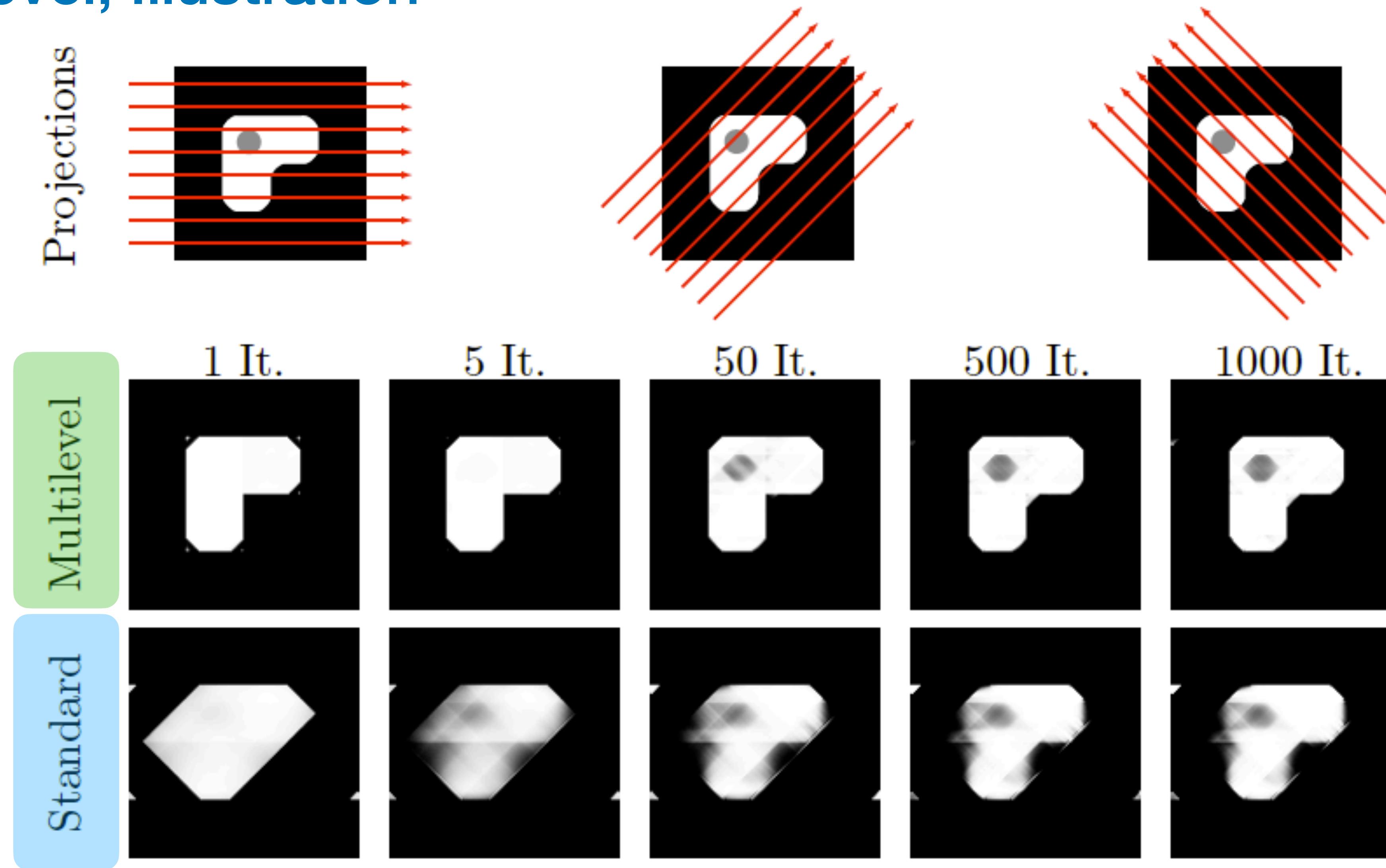
Retraction, Illustration



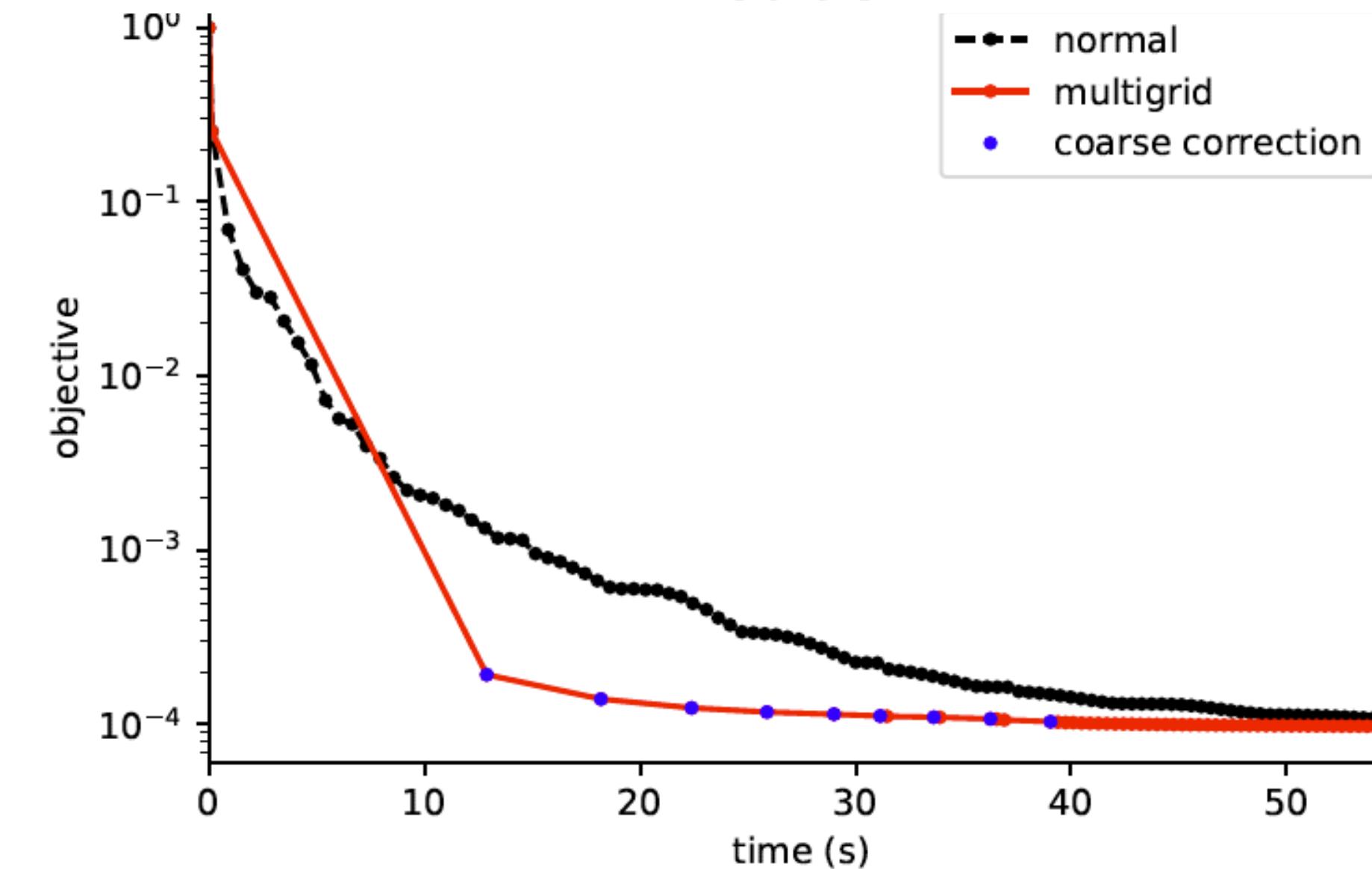
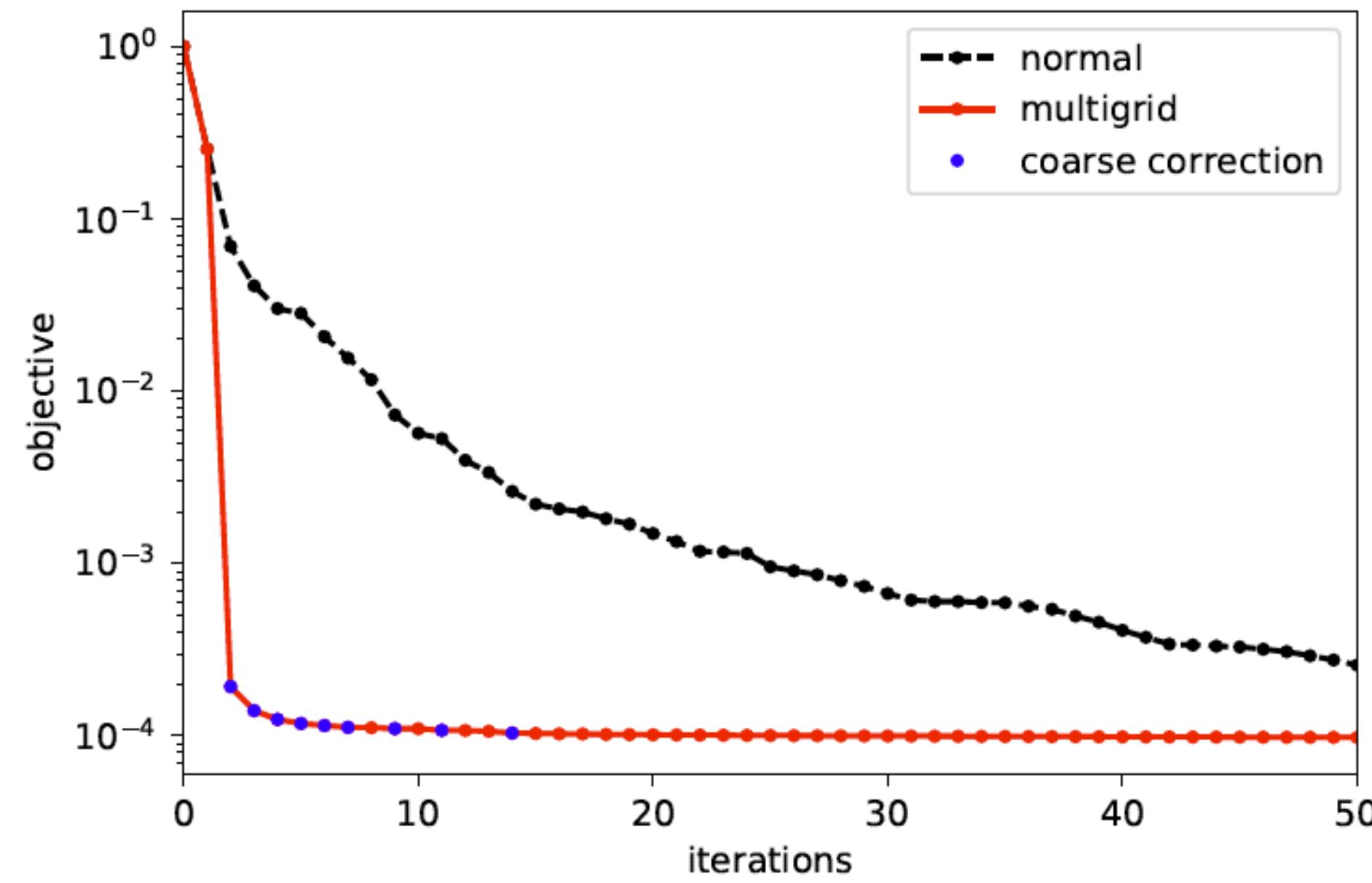
Two Grid Approach, Coarse Model



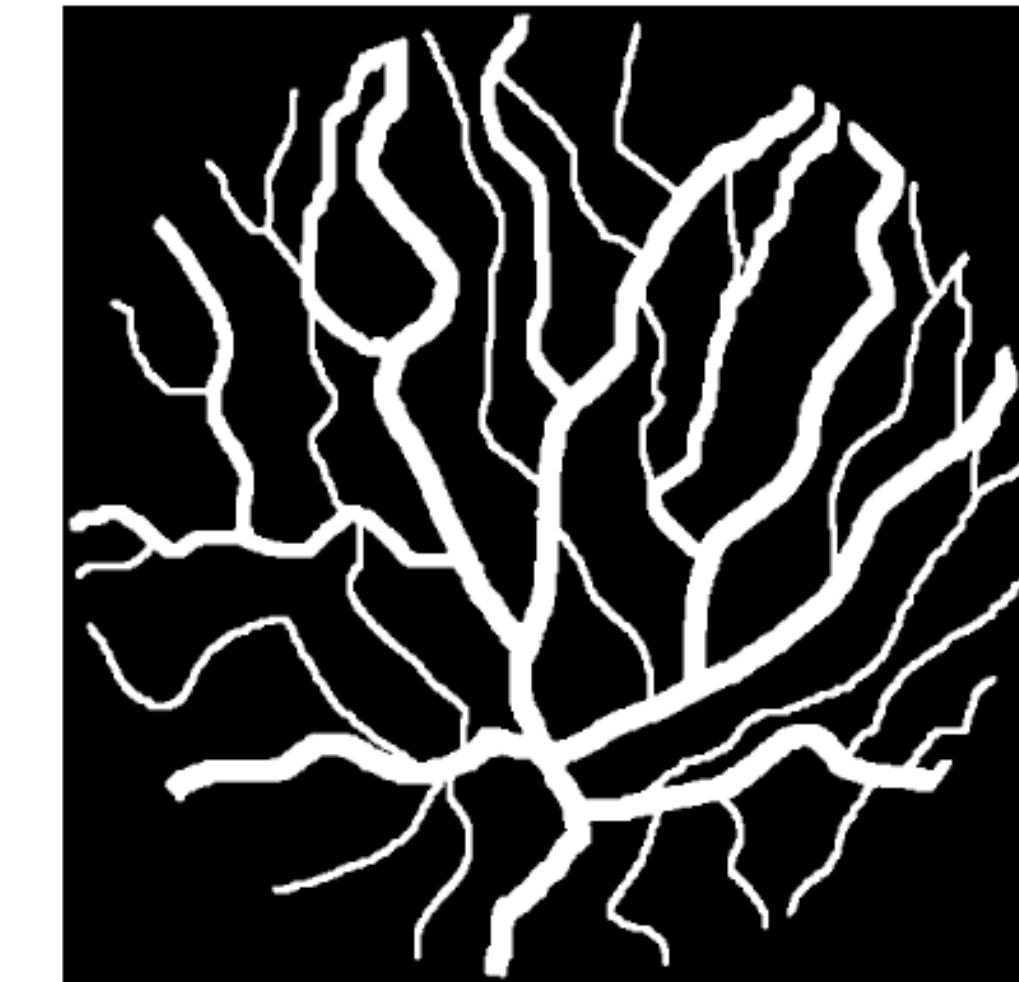
Multilevel, Illustration



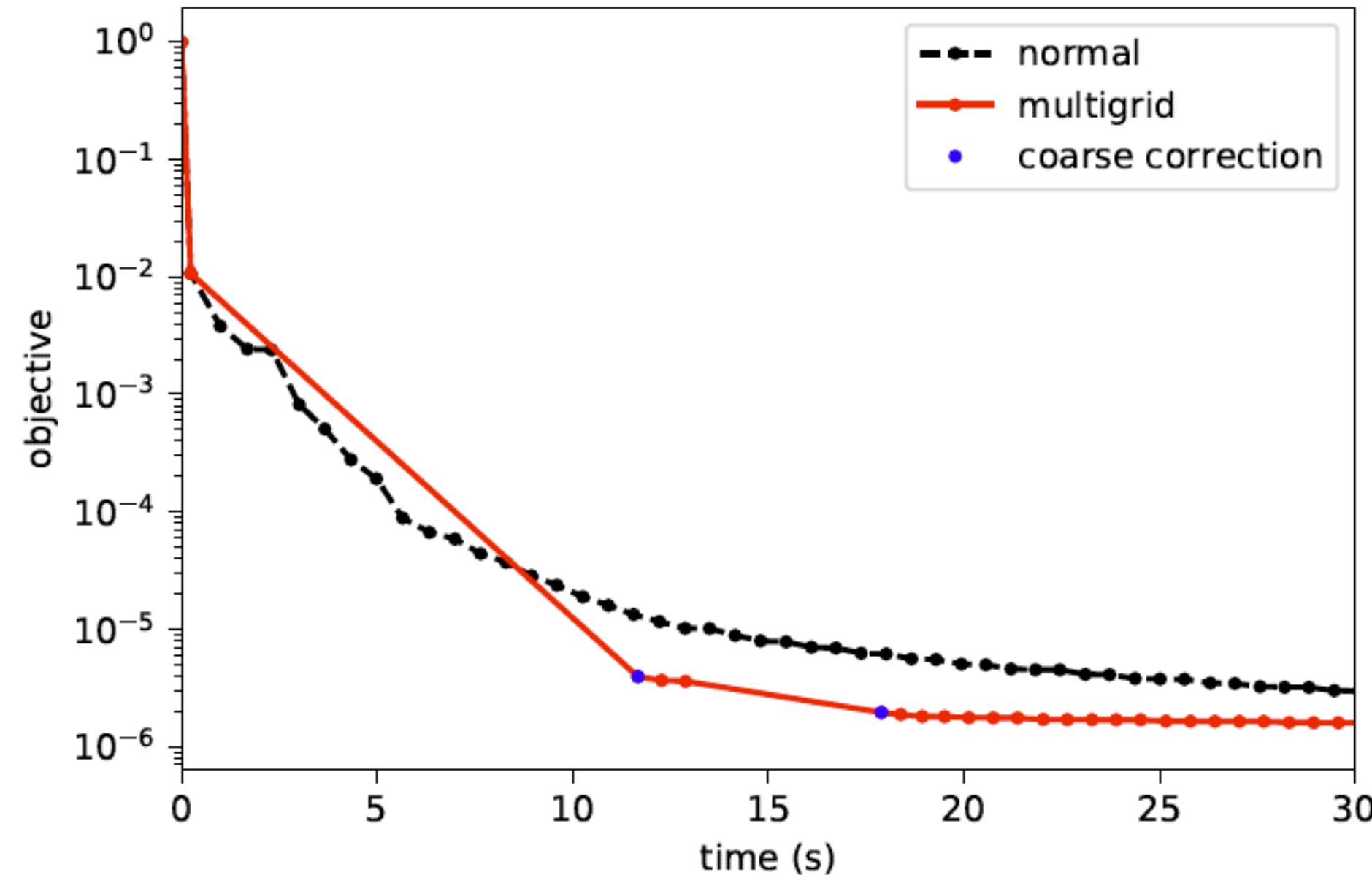
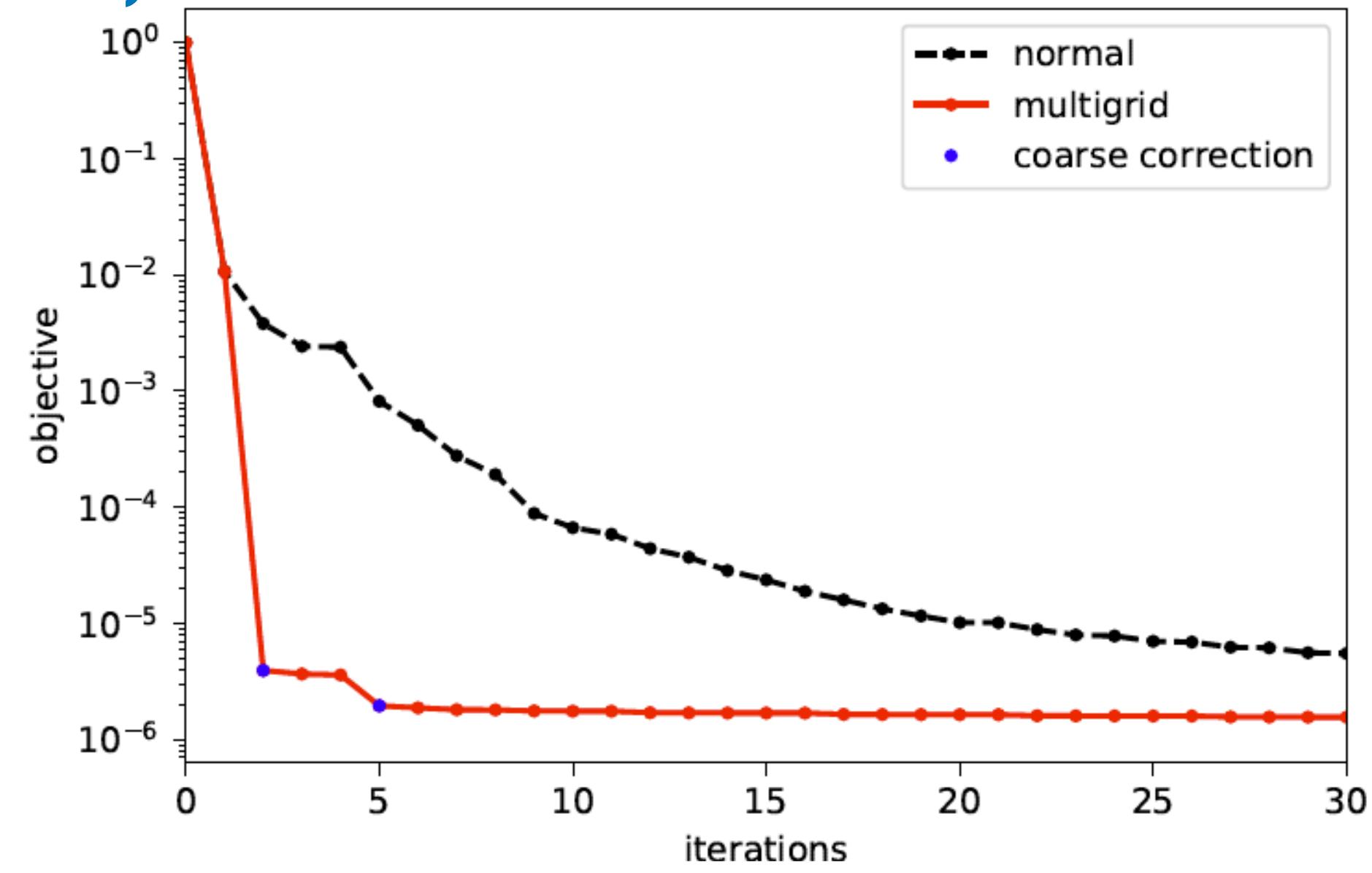
Multilevel, Numerical Results



1024 x 1024



Multilevel, Numerical Results



1024 x 1024



Conclusion

- *Multilevel / multigrid optimization* approach
 - *Coarse model*: efficient descent direction computation
 - *Geometry* takes into account constraints
 - *Recursive procedure*: more levels can be used
-

- *State-dependent* restriction and prolongation
- Coarse models for *non-convex* problems

References I

U. Trottenberg, C. Oosterlee, A. Schüller, Multigrid, Academic Press, 2001

S. G. Nash, A multigrid approach to discretized optimization problems. Optimization Methods and Software, 2000

S. Gratton, A. Sartenaer, P. L. Toint, Recursive trust-region methods for multiscale nonlinear optimization. SIAM Journal on Optimization, 2008

Z. Wen, D. Goldfarb, A line search multigrid method for large-scale nonlinear optimization. SIAM Journal on Optimization, 2009

S. Gratton, M. Mouffe, P. L. Toint, M. Weber-Mendonca, A recursive formula-trust-region method for bound-constrained nonlinear optimization. IMA Journal of Numerical Analysis, 2008

M. Kocvara and S. Mohammed, A first-order multigrid method for bound-constrained convex optimization. Optimization Methods and Software, 2016

P. Parpas, A multilevel proximal gradient algorithm for a class of composite optimization problems. SIAM J. Sci. Comp., 2017

References II

- S. Roux, H. Leclerc, F. Hild, **Efficient Binary Tomographic Reconstruction**, J Math Imaging Vis, 2014
- A. Dabravolski, K. J. Batenburg, J. Sijbers, **A Multiresolution Approach to Discrete Tomography Using DART**, PLoS One. 2014
- J. Plier, **Theoretical and Numerical Approaches to Co-/Sparse Recovery in Discrete Tomography**, PhD Thesis, HeiDOK, 2020
- J. Plier, F. Savarino, M. Kočvara, S. Petra, **First-Order Geometric Multilevel Optimization for Discrete Tomography**, SSVM 2021
- S. I. Amari, H. Nagaoka, **Methods of Information Geometry**, Amer. Math. Soc. and Oxford Univ. Press, 2000
- P. A. Absil, R. Mahony, R. Sepulchre, **Optimization Algorithms on Matrix Manifolds**, Princeton University Press, 2008
- F. Alvarez, J. Bolte, O. Brahic, **Hessian Riemannian gradient flows in convex programming**, SIAM J Control Optim, 2004

First Order Geometric Multilevel Optimization For Discrete Tomography

Stefania Petra
Mathematical Imaging Group
Heidelberg University

**Workshop on Optimization and Operator Theory dedicated to
Professor Lev Bregman's 80th
TECHNION, 15-17 Nov. 2021**

joint work with Jan Plier¹, Fabrizio Savarino², Michal Kočvara³

¹HITS, Heidelberg

²IPA, Heidelberg University

³School of Mathematics - University of Birmingham

