Plug-In SGD: Image Reconstruction in the Age of Machine Learning

Ulugbek S. Kamilov
Computational Imaging Group (CIG)
Washington University, St. Louis, USA
cigroup.wustl.edu • @wustlcig • kamilov@wustl.edu
Imaging technology is going through a paradigm shift with computation at its core
Imaging technology is going through a paradigm shift with computation at its core

**Past:** Can I see?

![Diagram](image)

- **Input**
- **Imaging system**
- **Output**
Imaging technology is going through a paradigm shift with computation at its core

**Past:** Can I see?

**Present:** Can I see better?

Inverse problem is well posed if $c_0 > 0$ s.t., for all $s \in X$, $c_0 \|s\|^k_H s^k$

Inverse problems in bio-imaging

Noise $n$

Linear forward model $s = H(s) + n$

Problem: recover $s$ from noisy measurements $y$

Backprojection (poor man's solution): $s \approx H^T y$

The easy scenario
Imaging technology is going through a paradigm shift with computation at its core

Past: Can I see?

Present: Can I see better?

Future: Can I see more?
Today we will talk about

- **Forward models in imaging**
  Relating the unknowns to the measured data

- **Notions of ill-posedness and regularization**
  When measurements are not enough

- **Optimization at large scales**
  When analytical solutions are not enough

- **Plug-and-Play Priors (PnP) at large scales**
  When traditional optimization is not enough
Today we will talk about

- **Forward models in imaging**
  Relating the unknowns to the measured data

- **Notions of ill-posedness and regularization**
  When measurements are not enough

- **Optimization at large scales**
  When analytical solutions are not enough

- **Plug-and-Play Priors (PnP) at large scales**
  When traditional optimization is not enough
Forward model relates the unknown object to the observed data
Forward model relates the unknown object to the observed data

\[
\begin{align*}
  y &= Hf + e \\
  \text{Imaging operator}
\end{align*}
\]
Forward model relates the unknown object to the observed data (assuming noise is negligible).

Inverse problem is well posed if \[ c_0 > 0 \]

s.t., for all \( s \in X \),

\[ \| c_0 s \| <= \| Hs \| \]

Inverse problems in bio-imaging

The backward problem: recover \( s \) from noisy measurements \( y \).

Backprojection (poor man’s solution):

\[ s \triangleq H^{-1} y \]

The easy scenario

Part 1: Setting up the problem (assuming noise is negligible)

\[ y = Hf + e \]
Forward model relates the unknown object to the observed data.

**Forward problem:** generate $y$ from $f$

$$y = Hf + e$$

**Inverse problem:** recover $f$ from $y$
Forward model relates the unknown object to the observed data (assuming noise is negligible).

Inverse problem is well posed if
\[ c_0 > 0 \text{ s.t., for all } s \in X, \quad c_0 k_s k_H s_k \]

Inverse problems in bio-imaging

Linear forward model $y = Hf + e$

Problem: recover $f$ from noisy measurements $y$.

Backprojection (poor man's solution): $f \mapsto H^T y$

The easy scenario

Part 1: Setting up the problem

Question: Which problem is harder to solve?
Forward models can be represented as integrals
Forward models can be represented as integrals
Forward models can be represented as integrals (assuming noise is negligible)

Inverse problem is well posed if

\[ c_0 > 0 \text{ s.t., for all } s \in X, \quad c_0 \| s \| H < 1 \]

Inverse problems in bio-imaging

Part 1: Setting up the problem

Linear forward model

\[ y = H f + n \]

Problem: recover \( s \) from noisy measurements

Backprojection (poor man's solution):

\[ s \mapsto H^T y \]

Forward models can be represented as integrals

Unknown molecular/anatomical map: \( f(r), \quad r = (x, y, z, t) \in \mathbb{R}^d \)

defined over a continuum in space-time

Forward models can be represented as integrals

Unknown molecular/anatomical map: \( f(r), \quad r = (x, y, z, t) \in \mathbb{R}^d \)

Space of finite-energy functions: \( f \in L_2(\mathbb{R}^d) \)

Forward models can be represented as integrals

**Unknown molecular/anatomical map:**  \( f(r), \quad r = (x, y, z, t) \in \mathbb{R}^d \)

**Space of finite-energy functions:**  \( f \in L_2(\mathbb{R}^d) \)

**Imaging operator:**  \( H : s \mapsto y = (y_1, \ldots, y_m) = H\{f\} \)

from continuum to finite dimensional:  \( H : L_2(\mathbb{R}^d) \to \mathbb{R}^m \)
Forward models can be represented as integrals

Unknown molecular/anatomical map: \( f(r), \quad r = (x, y, z, t) \in \mathbb{R}^d \)

Space of finite-energy functions: \( f \in L_2(\mathbb{R}^d) \)

Imaging operator: \( \mathbf{H} : s \mapsto \mathbf{y} = (y_1, \ldots, y_m) = \mathbf{H}\{f\} \)

Linearity assumption: \( \forall \alpha_1, \alpha_2 \in \mathbb{R}, \quad \forall f_1, f_2 \in L_2(\mathbb{R}^d) \)

\[
\mathbf{H}\{\alpha_1 f_1 + \alpha_2 f_2\} = \alpha_1 \mathbf{H}\{f_1\} + \alpha_2 \mathbf{H}\{f_2\}
\]
Forward models can be represented as integrals

Unknown molecular/anatomical map: \[ f(r), \quad r = (x, y, z, t) \in \mathbb{R}^d \]

Space of finite-energy functions: \[ f \in L_2(\mathbb{R}^d) \]

Imaging operator: \[ H : s \mapsto y = (y_1, \ldots, y_m) = H\{f\} \]

Linearity assumption: \[ \forall \alpha_1, \alpha_2 \in \mathbb{R}, \quad \forall f_1, f_2 \in L_2(\mathbb{R}^d) \]
\[ H\{\alpha_1 f_1 + \alpha_2 f_2\} = \alpha_1 H\{f_1\} + \alpha_2 H\{f_2\} \]

\[ \Rightarrow \quad [y]_m = y_m = \langle h_m, f \rangle = \int_{\mathbb{R}^d} h_m(r) f(r) \, dr \]

by the Riesz representation theorem

Forward models can be represented as integrals

Unknown molecular/anatomical map: \( f(r), \quad r = (x, y, z, t) \in \mathbb{R}^d \)

Space of finite-energy functions: \( f \in L_2(\mathbb{R}^d) \)

Imaging operator: \( H : s \mapsto y = (y_1, \ldots, y_m) = H\{f\} \)

Linearity assumption: \( \forall \alpha_1, \alpha_2 \in \mathbb{R}, \quad \forall f_1, f_2 \in L_2(\mathbb{R}^d) \)

\[ H\{\alpha_1 f_1 + \alpha_2 f_2\} = \alpha_1 H\{f_1\} + \alpha_2 H\{f_2\} \]

\[ \Rightarrow [y]_m = y_m = \langle h_m, f \rangle = \int_{\mathbb{R}^d} h_m(r) f(r) \, dr \]

impulse response of \( m \)th detector

Example imaging operator:
Fourier transform is extensively used in MRI
Example imaging operator:
Fourier transform is extensively used in MRI

“Images are obviously made of sine waves…”
Example imaging operator:
Fourier transform is extensively used in MRI

**Fourier transform:** \( \mathcal{F} : L_2(\mathbb{R}^d) \rightarrow L_2(\mathbb{R}^d) \)

\[
\hat{f}(\omega) = \mathcal{F}\{f\} = \int_{\mathbb{R}^d} f(r) e^{-i\langle \omega, r \rangle} \, dr
\]
Example imaging operator:
Fourier transform is extensively used in MRI

**Fourier transform:** $\mathcal{F} : L_2(\mathbb{R}^d) \rightarrow L_2(\mathbb{R}^d)$

$$\hat{f}(\omega) = \mathcal{F}\{f\} = \int_{\mathbb{R}^d} f(r) e^{-i\langle \omega, r \rangle} \, dr$$

**Inverse Fourier transform (reconstruction formula)**

$$f(r) = \mathcal{F}^{-1}\{f\} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\omega) e^{i\langle \omega, r \rangle} \, d\omega \quad \text{(a.e.)}$$
Example imaging operator: Fourier transform is extensively used in MRI

### Fourier transform

**Fourier transform:** $\mathcal{F} : L_2(\mathbb{R}^d) \rightarrow L_2(\mathbb{R}^d)$

$$\hat{f}(\omega) = \mathcal{F}\{f\} = \int_{\mathbb{R}^d} f(r) e^{-j \langle \omega, r \rangle} \, dr$$

### Inverse Fourier transform (reconstruction formula)

$$f(r) = \mathcal{F}^{-1}\{f\} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\omega) e^{j \langle \omega, r \rangle} \, d\omega$$

### As a measurement function

As a measurement function:

$$h_m(r) = e^{-j \langle \omega_m, r \rangle} \quad \text{(complex sinusoid)}$$

$$y_m = \langle h_m, f \rangle = \int_{\mathbb{R}^d} h_m(r) f(r) \, dr$$
Example imaging operator:
Fourier transform is extensively used in MRI
Example imaging operator:
Fourier transform is extensively used in MRI
Example imaging operator:
Fourier transform is extensively used in MRI

Linear forward model for MRI

\[ \hat{s}(\omega_m) = \int_{\mathbb{R}^3} s(r) e^{-j\langle \omega_m, r \rangle} \, dr \]

Sampling of Fourier transform

\[ r = (x, y, z) \quad \omega = (\omega_x, \omega_y, \omega_z) \]
Example imaging operator: Fourier transform is extensively used in MRI

Linear forward model for MRI

\[ \hat{s}(\omega_m) = \int_{\mathbb{R}^3} s(r) e^{-j\langle \omega_m, r \rangle} \, dr \]

sampling of Fourier transform

\[ r = (x, y, z) \quad \omega = (\omega_x, \omega_y, \omega_z) \]

(A) Completely sampled
(B) Uniformly under-sampled
(C) Incoherently under-sampled
(D) Variable density incoherently under-sampled

[Source]
Example imaging operator: Fourier transform is extensively used in MRI

**Linear forward model for MRI**

\[ \hat{s}(\omega_m) = \int_{\mathbb{R}^3} s(r) e^{-j\langle \omega_m, r \rangle} \, dr \]

**Extended forward model with coil sensitivity**

\[ \hat{s}_w(\omega_m) = \int_{\mathbb{R}^3} w(r) s(r) e^{-j\langle \omega_m, r \rangle} \, dr \]
Example imaging operator:
Radon transform is extensively used in tomography
Example imaging operator:
Radon transform is extensively used in tomography

Example imaging operator:
Radon transform is extensively used in tomography

Projection geometry: \[ r = t\theta + r\theta^\perp, \quad \theta = (\cos \theta, \sin \theta) \]
Example imaging operator:
Radon transform is extensively used in tomography

Projection geometry: \( r = t\theta + r\theta^\perp, \ \ \theta = (\cos \theta, \sin \theta) \)

Radon transform computes line integrals of the object:

\[
R_\theta \{ f(r) \}(t) = \int_{\mathbb{R}} f(t\theta + r\theta^\perp) \, dr \\
= \int_{\mathbb{R}^2} f(r) \delta(t - \langle r, \theta \rangle) \, dr
\]
Example imaging operator: Radon transform is extensively used in tomography

**Projection geometry:** \[ r = t\theta + r\theta^\perp, \quad \theta = (\cos \theta, \sin \theta) \]

Radon transform computes line integrals of the object:

\[
R_\theta \{ f(r) \}(t) = \int_{\mathbb{R}} f(t\theta + r\theta^\perp) \, dr \\
= \int_{\mathbb{R}^2} f(r)\delta(t - \langle r, \theta \rangle) \, dr
\]

image and its sinogram

Example imaging operator: Radon transform is extensively used in tomography

Projection geometry: \( r = t\theta + r\theta^\perp, \quad \theta = (\cos \theta, \sin \theta) \)

Radon transform computes line integrals of the object:

\[
\mathcal{R}_\theta \{ f(r) \} (t) = \int_{\mathbb{R}} f(t\theta + r\theta^\perp) \, dr \\
= \int_{\mathbb{R}^2} f(r) \delta(t - \langle r, \theta \rangle) \, dr
\]

As a measurement function:

\[
h_m(r) = \delta(t_m - \langle r, \theta_m \rangle)
\]

Central slice theorem relates projections to the Fourier transform of the object
Central slice theorem relates projections to the Fourier transform of the object

Radon transform: \( p_\theta(t) = R_\theta \{ f \}(t, \theta) \)
Central slice theorem relates projections to the Fourier transform of the object

Radon transform: \( p_\theta(t) = R_\theta\{f\}(t, \theta) \)

1D and 2D Fourier relationships:

1D Fourier of data
\( \hat{p}_\theta(\omega) = \mathcal{F}_{1D}\{p_\theta\}(\omega) \)

2D Fourier of image
\( \hat{f}(\omega) = \mathcal{F}_{2D}\{f\}(\omega) = \hat{f}_{pol}(\omega, \theta) \)

Central slice theorem relates projections to the Fourier transform of the object

Radon transform: \[ p_\theta(t) = R_\theta \{ f \}(t, \theta) \]

1D and 2D Fourier relationships:
\[ \hat{p}_\theta(\omega) = \mathcal{F}_{1D}\{p_\theta\}(\omega) \]
\[ \hat{f}(\omega) = \mathcal{F}_{2D}\{f\}(\omega) = \hat{f}_{\text{pol}}(\omega, \theta) \]

Central-slice theorem relates projections to Fourier sampling:
\[ \hat{p}_\theta(\omega) = \hat{f}(\omega \cos \theta, \omega \sin \theta) = \hat{f}_{\text{pol}}(\omega, \theta) \]

Establishes Fourier relationship between data and image

Central slice theorem relates projections to the Fourier transform of the object

Radon transform:  \( p_\theta(t) = R_\theta \{ f \}(t, \theta) \)

1D and 2D Fourier relationships:

\[ \hat{p}_\theta(\omega) = \mathcal{F}_{1D} \{ p_\theta \}(\omega) \]
\[ \hat{f}(\omega) = \mathcal{F}_{2D} \{ f \}(\omega) = \hat{f}_{\text{pol}}(\omega, \theta) \]

Central-slice theorem relates projections to Fourier sampling:

\[ \hat{p}_\theta(\omega) = \hat{f}(\omega \cos \theta, \omega \sin \theta) = \hat{f}_{\text{pol}}(\omega, \theta) \]

Establishes Fourier relationship between data and image

Proof for angle zero:

\[ \hat{f}(\omega, 0) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j\omega x} \, dx \, dy = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} f(x, y) \, dy \right) e^{-j\omega x} \, dx = \hat{p}_0(x) \]
Central slice theorem relates projections to the Fourier transform of the object

Radon transform: \( p_\theta(t) = R_\theta \{ f \}(t, \theta) \)

1D and 2D Fourier relationships:
\[
\hat{p}_\theta(\omega) = \mathcal{F}_\text{1D} \{ p_\theta \}(\omega) \\
\hat{f}(\omega) = \mathcal{F}_\text{2D} \{ f \}(\omega) = \hat{f}_\text{pol}(\omega, \theta)
\]

Central-slice theorem relates projections to Fourier sampling:
\[
\hat{p}_\theta(\omega) = \hat{f}(\omega \cos \theta, \omega \sin \theta) = \hat{f}_\text{pol}(\omega, \theta)
\]

Proof for angle zero:
\[
\hat{f}(\omega, 0) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j\omega x} \, dx \, dy = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} f(x, y) \, dy \right) e^{-j\omega x} \, dx = \hat{p}_0(x)
\]

Question: How to generalize to other angles?
Most imaging systems can be characterized with a forward model
Most imaging systems can be characterized with a forward model

<table>
<thead>
<tr>
<th>Modality</th>
<th>Radiation</th>
<th>Forward model</th>
<th>Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D or 3D tomography</td>
<td>coherent x-ray</td>
<td>$y_i = R_{\theta_i} x$</td>
<td>parallel, cone beam, spiral sampling</td>
</tr>
<tr>
<td>3D deconvolution microscopy</td>
<td>fluorescence</td>
<td>$y = H x$</td>
<td>brightfield, confocal, light sheet</td>
</tr>
<tr>
<td>structured illumination microscopy (SIM)</td>
<td>fluorescence</td>
<td>$y_i = H W_i x$</td>
<td>full 3D reconstruction, non-sinusoidal patterns</td>
</tr>
<tr>
<td>Positron Emission Tomography (PET)</td>
<td>gamma rays</td>
<td>$y_i = H_{\theta_i} x$</td>
<td>list mode with time-of-flight</td>
</tr>
<tr>
<td>Magnetic resonance imaging (MRI)</td>
<td>radio frequency</td>
<td>$y = F x$</td>
<td>uniform or non-uniform sampling in k space</td>
</tr>
<tr>
<td>Cardiac MRI (parallel, non-uniform)</td>
<td>radio frequency</td>
<td>$y_{t,i} = F_t W_i x$</td>
<td>gated or not, retrospective registration</td>
</tr>
<tr>
<td>Optical diffraction tomography</td>
<td>coherent light</td>
<td>$y_i = W_i F_i x$</td>
<td>with holography or grating interferometry</td>
</tr>
</tbody>
</table>
Most imaging systems can be characterized with a forward model

<table>
<thead>
<tr>
<th>Modality</th>
<th>Radiation</th>
<th>Forward model</th>
<th>Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D or 3D tomography</td>
<td>coherent x-ray</td>
<td>$y_i = R_{\theta_i} x$</td>
<td>parallel, cone beam, spiral sampling</td>
</tr>
<tr>
<td>3D deconvolution microscopy</td>
<td>fluorescence</td>
<td>$y = H x$</td>
<td>brightfield, confocal, light sheet</td>
</tr>
<tr>
<td>structured illumination microscopy</td>
<td>fluorescence</td>
<td>$y_i = H W_i x$</td>
<td>full 3D reconstruction, non-sinusoidal patterns</td>
</tr>
<tr>
<td>Positron Emission Tomography (PET)</td>
<td>gamma rays</td>
<td>$y_i = H_{\theta_i} x$</td>
<td>list mode with time-of-flight</td>
</tr>
<tr>
<td>Magnetic resonance imaging (MRI)</td>
<td>radio frequency</td>
<td>$y = F x$</td>
<td>uniform or non-uniform sampling in k space</td>
</tr>
<tr>
<td>Cardiac MRI (parallel, non-uniform)</td>
<td>radio frequency</td>
<td>$y_{t,i} = F_t W_i x$</td>
<td>gated or not, retrospective registration</td>
</tr>
<tr>
<td>Optical diffraction tomography</td>
<td>coherent light</td>
<td>$y_i = W_i F_i x$</td>
<td>with holography or grating interferometry</td>
</tr>
</tbody>
</table>
Discretization: Continuous domain formalism easily reduces to a noisy linear system
Discretization: Continuous domain formalism easily reduces to a noisy linear system.

Representation with basis functions:

\[ f(r) = \sum_{k \in \Omega} f[k] \beta_k(r) \]

Question: What type of representation is offered by \textbf{sinc}?
Discretization: Continuous domain formalism easily reduces to a noisy linear system

Representation with basis functions:

$$f(r) = \sum_{k \in \Omega} f[k] \beta_k(r)$$

Signal vector:

$$f = \{ f[k] \}_{k \in \Omega} \in \mathbb{R}^n$$

Discretization: Continuous domain formalism easily reduces to a noisy linear system

Representation with basis functions:
\[ f(r) = \sum_{k \in \Omega} f[k] \beta_k(r) \]

Signal vector: \[ f = \{f[k]\}_{k \in \Omega} \in \mathbb{R}^n \]

Discretized measurement model:
\[ y_i = \int_{\mathbb{R}^d} f(r) h_i(r) \, dr + e_i = \langle f, h_i \rangle + e_i, \quad (i = 1, \ldots, m) \]

Question: What are the sources of noise?

Discretization: Continuous domain formalism easily reduces to a noisy linear system

Representation with basis functions:

\[ f(r) = \sum_{k \in \Omega} f[k] \beta_k(r) \]

Signal vector:

\[ f = \{ f[k] \}_{k \in \Omega} \in \mathbb{R}^n \]

Discretized measurement model:

\[ y_i = \int_{\mathbb{R}^d} f(r) h_i(r) \, dr + e_i = \langle f, h_i \rangle + e_i, \quad (i = 1, \ldots, m) \]

\[ \Rightarrow \quad y = Hf + e \]

\[ [H]_{i,k} = \langle h_i, \beta_k \rangle = \int_{\mathbb{R}^d} h_m(r) \beta_k(r) \, dr \]

linear system of equations

To conclude “forward models”

Many imaging problems reduce to solving large and noisy linear systems

\[ y = Hf + e \]

Setting up the right forward model is a big step towards being able to form high quality images
Today we will talk about

- Forward models in imaging
  Relating the unknowns to the measured data

- Notions of ill-posedness and regularization
  When measurements are not enough

- Optimization at large scales
  When analytical solutions are not enough

- Plug-and-Play Priors (PnP) at large scales
  When traditional optimization is not enough
What makes imaging inverse problems difficult?
What makes imaging inverse problems difficult?

Inverse problem is well posed if \[ c_0 > 0 \] s.t., for all \( s \in X \), \[ c_0 \| s \|_H \geq 0 \]

Inverse problems in bio-imaging

noise

Linear forward model

\[ y = Hf + e \]

Problem: recover \( f \) from noisy measurements \( y \)
What makes imaging inverse problems difficult?

Problem: recover \( f \) from noisy measurements \( y \)

Question: Why can’t we simply compute the inverse \( f = H^{-1}y \)?
What makes imaging inverse problems difficult?

Inverse problem is well posed if

\[ \exists c_0 > 0 \text{ s.t., for all } s \in \mathcal{X}, \quad c_0 \| s \|_H \leq \| y \|_2 \]

Problem: recover \( f \) from noisy measurements \( y \)

Question: Why can’t we simply compute the inverse \( f = H^{-1}y \)?

1) Difficult to invert the matrix as it is non-square or too large
What makes imaging inverse problems difficult?

Problem: recover \( f \) from noisy measurements \( y \)

Question: Why can’t we simply compute the inverse \( f = H^{-1}y \)?

1) Difficult to invert the matrix as it is non-square or too large
2) Measurements do not uniquely describe the object
What makes imaging inverse problems difficult?

Problem: recover $f$ from noisy measurements $y$

Question: Why can’t we simply compute the inverse $f = H^{-1}y$?

1) Difficult to invert the matrix as it is non-square or too large
2) Measurements do not uniquely describe the object
What makes imaging inverse problems difficult?

Problem: recover $f$ from noisy measurements $y$

Question: Why can’t we simply compute the inverse $f = H^{-1}y$?

1) Difficult to invert the matrix as it is non-square or too large
2) Measurements do not uniquely describe the object
3) Noise amplification (related but not equal to 2)
Regularization framework enables the selection of the most suitable solution among alternatives
Regularization framework enables the selection of the most suitable solution among alternatives

Consider a noisy linear system with noise of bounded norm
Regularization framework enables the selection of the most suitable solution among alternatives.

Consider a noisy linear system with noise of bounded norm

\[ y = Hf + e \]  such that  \[ \|y - Hf\|_2^2 \leq \sigma^2 \]
Regularization framework enables the selection of the most suitable solution among alternatives

Consider a noisy linear system with noise of bounded norm

\[ y = Hf + e \text{ such that } \|y - Hf\|_2^2 \leq \sigma^2 \]

We consider a constrained optimization problem

**minimize** $\mathcal{R}(f)$ **subject to** $\|Hf - y\|_2^2 \leq \sigma^2$

- The “regularizer” picks the solution which we think is best
- Allows us to infuse prior knowledge into the problem
Regularization framework enables the selection of the most suitable solution among alternatives.

Consider a noisy linear system with noise of bounded norm

\[ y = Hf + e \quad \text{such that} \quad \|y - Hf\|_2^2 \leq \sigma^2 \]

We consider a constrained optimization problem

\[ \text{minimize } \mathcal{R}(f) \text{ subject to } \|Hf - y\|_2^2 \leq \sigma^2 \]

Examples when no noise

Elad, “Sparse and Redundant Representations,” 2010
Question: How to regularize in imaging?
Question: How to regularize in imaging?

\[
\minimize R(f) \text{ subject to } \|Hf - y\|^2_{\ell_2} \leq \sigma^2
\]
Question: How to regularize in imaging?

Classical approach: Tikhonov regularization

\[ R(f) = \| Df \|_{\ell_2}^2 \]

Assumption: image is smooth

minimize \( R(f) \) subject to \( \| Hf - y \|_{\ell_2}^2 \leq \sigma^2 \)

Andrey N. Tikhonov (1906-1993)
Question: How to regularize in imaging?

Classical approach: Tikhonov regularization

\[ \mathcal{R}(f) = \|Df\|_{\ell_2}^2 \quad \Rightarrow \quad \hat{f}_{\text{Tikh}} = (D^T D)^{-1} H^T \left[ H(D^T D)^{-1} H^T \right]^{-1} y \]

unique closed-form solution

minimize \( \mathcal{R}(f) \) subject to \( \|Hf - y\|_{\ell_2}^2 \leq \sigma^2 \)
Question: How to regularize in imaging?

**Classical approach: Tikhonov regularization**

$$\mathcal{R}(f) = \|Df\|_{\ell^2}^2 \Rightarrow \hat{f}_{\text{Tikh}} = (D^T D)^{-1} H^T \left[ H(D^T D)^{-1} H^T \right]^{-1} y$$

**Assumption:**
image is smooth

Question: Is image smoothness a reasonable assumption?
Question: How to regularize in imaging?

Classical approach: Tikhonov regularization

\[ R(f) = \|Df\|_2^2 \quad \Rightarrow \quad \hat{f}_{\text{Tikh}} = (D^TD)^{-1}H^T [H(D^TD)^{-1}H^T]^{-1} y \]

Modern approach: Transform-domain sparsity

\[ \text{minimize } R(f) \text{ subject to } \|Hf - y\|_2^2 \leq \sigma^2 \]
Question: How to regularize in imaging?

Classical approach: Tikhonov regularization

\[ R(f) = \|Df\|_2^2 \implies \hat{f}_{\text{Tikh}} = (D^TD)^{-1}H^T[H(D^TD)^{-1}H^T]^{-1}y \]

Modern approach: Transform-domain sparsity

minimize \( R(f) \) subject to \( \|Hf - y\|_2^2 \leq \sigma^2 \)
Question: How to regularize in imaging?

Classical approach: Tikhonov regularization

\[ \mathcal{R}(f) = \|Df\|_{\ell_2}^2 \quad \Rightarrow \quad \hat{f}_{\text{Tikh}} = (D^T D)^{-1} H^T \left[ H(D^T D)^{-1} H^T \right]^{-1} y \]

Modern approach: Transform-domain sparsity
Question: How to regularize in imaging?

Classical approach: Tikhonov regularization

\[ \mathcal{R}(f) = \|Df\|_{\ell_2}^2 \implies \hat{f}_{\text{Tikh}} = (D^T D)^{-1} H^T [H (D^T D)^{-1} H^T]^{-1} y \]

Modern approach: Transform-domain sparsity

\[ \mathcal{R}(f) = \|Df\|_{\ell_0} = \# \{ i : [Df]_i \neq 0 \} \quad \text{intractable nonconvex optimization} \]
Question: How to regularize in imaging?

Classical approach: Tikhonov regularization

\[ \mathcal{R}(f) = \|Df\|_{\ell_2}^2 \Rightarrow \hat{f}_{\text{Tikh}} = (D^T D)^{-1} H^T \left[ H (D^T D)^{-1} H^T \right]^{-1} y \]

Modern approach: Transform-domain sparsity

\[ \mathcal{R}(f) = \|Df\|_{\ell_1} \]

- convex (but nondifferentiable)
- promotes sparsity
To conclude “regularization”

Many imaging problems are ill-posed: there are infinitely many solutions

\[ y = Hf + e \]

Regularization is a strategy to select the solution that “makes sense”

\[
\begin{align*}
\text{minimize } \mathcal{R}(f) \text{ subject to } & \|Hf - y\|_2^2 \leq \sigma^2 \\
\end{align*}
\]

Classical image regularizers are linear, but increasingly they are nonlinear

\[
\begin{align*}
(20\text{th}) & \quad \mathcal{R}(f) = \|Df\|_2^2 & \Rightarrow & \quad \mathcal{R}(f) = \|Df\|_1 \\
(21\text{st})
\end{align*}
\]
Today we will talk about

- Forward models in imaging
  Relating the unknowns to the measured data

- Notions of ill-posedness and regularization
  When measurements are not enough

- **Optimization at large scales**
  When analytical solutions are not enough

- Plug-and-Play Priors (PnP) at large scales
  When traditional optimization is not enough
Proximal operator corresponds to image denoising
Proximal operator corresponds to image denoising

A more convenient formulation

\[
\min \ R(f) \text{ subject to } \| Hf - y \|_{\ell_2}^2 \leq \sigma^2 \quad \Leftrightarrow \quad \min_f \left\{ \frac{1}{2} \| y - Hf \|_{\ell_2}^2 + \lambda R(f) \right\}
\]

constrained optimization

unconstrained optimization
Proximal operator corresponds to image denoising

A more convenient formulation

\[
\min \mathcal{R}(f) \text{ subject to } \|Hf - y\|_2^2 \leq \sigma^2 \iff \min_f \left\{ \frac{1}{2}\|y - Hf\|_2^2 + \lambda \mathcal{R}(f) \right\}
\]

Image denoising corresponds to
identity measurement matrix

\[
\min_{f} \left\{ \frac{1}{2}\|y - f\|_2^2 + \lambda \mathcal{R}(f) \right\}
\]

Question: Can you comment on convexity?
Proximal operator corresponds to image denoising

A more convenient formulation

\[
\min \mathcal{R}(f) \text{ subject to } \|Hf - y\|_{\ell_2}^2 \leq \sigma^2 \iff \min_f \left\{ \frac{1}{2}\|y - Hf\|_{\ell_2}^2 + \lambda \mathcal{R}(f) \right\}
\]

Image denoising corresponds to identity measurement matrix

For a convex regularizer, the objective is strongly convex
\[
\Rightarrow \text{there is a unique minimizer}
\]
Proximal operator corresponds to image denoising

A more convenient formulation

$$\min \mathcal{R}(f) \text{ subject to } \|Hf - y\|_2^2 \leq \sigma^2 \iff \min_f \left\{ \frac{1}{2} \|y - Hf\|_2^2 + \lambda \mathcal{R}(f) \right\}$$

Image denoising corresponds to identity measurement matrix

$$\min_f \left\{ \frac{1}{2} \|y - f\|_2^2 + \lambda \mathcal{R}(f) \right\}$$

We can thus define the prox operator that solves the denoising problem

$$\text{prox}_{\lambda \mathcal{R}}(y) \triangleq \arg\min_f \left\{ \frac{1}{2} \|y - f\|_2^2 + \lambda \mathcal{R}(f) \right\}$$
Proximal operator corresponds to image denoising

Some examples of pointwise proximals

FISTA and ADMM are two popular algorithms for large-scale and nonsmooth optimization.
FISTA and ADMM are two popular algorithms for large-scale and nonsmooth optimization

Consider the objective function

\[ C(f) = D(f) + R(f) \]

where

\[ D(f) \triangleq \frac{1}{2} \| Hf - y \|_{\ell_2}^2 \]

data fit + regularizer
FISTA and ADMM are two popular algorithms for large-scale and nonsmooth optimization.

Consider the objective function

\[ C(f) = D(f) + R(f) \quad \text{where} \quad D(f) \triangleq \frac{1}{2} \|Hf - y\|_2^2 \]

Fast iterative shrinkage/thresholding algorithm (FISTA) vs. Alternating direction method of multipliers (ADMM)

\[
\begin{align*}
    z^k &\leftarrow s^{k-1} - \gamma \nabla D(s^{k-1}) \\
    f^k &\leftarrow \text{prox}_{\gamma R}(z^k) \\
    s^k &\leftarrow f^k + \left(\frac{(q_{k-1} - 1)}{q_k}\right)(f^k - f^{k-1})
\end{align*}
\]

ISTA: \( q_k = 1 \implies O(1/t) \)

FISTA: specific \( q_k \implies O(1/t^2) \)

\[
\begin{align*}
    z^k &\leftarrow \text{prox}_{\gamma D}(f^{k-1} - s^{k-1}) \\
    f^k &\leftarrow \text{prox}_{\gamma R}(z^k + s^{k-1}) \\
    s^k &\leftarrow s^{k-1} + (z^k - f^k)
\end{align*}
\]

ADMM fast practical convergence
FISTA and ADMM are two popular algorithms for large-scale and nonsmooth optimization.

Consider the objective function

\[ C(f) = D(f) + R(f) \quad \text{where} \quad D(f) \triangleq \frac{1}{2}\|Hf - y\|_2^2 \]

Fast iterative shrinkage/thresholding algorithm (FISTA) vs. Alternating direction method of multipliers (ADMM)

\[
\begin{align*}
    z^k &\leftarrow s^{k-1} - \gamma \nabla D(s^{k-1}) \\
    f^k &\leftarrow \text{prox}_{\gamma R}(z^k) \\
    s^k &\leftarrow f^k + ((q_{k-1} - 1)/q_k)(f^k - f^{k-1})
\end{align*}
\]

\[
\begin{align*}
    z^k &\leftarrow \text{prox}_{\gamma D}(f^{k-1} - s^{k-1}) \\
    f^k &\leftarrow \text{prox}_{\gamma R}(z^k + s^{k-1}) \\
    s^k &\leftarrow s^{k-1} + (z^k - f^k)
\end{align*}
\]

Question: Which one is computationally more efficient?
FISTA and ADMM are two popular algorithms for large-scale and nonsmooth optimization.

Consider the objective function

\[ C(f) = D(f) + R(f) \quad \text{where} \quad D(f) \triangleq \frac{1}{2} \| Hf - y \|_{l^2}^2 \]

Fast iterative shrinkage/thresholding algorithm (FISTA) vs. Alternating direction method of multipliers (ADMM)

\[
\begin{align*}
    z^k & \leftarrow s^{k-1} - \gamma \nabla D(s^{k-1}) \\
    f^k & \leftarrow \text{prox}_{\gamma R}(z^k) \\
    s^k & \leftarrow f^k + \left(\frac{q_k - 1}{q_k}\right)(f^k - f^{k-1})
\end{align*}
\]

\[
\begin{align*}
    z^k & \leftarrow \text{prox}_{\gamma D}(f^{k-1} - s^{k-1}) \\
    f^k & \leftarrow \text{prox}_{\gamma R}(z^k + s^{k-1}) \\
    s^k & \leftarrow s^{k-1} + (z^k - f^k)
\end{align*}
\]

Per-iteration complexity of ADMM is generally higher

\[
\nabla D(f) = H^T(Hf - y)
\]

\[
\text{prox}_{\gamma D}(f) = \left[ I + \gamma H^TH \right]^{-1}(f + \gamma H^Ty)
\]

requires matrix inversion
To conclude “optimization”

Many imaging problems are ill-posed: there are infinitely many solutions

\[ y = Hf + e \]

Regularization is a strategy to select the solution that “makes sense”

minimize \( \mathcal{R}(f) \) subject to \( \|Hf - y\|_{\ell^2}^2 \leq \sigma^2 \)

Classical image regularizers are linear, but increasingly they are nonlinear

(20th) \[ \mathcal{R}(f) = \|Df\|_{\ell^2}^2 \Rightarrow \mathcal{R}(f) = \|Df\|_{\ell^1} \] (21st)
Today we will talk about

- Forward models in imaging
  Relating the unknowns to the measured data

- Notions of ill-posedness and regularization
  When measurements are not enough

- Optimization at large scales
  When analytical solutions are not enough

- **Plug-and-Play Priors (PnP) at large scales**
  When traditional optimization is not enough
Deep learning is currently getting the best performance for image reconstruction
Deep learning is currently getting the best performance for image reconstruction.
Deep learning is currently getting the best performance for image reconstruction.

**Biomedical Dataset**

Figures 5 and 6 and Table II show the results for the biomedical dataset. In Figure 5, again, the sparse-view FBP contains line artifacts. Both TV and the proposed method remove streaking artifacts satisfactorily; however, the TV reconstruction shows the cartoon-like artifacts that are typical of TV reconstructions. This trend is also observed in severe case (x20) in Fig. 6. Quantitatively, the proposed method outperforms the TV method.

**Figure 5.** Reconstructed images of biomedical dataset from 143 views using FBP, TV regularized convex optimization [13], and the FBPConvNet.

**Table II**

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Methods</th>
<th>avg. SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FBP</td>
<td>24.97</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>36.15</td>
</tr>
<tr>
<td></td>
<td>50 views (x20)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FBP</td>
<td>13.52</td>
</tr>
<tr>
<td></td>
<td>TV</td>
<td>25.20</td>
</tr>
<tr>
<td></td>
<td>Proposed</td>
<td>28.83</td>
</tr>
</tbody>
</table>

**Figure 6:** Reconstruction results by TV, PR-FOCUSS and the proposed method which is trained using 1 MR slice for the 36 view input image. The NMSE values are written at the corner.

Han et al., 2017

**X-Ray CT**

Jin et al., 2016

**MRI**
We design the ScaDec model based on the popular U-Net architecture [45], which was recently applied to various image reconstruction tasks such as X-ray CT [27, 31]. Fig. 3 shows a detailed diagram of the proposed ConvNet architecture. There are two key properties that recommend multi-resolution decomposition based on the max-pooling and the up-convolution. This means that given a fixed size convolution kernel (FB-NN, column 2) and the total variation [17] (FB-TV, column 4), the non-linear method (LS-TV, column 5), the reconstruction by using BM3D as a plug-and-play prior (LS-BM3D, column 6). The values above images show the SNR (dB) of the reconstruction. The first effective receptive field of the network increases in our case), the effective receptive field of the network increases different scattering scenario,

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground truth</td>
<td>FBP SNR 24.06</td>
<td>TV SNR 29.64</td>
<td>FBPConvNet SNR 35.38</td>
</tr>
</tbody>
</table>

**X-Ray CT**  
Jin et al., 2016

<table>
<thead>
<tr>
<th>Weak</th>
<th>14.64 dB</th>
<th>14.64 dB</th>
<th>22.32 dB</th>
<th>22.32 dB</th>
<th>26.56 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>8.90 dB</td>
<td>14.75 dB</td>
<td>10.28 dB</td>
<td>22.47 dB</td>
<td>26.15 dB</td>
</tr>
</tbody>
</table>

**Diffraction Tomography**  
Sun et al., 2018

**MRI**  
Han et al., 2017

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground truth</td>
<td>X: Input (36)</td>
<td>Total variation</td>
<td></td>
</tr>
<tr>
<td>PR-FOCUSS</td>
<td>Proposed (1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**D:+PRWV**  
G

**OPTICS**  
14684
A well established deep learning pipeline: first backproject then denoise with a ConvNet
A well established deep learning pipeline: first backproject then denoise with a ConvNet

Data processing pipeline

- **data**
- **H^H** ("backproject")
- **noisy image**
- **denoising CNN**
- **output**
Question: What are some of the key limitations of this approach?
A well established deep learning pipeline: first backproject then denoise with a ConvNet

Data processing pipeline

1) Implicit dependence of CNN on the forward model

Hard to decouple the individual contributions of $D$ and $R$
A well established deep learning pipeline: first backproject then denoise with a ConvNet

Data processing pipeline

1) Implicit dependance of CNN on the forward model

2) Consistency with the measured data is unclear

No explicit measure of the deviation from the data
A well established deep learning pipeline: first backproject then denoise with a ConvNet

Data processing pipeline

1) Implicit dependance of CNN on the forward model
2) Consistency with the measured data is unclear
3) Difficult to impose nontrivial hard constraints on the image

Example: We absolutely need the image gradient to be smaller than epsilon
A well established deep learning pipeline: first backproject then denoise with a ConvNet

Data processing pipeline

1) Implicit dependance of CNN on the forward model
2) Consistency with the measured data is unclear
3) Difficult to impose nontrivial hard constraints on the image
4) Not principled: how to select the right architecture?

Variations in the problem are not explicitly linked to model parameters
A well established deep learning pipeline: first backproject then denoise with a ConvNet

Data processing pipeline

1) Implicit dependance of CNN on the forward model
2) Consistency with the measured data is unclear
3) Difficult to impose nontrivial hard constraints on the image
4) Not principled: how to select the right architecture?
5) Difficult to generalize to nonlinear forward models

What happens if there is no backprojection?
A well established deep learning pipeline: first backproject then denoise with a ConvNet

1) Implicit dependance of CNN on the forward model
2) Consistency with the measured data is unclear
3) Difficult to impose nontrivial hard constraints on the image
4) Not principled: how to select the right architecture?
5) Difficult to generalize to nonlinear forward models
Treating the denoising CNN as a proximal operator allows to separate the prior from the forward model.
Treating the denoising CNN as a proximal operator allows to separate the prior from the forward model.
Treating the denoising CNN as a proximal operator allows to separate the prior from the forward model.

Train a CNN to denoise for various noise levels

$$\approx \text{prox}_{\gamma R}(y)$$

Treating the denoising CNN as a proximal operator allows to separate the prior from the forward model.

Train a CNN to denoise for various noise levels.

Use the trained CNN as a Plug-and-Play Prior (PnP)

\[ z^k \leftarrow s^{k-1} - \gamma \nabla D(s^{k-1}) \]
\[ f^k \leftarrow \text{denoise}_\sigma(z^k) \]
\[ s^k \leftarrow f^k + ((q_{k-1} - 1)/q_k)(f^k - f^{k-1}) \]

\text{PnP-FISTA}

\[ z^k \leftarrow \text{prox}_{\gamma D}(f^{k-1} - s^{k-1}) \]
\[ f^k \leftarrow \text{denoise}_\sigma(z^k + s^{k-1}) \]
\[ s^k \leftarrow s^{k-1} + (z^k - f^k) \]

\text{PnP-ADMM}

Plug-and-Play Priors (PnP) approach has been shown to yield state-of-the-art results
Plug-and-Play Priors (PnP) approach has been shown to yield state-of-the-art results

<table>
<thead>
<tr>
<th>Method</th>
<th>Average PSNR (dB) over 10 images</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV</td>
<td>29.22</td>
</tr>
<tr>
<td>IDD-BM3D</td>
<td>30.92</td>
</tr>
<tr>
<td>ASDS-Reg</td>
<td>30.11</td>
</tr>
<tr>
<td>NCSR</td>
<td>31.09</td>
</tr>
<tr>
<td>PnP</td>
<td>31.33</td>
</tr>
</tbody>
</table>

Romano et al., “The Little Engine That Could: Regularization by Denoising,” 2017
Plug-and-Play Priors (PnP) approach has been shown to yield state-of-the-art results.

<table>
<thead>
<tr>
<th>Method</th>
<th>Average PSNR (dB) over 10 images</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV</td>
<td>29.22</td>
</tr>
<tr>
<td>IDD-BM3D</td>
<td>30.92</td>
</tr>
<tr>
<td>ASDS-Reg</td>
<td>30.11</td>
</tr>
<tr>
<td>NCSR</td>
<td>31.09</td>
</tr>
<tr>
<td>PnP</td>
<td>31.33</td>
</tr>
</tbody>
</table>

Romano et al., “The Little Engine That Could: Regularization by Denoising,” 2017
Plug-and-Play Priors (PnP) approach has been shown to yield state-of-the-art results.

<table>
<thead>
<tr>
<th>Method</th>
<th>Average PSNR (dB) over 10 images</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV</td>
<td>29.22</td>
</tr>
<tr>
<td>IDD-BM3D</td>
<td>30.92</td>
</tr>
<tr>
<td>ASDS-Reg</td>
<td>30.11</td>
</tr>
<tr>
<td>NCSR</td>
<td>31.09</td>
</tr>
<tr>
<td>PnP</td>
<td>31.33</td>
</tr>
</tbody>
</table>

(a) Ground Truth  
(b) Input 21.40dB  
(d) NCSR 30.03dB  
(e) $P^3$-TNRD 30.36dB
Can we say anything about convergence?
Can we say anything about convergence?


Can we say anything about convergence?

Result #1: When $\mathcal{D}(\cdot)$ is convex and $\nabla \text{denoise}_\sigma(\cdot)$ is a symmetric matrix with eigenvalues in $[0, 1]$, then $\text{denoise}_\sigma(\cdot)$ is a proximal operator.

Result #2: When both $\nabla \mathcal{D}(\cdot)$ and $\text{denoise}_\sigma(\cdot)$ are bounded operators, PnP-ADMM with damping converges to a fixed point.
Can we say anything about convergence?

**Result #1:** When $\mathcal{D}(\cdot)$ is convex and $\nabla\text{denoise}_\sigma(\cdot)$ is a symmetric matrix with eigenvalues in $[0, 1]$, then $\text{denoise}_\sigma(\cdot)$ is a proximal operator.

**Result #2:** When both $\nabla\mathcal{D}(\cdot)$ and $\text{denoise}_\sigma(\cdot)$ are bounded operators, PnP-ADMM with damping converges to a fixed point.

<table>
<thead>
<tr>
<th>Method</th>
<th>Avg 10%</th>
<th>Avg 20%</th>
<th>Avg 50%</th>
<th>Avg 70%</th>
<th>Avg 90%</th>
<th>Avg 95%</th>
<th>Avg 100%</th>
</tr>
</thead>
</table>

---


Can we say anything about convergence?
Can we say anything about convergence?

Can we say anything about convergence?

**Useful definitions**

\[ P(f) \triangleq \text{denoise}_\sigma(f - \gamma \nabla D(f)) \]

**gradient-denoiser operator**

\[ \text{fix}(P) \triangleq \{ f \in \mathbb{R}^n : f = P(f) \} \]

**its of fixed points**
Can we say anything about convergence?

**Useful definitions**

\[ P(f) \triangleq \text{denoise}_\sigma(f - \gamma \nabla D(f)) \quad \text{fix}(P) \triangleq \{ f \in \mathbb{R}^n : f = P(f) \} \]

#1: Let \( \text{denoise}_\sigma(\cdot) = \text{prox}_{\gamma R}(\cdot) \). Then, \( f^* \in \text{fix}(P) \) iff it minimizes \( C = D + R \)

#2: Run PnP-ISTA with a nonexpansive denoiser for \( t \geq 1 \) iterations. Then

\[
\min_{k \in \{1, \ldots, t\}} \left\{ \| f^{k-1} - P(f^{k-1}) \|_{\ell_2}^2 \right\} = O(1/t)
\]

#3: For nonexpansive denoisers, fixed points of PnP-ADMM coincide with \( \text{fix}(P) \)
PnP-SGD is an online extension useful when dealing with a large number of measurements
PnP-SGD is an online extension useful when dealing with a large number of measurements.

Consider the following data-fidelity term

\[
D(f) = \frac{1}{2I} \sum_{i=1}^{I} \|y_i - H_i f\|_2^2 \Rightarrow \nabla D(f) = \frac{1}{I} \sum_{i=1}^{I} H_i^T(H_i f - y)
\]

cost of computing the gradient is linear in the number of measurements.

PnP-SGD is an online extension useful when dealing with a large number of measurements.

Consider the following data-fidelity term

$$D(f) = \frac{1}{2I} \sum_{i=1}^{I} \| y_i - H_i f \|_{2}^{2} \quad \Rightarrow \quad \nabla D(f) = \frac{1}{I} \sum_{i=1}^{I} H_i^T (H_i f - y)$$

PnP-SGD can accelerate imaging by parallelizing the processing of each data item.

\[
\nabla \hat{D}(s^{k-1}) \leftarrow \text{minibatchGradient}(s^{k-1}, B)
\]
\[
z^k \leftarrow s^{k-1} - \gamma \nabla \hat{D}(s^{k-1})
\]
\[
f^k \leftarrow \text{denoise}_\sigma(z^k)
\]
\[
s^k \leftarrow f^k + \left(\frac{(q_{k-1} - 1)}{q_k}\right)(f^k - f^{k-1})
\]

use only $B$ measurements per iteration instead of $I$

PnP-SGD is an online extension useful when dealing with a large number of measurements

Consider the following data-fidelity term

\[ D(f) = \frac{1}{2I} \sum_{i=1}^{I} \| y_i - H_i f \|_{\ell^2}^2 \quad \Rightarrow \quad \nabla D(f) = \frac{1}{I} \sum_{i=1}^{I} H_i^T (H_i f - y) \]

PnP-SGD can accelerate imaging by parallelizing the processing of each data item

\[
\hat{\nabla} D(s^{k-1}) \leftarrow \text{minibatchGradient}(s^{k-1}, B) \\
S^k \leftarrow S^{k-1} - \gamma \hat{\nabla} D(s^{k-1}) \\
f^k \leftarrow \text{denoise}_\sigma(z^k) \\
s^k \leftarrow f^k + \left( (q_{k-1} - 1)/q_k \right) (f^k - f^{k-1})
\]

PnP-SGD converges to the same set of fixed points as batch PnP algorithms

PnP-SGD converges to the same set of fixed points as batch PnP algorithms

#4: Run PnP-SGD for $t \geq 1$ iterations under some mild assumptions. Then

$$\mathbb{E} \left[ \min_{k \in \{1, \ldots, t\}} \left\{ \| f^{k-1} - P(f^{k-1}) \|_2^2 \right\} \right] \leq C \left[ \frac{\gamma^2 \nu^2}{B} + \frac{2\gamma \nu}{\sqrt{B}} \| f^0 - f^* \|_{\ell_2} + \frac{\| f^0 - f^* \|_{\ell_2}^2}{t} \right]$$

Convergence in expectation. $C$ is a constant. Note the case when $B = t$

PnP-SGD converges to the same set of fixed points as batch PnP algorithms

#4: Run PnP-SGD for \( t \geq 1 \) iterations under some mild assumptions. Then

\[
\mathbb{E} \left[ \min_{k \in \{1, \ldots, t\}} \left\{ \|f^{k-1} - P(f^{k-1})\|_2^2 \right\} \right] \leq C \left[ \frac{\gamma^2 \nu^2}{B} + \frac{2\gamma \nu}{\sqrt{B}} \|f^0 - f^*\|_{\ell_2} + \frac{\|f^0 - f^*\|_{\ell_2}^2}{t} \right]
\]

For many measurements PnP-SGD converges faster than batch algorithms.

For the same measurement budget, PnP-SGD gets much higher quality results.
Conclusion

Image reconstruction is a fascinating research area that brings together physics, signal processing, nonlinear optimization, and machine learning.

We are increasingly reliant on implicit regularization using nonlinear operators, such as deep neural networks or nonlinear filters.

Plug-In SGD is a theoretically sound algorithm that can regularize at large-scales using nonlinear operators.

Contact Info

Prof. Ulugbek Kamilov
Computational Imaging Group (CIG)
Washington University in St. Louis
Email: kamilov@wustl.edu
Web: http://cigroup.wustl.edu
Twitter: @wustlcig
Judea Pearl won the Turing Award in 2011 for fundamental contributions to artificial intelligence
Judea Pearl won the Turing Award in 2011 for fundamental contributions to artificial intelligence.
Judea Pearl won the Turing Award in 2011 for fundamental contributions to artificial intelligence

We live in an era that presumes Big Data to be the solution to all our problems (...) But I hope with this book to convince you that data are profoundly dumb. Data can tell you that the people who took a medicine recovered faster than those who did not take it, but they can’t tell you why.
We live in an era that presumes Big Data to be the solution to all our problems. Courses in data science are proliferating in our universities, and jobs for data scientists are lucrative in the companies that participate in the data economy.

But I hope with this book to convince you that data are profoundly dumb. Data can tell you that the people who took a medicine recovered faster than those who did not take it, but they can't tell you why.

The belief that data can tell the full story is a misconception. To produce truly useful insights, data must be combined with models that infuse what we know about the problem.