The best of both worlds: synthesis-based acceleration for physics-driven cosparse regularization

Srđan Kitić
Technicolor R&I, Rennes.

iTwist 2016, Aalborg

with Nancy Bertin and Rémi Gribonval
PANAMA team, Inria/IRISA, Rennes

technicolor

Table of contents

Digression: cosparse and sparse data models

Sparse analysis (*cosparse*)

$$\mathbf{x}_a \in \cup_{\#\Lambda \geq \ell} \mathtt{null}(\mathbf{A}_\Lambda)$$

- "Descriptive" model.
- The analysis operator: $\mathbf{A} \in \mathbb{R}^{p \times n}$, $p \geq n$
- Recovering the *cosupport* Λ:

$$\min_{\mathbf{x}_a} \|\mathbf{A}\mathbf{x}_a\|_0 \text{ subject to } \mathbf{M}\mathbf{x}_a \approx \mathbf{y}$$

- Tractable approximations: convex relaxations, GAP, AIHT, ACoSaMP..

Sparse synthesis

$$\mathbf{x}_s \in \cup_{\#\Gamma \leq k} \mathtt{range}(\mathbf{D}_\Gamma)$$

- "Constructive" model.
- The (synthesis) dictionary: $\mathbf{D} \in \mathbb{R}^{n \times d}$, $n \leq d$
- Recovering the support Γ:

$$\min_{\mathbf{x}_s} \|\mathbf{x}_s\|_0 \text{ subject to } \mathbf{M}\mathbf{D}\mathbf{x}_s \approx \mathbf{y}$$

- Tractable approximations: convex relaxations, MP, OMP, IHT, CoSaMP..

technicolor

Digression: cosparse and sparse data models

Sparse analysis (*cosparse*)

$$\mathbf{x}_a \in \cup_{\#\Lambda \geq \ell} \mathtt{null}(\mathbf{A}_\Lambda)$$

- "Descriptive" model.
- The analysis operator: $\mathbf{A} \in \mathbb{R}^{p \times n}$, $p \geq n$
- Recovering the *cosupport* Λ:

$$\min_{\mathbf{x}_a} \|\mathbf{A}\mathbf{x}_a\|_0 \text{ subject to } \mathbf{M}\mathbf{x}_a \approx \mathbf{y}$$

- Tractable approximations: convex relaxations, GAP, AIHT, ACoSaMP...

Sparse synthesis

$$\mathbf{x}_s \in \cup_{\#\Gamma \leq k} \mathtt{range}(\mathbf{D}_\Gamma)$$

- "Constructive" model.
- The (synthesis) dictionary: $\mathbf{D} \in \mathbb{R}^{n \times d}$, $n \leq d$
- Recovering the support Γ:

$$\min_{\mathbf{x}_s} \|\mathbf{x}_s\|_0 \text{ subject to } \mathbf{M}\mathbf{D}\mathbf{x}_s \approx \mathbf{y}$$

- Tractable approximations: convex relaxations, MP, OMP, IHT, CoSaMP...

technicolor

Digression: cosparse and sparse data models

Sparse analysis (*cosparse*)

$$\mathbf{x}_a \in \cup_{\#\Lambda \geq \ell} \mathtt{null}(\mathbf{A}_\Lambda)$$

- "Descriptive" model.
- The analysis operator: $\mathbf{A} \in \mathbb{R}^{p \times n}$, $p \geq n$
- Recovering the *cosupport* Λ:

$$\min_{\mathbf{x}_a} \|\mathbf{A}\mathbf{x}_a\|_0 \text{ subject to } \mathbf{M}\mathbf{x}_a \approx \mathbf{y}$$

- Tractable approximations: convex relaxations, GAP, AIHT, ACoSaMP..

Sparse synthesis

$$\mathbf{x}_s \in \cup_{\#\Gamma \leq k} \mathtt{range}(\mathbf{D}_\Gamma)$$

- "Constructive" model.
- The (synthesis) dictionary: $\mathbf{D} \in \mathbb{R}^{n \times d}$, $n \leq d$
- Recovering the support Γ:

$$\min_{\mathbf{x}_s} \|\mathbf{x}_s\|_0 \text{ subject to } \mathbf{M}\mathbf{D}\mathbf{x}_s \approx \mathbf{y}$$

- Tractable approximations: convex relaxations, MP, OMP, IHT, CoSaMP..

Equivalent *only* if $\mathbf{D} = \mathbf{A}^{-1}$.

technicolor

Digression: cosparse and sparse data models

Sparse analysis (*cosparse*)

$$\mathbf{x}_a \in \cup_{\#\Lambda \geq \ell} \mathtt{null}(\mathbf{A}_\Lambda)$$

- "Descriptive" model.
- The analysis operator: $\mathbf{A} \in \mathbb{R}^{\mathsf{p} \times \mathsf{n}}$, $\mathsf{p} \geq \mathsf{n}$
- Recovering the *cosupport* Λ:

$$\min_{\mathbf{x}_a} \|\mathbf{A}\mathbf{x}_a\|_0 \text{ subject to } \mathbf{M}\mathbf{x}_a \approx \mathbf{y}$$

- Tractable approximations: convex relaxations, GAP, AIHT, ACoSaMP..

Analysis BP

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{M}\mathbf{x}$$

Sparse synthesis

$$\mathbf{x}_s \in \cup_{\#\Gamma \leq \mathsf{k}} \mathtt{range}(\mathbf{D}_\Gamma)$$

- "Constructive" model.
- The (synthesis) dictionary: $\mathbf{D} \in \mathbb{R}^{\mathsf{n} \times \mathsf{d}}$, $\mathsf{n} \leq \mathsf{d}$
- Recovering the support Γ:

$$\min_{\mathbf{x}_s} \|\mathbf{x}_s\|_0 \text{ subject to } \mathbf{M}\mathbf{D}\mathbf{x}_s \approx \mathbf{y}$$

- Tractable approximations: convex relaxations, MP, OMP, IHT, CoSaMP..

Synthesis BP

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \text{ subject to } \mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{z}$$

technicolor

Physics-driven linear inverse problem

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \mathbf{y} = M(x) + \mathbf{e}$$

x : a physical field

M : a spatial subsampling operator

$\mathbf{e}$: an instrumental/environmental noise

- Goal: *find signal x given array measurements $\mathbf{y}$*.
- "Passive" mode: no control over the generative process.
- Irreversible without a model of x.

technicolor

Physics-driven linear inverse problem

Signal x is governed by a linear PDE, *e.g.*:

- Electrodynamics, optics: Maxwell's equations
- Sound propagation: the acoustic wave equation
- Thermodynamics: heat equation
- Electrostatics, mechanics: Poisson's equation, Laplace equation
- Nuclear magnetic resonance: Bloch's equations
- Finance: Black-Scholes equation etc.

technicolor

Physics-driven signal representations

Linear PDE ($\boldsymbol{\omega} \in \Omega$):

$$\sum_{|\mathbf{k}|\leq\xi} c(\mathbf{k},\boldsymbol{\omega}) D^{\mathbf{k}} x(\boldsymbol{\omega}) = z(\boldsymbol{\omega})$$

including boundary conditions!

$$A x(\boldsymbol{\omega}) = z(\boldsymbol{\omega})$$
$$\downarrow$$
$$\mathbf{A}\mathbf{x} = \mathbf{z}$$

Superposition principle:

$$x(\boldsymbol{\omega}) = \int_{\Omega} g(\boldsymbol{\omega},\mathbf{s})\delta(\boldsymbol{\omega}-\mathbf{s}) z(\mathbf{s}) d\mathbf{s}$$

The Green's function: $Ag(\boldsymbol{\omega},\mathbf{s}) = \delta(\mathbf{s}-\boldsymbol{\omega})$

$$x(\boldsymbol{\omega}) = D z(\boldsymbol{\omega})$$
$$\downarrow$$
$$\mathbf{x} = \mathbf{D}\mathbf{z}.$$

Very often: $\mathbf{s} \in A \Leftrightarrow$ sources, sinks...

$$z(\boldsymbol{\omega}) = \sum_{\mathbf{s}} b(\mathbf{s})\delta(\boldsymbol{\omega}-\mathbf{s})$$

$\mathbf{z}$ is sparse, $\mathbf{x}$ is cosparse!

technicolor

Physics-driven signal representations

Linear PDE ($\boldsymbol{\omega} \in \Omega$):

$$\sum_{|\mathbf{k}| \leq \xi} c(\mathbf{k}, \boldsymbol{\omega}) D^{\mathbf{k}} x(\boldsymbol{\omega}) = z(\boldsymbol{\omega})$$

including boundary conditions!

$$Ax(\boldsymbol{\omega}) = z(\boldsymbol{\omega})$$
$$\downarrow$$
$$\mathbf{A}\mathbf{x} = \mathbf{z}$$

Superposition principle:

$$x(\boldsymbol{\omega}) = \int_{\Omega} g(\boldsymbol{\omega}, \mathbf{s}) \delta(\boldsymbol{\omega} - \mathbf{s}) z(\mathbf{s}) d\mathbf{s}$$

The Green's function: $Ag(\boldsymbol{\omega}, \mathbf{s}) = \delta(\mathbf{s} - \boldsymbol{\omega})$

$$x(\boldsymbol{\omega}) = Dz(\boldsymbol{\omega})$$
$$\downarrow$$
$$\mathbf{x} = \mathbf{D}\mathbf{z}.$$

Very often: $\mathbf{s} \in A \Leftrightarrow$ sources, sinks...

$$z(\boldsymbol{\omega}) = \sum_{\mathbf{s}} b(\mathbf{s}) \delta(\boldsymbol{\omega} - \mathbf{s})$$

$\mathbf{z}$ is sparse, $\mathbf{x}$ is cosparse!

technicolor

Physics-driven signal representations

Linear PDE ($\boldsymbol{\omega} \in \Omega$):

$$\sum_{|\mathbf{k}|\leq\xi} c(\mathbf{k},\boldsymbol{\omega}) D^{\mathbf{k}} x(\boldsymbol{\omega}) = z(\boldsymbol{\omega})$$

including boundary conditions!

$$Ax(\boldsymbol{\omega}) = z(\boldsymbol{\omega})$$
$$\downarrow$$
$$\mathbf{A}\mathbf{x} = \mathbf{z}$$

Superposition principle:

$$x(\boldsymbol{\omega}) = \int_{\Omega} g(\boldsymbol{\omega},\mathbf{s})\delta(\boldsymbol{\omega}-\mathbf{s})z(\mathbf{s})d\mathbf{s}$$

The Green's function: $Ag(\boldsymbol{\omega},\mathbf{s}) = \delta(\mathbf{s}-\boldsymbol{\omega})$

$$x(\boldsymbol{\omega}) = Dz(\boldsymbol{\omega})$$
$$\downarrow$$
$$\mathbf{x} = \mathbf{D}\mathbf{z}.$$

Very often: $\mathbf{s} \in A \Leftrightarrow$ sources, sinks...

$$z(\boldsymbol{\omega}) = \sum_{\mathbf{s}} b(\mathbf{s})\delta(\boldsymbol{\omega}-\mathbf{s})$$

$\mathbf{z}$ is sparse, $\mathbf{x}$ is cosparse!

technicolor

Physics-driven signal representations

$\mathbf{A}$ encodes a PDE:

- Discretization locally supported: $\mathtt{nnz}(\mathbf{A}) = O(\mathsf{n})$
- Generally, A unbounded, thus $\|\mathbf{A}\|_2 \xrightarrow{\mathsf{n}\to\infty} \infty$.
- If $\mathbf{A} = \tau(\mathsf{n})\bar{\mathbf{A}}$, with $\bar{\mathbf{A}}_{\mathsf{i},\mathsf{j}}$ independent of n, then: $\|\bar{\mathbf{A}}\|_2^2 \leq \|\bar{\mathbf{A}}\|_1 \|\bar{\mathbf{A}}\|_\infty < \infty$.

$\mathbf{D}$ encodes impulse respones (K. et al., 2016):

- Discretized eigenfunctions of A (generally dense): $\mathtt{nnz}(\mathbf{D}) = O(\mathsf{n}^2)$.
- Fast multiplication in *restricted regimes*.
- $\|\mathbf{D}\|_2$ also generally unbounded (unless $\mathtt{dom}(D) \subseteq H^{\mathsf{k}}(\Omega)$): $\|\mathbf{D}\|_2 \xrightarrow{\mathsf{n}\to\infty} \infty$.

technicolor

Physics-driven basis pursuits

Analysis BP

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{M}\mathbf{x}$$

Synthesis BP

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \text{ subject to } \mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{z}$$

Main difficulties:

- Potentially huge number of variables.
- Condition number increases with problem size.
- Non-smooth optimization, no strict convexity.
- Composite linear term in Analysis BP.

Apply the preconditioned ADMM algorithm (Chambolle and Pock, 2011).

The Chambolle-Pock algorithm

- Solves a generic (convex) saddle-point problem:

$$\min_{\mathbf{v}} f_1(\mathbf{K}\mathbf{v}) + f_2(\mathbf{v}) = \min_{\mathbf{v}} \max_{\mathbf{b}} \langle \mathbf{K}\mathbf{v}, \mathbf{b} \rangle - f_1^*(\mathbf{b}) + f_2(\mathbf{v}) = \min_{\mathbf{v}} \max_{\mathbf{b}} \mathcal{L}(\mathbf{v}, \mathbf{b})$$

- Provided $\mu\sigma\|\mathbf{K}\|_2^2 < 1$, iterate:

$$\mathbf{b}^{(\mathrm{i}+1)} = \mathrm{prox}_{\sigma f_1^*}\left(\mathbf{b}^{(\mathrm{i})} + \sigma\mathbf{K}\bar{\mathbf{v}}^{(\mathrm{i})}\right)$$

$$\mathbf{v}^{(\mathrm{i}+1)} = \mathrm{prox}_{\mu f_2}\left(\mathbf{v}^{(\mathrm{i})} - \mu\mathbf{K}^{\mathrm{H}}\mathbf{b}^{(\mathrm{i}+1)}\right)$$

$$\bar{\mathbf{v}}^{(\mathrm{i}+1)} = \mathbf{v}^{(\mathrm{i}+1)} + \theta\left(\mathbf{v}^{(\mathrm{i}+1)} - \mathbf{v}^{(\mathrm{i})}\right)$$

- Optimal asymptotic rate (Nesterov, 2005) ($\mathbf{v}^*, \mathbf{b}^*$ fixed points):

$$\mathcal{L}(\mathbf{v}^{(\mathrm{k})}, \mathbf{b}^*) - \mathcal{L}(\mathbf{v}^*, \mathbf{b}^{(\mathrm{k})}) \leq \frac{1}{\mathrm{k}}\left(\frac{1}{\mu}\|\mathbf{v}^* - \mathbf{v}^{(0)}\|_2^2 + \frac{1}{\sigma}\|\mathbf{b}^* - \mathbf{b}^{(0)}\|_2^2 - \langle \mathbf{K}(\mathbf{v}^* - \mathbf{v}^{(0)}), \mathbf{b}^* - \mathbf{b}^{(0)} \rangle\right)$$

Convergence rate

$O(\|\mathbf{K}\|_2^2/\mathrm{k})$, dependent on $\|\mathbf{v}^* - \mathbf{v}^{(0)}\|_2$.

technicolor

The Chambolle-Pock algorithm

- Solves a generic (convex) saddle-point problem:

$$\min_{\mathbf{v}} f_1(\mathbf{K}\mathbf{v}) + f_2(\mathbf{v}) = \min_{\mathbf{v}} \max_{\mathbf{b}} \langle \mathbf{K}\mathbf{v}, \mathbf{b}\rangle - f_1^*(\mathbf{b}) + f_2(\mathbf{v}) = \min_{\mathbf{v}} \max_{\mathbf{b}} \mathcal{L}(\mathbf{v}, \mathbf{b})$$

- Provided $\mu\sigma\|\mathbf{K}\|_2^2 < 1$, iterate:

$$\mathbf{b}^{(\mathrm{i}+1)} = \mathrm{prox}_{\sigma f_1^*}\left(\mathbf{b}^{(\mathrm{i})} + \sigma\mathbf{K}\bar{\mathbf{v}}^{(\mathrm{i})}\right)$$
$$\mathbf{v}^{(\mathrm{i}+1)} = \mathrm{prox}_{\mu f_2}\left(\mathbf{v}^{(\mathrm{i})} - \mu\mathbf{K}^{\mathsf{H}}\mathbf{b}^{(\mathrm{i}+1)}\right)$$
$$\bar{\mathbf{v}}^{(\mathrm{i}+1)} = \mathbf{v}^{(\mathrm{i}+1)} + \theta\left(\mathbf{v}^{(\mathrm{i}+1)} - \mathbf{v}^{(\mathrm{i})}\right)$$

- Optimal asymptotic rate (Nesterov, 2005) ($\mathbf{v}^*, \mathbf{b}^*$ fixed points):

$$\mathcal{L}(\mathbf{v}^{(\mathrm{k})}, \mathbf{b}^*) - \mathcal{L}(\mathbf{v}^*, \mathbf{b}^{(\mathrm{k})}) \leq \frac{1}{\mathrm{k}}\left(\frac{1}{\mu}\|\mathbf{v}^* - \mathbf{v}^{(0)}\|_2^2 + \frac{1}{\sigma}\|\mathbf{b}^* - \mathbf{b}^{(0)}\|_2^2 - \langle \mathbf{K}(\mathbf{v}^* - \mathbf{v}^{(0)}), \mathbf{b}^* - \mathbf{b}^{(0)}\rangle\right)$$

Convergence rate

$O(\|\mathbf{K}\|_2^2/\mathrm{k})$, dependent on $\|\mathbf{v}^* - \mathbf{v}^{(0)}\|_2$.

technicolor

The Chambolle-Pock algorithm

- Solves a generic (convex) saddle-point problem:

$$\min_{\mathbf{v}} f_1(\mathbf{K}\mathbf{v}) + f_2(\mathbf{v}) = \min_{\mathbf{v}} \max_{\mathbf{b}} \langle \mathbf{K}\mathbf{v}, \mathbf{b} \rangle - f_1^*(\mathbf{b}) + f_2(\mathbf{v}) = \min_{\mathbf{v}} \max_{\mathbf{b}} \mathcal{L}(\mathbf{v}, \mathbf{b})$$

- Provided $\mu\sigma\|\mathbf{K}\|_2^2 < 1$, iterate:

$$\begin{aligned}
\mathbf{b}^{(\mathsf{i}+1)} &= \mathsf{prox}_{\sigma f_1^*}\left(\mathbf{b}^{(\mathsf{i})} + \sigma \mathbf{K}\bar{\mathbf{v}}^{(\mathsf{i})}\right)\\
\mathbf{v}^{(\mathsf{i}+1)} &= \mathsf{prox}_{\mu f_2}\left(\mathbf{v}^{(\mathsf{i})} - \mu \mathbf{K}^{\mathsf{H}}\mathbf{b}^{(\mathsf{i}+1)}\right)\\
\bar{\mathbf{v}}^{(\mathsf{i}+1)} &= \mathbf{v}^{(\mathsf{i}+1)} + \theta\left(\mathbf{v}^{(\mathsf{i}+1)} - \mathbf{v}^{(\mathsf{i})}\right)
\end{aligned}$$

- Optimal asymptotic rate (Nesterov, 2005) ($\mathbf{v}^*, \mathbf{b}^*$ fixed points):

$$\begin{aligned}
&\mathcal{L}(\mathbf{v}^{(\mathsf{k})}, \mathbf{b}^*) - \mathcal{L}(\mathbf{v}^*, \mathbf{b}^{(\mathsf{k})})\\
&\quad \leq \frac{1}{\mathsf{k}}\left(\frac{1}{\mu}\|\mathbf{v}^* - \mathbf{v}^{(0)}\|_2^2 + \frac{1}{\sigma}\|\mathbf{b}^* - \mathbf{b}^{(0)}\|_2^2 - \langle \mathbf{K}(\mathbf{v}^* - \mathbf{v}^{(0)}), \mathbf{b}^* - \mathbf{b}^{(0)}\rangle\right)
\end{aligned}$$

Convergence rate

$O(\|\mathbf{K}\|_2^2/\mathsf{k})$, dependent on $\|\mathbf{v}^* - \mathbf{v}^{(0)}\|_2$.

technicolor

CP algorithm and physics-driven BP regularization

Synthesis	Analysis
$\min_{\mathbf{z}} \max_{\mathbf{h}} \langle \mathbf{MDz} - \mathbf{y}, \mathbf{h} \rangle + \|\mathbf{z}\|_1$	$\min_{\mathbf{x}} \max_{\mathbf{q}} \langle \mathbf{Ax}, \mathbf{q} \rangle - \|\mathbf{q}\|_1^* + \chi_{\mathbf{M}\cdot=\mathbf{y}}(\mathbf{x})$

Synthesis

- ✗ Iteration / storage cost: $O(\mathrm{mn})$
- ✗ Convergence rate $\propto \|\mathbf{MD}\|_2^{-2}$:

  $\|\mathbf{MD}\|_2 \xrightarrow{\mathrm{m}\to\mathrm{n}} \|\mathbf{D}\|_2$
- ✓ Initialization: $\mathbf{z}^{(0)} = \mathbf{0}$

  $\|\mathbf{z}^* - \mathbf{z}^{(0)}\|_2$ small for *sparse* $\mathbf{z}^*$.

Analysis

- ✓ Iteration / storage cost: $O(\mathrm{n})$
- ✓ Convergence rate $\propto \|\mathbf{A}\|_2^{-2}$ or $\propto \|\bar{\mathbf{A}}\|_2^{-2}$:

  $\min_{\mathbf{x}} \|\mathbf{Ax}\|_1$ subject to $\mathbf{y} = \mathbf{Mx}$

  $\Leftrightarrow$

  $\min_{\mathbf{x}} \|\bar{\mathbf{A}}\mathbf{x}\|_1$ subject to $\mathbf{y} = \mathbf{Mx}$
- ✗ Initialization: $\mathbf{x}^{(0)} = ?$

CP algorithm and physics-driven BP regularization

Synthesis

$$\min_{\mathbf{z}} \max_{\mathbf{h}} \langle \mathbf{MDz} - \mathbf{y}, \mathbf{h} \rangle + \|\mathbf{z}\|_1$$

- ✗ Iteration / storage cost: $O(\mathsf{mn})$
- ✗ Convergence rate $\propto \|\mathbf{MD}\|_2^{-2}$:

  $\|\mathbf{MD}\|_2 \xrightarrow{\mathsf{m}\to\mathsf{n}} \|\mathbf{D}\|_2$

- ✓ Initialization: $\mathbf{z}^{(0)} = \mathbf{0}$

  $\|\mathbf{z}^* - \mathbf{z}^{(0)}\|_2$ small for *sparse* $\mathbf{z}^*$.

Analysis

$$\min_{\mathbf{x}} \max_{\mathbf{q}} \langle \mathbf{Ax}, \mathbf{q} \rangle - \|\mathbf{q}\|_1^* + \chi_{\mathbf{M}\cdot=\mathbf{y}}(\mathbf{x})$$

- ✓ Iteration / storage cost: $O(\mathsf{n})$
- ✓ Convergence rate $\propto \|\mathbf{A}\|_2^{-2}$ or $\propto \|\bar{\mathbf{A}}\|_2^{-2}$:

  $\min_{\mathbf{x}} \|\mathbf{Ax}\|_1$ subject to $\mathbf{y} = \mathbf{Mx}$

  $\Leftrightarrow$

  $\min_{\mathbf{x}} \|\bar{\mathbf{A}}\mathbf{x}\|_1$ subject to $\mathbf{y} = \mathbf{Mx}$

- ✗ Initialization: $\mathbf{x}^{(0)} = \mathbf{?}$

technicolor

The best of both worlds

Multiscale optimization:

1. Build a multiscale pyramid: $n_0 < n_1 < n_2 < \ldots n_r \ldots < n$
2. Solve the synthesis BP at $r = 0$ with $\mathbf{z}_0^{(0)} = \mathbf{0}$ and compute $\mathbf{x}_0 = \mathbf{D}_0 \mathbf{z}_0$
3. Interpolate $\mathbf{x}_r$ to $\tilde{\mathbf{x}}_{r+1}$
4. Solve the analysis CP with $\mathbf{x}_{r+1}^{(0)} = \tilde{\mathbf{x}}_{r+1}$ to obtain $\mathbf{x}_{r+1}$
5. If $n_{r+1} < n$, go to 3.

Advantages:

- Exploits sparse initialization.
- Linear per-iteration and memory cost (except at the coarsest level).
- Convergence rate improved due to $\|\mathbf{A}_r\|_2 < \|\mathbf{A}_{r+1}\|_2$.
- Potential speed-up with m increasing (Oymak et al., 2015).

technicolor

The best of both worlds

Multiscale optimization:

1. Build a multiscale pyramid: $\mathsf{n}_0 < \mathsf{n}_1 < \mathsf{n}_2 < \dots \mathsf{n}_\mathsf{r} \dots < \mathsf{n}$
2. Solve the synthesis BP at $\mathsf{r} = 0$ with $\mathbf{z}_0^{(0)} = \mathbf{0}$ and compute $\mathbf{x}_0 = \mathbf{D}_0 \mathbf{z}_0$
3. Interpolate $\mathbf{x}_\mathsf{r}$ to $\tilde{\mathbf{x}}_{\mathsf{r}+1}$
4. Solve the analysis CP with $\mathbf{x}_{\mathsf{r}+1}^{(0)} = \tilde{\mathbf{x}}_{\mathsf{r}+1}$ to obtain $\mathbf{x}_{\mathsf{r}+1}$
5. If $\mathsf{n}_{\mathsf{r}+1} < \mathsf{n}$, go to 3.

Advantages:

- Exploits sparse initialization.
- Linear per-iteration and memory cost (except at the coarsest level).
- Convergence rate improved due to $\|\mathbf{A}_\mathsf{r}\|_2 < \|\mathbf{A}_{\mathsf{r}+1}\|_2$.
- Potential speed-up with m increasing (Oymak et al., 2015).

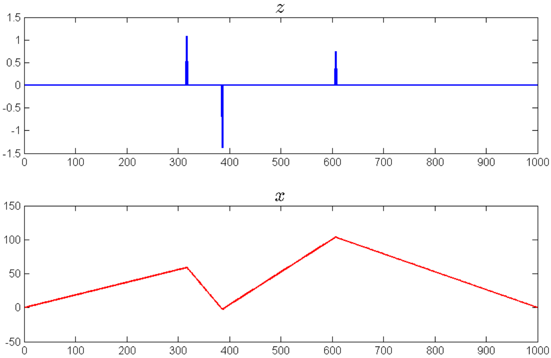
technicolor

Numerical results

"Simple" problem: $\frac{d^2}{d\mathrm{r}^2}x = z, \quad \mathrm{r} \in [0, \phi], \quad x(0) = x(\phi) = 0$

Numerical results

"Simple" problem: $\frac{d^2}{d\mathrm{r}^2}x = z, \quad \mathrm{r} \in [0, \phi], \quad x(0) = x(\phi) = 0$

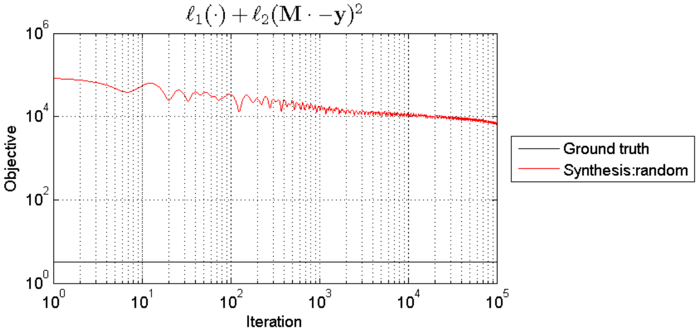


FDM, $\mathsf{n} = 1000$, $\mathsf{k} = 3$, $\mathsf{m} = 50$

technicolor

Numerical results

"Simple" problem: $\frac{d^2}{dr^2}x = z, \quad r \in [0, \phi], \quad x(0) = x(\phi) = 0$

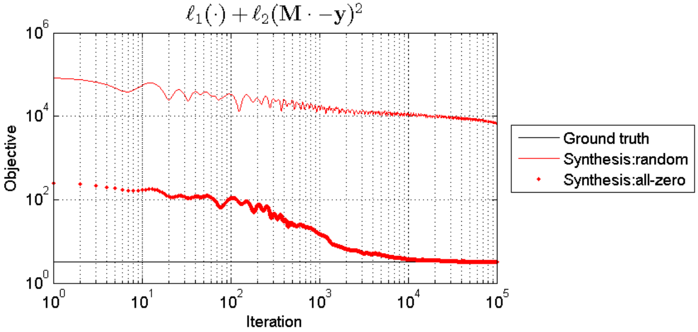


Synthesis with random initialization

technicolor

Numerical results

"Simple" problem: $\frac{d^2}{d\mathrm{r}^2}x = z, \quad \mathrm{r} \in [0, \phi], \quad x(0) = x(\phi) = 0$

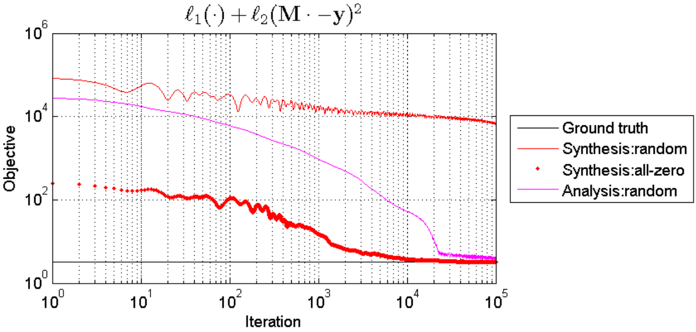


Synthesis with all-zero initialization - but we *cannot* solve it in the general case!

technicolor

Numerical results

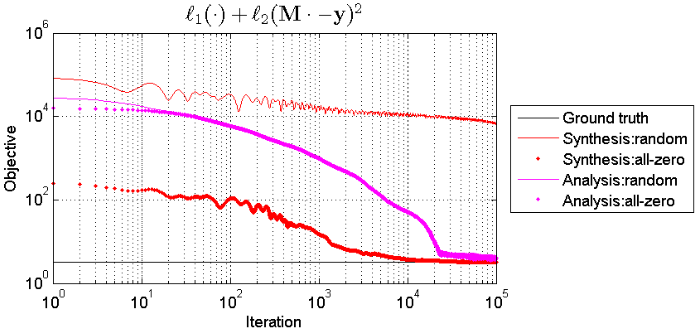
"Simple" problem: $\frac{d^2}{d\mathrm{r}^2}x = z, \quad \mathrm{r} \in [0, \phi]\,, \quad x(0) = x(\phi) = 0$



Analysis with random initialization

Numerical results

“Simple” problem: $\frac{d^2}{d\mathrm{r}^2}x = z, \quad \mathrm{r} \in [0, \phi], \quad x(0) = x(\phi) = 0$

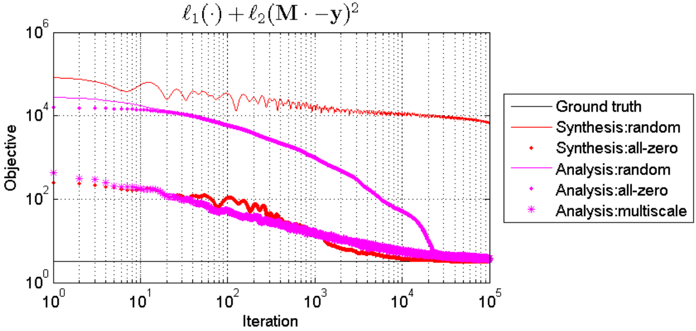


Analysis with all-zero initialization

technicolor

Numerical results

"Simple" problem: $\frac{d^2}{dr^2}x = z, \quad \mathrm{r} \in [0, \phi], \quad x(0) = x(\phi) = 0$

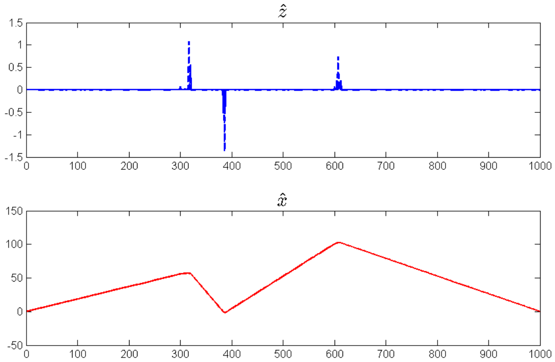


Analysis with *coarse* synthesis initialization: $\mathrm{n}_1 = 250, \; \mathrm{n} = 1000$

technicolor

Numerical results

"Simple" problem: $\frac{d^2}{d\mathrm{r}^2}x = z, \quad \mathrm{r} \in [0, \phi], \quad x(0) = x(\phi) = 0$



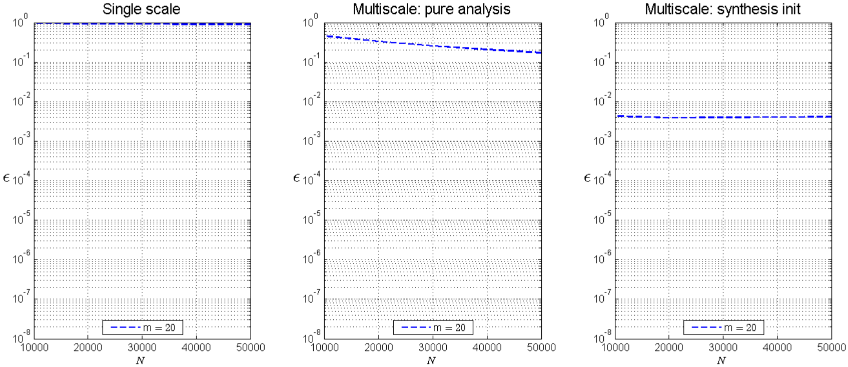
Final estimate

Numerical results

- $n = 10000$, $k = 5$, varying m.
- Three experimental regimes:
  1. Single (large) scale analysis,
  2. Five-scale pure analysis,
  3. Five-scale: coarse synthesis + four-scale analysis.
- Relative error criterion: $\epsilon = \|\mathbf{x}^* - \hat{\mathbf{x}}\|^2 / \|\mathbf{x}^*\|^2$
- *Fixed iteration budget*:
  - Iterations for the single scale: N
  - Iterations for the multiscale: $\sum N_r = N$.
  - Far fewer computations with multiscale, since $N_r > N_{r+1}$, $r = 1 \ldots 5$.

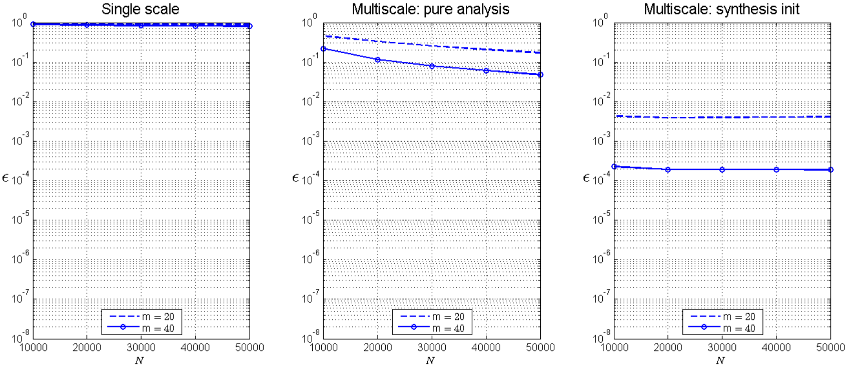
technicolor

Numerical results



N : total iteration budget

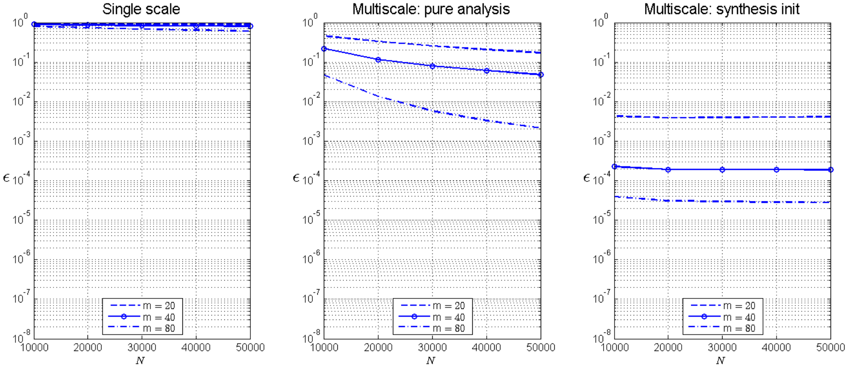
Numerical results



N : total iteration budget

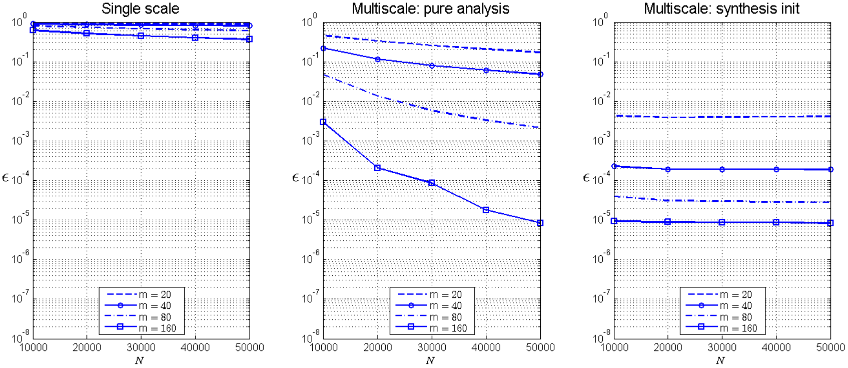
technicolor

Numerical results



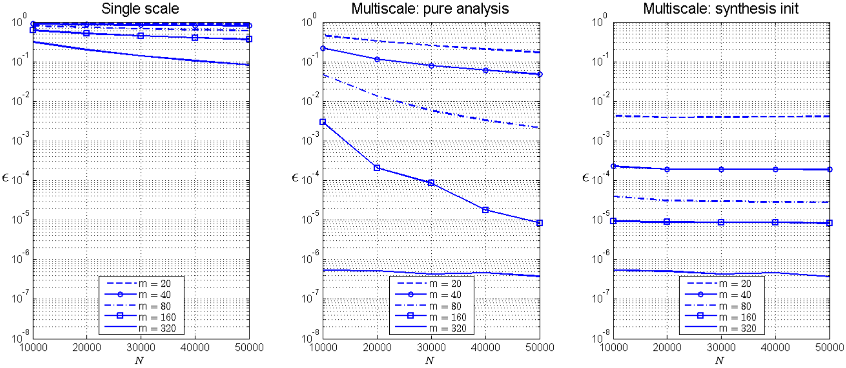
N : total iteration budget

Numerical results



N : total iteration budget

Numerical results

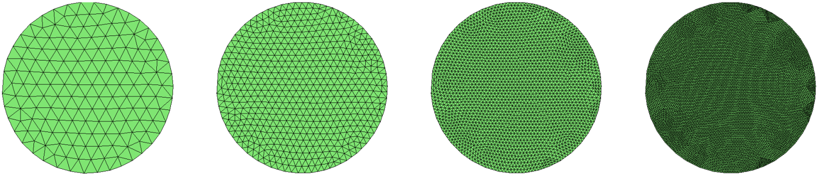


N : total iteration budget

Numerical results

$$\Delta x(\mathrm{r}) = z, \ \mathrm{r} \in \Omega \setminus \partial\Omega$$

FEM multiscale: $\mathsf{n}_1 < 200, \ 9 \cdot 10^3 < \mathsf{n}$.

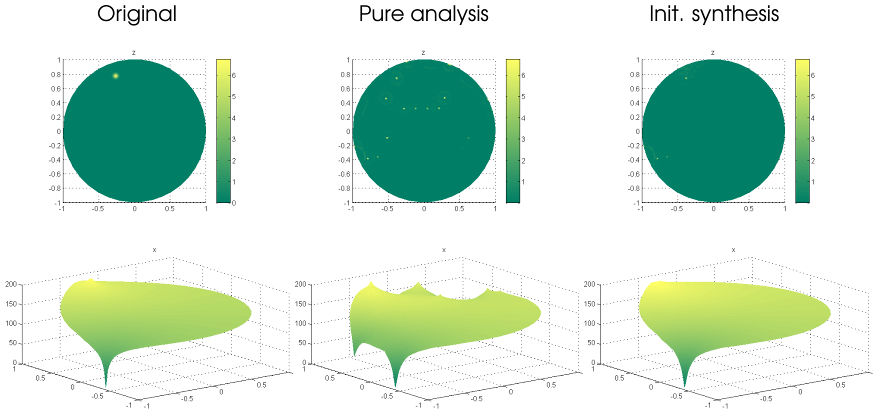


Off-the-grid measurements : calibration error.

Off-the-grid sources : "support leakage".

technicolor

Numerical results



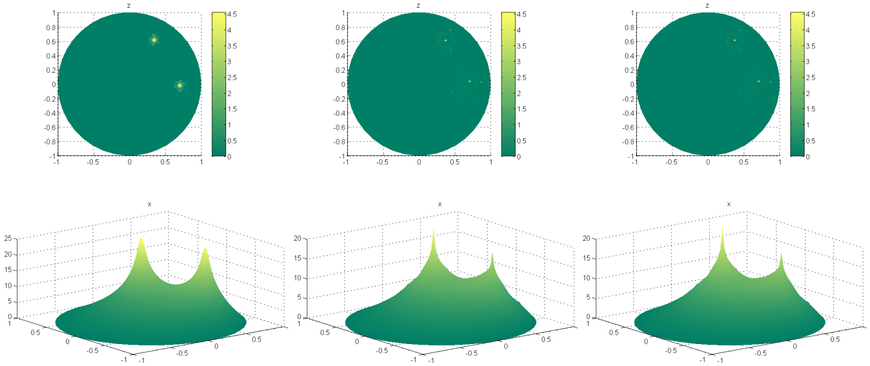
$$x(\mathrm{r}_1) = 1,\ \mathrm{r}_1 \in \partial\Omega_1 \quad \frac{\partial}{\partial\vec{n}}x(\mathrm{r}_2) = 0,\ \mathrm{r}_2 \in \partial\Omega_2$$

$$\mathsf{k} = 1,\ N = 2000, \mathsf{m} = 25$$

technicolor

Numerical results

Original | Pure analysis | Init. synthesis



$$x(\mathrm{r}) = 0, \ \mathrm{r} \in \partial\Omega$$

$$\mathsf{k} = 2, \ N = 2000, \ \mathsf{m} = 100$$

technicolor

Conclusion

- Low-cost cosparse multiscale optimization.
- Sparse solutions: synthesis-based initialization is simple and effective.
- Applicable to variety of physics-driven problems.

- Generalization to other types of algorithms/regularizers.
- Significance of interpolation/discretization method?
- CP stepsize selection?
- Verification on the real-world problems/data.

Thank you!

References

(Chambolle and Pock, 2011) Chambolle, A. and Pock, T. (2011).
A first-order primal-dual algorithm for convex problems with applications to imaging.
Journal of Mathematical Imaging and Vision, 40(1):120–145.

(K. et al., 2016) K., S., Albera, L., Bertin, N., and Gribonval, R. (2016).
Physics-driven inverse problems made tractable with cosparse regularization.
IEEE Transactions on Signal Processing, 64(2):335–348.

(Nesterov, 2005) Nesterov, Y. (2005).
Smooth minimization of non-smooth functions.
Mathematical programming, 103(1):127–152.

(Oymak et al., 2015) Oymak, S., Recht, B., and Soltanolkotabi, M. (2015).
Sharp time–data tradeoffs for linear inverse problems.
arXiv preprint arXiv:1507.04793.

technicolor