A Proximal Algorithm for Total Variation-Based Phase Retrieval

Anqi Fu
Department of Electrical Engineering
Stanford University
350 Serra Mall, Stanford, CA 94305
Email: anqif@stanford.edu

***Abstract*—In this paper, I study the problem of phase retrieval with total variation (TV) smoothing and denoising. I construct an optimization model with an anisotropic TV penalty term, then show how to solve it using a proximal gradient method. This iterative-shrinkage-thresholding phase retrieval algorithm (ISTPRA) combines fixed-point iterations with a fast gradient solver for the convex subproblem. I compare my algorithm's performance to several alternating projection methods. My results show that ISTPRA achieves a lower reconstruction error, while remaining robust to noise.**

I. Introduction

In many science and engineering applications, we can only measure the magnitude of a signal's Fourier transform and not its phase. For example, in optical settings, devices like CCD cameras capture just the photon flux, so information about the image is lost. This information is often more important than the Fourier magnitude. As an illustration, in Fig. 1, I swap the Fourier phases of two images and display the results. The output images bear a clear resemblance to the input with the corresponding phase, rather than magnitude. Thus, ignoring the phase and simply performing an inverse Fourier transform will lead to a poor reconstruction of the signal.

Our challenge is to accurately recover a signal from measurements of its Fourier magnitude. This problem, known as phase retrieval, arises in a wide range of fields including optics, X-ray crystallography, electron microscopy, and astronomy. The classical approach is to apply an alternating projection algorithm, such as hybrid input-output (HIO), that iteratively imposes time and frequency domain constraints [1]. More recent work has focused on constructing a formal model that incorporates prior information on the signal [2]. One popular model is PhaseLift, which casts phase retrieval as a semidefinite program (SDP) solvable by standard convex optimization techniques [3]. Other research has explored the effects of imposing sparsity constraints [4] [5].

In this paper, I develop an optimization model for phase retrieval with total variation regularization. Total variation has been shown to be an excellent penalty function for denoising signals [6]. Furthermore, it is convex and allows us to employ proximal gradient methods to solve the optimization problem. I derive a proximal algorithm based on the monotone fast iterative-shrinkage-thresholding algorithm (MFISTA) by [14]. With an appropriate initialization, my algorithm converges within 20 iterations and recovers the signal with 14% greater accuracy than classical projection methods. These results are robust to noise.

This paper is organized as follows. In section II, I consider phase retrieval as a feasibility problem. I examine three projection algorithms, comparing their performance and discussing their limitations. This leads me to section III, where I reformulate phase retrieval as an optimization problem with total variation regularization. I show that this problem can be solved using the proximal gradient method and propose a new algorithm, ISTPRA. In section IV, I test my algorithm on a set of images, analyzing its accuracy and performance with noise. Finally, section V concludes.

II. Feasibility Problem

A. Problem Formulation

Let $x \in \mathbb{R}^{m \times n}$ denote the image to be recovered and $b = |\mathcal{F}x| \in \mathbb{R}_+^{m \times n}$ its Fourier magnitude, where $\mathcal{F}$ is the 2D discrete Fourier transform (DFT) operator. We know x has non-zero support on a bounded set E and define the set of such images as

$$S := \{x_{i,j} | x_{i,j} = 0 \text{ for } (i,j) \notin E\} \tag{1}$$

Given an observation b, we are interested in images whose Fourier magnitudes match b, i.e.

$$B := \{x_{i,j} | |\mathcal{F}x| = b\} \tag{2}$$

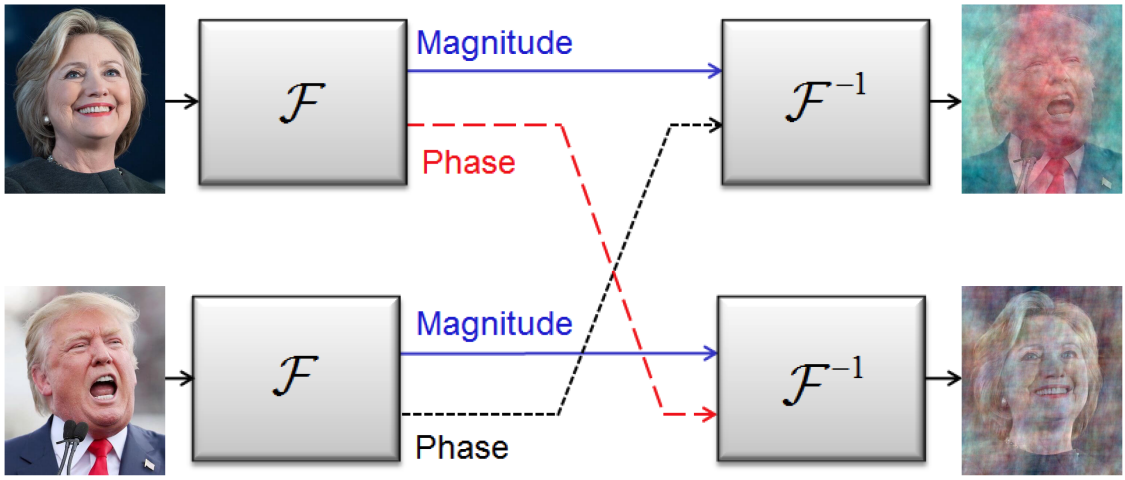


Fig. 1: An example showing the importance of the Fourier phase in reconstructing an image (adapted from [7]).

Define the projection operators

$$\mathcal{P}_S(x) = \begin{cases} x_{i,j} & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

and

$$\mathcal{P}_B(x) = \mathcal{F}^{-1}(\hat{y}), \hat{y}_{i,j} = \begin{cases} b_{i,j} \frac{(\mathcal{F}x)_{i,j}}{|(\mathcal{F}x)_{i,j}|} & \text{if } (\mathcal{F}x)_{i,j} \neq 0 \\ (\mathcal{F}x)_{i,j} & \text{otherwise} \end{cases} \tag{4}$$

The operator $\mathcal{P}_B$ constructs a modified signal $\hat{y}$ by combining the known Fourier magnitude b with the phase of $\mathcal{F}x$ at every pixel where the DFT of x is non-zero. If the DFT is zero, the phase is taken to be zero. $\mathcal{P}_B$ then takes the inverse DFT of $\hat{y}$ to obtain the projection of x into B.

With this notation, I can formulate phase retrieval as a feasibility problem

$$\text{find } x \in S \cap B \tag{5}$$

B. Projection Algorithms

Problem 5 is usually solved by a projection algorithm. Here I describe three: error reduction (ER) [8], hybrid projection reflection (HPR) [9], and relaxed averaged alternating reflection (RAAR) [10]. The algorithm proceeds iteratively. At every iteration k, an approximation of the signal x^k is generated by the update rule [1]

Algorithm	Formula
ER	$x^k = \mathcal{P}_S\mathcal{P}_B(x^{k-1})$
HPR	$x^k = ((1+\beta)\mathcal{P}_{S_+}\mathcal{P}_B + I - \mathcal{P}_{S_+} - \beta\mathcal{P}_B)(x^{k-1})$
RAAR	$x^k = (2\beta\mathcal{P}_{S_+}\mathcal{P}_B + \beta I - \beta\mathcal{P}_{S_+} + (1-2\beta)\mathcal{P}_B)(x^{k-1})$

(a) Padded Image (b) Initial Value

Fig. 2: The test image was zero-padded and x_0 initialized to the mask corresponding to the non-zero support.

where $S_+ := S \cap \{x_{i,j} | x_{i,j} \geq 0 \text{ for all } (i,j) \in E\}$[1] and $\beta > 0$ controls the step size. The updates continue until a stopping criteria is reached, typically when

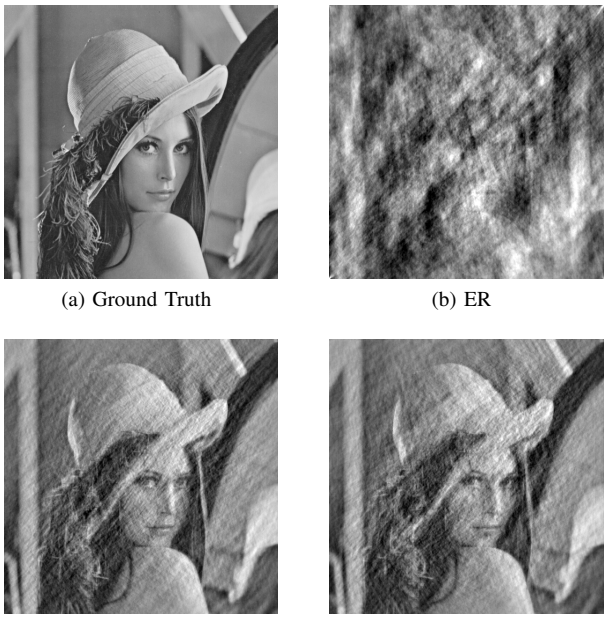
$$\Delta^k := \|x^k - x^{k-1}\|_F^2 < \epsilon \tag{6}$$

for a user-specified tolerance $\epsilon > 0$.

C. Experimental Comparison

I implement ER, HPR, and RAAR in MATLAB and compare their performance on the 512×512 Lena image. For test purposes, I zero-pad the image so the DFT oversamples by a factor of two before computing its Fourier magnitude. My non-zero support E contains all pixels outside the padding region, and I initialize each algorithm to its corresponding mask (see Fig. 2). At

[1]HPR is simply a restriction of hybrid input-output (HIO) to the set of non-negative signals with support E.



(a) Ground Truth (b) ER

(c) HPR ($\beta = 0.95$) (d) RAAR ($\beta = 0.95$)

Fig. 3: A comparison of different projection algorithms. HPR gives the most accurate reconstruction with a normalized error of 0.1258.

iteration k, I compute x^k, Δ^k, and the normalized error from the true image v,

$$\text{err}(x^k, v) := \frac{\|x^k - v\|_F}{\|v\|_F} \tag{7}$$

calculated over $(i, j) \in E$. The updates terminate when $\Delta^k < 10^{-6}$ or $k > 200$. My results are shown in Fig. 3.

I find that $\beta = 0.95$ produces the best reconstructions by HPR and RAAR, with err(x^*, v) equal to 0.1258 and 0.1372, respectively. ER performs far worse, attaining an error of 0.3474 upon termination at 44 iterations. The recovered ER image is unrecognizable as Lena: only a faint outline of the hat and shoulder survive, while all other features are lost. In contrast, HPR and RAAR recover the full human figure, including details of the face and hair as well as background shapes. However, both images exhibit a distinct grainy texture.

This graininess is the main contributor to the reconstruction error. Although Lena is recognizable, images (c) and (d) in Fig. 3 suffer from a rough, noisy pattern that obscures fine edges, such as the folds of the hat. In some applications, noise may be a serious issue, especially if we expect the image to fulfill certain smoothness constraints.

III. Total Variation-Based Phase Retrieval

A. Optimization Problem

In order to balance Fourier accuracy with other signal constraints, I reformulate phase retrieval as an optimization problem

$$\min_{x \in C} \{c(\mathcal{F}x, b) + \lambda r(x)\} \tag{8}$$

The cost function $c : \mathbb{R}_+^{m \times n} \times \mathbb{R}_+^{m \times n} \to \mathbb{R}$ penalizes deviations of the DFT of x from the known Fourier magnitude b. Typically in the literature, $c(x, y)$ is taken to be the squared distance metric $\|x - y\|^2$. The regularization function $r : \mathbb{R}^{m \times n} \to \mathbb{R}$ incorporates additional information on x, e.g. smoothness, which contributes to the total cost via the weight $\lambda \geq 0$. As λ increases, the objective places more importance on regularizing x.

For smoothing and denoising, one popular choice for $r(x)$ is the l_1-based anisotropic total variation,

$$\begin{aligned}\text{TV}(x) = &\sum_{i=1}^{m-1}\sum_{j=1}^{n-1}\{|x_{i,j} - x_{i+1,j}| + |x_{i,j} - x_{i,j+1}|\} \\ &+ \sum_{i=1}^{m-1}|x_{i,n} - x_{i+1,n}| + \sum_{j=1}^{n-1}|x_{m,j} - x_{m,j+1}|\end{aligned}$$

which we can write compactly as $\text{TV}(x) = \|\nabla x\|_1$, where ∇ is the 2D first difference operator. The total variation captures the sum of the absolute difference between neighboring row/column values. As a regularizer, it penalizes large changes in neighboring pixel intensities, making it useful for removing noise. Furthermore, unlike techniques such as median filtering, TV regularization preserves edges so that only flat regions of an image are smoothed [11] [12].

The phase retrieval problem becomes

$$\min_{x \in C} \left\{\||\mathcal{F}x| - b\|_F^2 + 2\lambda \text{TV}(x)\right\} \tag{9}$$

Here I assume the signal lies in a known convex set C. For example, if x is an image with pixel values in the range $[l, u]$, then C is the box

$$B_{l,u} = \{x : l \leq x_{i,j} \leq u \text{ for all } i, j\}$$

The unconstrained case corresponds to $C = \mathbb{R}^{m \times n}$. A number of algorithms exist to solve problems of the form 9, which derive from the proximal gradient method.

B. Proximal Gradient Method

Consider problems of the form

$$\min_x \{F(x) \equiv f(x) + g(x)\} \tag{10}$$

where $g : \mathbb{R}^{m\times n} \to (-\infty, \infty]$ is a proper closed, convex function and $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ is continuously differentiable. The function g has a proximal map

$$\text{prox}_{tg}(v) := \underset{x}{\text{argmin}} \left\{ \frac{1}{2t}\|x - v\|^2 + g(x) \right\}$$

and a common approach to solving 10 is via the proximal gradient method, which iteratively updates

$$x_{k+1} := \text{prox}_{t_k g}(x_k - t_k \nabla f(x_k))$$

See Parikh and Boyd [13] for a thorough treatment.

Various extensions of this method exist in the literature; for the purposes of this paper, I focus on the monotone fast iterative-shrinkage-thresholding algorithm (MFISTA) presented in [14]. This algorithm has nice convergence properties and is shown to work well for convex TV-based denoising and deblurring. Although my cost function is non-convex, we will see that MFISTA produces a fairly accurate result given the appropriate initialization.

For ease of notation, define the proximal map

$$\begin{aligned} p_L(y) &:= \text{prox}_{\frac{1}{L}g}\left(y - \frac{1}{L}\nabla f(y) \right) \\ &= \underset{x}{\text{argmin}} \left\{ \frac{L}{2} \left\| x - \left(y - \frac{1}{L}\nabla f(y) \right) \right\|^2 + g(x) \right\} \end{aligned} \tag{11}$$

MFISTA(x_0, L, K)

Input: x_0 initial value, $L > 0$ step size, K iterations.
Initialize: $y_1 = x_0, t_1 = 1$.
Step k: $(1 \le k \le K)$

$$\begin{aligned} z_k &= p_L(y_k) \\ x_k &= \text{argmin}\{F(x) : x = z_k, x_{k-1}\} \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2} \\ y_{k+1} &= x_k + \frac{t_k}{t_{k+1}}(z_k - x_k) + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}) \end{aligned}$$

Output: $x^* = x_K$.

C. Phase Retrieval Algorithm

Problem 9 is a specific case of 10 with

$$\begin{aligned} f(x) &= \||\mathcal{F}x| - b\|_F^2 \\ g(x) &= 2\lambda \text{TV}(x) + I_C(x) \end{aligned}$$

where

$$I_C(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C \end{cases}$$

is the indicator function of inclusion in C. In order to use MFISTA, we need the following result.

Lemma 3.1: The gradient of $f(x) \equiv \||\mathcal{F}x| - b\|_F^2$ is

$$\nabla f(x) = 2\left(x - \mathcal{F}^{-1}\left\{ b\frac{\mathcal{F}x}{|\mathcal{F}x|} \right\} \right) = 2(I - \mathcal{P}_B)(x) \tag{12}$$

Proof See section 4.3 of [15].

Plugging into 11 and factoring out $\frac{L}{2}$ from the objective, I find the proximal map is

$$\underset{x\in C}{\text{argmin}} \left\{ \left\| x - \left(y - \frac{2}{L}(I - \mathcal{P}_B)(y) \right) \right\|_F^2 + 2\frac{2\lambda}{L}\text{TV}(x) \right\}$$

To compute this mapping, I first consider the simpler TV-based denoising problem

$$\min_{x\in C} \left\{ \|x - d\|_F^2 + 2\alpha \text{TV}(x) \right\} \tag{13}$$

This is a convex optimization problem, which can be solved by deriving its dual and applying a gradient-based method. Beck and Teboulle provide a full proof in section IV of [14]. Here I merely reproduce their solution. Define the following notation:

- $\mathcal{P}_P(p, q) = (r, s)$ where $r \in \mathbb{R}^{(m-1)\times n}$ and $s \in \mathbb{R}^{m\times(n-1)}$ is the projection
$$r_{i,j} = \frac{p_{i,j}}{\max(1, |p_{i,j}|)}, s_{i,j} = \frac{q_{i,j}}{\max(1, |q_{i,j}|)}$$
- $\mathcal{P}_C$ is the orthogonal projection operator onto C.
- $\mathcal{L} : \mathbb{R}^{(m-1)\times n} \times \mathbb{R}^{m\times(n-1)} \to \mathbb{R}^{m\times n}$ is the operator
$$\mathcal{L}(p, q)_{i,j} = p_{i,j} + q_{i,j} - p_{i-1,j} - q_{i,j-1}$$
where we assume $p_{0,j} = p_{m,j} = q_{i,0} = q_{i,n} = 0$ for all $i = 1, \ldots, m, j = 1, \ldots, n$.
- $\mathcal{L}^* : \mathbb{R}^{m\times n} \to \mathbb{R}^{(m-1)\times n} \times \mathbb{R}^{m\times(n-1)}$ is the adjoint of $\mathcal{L}$ given by $\mathcal{L}^*(x) = (p, q)$, where $p \in \mathbb{R}^{(m-1)\times n}$ and $q \in \mathbb{R}^{m\times(n-1)}$ are
$$\begin{aligned} p_{i,j} &= x_{i,j} - x_{i+1,j}, i = 1, \ldots, m-1, j = 1, \ldots, n \\ q_{i,j} &= x_{i,j} - x_{i,j+1}, i = 1, \ldots, m, j = 1, \ldots, n-1 \end{aligned}$$

The dual of the denoising problem is

$$\min_{(p,q)\in P} \{ -\|H_C(d - \alpha\mathcal{L}(p, q))\|_F^2 + \|d - \alpha\mathcal{L}(p, q)\|_F^2 \} \tag{14}$$

where $H_C(x) := x - \mathcal{P}_C(x)$. We can recover the solution of the primal problem from the dual variables with

$$x = P_C(d - \alpha\mathcal{L}(p, q)) \tag{15}$$

This observation gives us the fast gradient projection (FGP) algorithm.

FGP(d, β, N, η)

Input: d observed image, α regularization weight, N iterations, $\eta > 0$ tolerance.
Initialize: $(r_1, s_1) = (p_0, q_0) = (0, 0), t_1 = 1$.
Step k: $(1 \leq k \leq N)$

$$y_k = \mathcal{P}_C(d - \alpha\mathcal{L}(r_k, s_k))$$
$$(p_k, q_k) = \mathcal{P}_P\left((r_k, s_k) + \frac{1}{8\alpha}\mathcal{L}^*(y_k)\right)$$
$$x_k = \mathcal{P}_C(d - \alpha\mathcal{L}(p_k, q_k))$$
$$\epsilon_k = \frac{\|x_k - x_{k-1}\|_F}{\|x_k\|_F}$$

If $\epsilon_k < \eta$, terminate.

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$
$$r_{k+1} = p_k + \frac{t_k - 1}{t_{k+1}}(p_k - p_{k-1})$$
$$q_{k+1} = q_k + \frac{t_k - 1}{t_{k+1}}(q_k - q_{k-1})$$

Output: $x^* = \mathcal{P}_C(d - \alpha\mathcal{L}(p_N, q_N))$.

If we define the solution to 13 as $x^* := D_C(d, \alpha)$, then the proximal map of 9 is

$$p_L(y) = D_C\left(y - \frac{2}{L}(I - \mathcal{P}_B)(y), \frac{2\lambda}{L}\right) \tag{16}$$

Thus, to apply MFISTA to our phase retrieval problem, we must solve a TV-based denoising subproblem at every iteration. The resulting iterative-shrinkage-thresholding phase retrieval algorithm (ISTPRA) is

ISTPRA$(x_0, L, M, \lambda, \eta)$

Input: x_0 initial value, $L > 0$ step size, M iterations, $\lambda > 0$ regularization, $\eta > 0$ tolerance.
Initialize: $y_1 = x_0, t_1 = 1$.
Step k: $(1 \leq k \leq M)$

$$z_k = \textbf{FGP}\left(y_k - \frac{2}{L}(I - \mathcal{P}_B)(y_k), \frac{2\lambda}{L}, M, \eta\right)$$
$$x_k = \underset{x \in \{z_k, x_{k-1}\}}{\operatorname{argmin}} \left\{\||\mathcal{F}x| - b\|_F^2 + 2\lambda \mathrm{TV}(x)\right\}$$
$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$
$$y_{k+1} = x_k + \frac{t_k}{t_{k+1}}(z_k - x_k) + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1})$$

Output: $x^* = x_K$.

Since our objective is non-convex, a warm start is crucial to reconstruction accuracy. In the next section, I test ISTPRA with an initial x_0 determined by the projection algorithms discussed in II-B.

IV. Results

A. Regularization Weights

I implement my algorithm in MATLAB and analyze its performance on the 512×512 Lena image, oversampling the DFT by a factor of two as before. My parameters are $L = 3, M = 20, \eta = 10^{-4}, C = B_{0,1}$, and domain E containing the set of pixels outside the zero-padded region. I take as my initial x_0 the HPR result from section II-C, as this produced the lowest error. The recovered images for a few λ are shown in Fig. 4. At $\lambda = 0.05$, the image remains grainy and textured. As I increase the regularization weight, the graininess begins to smooth out so features like Lena's shoulder resemble the true image. However, when $\lambda = 0.20$, details such as the eyes and strands of hair start to blur, indicating that my total variation penalty is too high.

To determine the optimal λ, I run ISTPRA with a range of TV weights and calculate the reconstruction error for each output. Fig. 5 shows that the minimum error is attained at $\lambda^* = 0.1$ with $\mathrm{err}(x^*, v) = 0.1075$, a decrease of about 14.55% from the HPR error of 0.1258 (Table I). The corresponding image (b) in Fig. 4 retains nearly all fine features while significantly reducing the diagonal texture. Compared to the results of alternating projection, it is a more faithful recovery of the true Lena.

(a) $\lambda = 0.05$ (b) $\lambda = 0.10$ (c) $\lambda = 0.20$

Fig. 4: ISTPRA for three λ values. As λ increases, the image is smoothed further at the expense of textured details.

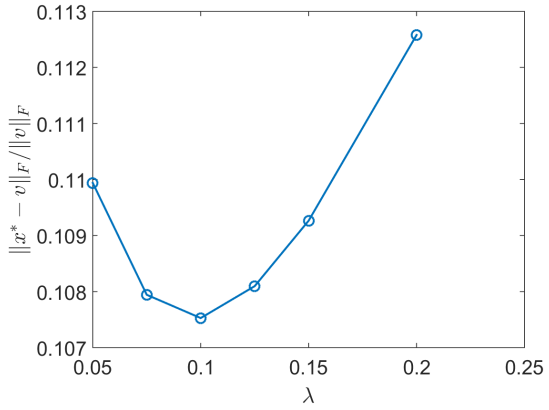


Fig. 5: Lena reconstruction error for various TV weights. The minimum error is attained at $\lambda^* = 0.1$.

B. Noisy Signal Reconstruction

Given the success of TV-based denoising models, I expect my algorithm to perform well on signals that have been corrupted by noise. To test this, I repeat the process from the previous section on the 256×256 Cameraman image, with and without additive Gaussian noise of $\mu = 0, \sigma = 0.08$. The reconstruction error is calculated with respect to the original (clean) image. I keep the same parameters as Lena, changing only $L = 4$ and the domain to the Cameraman mask. Of the three projection algorithms, HPR achieves the lowest error in both cases, so I use its output for ISTPRA.

As one would expect, noise gives rise to uniformly higher errors (Table I). HPR attains a minimum error of $\text{err}(x^*, v) = 0.2781$ with $\beta = 0.86$ on the noisy image, which exceeds the 0.2013 from the clean image. The

TABLE I: Reconstruction Error

Method	Lena	Camera (Clean)	Camera (Noisy)
ER	0.3474	0.4837	0.5670
HPR	0.1258	0.2013	0.2781
RAAR	0.1372	0.2662	0.3410
ISTPRA	0.1075	0.1688	0.2094

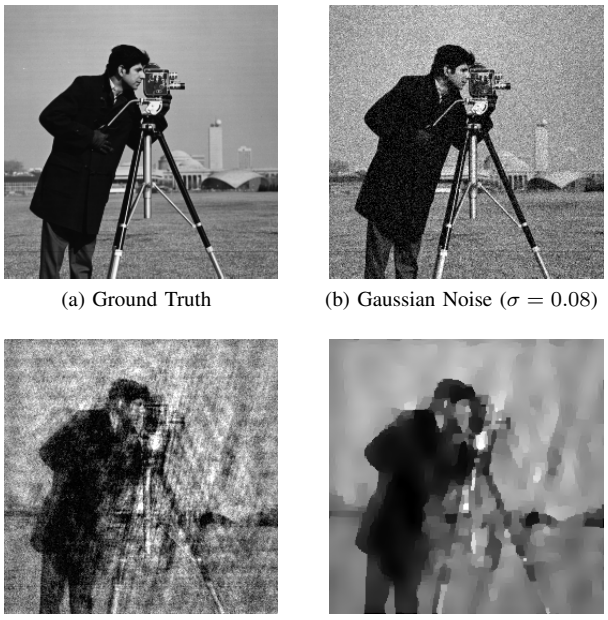
image recovered from the noisy input is shown in Fig. 6 (c). Due to the added noise, HPR produces a picture with significantly more graininess, white lines obscuring much of the background; the black coat and camera stand are both roughly textured.

After running ISTPRA with an optimal $\lambda^* = 0.20$, the final image (d) is nearly as smooth as the original (a). The coat is completely black and most of the noisy lines have been eliminated from the background. Some diagonal artifacts still remain around the center, but their edges are fainter. However, due to the large TV penalty, the fine details of the camera and the man's face are lost and the buildings in the distance are barely distinguishable. There is a clear trade-off between signal clarity and denoising.

Nevertheless, the ISTPRA error is 0.2094, a decrease of 24.7% from the HPR error. This is in contrast to a drop of 16.15% for the clean image input. Thus, we see that TV regularization improves reconstruction accuracy by a greater amount in the presence of Gaussian noise.

V. Conclusion

I introduced a new method for phase retrieval with total variation regularization. My method is based on iteratively solving the proximal map using fast gradient



(a) Ground Truth (b) Gaussian Noise ($\sigma = 0.08$)

(c) HPR ($\beta = 0.86$) (d) ISTPRA ($\lambda = 0.20$)

Fig. 6: Phase retrieval in the presence of noise. HPR is performed on the Fourier magnitude of the noisy image, and its result is used to initialize ISTPRA.

projections. I presented several examples of image recovery, which demonstrate that my algorithm is more accurate and robust than classical projection schemes like ER, HPR, and RAAR. Furthermore, my model can be readily extended to other convex loss functions.

Future work will focus on incorporating more regularizers, developing a model that handles non-Gaussian priors, and experimenting with alternate proximal algorithms such as primal-dual splitting [18].

ACKNOWLEDGMENT

The author would like to thank Felix Heide, Gordon Wetzstein, and Stephen Boyd for their research advice and guidance.

REFERENCES

[1] Z. Wen, C. Yang, X. Liu, and S. Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," *Inv. Probl.*, vol. 28, no. 11, 2012.

[2] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval: An overview of recent developments," *arXiv preprint*, Oct. 2015. [Online]. Available: https://arxiv.org/pdf/1510.07713.pdf

[3] E. J. Candes, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Comm. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.

[4] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Convex algorithms and limitations," in *IEEE 10th Int. Symp. Info. Theory*, Jul. 2013, pp. 1022–1026.

[5] Y. L. Montagner, E. D. Angelini, and J. C. Olivo-Marin, "Phase retrieval with sparsity priors and application to microscopy video reconstruction," in *IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 604–607.

[6] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.

[7] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, 2015.

[8] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, 1982.

[9] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Hybrid projection–reflection method for phase retrieval," *J. Opt. Soc. Am. A*, vol. 20, no. 6, pp. 1025–1034, 2003.

[10] D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," *Inv. Probl.*, vol. 21, no. 1, p. 37, 2004.

[11] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1, pp. 259–268, 1992.

[12] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inv. Probl.*, vol. 19, no. 6, p. S165, 2003.

[13] N. Parikh and S. Boyd, "Proximal algorithms," *Found. and Trends in Opt.*, vol. 1, no. 3, pp. 127–239, 2014.

[14] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[15] E. Osherovich, "Numerical methods for phase retrieval," Ph.D. dissertation, Technion-IIT, Haifa, Dec. 2011. [Online]. Available: https://arxiv.org/pdf/1203.4756v1.pdf

[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

[17] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1–2, pp. 89–97, 2004.

[18] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[19] W. Zuo and Z. Lin, "A generalized accelerated proximal gradient approach for total-variation-based image restoration," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2748–2759, 2011.