Essay

Henrique Medeiros Dos Reis

January 22, 2024

Introduction:Navigating the Challenge

Addressing the inverse problem in medical imaging is a multifaceted challenge, and an initial exploration involves framing it as a least squares problem. The quest for reconstructing an image $\mathbf{x}$ from linear noisy measurements $\mathbf{y}$ encounters a hurdle in the form of the forward model matrix A having a nontrivial null space, particularly in scenarios where $m < n$. This intricacy gives rise to the existence of infinitely many solutions, prompting the need for more sophisticated approaches in the quest for accurate image reconstruction.

Computed Tomography (CT) scanning is a medical imaging technique that plays a crucial role in non-invasive diagnostics. It involves the use of X-ray technology to create detailed cross-sectional images of the inside of the body. During a CT scan, a series of X-ray beams are directed through the body at different angles, and detectors measure the amount of radiation absorbed. A computer processes this data to construct detailed, three-dimensional images of the internal structures, offering a comprehensive view of organs, tissues, and bones. CT scans are particularly valuable for identifying abnormalities, tumors, fractures, and internal injuries. The importance of CT scans lies in their ability to provide precise and detailed information about the anatomy, aiding clinicians in accurate diagnosis and treatment planning.

However, despite the diagnostic power of CT scans, challenges persist in the imaging process. One significant challenge is the trade-off between image quality and radiation exposure. While CT scans provide detailed anatomical information, repeated exposure to ionizing radiation can pose risks to patients. Striking a balance between obtaining high-quality images and minimizing radiation dosage remains an ongoing concern in medical imaging. Additionally, artifacts and noise can impact the fidelity of CT images, complicating the task of extracting accurate information. Furthermore, the inherently sparse nature of certain objects or structures within the scanned volume can pose challenges for reconstruction algorithms. In this study, the focus on sparse image reconstruction addresses these challenges by enhancing the efficiency and accuracy of image reconstruction, offering potential solutions to improve the overall utility of CT scans while mitigating associated risks and limitations.

In the context of this paper, all simulations and results will be based on a dataset that includes a CT scan of a walnut. This specific choice allows for a controlled and well-defined imaging scenario, enabling a focused exploration of sparse image reconstruction techniques. By leveraging the inherent characteristics of a CT scan, such as varying densities and structural details within the walnut, the research aims to showcase the efficacy of optimization algorithms in reconstructing sparse images. This approach not only provides a practical and tangible context for the study but also aligns with the broader goal of advancing imaging methodologies for various applications.

Linear Inverse Problems in Medical Imaging

Many problems in medical imaging can be formulated as linear inverse problems, characterized by the equation:

$$\mathbf{y} = A\mathbf{x} + \mathbf{n}$$

Here, $\mathbf{y} \in \mathbb{R}^m$ represents the measurements, $A \in \mathbb{R}^{m \times n}$ is the forward model, $\mathbf{x} \in \mathbb{R}^n$ is the vectorized image, and $\mathbf{n} \in \mathbb{R}^m$ is the noise vector.

The primary objective in these problems is to reconstruct the image $\mathbf{x}$ given the linear noisy measurements $\mathbf{y}$ and some prior knowledge of the matrix A.

This naturally raises the question: How can we effectively approach this task? A logical starting point is a naive approach, wherein the problem is transformed into a least squares formulation:

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|^2$$

This least squares problem has a well-known solution $\mathbf{x}_{ls} = (A^\top A)^{-1}A^\top \mathbf{y}$. However, in the context of image problems, where $A \in \mathbb{R}^{m\times n}$ and $m < n$, the matrix A typically possesses a nontrivial null space. Consequently, this introduces a challenge, as it leads to infinitely many solutions.

Addressing the Issue: Introducing Regularization

To mitigate the challenges arising from the nontrivial null space in image reconstruction, one common strategy is to introduce regularization. A commonly used approach is to minimize the following objective function:

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|^2 + \lambda\|\mathbf{x}\|^2$$

The regularization term $\lambda\|\mathbf{x}\|^2$ encourages a unique solution in linear inverse problems, but it may not always yield a useful result. It tends to over-smooth the solution, potentially losing important details in the image. Additionally, it treats all components equally and might not adequately address the presence of a nontrivial null space in the matrix A. The parameter λ introduces a trade-off that is sensitive to the specific characteristics of the data, making it challenging to choose an appropriate value. Consequently, while this regularization term provides uniqueness, its limitations prompt exploration of alternative approaches tailored to the nuances of the imaging problem.

Seeking a New Approach

In the 1990s, alternative methods gained popularity, particularly those promoting sparsity as an alternative to traditional regularization techniques. One notable approach is sparse linear regression, formulated as:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|A\mathbf{x}-\mathbf{y}\|^2 + \lambda \mathrm{nnz}(\mathbf{x})$$

Here, $\mathrm{nnz}(\mathbf{x})$ represents a function outputting the number of non-zero entries in the vector $\mathbf{x}$. The underlying idea is to impose a "penalty" for a high number of nonzero entries.

However, this strategy introduces a new challenge—the objective function is computationally demanding, constituting an NP-hard problem. As a result, optimizing this function becomes a non-trivial task in terms of computational complexity.

To address the optimization challenges posed by the non-smooth function $nnz(\mathbf{x})$, a "convex relaxation" is necessary. Convexity refers to the property that a function lies below its chords, ensuring a unique global minimum.

A convex relaxation, denoted as $u(\mathbf{x})$, is introduced, satisfying the conditions of convexity ($u(\mathbf{x})$ is convex) and $u(\mathbf{x}) \leq nnz(\mathbf{x})$ for all $\mathbf{x}$.

The convex envelope $\tilde{f}$ is then defined as the "highest" convex relaxation of $nnz(\mathbf{x})$, ensuring $nnz(\mathbf{x}) \geq \tilde{f}(\mathbf{x}) \geq u(\mathbf{x})$.

A natural choice for the convex envelope is the L-1 norm, $\|\mathbf{x}\|_1$, given by:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i| = \sum_{i:x_i\neq 0} |x_i| \leq \sum_{i:x_i\neq 0} 1 = nnz(\mathbf{x})$$

This leads to a modified optimization problem:

$$\min_{\mathbf{x}} \|A\mathbf{x}-\mathbf{y}\|^2 + \lambda\|\mathbf{x}\|_1$$

Although this problem is convex, it remains non-smooth. To navigate this, Proximal Gradient Descent is employed as an effective optimization technique.

Proximal Gradient Descent

The Proximal Gradient Descent algorithm aims to minimize the sum of two functions $f(\mathbf{x})$ and $g(\mathbf{x})$. This is achieved through the following iterative procedure:

Algorithm 1 Proximal Gradient Descent

for $k \leftarrow 0, 1, 2, ...$ **do**

  $\mathbf{z}_{k+1} \leftarrow \mathbf{x}_k - \tau\nabla f(\mathbf{x}_k)$ ▷ Gradient Step with respect to f

  $\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} g(\mathbf{x}) + \frac{1}{2\tau}\|\mathbf{x} - \mathbf{z}_{k+1}\|^2$ ▷ Denoted as $prox_g(\mathbf{z}_k, \tau)$

end for

Although in each step, we optimize a different function, the optimization tasks are generally computationally manageable, making the algorithm practical.

For the specific problem at hand, where $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$, the minimization task reduces to finding $\arg\min_{\mathbf{x}} \sum_{i=1}^{n} \left(|x_i| + \frac{\tau}{2}|x_i - z_{ki}|^2\right)$. The solution for each entry is denoted as $\sigma(z_{ki})$ and can be expressed as:

$$\sigma(z_{ki}) = \begin{cases} s + \mu, & \text{if } s < -\mu \\ 0, & \text{if } |s| \leq \mu \\ s - \mu, & \text{if } s > \mu \end{cases}$$

Results: Proximal Gradient Descent

Figure 1: Proximal Gradient Descent, 10 iterations, run-time: 6.9698s

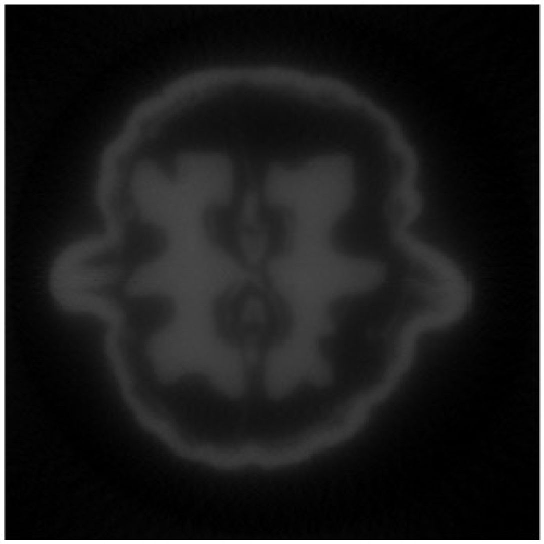


Figure 2: Proximal Gradient Descent, 100 iterations, run-time: 11.5669s

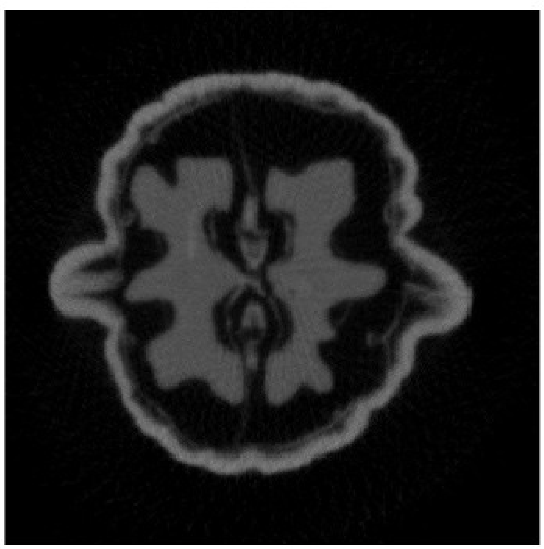


Figure 3: COST

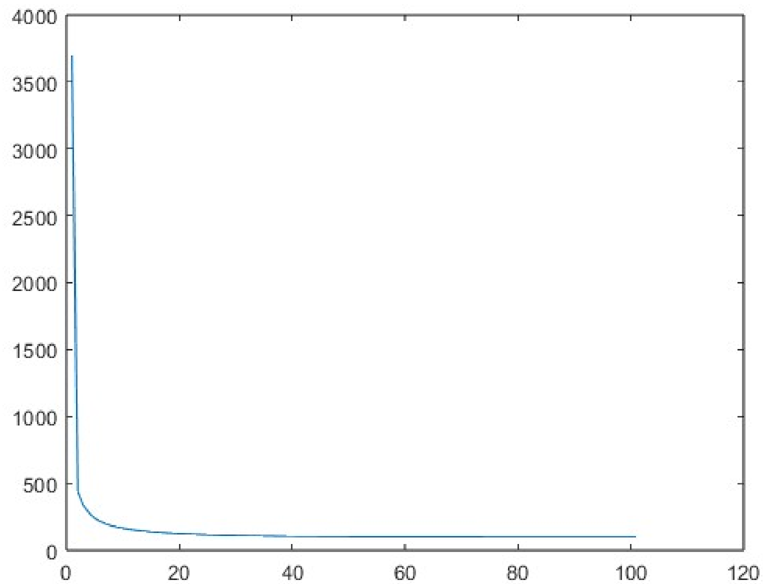
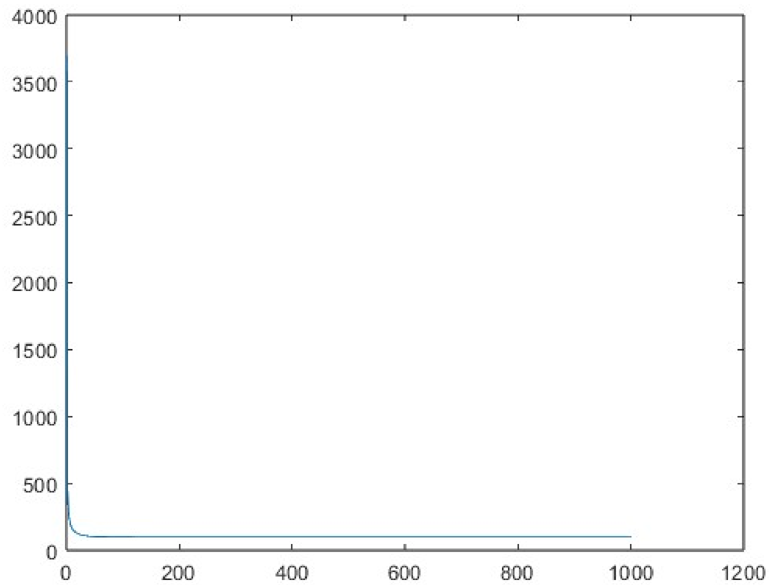


Figure 4: Proximal Gradient Descent, 1000 iterations, run-time: 48.1285s

Figure 5: COST



While proximal gradient methods provide an effective optimization framework, challenges arise in reconstruction of real-world images. Our experiments applying CP to recover a sample image of a walnut indicate slow convergence even after 1000 iterations.

The image remains blurry and artifacts persist throughout the whole image, suggesting the algorithm struggles to resolve finer-scale details. Tracing the evolution of image over iterations reveals steady but slow improvement, not approaching convergence after 1000 iterations, and yet still far from the true image.

Clearly, proximal gradient descent alone lacks efficiency for high-fidelity recovery of complex images. More informed regularization strategies and faster optimization schemes are needed. Possible directions include combining proximal gradients with acceleration techniques like momentum, exploring learned regularizers to improve generalization, and applying proximal methods in tandem with learning-based approaches. Analyzing the limitations of current proximal algorithms provides insights to guide development of enhanced optimization frameworks for image reconstruction.

FISTA: Fast Iterative Shrinkage-Thresholding Algorithm

Proximal gradient descent can be slow when dealing with matrices A with large condition numbers, a common scenario in imaging applications such as deblurring. FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) introduces a "momentum" update to accelerate the convergence of the algorithm.

Algorithm 2 FISTA

$L \leftarrow L(f)$ ▷ A Lipschitz constant of ∇f
$\mathbf{y}_1 \leftarrow \mathbf{x}_0 \in \mathbb{R}^n$, $t_1 \leftarrow 1$
for $k \leftarrow 1, 2, 3, ...$ **do**
  $\mathbf{z}_k \leftarrow \mathbf{y}_k - \tau \nabla f(\mathbf{y}_k)$
  $\mathbf{x}_k \leftarrow \max(|\mathbf{z}_k| - \lambda\tau, 0) \cdot \mathrm{sign}(\mathbf{z}_k)$ ▷ applied entry-wise
  $t_{k+1} \leftarrow \frac{1+\sqrt{1+4t_k^2}}{2}$
  $\mathbf{y}_{k+1} \leftarrow \mathbf{x}_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\mathbf{x}_k - \mathbf{x}_{k-1})$
end for

Here, L represents the Lipschitz constant of ∇f. The algorithm iteratively updates $\mathbf{z}_k$ using the proximal gradient step, applies the entry-wise soft thresholding to obtain $\mathbf{x}_k$, and incorporates a momentum term in $\mathbf{y}_{k+1}$ to accelerate convergence. The variable t_k tracks the momentum term, enhancing the efficiency of the FISTA algorithm.

Results: Proximal Gradient Descent

Figure 6: Proximal Gradient Descent, 10 iterations, run-time: 6.9698s

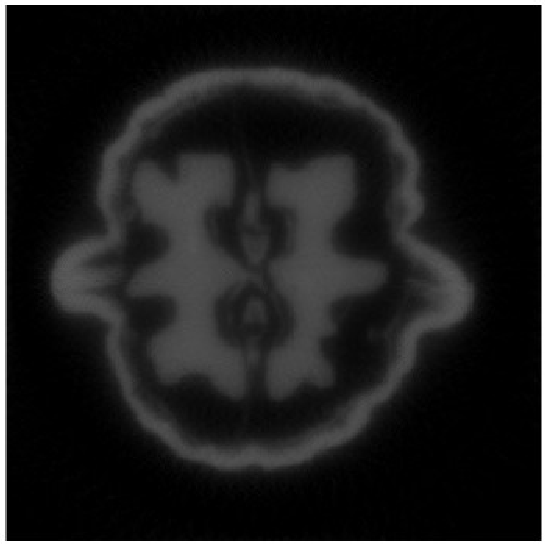


Figure 7: Proximal Gradient Descent, 100 iterations, run-time: 11.5669s

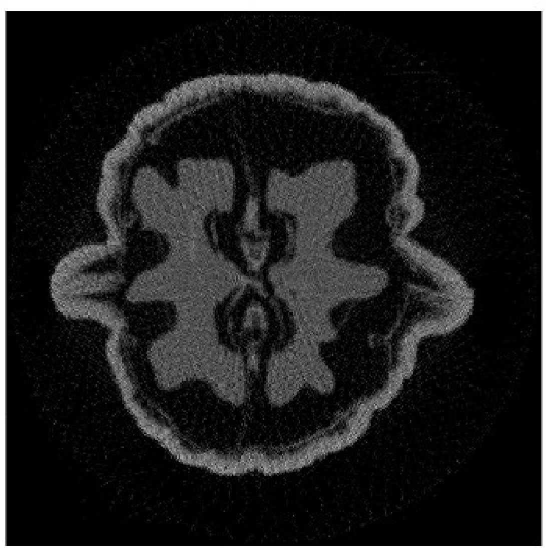
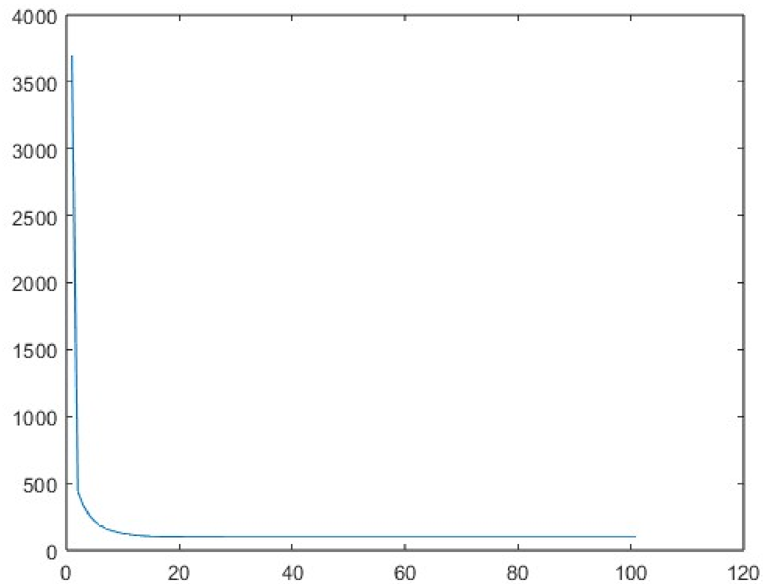


Figure 8: COST



While proximal gradient methods provide an effective optimization framework, challenges arise in reconstruction of real-world images. Our experiments applying CP to recover a sample image of a walnut indicate slow convergence even after 1000 iterations.

The image remains blurry and artifacts persist throughout the whole image, suggesting the algorithm struggles to resolve finer-scale details. Tracing the evolution of image over iterations reveals steady but slow improvement, not approaching convergence after 1000 iterations, and yet still far from the true image.

Clearly, proximal gradient descent alone lacks efficiency for high-fidelity recovery of complex images. More informed regularization strategies and faster optimization schemes are needed. Possible directions include combining proximal gradients with acceleration techniques like momentum, exploring learned regularizers to improve generalization, and applying proximal methods in tandem with learning-based approaches. Analyzing the limitations of current proximal algorithms provides insights to guide development of enhanced optimization frameworks for image reconstruction.

Sparse Image Reconstruction with Linear Combination

When promoting sparsity in image reconstruction, a linear combination of $\mathbf{x}$ is often needed. This leads to the following model and algorithm:

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|^2 + \lambda\|W\mathbf{x}\|_1$$

Algorithm 3 Proximal Gradient Descent with Linear Combination

---

for $k \leftarrow 0, 1, 2, ...$ **do**

$\quad \mathbf{z}_{k+1} \leftarrow \mathbf{x}_k - \tau\nabla f(\mathbf{x}_k)$ $\qquad \triangleright$ Gradient Step with respect to f

$\quad \mathbf{x}_{k+1} \leftarrow \arg\min_{\mathbf{x}} \lambda\|W\mathbf{x}\|_1 + \frac{1}{2\tau}\|\mathbf{x} - \mathbf{z}_{k+1}\|^2$ $\qquad \triangleright$ Denoted as $prox_g(\mathbf{z}_k, \tau)$

end for

It is also possible to apply the FISTA algorithm to this formulation.

To address the optimization challenge in the second step, where $\|W\mathbf{x}\|_1$ is not straightforward to optimize, a change of variables is introduced. Assuming W is invertible, let $\mathbf{c} = W\mathbf{x} \Leftrightarrow \mathbf{x} = W^{-1}\mathbf{c}$. The cost function becomes:

$$\min_{\mathbf{c}} \frac{1}{2}\|AW^{-1}\mathbf{c} - \mathbf{y}\|^2 + \lambda\|\mathbf{c}\|_1$$

This is then solved using Proximal Gradient Descent or FISTA, replacing A with $\tilde{A} = AW^{-1}$ and $\mathbf{x}$ with $\mathbf{c}$.

The choice of the matrix W is problem-dependent. In this context, the Haar Wavelet Transform is proposed, providing an orthogonal matrix W that facilitates inversion and performs signal compression.

Results: Proximal Gradient Descent

Figure 9: Proximal Gradient Descent, 10 iterations, run-time: 6.9698s

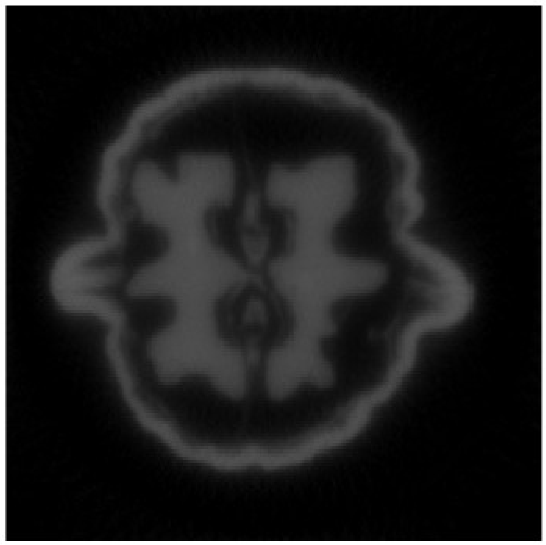


Figure 10: Proximal Gradient Descent, 100 iterations, run-time: 11.5669s

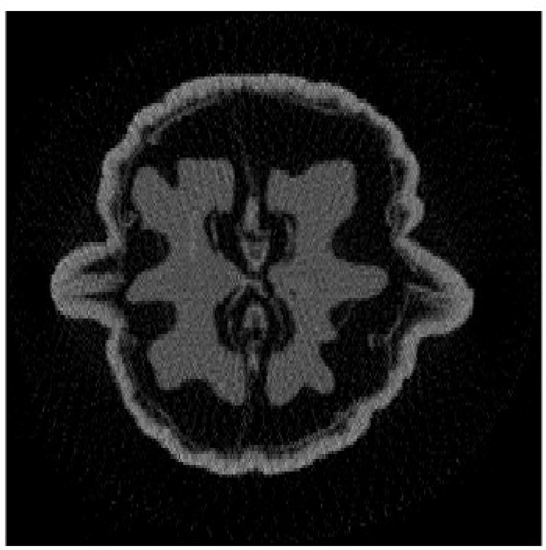
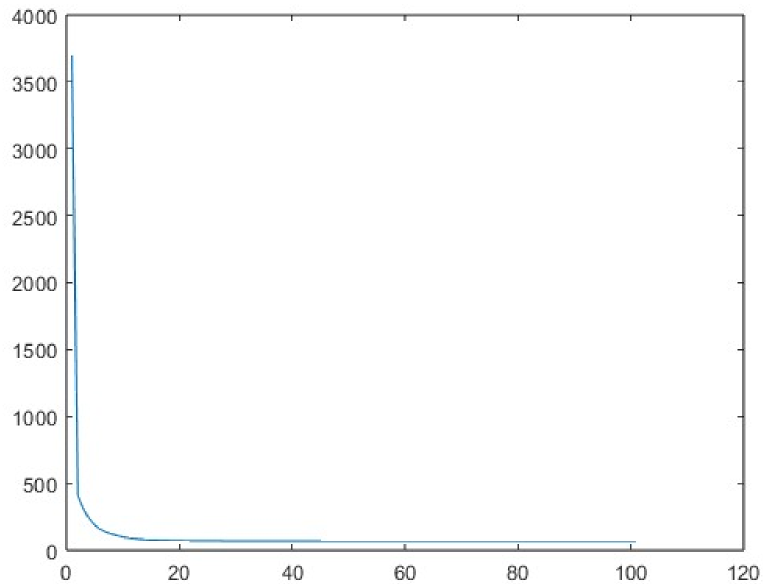


Figure 11: COST



While proximal gradient methods provide an effective optimization framework, challenges arise in reconstruction of real-world images. Our experiments applying CP to recover a sample image of a walnut indicate slow convergence even after 1000 iterations.

The image remains blurry and artifacts persist throughout the whole image, suggesting the algorithm struggles to resolve finer-scale details. Tracing the evolution of image over iterations reveals steady but slow improvement, not approaching convergence after 1000 iterations, and yet still far from the true image.

Clearly, proximal gradient descent alone lacks efficiency for high-fidelity recovery of complex images. More informed regularization strategies and faster optimization schemes are needed. Possible directions include combining proximal gradients with acceleration techniques like momentum, exploring learned regularizers to improve generalization, and applying proximal methods in tandem with learning-based approaches. Analyzing the limitations of current proximal algorithms provides insights to guide development of enhanced optimization frameworks for image reconstruction.

Sparse Image Reconstruction with Limited Differences

Enhancing the sparsity-promoting model, a modified matrix K is introduced, focusing solely on the vertical and horizontal differences. The adjusted cost function becomes:

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{y}\|^2 + \lambda\|K\mathbf{x}\|_1$$

The objective of matrix K is to encourage images that are not only sparse but also possess sparse derivatives, emphasizing edges.

However, solving this modified cost function poses a challenge since the matrix K is neither invertible nor has a pseudo-inverse. This renders conventional approaches presented earlier ineffective for this specific problem.

Enhancing Sparse Image Reconstruction:

A Saddle Point Approach

To further improve sparse image reconstruction, a novel approach leveraging concave-convex saddle point optimization is introduced. The modified cost function aims to balance sparsity and encourage images with sparse derivatives, particularly emphasizing edges:

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{m}\|^2 + \lambda\left(\max_{\|\mathbf{y}\|_\infty \leq 1} < K\mathbf{x}, \mathbf{y} >\right)$$

To overcome the challenge posed by the limited differences matrix K, the problem is reformulated as a concave-convex saddle point problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \max_{\mathbf{y}\in\mathcal{Y}} \frac{1}{2\lambda}\|A\mathbf{x}-\mathbf{m}\|_2^2 + \lambda < K\mathbf{x}, \mathbf{y} > -F^*(\mathbf{y})$$

where $F^*(\mathbf{y})$ imposes a constraint ensuring $\|\mathbf{y}\|_\infty \leq 1$. The Chambolle-Pock algorithm, tailored for finding the saddle point of two combined functions, is employed:

Algorithm 4 Chambolle-Pock

Choose $\sigma, \tau > 0$ such that $\sigma\tau\|K\|_2^2 \leq 1$
for $k \leftarrow 1, 2, ...$ **do**
  $\mathbf{y}_{k+1} \leftarrow \text{clip}(\mathbf{y}_k + \sigma K\bar{\mathbf{x}}_k, 1)$ ▷ Projection onto $\mathcal{Y}$
  $\mathbf{x}_k \leftarrow \left(I + \frac{\tau}{\lambda}A^\top A\right)^{-1}\left(\mathbf{x}_k + \frac{\tau}{\lambda}A^\top\mathbf{m}\right)$ ▷ Linear system solution
  $\mathbf{x}_{\bar{k}+1} \leftarrow 2\mathbf{x}_{k+1} - \mathbf{x}_k$ ▷ Update momentum term
end for

This approach promotes images characterized by rapid transitions between regions with substantial value differences. The combination of concave-convex optimization, the use of the limited differences matrix K, and the Chambolle-Pock algorithm presents a robust method for enhancing sparse image reconstruction, particularly in scenarios with edges and sparse derivatives.

Balancing Updates: Alternating Direction Method of Multipliers (ADMM)

The Chambolle-Pock (CP) algorithm, while effective for sparse image reconstruction, faces a notable challenge in the computational intensity of solving the linear system for the proximal gradient of the total variation term (G). Although this step efficiently brings the solution closer to convergence in a specific direction, the proximal gradient of the data fit term (F^*) is comparatively quicker to compute but lacks the pace to keep up with updates in the variable $\mathbf{x}$. This asymmetry results in the updates primarily moving towards one direction and lagging in the other.

To address this imbalance and enhance the overall convergence speed, the Alternating Direction Method of Multipliers (ADMM) is introduced. The core idea is to concatenate the CT matrix and the convolutions matrix into a tall matrix $B = \begin{bmatrix} A \\ K \end{bmatrix}$. The cost function is then redefined as:

$$\min_{\mathbf{x}\in\mathcal{X}} F^*(B\mathbf{x}) + G(\mathbf{x})$$

where F^* combines both the data fit and total variation terms, and $G(\mathbf{x}) = 0$. The algorithm undergoes subtle adjustments to accommodate this transformation:

Algorithm 5 ADMM

$L \leftarrow \|B\|_2^2$, $\tau \leftarrow 1/L, \sigma \leftarrow 1/L, \theta \leftarrow 1$
for $k \leftarrow 1, 2, ...$ **do**
  $\mathbf{p}_{k+1} \leftarrow (\mathbf{p}_k + \sigma(A\mathbf{x} - \mathbf{m}))/(1+\sigma)$ ▷ Update for $\mathbf{p}$
  $\mathbf{y}_{k+1} \leftarrow \text{clip}(\mathbf{y}_k, \lambda)$ ▷ Projection onto $\mathcal{Y}$
  $\mathbf{x}_k \leftarrow \mathbf{x}_k - \tau A^\top \mathbf{p}_{k+1} - \tau K^\top(\mathbf{y}_{k+1})$ ▷ Update for $\mathbf{x}$
  $\bar{\mathbf{x}_{k+1}} \leftarrow \mathbf{x}_{k+1} + \theta(\mathbf{x}_{k+1} - \bar{\mathbf{x}}_k)$ ▷ Update momentum term
end for

The ADMM algorithm ensures a more balanced and synchronized update strategy in both maximizing and minimizing directions, facilitating a more uniform convergence. Despite these adjustments, ADMM converges to the exact solution as CP, but with improved efficiency and equilibrium in update rates.

Deep Learning:

Letting the Machine Learn the filters:

In traditional optimization approaches to image processing, we manually define the regularization term $R(\mathbf{x})$ in an objective function like:

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 + R(\mathbf{x})$$

The regularization term acts as a prior that enforces desired properties on the solution ($\mathbf{x}$, such as smoothness or sparsity. Manually designing regularization terms is effectively designing denoising filters by hand.

Recent advances in deep learning allow us to learn these filters automatically. Instead of hand-engineering a regularization term, we can train a neural network as a denoising operator. The network learns an implicit image prior from training data, removing the need to explicitly define a regularization term.

These learned denoisers can act as plug-and-play priors in optimization methods. By replacing a hand-designed regularization term with a learned denoising neural network, we obtain a flexible framework for image recovery. The network is trained once on appropriate data and then plugged into various optimization algorithms as needed.

Plug-and-Play Priors: Flexibility, Limitations, and Requirement

Plug-and-play priors provide a flexible way to integrate learned denoisers into optimization methods. However, there are some limitations:

The approach is ad-hoc - there are no guarantees on convergence. Proper training data and network design are critical. The denoiser is treated as a black box. There is limited interpretability of the network's learned prior. Training data must closely match the target recovery application. Learning a suitable prior requires a large dataset of relevant images.

To successfully apply plug-and-play priors:

A large, representative training dataset is needed. Images should match the target recovery application. Careful network design and training is critical to learn a useful prior. The choice of network architecture affects results. Convergence is not guaranteed. Empirical validation on test data is needed to ensure stability. In summary, with good data and design, plug-and-play neural network priors provide flexibility. But caution is needed, as results can be unpredictable for poorly matched data or networks. Explicit integration of domain knowledge can help regularize the learned prior.

Conclusion

In the pursuit of enhancing sparse image reconstruction, we explored two powerful optimization algorithms—Chambolle-Pock (CP) and Alternating Direction Method of Multipliers (ADMM). While CP demonstrated effectiveness, it faced challenges in the time-consuming linear system solution for the proximal gradient of the total variation term. The imbalance in update speeds between the data fit and total variation terms prompted the introduction of ADMM.

ADMM addresses this imbalance by concatenating the CT matrix and the convolutions matrix into a unified tall matrix, reshaping the cost function for more synchronized updates. The algorithm achieves this by combining the advantages of CP's efficient linear system solution and the quick computation of the total variation term's proximal gradient. The resulting convergence is more uniform, moving at the same rate in both maximizing and minimizing directions.

In conclusion, ADMM presents a refined and balanced approach to sparse image reconstruction, offering improved computational efficiency without sacrificing the accuracy of the solution. The exploration of these optimization techniques contributes to the evolving landscape of image reconstruction methodologies, providing valuable insights for future research and applications.

While optimization algorithms like CP and ADMM advance iterative reconstruction, deep learning promises a paradigm shift. Neural networks learn end-to-end mappings from sparse, noisy data to high-quality images. Once trained, inference is extremely fast, requiring only a single feedforward pass through the network.

However, deep learning methods lack interpretability. Plug-and-play techniques seek to get the best of both worlds - the speed of learned models with the transparency of optimization. These methods integrate deep denoisers as priors within iterative algorithms. The denoiser learns an implicit image model from training data, acting as a regularization term.

Initial results with plug-and-play are promising. Learned regularizers improve reconstruction quality compared to hand-designed priors. While challenges remain in terms of stability and convergence guarantees, plug-and-play methods offer an intuitive bridge between deep learning and traditional optimization. Combining the strengths of both approaches is an exciting direction for future sparse image reconstruction research.

As deep learning capabilities grow alongside optimization algorithms, the fusion of these disciplines has immense potential to push image reconstruction performance even further. Both fields still have much to contribute in the pursuit of high-fidelity, efficient medical imaging.

Bibliography

https://arxiv.org/abs/2005.06001

https://www.winlab.rutgers.edu/~crose/322_html/convexity2.pdf

https://web.stanford.edu/~boyd/papers/pdf/prox_algs.pdf

https://see.stanford.edu/materials/lsoeldsee263/08-min-norm.pdf

PDFFISTA

https://www.cis.upenn.edu/~cis5150/cis515-20-sl-Haar.pdf

https://medium.com/@koushikc2000/2d-discrete-wavelet-transformation-and-its-applicati