Deep Learning-based Solvability of Underdetermined Inverse problems in medicines

Learn $f(data) = \text{useful output}$

Jan 2020@ IPAM

Jin Keun Seo
with Chang Min Hyun & Seong Hyeon Baek
Yonsei Univ., Korea

lunar New Year 2020
This talk is based on joint work with my PhD students.

In this talk, many of my personal opinions (not rigorous) are included to give an exaggerated emphasis on deep learning.

Learn \( f(b) = \text{useful output} \)
ill-posed inverse problems

Hadamard's well-posedness (excluding existence)

\[ Ay = b \]

is well-posed if the following two conditions hold:

1) for each \( b \), \( Ay = b \) has a unique solution;
2) the solution is stable under perturbation of \( b \).

Whether or not a problem is well-posed may be dependent on how the solution is expressed.

Many problems are ill-posed because we are overly ambitious or lacking in expressiveness.
Conventional CT and MRI data collections are designed for the corresponding forward matrix $A$ to be well-expressed & to be reasonably complete.

# of equations (data) $\approx$ # of unknowns (pixels of image)
The classical principle that make problems well-posed is:

\[ \# \text{ of equations (number of samples)} \approx \# \text{ of unknowns (number of pixels of image)}. \]

**Tomography with Nyquist Sampling**

- **MRI** measures approximately an image's Fourier transform. Nyquist sampling is required for the analytic reconstruction.

  \[ \# \text{ pixels in image } \approx \# \text{ samplings in } k\text{-space} \]

- **CT** measures approximately an image’s Radon transform. According to Nyquist sampling & Fourier slice theorem,

  \[ \sqrt{\# \text{ pixels in image}} \approx \# \text{ projection angles} \]
Why do we pay attention to underdetermined problems (fewer equations than unknowns) in CT & MRI?

It is because of the great needs to reduce radiation dose in CT & data acquisition time in MRI.

The goal is to make
\[
\frac{\# \text{ of equations}}{\# \text{ of unknowns}} \quad \text{as small as possible.}
\]

\[
\dim \{ z : Az = 0 \} = \# \text{unknowns} - \# \text{equations}
\]
Solving \( Ay = b \) is to find the reconstruction map \( f : b \rightarrow y = A_{full}^{-1}b_{full} \).

Is it possible to solve it?

- \( b_{full} \) denotes the "fully sampled" data (e.g., sinogram in CT and k-space data in MRI).
- \( b = S_{ub}b_{full} \) where \( S_{ub} \) denotes a subsampling operator.

\( A = S_{ub} \)

\( A_{full} \) is discrete Fourier transform in MRI & Radon transform in CT.
Undersampled MRI problem

Is it possible?

$x = A^\dagger b$

$y = A_{\text{full}}^{-1} b_{\text{full}}$

$f: x \to y$

Subsampling (30%)

$A^\dagger$: Pseudo-Inverse of $A$. 

Full sampling
Without imposing prior knowledge on the solution, this problem has infinitely many solutions.

Need to choose one out of infinitely many images in $N_b(A): = \{ z : Az = b \}$.

$dim \, N_b(A) = \# \text{columns} - \# \text{rows}$
Is it possible to find $f: A^\dagger b \rightarrow y = A_{full}^{-1} b_{full}$?

✓ Solving $Ay = b$ depends on an appropriate use of a priori information about medical CT or MRI images as solutions.
✓ We need to consider a constraint problem:

$$Ay = b \text{ subject to } y \in M \text{ (Solution Manifold)}$$

Example: sparse view CT

$$= A_{full}^{-1} b_{full} \neq A^\dagger b =$$

$\in M$
Example 1

Sparse View CT

Well-expressed

Well-expressed

Similar noise patterns regardless of images

It is capable of learning

\[ f(A^\dagger Ay) = y \quad \forall \ y \in \text{Image Manifold} \]
Example 2

Local CT

Subsampling
\[ S_{\text{sub}} \]

\[ b = S_{\text{sub}} b_{\text{full}} \]

Zero filling
\[ \mathcal{S}_{\text{sub}}^* \]

\[ S_{\text{sub}}^* b = \mathcal{S}_{\text{sub}}^* S_{\text{sub}} b_{\text{full}} \]

Well-expressed
\[ A_{\text{full}}^{-1} y = b_{\text{full}} \]

\[ y = A_{\text{full}}^{-1} b_{\text{full}} \chi_{\Omega_{\text{ROI}}} \]

\[ x - y = \frac{1}{2\pi^2} \int_{t_{\theta}(a) \neq \Omega_{\text{ROI}}} \mathcal{R}_{\theta} \left( \frac{\partial}{\partial a} b(t_{\theta}(a')) \right) \frac{a - a'}{a - a'} da' \]

It is capable of learning because \( x - y \) is analytic

\[ f \]

\[ x = A_{\text{full}}^\dagger b = \mathcal{R}_{\text{sub}}^{-1} S_{\text{sub}}^* b \]
Dental CBCT: Need to develop a reconstruction method that addresses the problems caused by "Offset detector, FOV truncation, Low X-ray dose".

Avoid methods having many iteration steps!
Underdetermined MRI

Example 3

Well-expressed

$$A_{\text{full}} y = b_{\text{full}}$$

$$y = A_{\text{full}}^{-1} b_{\text{full}}$$

$$b_{\text{full}}$$

$$S_{\text{sub}} b = S_{\text{sub}} S_{\text{sub}} b_{\text{full}}$$

Subsampling

$$b = S_{\text{sub}} b_{\text{full}}$$

Zero filling

$$S_{\text{sub}}^* b = S_{\text{sub}}^* S_{\text{sub}} b_{\text{full}}$$

Violating Nyquist Sampling Rule

$$A y = b$$

ill-posed

$$f_b$$

$$A$$

$$A^\dagger$$

$$A_{\text{full}}$$

$$x = A^\dagger b$$

$$f$$

$$f(A^\dagger A y) = y \quad \forall \ y \in \text{Image Manifold}$$

Is it capable of learning?
Methods to solve the ill-posed problem $y = Ax$

✓ This is a highly nonlinear problem!
The degree of nonlinearity depends on the sampling of data $b$ and solution manifold.

✓ Methods to impose Prior Knowledge on the solution

Hand-made Sparse Sensing
- Use sparse representation of $y$
- Regularized data fitting method:
  $$f(x) = Wh, \ h = \arg\min_h ||AWh - x||_2^2 + ||h||_1$$
- Single data fidelity

Machine-made Deep Regression
- Use training data $\{y^{(n)}: n = 1, \ldots, N\}$ to get the prior knowledge.
- Deep Learning:
  $$f = \arg\min_{f \in \text{Neural Nets}} \sum_k ||y_k - f(x_k)||_2^2$$
- Group data fidelity
Comparison

- Hand-made Sparse Sensing

\[ f(x) = Wh, \ h = \arg \min_h ||AWh - x||_2^2 + \lambda ||h||_1 \]

versus

- Machine-made DL Approach

\[ f = \arg \min_{f \in \text{Neural Nets}} \sum_k ||y_k - f(x_k)||^2 \]
Comparison: Hand-Made vs Machine-Made

Test problem: Sparse View CT model with specially chosen $M_{\text{image}}$

In this sparse-view CT model, CS methods are known to work well.

Find $x = A^\dagger b$

Find $f: x \rightarrow y$

Performance Evaluation

Feature 1

Feature 2
For this sparse-view CT problem, we use a special solution manifold $M_{image}$ (assumed to be unknown).

Since this solution manifold is only 7 dimension, $Ay = b$ can be solvable only with 7 equations.
Comparison: Hand-Made vs Machine-Made

- CS and linear approaches eliminate the feature 1.

\[ f(x) = Dh, \quad h = \arg\min_h \|ADh - Ax\|_2^2 \]

\[ f(x) = Dh, \quad h = \arg\min_h \|ADh - Ax\|_2^2 + \lambda \|h\|_1 \]

\[ f(x) = \arg\min_y \|Ay - Ax\|_2^2 + \lambda \|\nabla y\|_1 \]

- Deep learning preserves the feature 1.

\[ f = \arg\min_{f \in \text{NN}} \frac{1}{n_{\text{data}}} \sum_{n=1}^{n_{\text{data}}} \|f(x^{(n)}) - y^{(n)}\|_2^2 \]
Man-Made vs Machine-Made

\[ x = A^\dagger b \]

\[ f : x \rightarrow y \]

PCA

Produce terrible outcome due to the use of insufficient orthogonal basis.

Total Variation

Remove key features.

Deep Learning

Keep key features.
Linear methods (PCA, Wavelet decomposition) may be unable to deal with the highly curved solution manifold.

Consider the vector space spanned by images \( \{ y^{(0)}, y^{(1)}, \ldots, y^{(N)} \} \) where \( y^{(k)} = k\pi/N \) rotation of image \( y^{(0)} \).

The middle image between \( y^{(0)} \) & \( y^{(1)} \) cannot be expressed properly by the space spanned by \( \{ y^{(0)}, y^{(1)}, \ldots, y^{(N)} \} \).
Can CS methods preserve the feature 1 while removing artifacts? May be NOT.

TV approach: $f_{TV}(x) = \arg\min_y \|Ay - x\|_2^2 + \lambda \|\nabla y\|_1$

It may not preserve some detailed structures that may contain crucial medical information.

Remove everything within this interval without exception.
TV approach: $f_{TV}(x) = \arg\min_y ||Ay - x||^2_{\ell^2} + \lambda||\nabla y||_{\ell^1}$

- The performance depends on the regularization parameter.
- Need several iterations to find a sparse expression.
• DL approach can selectively preserve the feature 1.
Deep Learning Approach

Use training data to learn both $f$ and image manifold such that

$$f(A^\dagger Ay) = y \quad \forall \ y \in \text{Image Manifold}.$$
One of DL’s most important advantages is to provide non-iterative reconstruction methods for highly non-linear problems.

What is learnable. What is NOT.

The necessary condition for learning $f$ is that

$$f(A^\dagger Ay) = y \quad \forall y \in \text{Image Manifold}.$$  

Use training data $\{y^{(n)}: n = 1, \ldots, N\}$ to get prior knowledge.
Impact of Training Data: It is critical to choose suitable training datasets to reflect the appropriate image priors, in order to preserve detailed features of the images.

\[ f(\text{anomaly}) = ? \]

<table>
<thead>
<tr>
<th>Training Data 1</th>
<th>Training Data 2</th>
<th>Training Data 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
<td><img src="image2.png" alt="Image 2" /></td>
<td><img src="image3.png" alt="Image 3" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Image 4" /></td>
<td><img src="image5.png" alt="Image 5" /></td>
<td><img src="image6.png" alt="Image 6" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image 7" /></td>
<td><img src="image8.png" alt="Image 8" /></td>
<td><img src="image9.png" alt="Image 9" /></td>
</tr>
</tbody>
</table>

Joint work with Hyungsuk Park
Observation: The reconstruction map $f: x = A^\dagger b \to y$ is learnable if $A$ satisfies the M-RIP (manifold restricted isometry property) condition.

\[ c\|y - y'\| \leq \|Ay - Ay'\| \leq \frac{1}{c}\|y - y'\| \quad \text{for all } y, y' \in M_{\text{image}} \]

- Necessary condition for learnability is

\[ \dim M_{\text{image}} \leq \text{Rank } A \]

- $M_{\text{image}}$ indicates the unknown solution Manifold
M-RIP condition: $N_b(A) \cap M_{image} = \{y\}$ (uniqueness & stability).

Solving Underdetermined Problem: $Ay = b$

- $M_{image}$ indicates a solution manifold that is assumed to be a good regression of MR head image data distributions $\{y^{(n)}: n = 1, \cdots, N\}$.

- $x = A^\dagger b$ is the minimum-norm solution which will be used to find the true solution $y$. 

\[\mathcal{N}_b(A) = \{z: Az = Ax = b\} = \{\cdots\} \]

\[\dim M_{image} \leq \mathcal{D}_{image} - \dim \mathcal{N}_b(A)\]
is a highly nonlinear problem!

\[ \mathbf{A} \mathbf{x} = \mathbf{b} \]

- \( \mathbf{A} \) is \((\# \text{ equations}) \times (\# \text{ unknowns})\) matrix
- \( M_{\text{image}} = \{ y : y = G(h), \ h \in K \} \), where \( G \) is a generator & \( K \) is a compact subset of \( \mathbb{R}^k \).

**Observation:** \( f : x = \mathbf{A}^\dagger \mathbf{b} \rightarrow y \) is nonlinear if \( \dim (\text{span}\{\partial_j G(h) : h \in K\}) > \# \text{ equations} \).

- The map \( f : x = \mathbf{A}^\dagger \mathbf{b} \rightarrow y \) can be viewed as an image restoration function with filling-in missing data in \( x \). Therefore, \( \nabla f(x) \) depends on the image structure in \( x \).
- The nonlinearity of \( f \) is affected by sampling and the degree of bending of the manifold \( M_{\text{image}} \).
Observation: \( f: x = A^+b \rightarrow y \) is nonlinear if \( \dim(\text{span}\{\partial_jG(h): h \in K\}) > \# \text{ equations} \).

Proof:

- \( f(x) = y \rightarrow f(A^+Ay) = y \rightarrow f(A^+AG(h)) = AG(h) \)
- If \( f: x = A^+b \rightarrow y \) is linear, \( B = \nabla f(\cdot) \) is a constant matrix &
  \[
  BA^+A\nabla G(h) = \nabla G(h) \text{ for all } h \in K.
  \]

\( \checkmark \) Hence, all \( \partial_jG(h) \in \text{Eigen}_1(BA^+A) \), the eigenspace of \( BA^+A \) corresponding to the eigenvalue 1.

\( \checkmark \) This is not possible if \( \dim(\text{span}\{\partial_jG(h): h \in K\}) > \# \text{ equations} \).

Message: The degree of nonlinearity depends on the sampling of data b &
the degree of bending of the solution manifold \( M_{image} \).
Example 1: Sparse View CT

Both deep learning and compressed sensing work very well for this kind of problems.

$L^1$ – Regularized data fitting technique

Assume that there exist $W$ such that $y = Wz$ with $z$ being sparse.

$$
\min E(z) := \|g(Wz) - x\|^2 + \lambda \|z\|_1
$$

$$
z = S_{\lambda\alpha}(z - \alpha \nabla \|g(Wz) - x\|^2)
$$

$$
S_{\lambda\alpha}(z) = \text{sign}(z) \max\{ |z| - \lambda \alpha, 0 \}
$$

By eliminating this simple noise structure.
Deep learning works well because of unique continuation of analytic function along the vertical direction.

Example 2: Local CT

\[
\Psi_{\theta_0}^{\text{out}} b(t_{\theta_0}(a)) = \frac{1}{2\pi^2} \int_{t_{\theta_0}(a') \notin \Omega_{\text{ROI}}} t_{\theta_0}(a') \frac{\partial}{\partial a'} b(t_{\theta_0}(a')) \frac{R_{\theta_0}^*}{a' - a} \, da'
\]
Inverse Fourier Transform

According to the Poisson summation formula, the discrete Fourier transform of $b = S_{sub} b_{full}$ (uniformly subsampled data with factor 4) produces the following four-folded image.

Example 3: Underdetermined MRI

$\mathbf{b}_{full}$
Full sampling

$b = S_{sub} b_{full}$
uniform subsampling with factor 4

$A^{-1}_{full} b_{full}$

Inverse Fourier Transform

$A^\dagger b$

Inverse Fourier Transform
If we use **uniform subsampling** $S_{sub}$ with factor 4,

it is difficult to learn $f$ s. t. $f(A^\dagger Ay) = y \quad \forall \, y \in \text{Image Manifold}$

- Fail to satisfy M-RIP condition.
- DL is NOT a magic.

**Null space**

$\psi \in \mathcal{N}_0(A^\dagger A)$

$y \in \mathcal{M}$

$x = A^\dagger Ay = A^\dagger A(y + \psi)$

**Poisson summation formula**
However, the result changes dramatically by adding only one line in k-space.

\[ b = S_{sub} b_{full} \]

\[ x = A^\dagger b \]

\( f(x) \)
Why does the learning effect dramatically change by adding only one line in k-space?

- Let $\mathbf{A}_1$ be the sensitivity matrix corresponding to uniform sampling with factor 4.
- Let $\mathbf{A}_2$ be the sensitivity matrix corresponding to the row just above the center.

$$\{ \mathbf{y} : \mathbf{A}_1^\dagger \mathbf{A}_1 \mathbf{y} = \text{image} \} = \{ \text{image}, \text{image}, \text{image}, \text{image} \}$$

Indistinguishable

$\mathbf{A}_2^\dagger \mathbf{A}_2$

Distinguishable

Inverse Fourier transform of the single line in k-space

Because of this rough signature, it is capable of learning $f$ s.t.

$$f(\mathbf{A}^\dagger \mathbf{A} \mathbf{y}) = \mathbf{y} \quad \forall \mathbf{y} \in \text{Image Manifold}$$
Let us consider learning ability issue:

Patch images vs full image

Learning ability

\[ f_\eta : x_\eta \rightarrow y_\eta \]

As \( \eta \) increases, the number of unknowns increases more rapidly than the number of equations.

\[ M_{image}^{\eta} = \{ y_\eta : y_\eta \text{ is a } 256 \times \eta \text{ image patch extracted from } y \in M_{image} \} \]

My personal opinion

- Dimension of the manifold \( M_{image}^{\eta} \) does not increase proportionally to \( \eta \).
- Hence, the learning ability about \( f_\eta : x_\eta \rightarrow y_\eta \) is gradually improved as \( \eta \) increases.
Experimental results demonstrate that the learning ability about $f_\eta : x_\eta \rightarrow y_\eta$ is gradually improved as $\eta$ increases.
Reasons for expecting $\text{dim } M_{image}^\eta$ to grow significantly slowly as $\eta$ increases.

$M_{image}^\eta = \{ y_\eta : y_\eta \text{ is a } 256 \times \eta \text{ image patch extracted from } y \in M_{image} \}$

✓ Assume that $M_{image}$ is the set of all the human head MR images.
✓ Then, all the images in $M_{image}$ possess a similar anatomical structure that consists of skull, gray matter, white matter, cerebellum, among others.
✓ In addition, every skull and tissue in the image have distinct features that can be represented nonlinearly by a relatively small number of latent variables, and so does for the entire image.
✓ Notably, the skull and tissues of the image are spatially interconnected, and even if a part of the image is missing, the missing part can be recovered with the help of the surrounding image information.
Challenging Issue: Low-dimensional representation of MR and CT images (high dimensional data: $512 \times 512 \times 400$ voxels)

**GAN** (Generative Adversarial Network)

**VAE** (Variational Autoencoder)

Given data distributions $\{y^{(n)}: n = 1, \ldots, N\}$ in medical images (e.g. dental CBCT data), can we find a low dimensional latent generator (decoder) $\Psi: h \rightarrow y$ and an encoder $\Phi: y \rightarrow h$ such that $\Psi \circ \Phi(y) \approx y$ for all $y \in M_{image}$.
One of challenging issues for solving an ill-posed problem is to find a low-dimensional representation.

5 Latent variables

\[ h = (h_1, \ldots, h_5) \]

\[ \Psi(h) \]

Generator / decoder
Latent variable

Disentangled expression with extracting the underlying explanatory axis
Electrical Impedance Tomography is known to be a highly ill-posed problem. 

$$A = \frac{db}{dy}$$ (sensitivity matrix)

$$\dim\{y \in \mathbb{R}^{16384} : Ay = 0\} > 16384 - 208$$

However, it can be well-posed if we give up excessive ambition or find a way to make a low dimensional expression.
Hand-made regularization techniques may not be effective for EIT imaging.

\[ \hat{\gamma} = \text{argmin} \| A \hat{\gamma} - \hat{V} \| + \lambda \text{Reg}(\hat{\gamma}) \]

\( (L^2, L^1, TV \text{ regularization}) \)

Despite myriads of profound theories of EIT over the past 40 years, there still are some problems for clinical use.
208 = \# of equations (data) \ll 16384 = \# of unknowns (pixels of image).

\[ A = \frac{db}{dy} \text{ (sensitivity matrix)} \]

\[ \dim\{y \in \mathbb{R}^{16384} : Ay = 0\} > 16384 - 208 \]

This can be well-posed if we can find a low dimensional representation of solutions.

Deep learning framework may provide a nonlinear regression on training data which acts as learning complex prior knowledge on the output.

\[ y = \Psi(h) \]
Low-dimensional latent representation produces $\mathcal{M}$ manifold.

- Interpolation between two points $h_i$ and $h_j$ in the latent space. Between the two given images, VAE can generate the interpolated image.

- Tangent vectors on manifold $\mathcal{M}$

Image $\dot{\gamma} = \Psi(h)$
What about low-dimensional representation of high dimensional images such as MR and CT images.

\[ \Psi(h) = \psi(h) \]

So far, my team has tried several kinds of GANs and VAE, but has not succeeded.
For high dimensional data, AEs suffer from image blurring and loss of small details.
GANs have shown remarkable success in generation of various realistic images. However, there exist some limitations in synthesizing high resolution medical data. The GAN's approach makes it difficult to deal with high-dimensional data because the generated image can be easily distinguished from the training data, which can lead to collapse or instability during training process.

<table>
<thead>
<tr>
<th>$\theta_{seven} = 100$</th>
<th>Generated 256×256 Head MR Image</th>
<th>Generated 512×512 Head CT Image</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VAE</strong></td>
<td><img src="image1.png" alt="Images" /></td>
<td><img src="image2.png" alt="Images" /></td>
</tr>
<tr>
<td><strong>GAN</strong></td>
<td><img src="image3.png" alt="Images" /></td>
<td><img src="image4.png" alt="Images" /></td>
</tr>
</tbody>
</table>
GANs have a remarkable ability to generate these images. AEs learn a bidirectional mapping (encoder and decoder), while GANs learn only the unidirectional mapping (decoding) in high dimensional medical images.

However, for high dimensional data, AEs suffer from image blurring and loss of small details. GANs have difficulties in encoding high dimensional images.

My personal opinion

AE can control this latent variables
Challenging Issue: Generalization

Training error $\approx 0$
$$\sum_{k} ||y_k - f(x_k)||^2 \approx 0$$

Memorize learning materials well

Test error $\approx 0$
$$y_{test} - f(x_{test}) \approx 0$$

Hope

Problems that do not appear in the tutorial also find the correct answer.
Example of Memorization without Generalization

Recently, several experiments regarding adversarial classifications (false positive output of cancer) have shown that deep neural networks (obtained via gradient descent-based error minimization procedure) are vulnerable to various noisy-like perturbations, resulting in incorrect output (that can be critical in medical environments).

Adversarial attacks against medical deep learning systems
The percentage represents the probability of Pneumothorax.
MNIST example of Memorization without Generalization

\[ \text{dist}_{\text{human}}(1, 1) = 0 \]

However, deep learning may provide

\[ f(1) = 1 \neq 8 = f(1) \]

\[ f(x) = \sigma \left( W^L \otimes (\sigma \circ P \circ \sigma (W^{L-1} \otimes (\cdots \sigma \circ P \circ \sigma (W^1 \otimes x + b_1) \cdots ) + b_{L-1}) + b_L \right) \]
Adversarial attacks against MNIST handwritten classification

We will focus on this 13th signal and analysis what it means.

Library of 16 features for 6

14th Mnist Data  Confuses between 6 and 0  Classified as 6  Classified as 6  Classified as 6  Classified as 1
These adversarial examples show that a well-trained function $f: x \rightarrow y$ works only in the immediate vicinity of a manifold, whereas producing incorrect results if the input deviates even slightly from the training data manifold.

In practice, the measured data is exposed to various noise sources such as machine dependent noise; therefore, the developed algorithm must be stable against the perturbations due to noise sources.

Hence, normalization of the input data is essential for improving robustness and generalizability of the deep learning network against adversarial attacks.

$dist_{\text{radiologist}}(x_1, x_2) = 0 \ & \ dist_{\text{radiologist}}(x_3, x_4) = 0$
Final Remark: Historically, our mathematicians have tried to find well-posed model by imposing appropriate constraints to solution spaces. In the simple Dirichlet problem, it took decades to find the appropriate space $W^{1,2}(\Omega)$. It can take decades to solve the challenging problems in DL.

In terms of M-RPI, note that $|| Au - Au'||_{H^1(\partial\Omega)} \approx || u - u'||_{H^1(\Omega)}$ for all $u, u' \in M = \{ u \in W^{1,2}(\Omega) : \nabla \cdot \nabla u = 0 \text{ in } \Omega \}$. 

The Dirichlet problem may not be well-posed without the constraint $W^{1,2}(\Omega)$.

- Without the constraint $M = W^{1,2}(\Omega)$, $Au = 0$ has infinitely many solutions in $C^\infty(\Omega)$: $u(r, \theta) = (r^{\frac{2n}{3}} - r^{-\frac{2n}{3}}) \sin \frac{2n}{3} \theta, \quad n = 0, 1, 2, \cdots$
- With the constraint $M = W^{1,2}(\Omega)$, $Au = 0$ has the unique solution $u = 0$. 

\[ \Omega = \{(r, \theta) : 0 < r < 1, 0 < \theta < \frac{3}{2} \pi\} \]
I hope that we will discuss various challenging issues during this meeting. Thank you!