



Acceleration of RED via vector extrapolation [☆]

Tao Hong ^a, Yaniv Romano ^b, Michael Elad ^a

^a Department of Computer Science, Technion – Israel Institute of Technology, Haifa 32000, Israel

^b Department of Statistics, Stanford University, Stanford, CA 94305, USA



ARTICLE INFO

Article history:

Received 4 December 2018

Revised 1 April 2019

Accepted 6 July 2019

Available online 9 July 2019

Keywords:

Inverse problem

RED – REgularization by Denoising

Fixed-point

Vector extrapolation

Acceleration

ABSTRACT

Models play an important role in inverse problems, serving as the prior for representing the original signal to be recovered. REgularization by Denoising (RED) is a recently introduced general framework for constructing such priors using state-of-the-art denoising algorithms. Using RED, solving inverse problems is shown to amount to an iterated denoising process. However, as the complexity of denoising algorithms is generally high, this might lead to an overall slow algorithm. In this paper, we suggest an accelerated technique based on vector extrapolation (VE) to speed-up existing RED solvers. Numerical experiments validate the obtained gain by VE, leading to substantial savings in computations compared with the original fixed-point method.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

Inverse problems in imaging address the reconstruction of clean images from their corrupted versions. The corruption can be a blur, loss of samples, downscale or a more complicated operator (e.g., CT and MRI), accompanied by a noise contamination. Roughly speaking, inverse problems are characterized by two main parts: the first is called the forward model, which formulates the relation between the noisy measurement and the desired signal, and the second is the prior, describing the log-probability of the destination signal.

In recent years, we have witnessed a massive advancement in a basic inverse problem referred to as image denoising [1–6]. Indeed, recent work goes as far as speculating that the performance obtained by leading image denoising algorithms is getting very close to the possible ceiling [7–9]. This motivated researchers to seek ways to exploit this progress in order to address general inverse problems. Successful attempts, as in [10–12], suggested an exhaustive manual adaptation of existing denoising algorithms, or the priors used in them, treating specific alternative missions. This line of work has a clear limitation, as it does not offer a flexible and general scheme for incorporating various image denoising

achievements for tackling other advanced image processing tasks. This led to the following natural question: is it possible to suggest a general framework that utilizes the abundance of high-performance image denoising algorithms for addressing general inverse problems? Venkatakrishnan et al. gave a positive answer to this question, proposing a framework called Plug-and-Play Priors (P^3) method [13–15]. Formulating the inverse problem as an optimization task and handling it via the Alternating Direction Method of Multipliers (ADMM) scheme [16], P^3 shows that the whole problem is decomposed into a sequence of image denoising sub-problems, coupled with simpler computational steps. The P^3 scheme provides a constructive answer to the desire to use denoisers within inverse problems, but it suffers from several key disadvantages: P^3 does not define a clear objective function, since the regularization used is implicit; Tuning the parameters in P^3 is extremely delicate; and since P^3 is tightly coupled with the ADMM, it has no flexibility with respect to the numerical scheme.

A novel framework named REgularization by Denoising (RED) [17] proposes an appealing alternative while overcoming all these flaws. The core idea in RED is the use of the given denoiser within an expression of regularization that generalizes a Laplacian smoothness term. The work in [17] carefully shows that the gradient of this regularization is in fact the denoising residual. This, in turn, leads to several iterative algorithms, all guaranteed to converge to the global minimum of the inverse problem's penalty function, while using a denoising step in each iteration.

[☆] This paper has been recommended for acceptance by Zicheng Liu.

E-mail addresses: hongtao@cs.technion.ac.il (T. Hong), yromano@stanford.edu (Y. Romano), elad@cs.technion.ac.il (M. Elad)

The idea of using a state-of-the-art denoising algorithm for constructing an advanced prior for general inverse problems is very appealing.¹ However, a fundamental problem still exists due to the high complexity of typical denoising algorithms, which are required to be activated many times in such a recovery process. Indeed, the evidence from the numerical experiments posed in [17] clearly exposes this problem, in all the three methods proposed, namely the steepest descent, the fixed-point (FP) strategy and the ADMM scheme. Note that the FP method is a parameter free and the most efficient among the three, and yet this approach too requires the activation of the denoising algorithms dozens of times for a completion of the recovery algorithm.

Our main contribution in this paper is to address these difficulties by applying vector extrapolation (VE) [18–20] to accelerate the FP algorithm shown in [17]. Our simulations illustrate the effectiveness of VE for this acceleration, saving more than 50% of the overall computations involved compared with the native FP method.

The rest of this paper is organized as follows. We review RED and its FP method in Section 2. Section 3 recalls the Vector Extrapolation acceleration idea. Several experiments on image deblurring and super-resolution, which follows the ones given in [17], show the effectiveness of VE, and these are brought in Section 4. We conclude our paper in Section 5.

2. REGularization by Denoising (RED)

This section reviews the framework of RED, which utilizes denoising algorithms as image priors [17]. We also describe its original solver based on the Fixed Point (FP) method.

2.1. Inverse problems as optimization tasks

From an estimation point of view, the signal \mathbf{x} is to be recovered from its measurements \mathbf{y} using the posterior conditional probability $P(\mathbf{x}|\mathbf{y})$. Using maximum a posterior probability (MAP) and the Bayes' rule, the estimation task is formulated as:

$$\begin{aligned} \mathbf{x}_{MAP}^* &= \arg \max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y}) \\ &= \arg \max_{\mathbf{x}} \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \\ &= \arg \max_{\mathbf{x}} P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) \\ &= \arg \min_{\mathbf{x}} -\log\{P(\mathbf{y}|\mathbf{x})\} - \log P(\mathbf{x}). \end{aligned}$$

The third equation is obtained by the fact that $P(\mathbf{y})$ does not depend on \mathbf{x} . The term $-\log\{P(\mathbf{y}|\mathbf{x})\}$ is known as the log-likelihood $\ell(\mathbf{y}, \mathbf{x})$. A typical example is

$$\ell(\mathbf{y}, \mathbf{x}) \triangleq -\log\{P(\mathbf{y}|\mathbf{x})\} = \frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 \quad (1)$$

referring to the case $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}$, where \mathbf{H} is any linear degradation operator and \mathbf{e} is a white mean zero Gaussian noise with variance σ^2 . Note that the expression $\ell(\mathbf{y}, \mathbf{x})$ depends on the distribution of the noise.² Now, we can write the MAP optimization problem as

$$\mathbf{x}_{MAP}^* = \arg \min_{\mathbf{x}} \ell(\mathbf{y}, \mathbf{x}) + \alpha R(\mathbf{x}) \quad (2)$$

where $\alpha > 0$ is a trade-off parameter to balance $\ell(\mathbf{y}, \mathbf{x})$ and $R(\mathbf{x})$. $R(\mathbf{x}) \triangleq -\log P(\mathbf{x})$ refers to the prior that describes the statistical nature of \mathbf{x} . This term is typically referred to as the regularization,

¹ Firstly, it enables utilizing the vast progress in image denoising for solving challenging inverse problems as explained above. Secondly, RED enables utilizing the denoiser as a black-box.

² White Gaussian noise is assumed throughout this paper.

as it is used to stabilize the inversion by emphasizing the features of the recovered signal. In the following, we will describe how RED activates denoising algorithms for composing $R(\mathbf{x})$. Note that Eq. (2) defines a wide family of inverse problems including, but not limited to, inpainting, deblurring, super-resolution, [21] and more.

2.2. RED and the fixed-point method

Define $f(\mathbf{x})$ as an abstract and differentiable denoiser.³ RED suggests applying the following form as the prior:

$$R(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T (\mathbf{x} - f(\mathbf{x})), \quad (3)$$

where \mathcal{T} denotes the transpose operator. The term $\mathbf{x}^T (\mathbf{x} - f(\mathbf{x}))$ is an image-adaptive Laplacian regularizer, which favors either a small residual $\mathbf{x} - f(\mathbf{x})$, or a small inner product between \mathbf{x} and the residual [17]. Plugged into Eq. (2), this leads the following minimization task:

$$\min_{\mathbf{x}} E(\mathbf{x}) \triangleq \ell(\mathbf{y}, \mathbf{x}) + \alpha \frac{1}{2} \mathbf{x}^T (\mathbf{x} - f(\mathbf{x})). \quad (4)$$

The prior $R(\mathbf{x})$ of RED is a convex function and easily differentiated if the following two conditions are met:

- Local Homogeneity: For any scalar c arbitrarily close to 1, we have $f(c\mathbf{x}) = cf(\mathbf{x})$.
- Strong Passivity: The Jacobian $\nabla_{\mathbf{x}} f(\mathbf{x})$ is stable in the sense that its spectral radius is upper bounded by one, $\rho(\nabla_{\mathbf{x}} f(\mathbf{x})) \leq 1$.

Surprisingly, the gradient of $E(\mathbf{x})$ is given by

$$\nabla_{\mathbf{x}} E(\mathbf{x}) = \nabla_{\mathbf{x}} \ell(\mathbf{y}, \mathbf{x}) + \alpha(\mathbf{x} - f(\mathbf{x})). \quad (5)$$

As discussed experimentally and theoretically in [17, Section 3.2], many of the state-of-the-art denoising algorithms satisfy the above-mentioned two conditions, and thus the gradient of (4) is simply evaluated through (5). As a consequence, $E(\mathbf{x})$ in Eq. (4) is a convex function if $\ell(\mathbf{x}, \mathbf{y})$ is convex, such as in the case of (1). In such cases any gradient-based algorithm can be utilized to address Eq. (4) leading to its global minimum.

Note that evaluating the gradient of $E(\mathbf{x})$ calls for one denoising activation, resulting in an expensive operation as the complexity of good-performing denoising algorithms is typically high. Because of the slow convergence speed of the steepest descent and the high complexity of ADMM, the work reported in [17] suggested using the FP method to handle the minimization task posed in Eq. (4). The development of the FP method is rather simple, relying on the fact that the global minimum of (4) should satisfy the first-order optimality condition, i.e., $\nabla_{\mathbf{x}} \ell(\mathbf{y}, \mathbf{x}) + \alpha(\mathbf{x} - f(\mathbf{x})) = \mathbf{0}$. For the FP method, we utilize the following iterative formula to solve this equation:

$$\nabla_{\mathbf{x}} \ell(\mathbf{y}, \mathbf{x}_{k+1}) + \alpha(\mathbf{x}_{k+1} - f(\mathbf{x}_k)) = \mathbf{0}. \quad (6)$$

The explicit expression of the above for $\ell(\mathbf{y}, \mathbf{x}) = \frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2$ is

$$\mathbf{x}_{k+1} = \left[\frac{1}{\sigma^2} \mathbf{H}^T \mathbf{H} + \alpha \mathbf{I} \right]^{-1} \left[\frac{1}{\sigma^2} \mathbf{H}^T \mathbf{y} + \alpha f(\mathbf{x}_k) \right], \quad (7)$$

where \mathbf{I} represents the identity matrix.⁴ The convergence of the FP method is guaranteed since

³ This denoising function admits a noisy image \mathbf{x} , and removes additive Gaussian noise form it, assuming a prespecified noise energy.

⁴ This matrix inversion is calculated in the Fourier domain for block-circulant \mathbf{H} , or using iterative methods for the more general cases.

$$\rho\left(\left[\frac{1}{\sigma^2}\mathbf{H}^T\mathbf{H} + \alpha\mathbf{I}\right]^{-1}\alpha\nabla_{\mathbf{x}}f(\mathbf{x}_k)\right) < 1.$$

Although the FP method is more efficient than the steepest descent and the ADMM, it still needs hundreds of iterations, which means hundreds of denoising activations, to reach the desired minimum. This results in a high complexity algorithm which we aim to address in this work. In the next section, we introduce an accelerated technique called Vector Extrapolation (VE) to substantially reduce the amount of iterations in the FP method.

3. Proposed method via vector extrapolation

We begin this section by introducing the philosophy of VE in linear and nonlinear systems and then discuss three variants of VE⁵, i.e., Minimal Polynomial Extrapolation (MPE), Reduced Rank Extrapolation (RRE) and Singular Value Decomposition Minimal Polynomial Extrapolation (SVD-MPE) [20,22]. Efficient implementation of these three variants is also discussed. We end this section by embedding VE in the FP method for RED, offering an acceleration of this scheme. Finally, we discuss the convergence and stability properties of VE.

3.1. VE in linear and nonlinear systems

Consider a vector set $\{\mathbf{x}_i \in \mathfrak{R}^N\}$ generated via a linear process,

$$\mathbf{x}_{i+1} = \mathbf{A}\mathbf{x}_i + \mathbf{b}, i = 0, 1, \dots, \quad (8)$$

where $\mathbf{A} \in \mathfrak{R}^{N \times N}$, $\mathbf{b} \in \mathfrak{R}^N$ and \mathbf{x}_0 is the initial vector. If $\rho(\mathbf{A}) < 1$, a limit point \mathbf{x}^* exists, being the FP of (8), $\mathbf{x}^* = \mathbf{A}\mathbf{x}^* + \mathbf{b}$. We turn to describe how VE works on such linear systems [23]. Denote $\mathbf{u}_i = \mathbf{x}_{i+1} - \mathbf{x}_i, i = 0, 1, \dots$, and define the defective vector \mathbf{e}_i as

$$\mathbf{e}_i = \mathbf{x}_i - \mathbf{x}^*, i = 0, 1, \dots \quad (9)$$

Subtracting \mathbf{x}^* from both sides of (8) and utilizing the fact that \mathbf{x}^* is the FP, we have $\mathbf{e}_{i+1} = \mathbf{A}\mathbf{e}_i$ resulting in

$$\mathbf{e}_{i+1} = \mathbf{A}^{i+1}\mathbf{e}_0. \quad (10)$$

We define a new extrapolated vector $\mathbf{x}_{(m,\kappa)}$ as a weighted average of the form

$$\mathbf{x}_{(m,\kappa)} = \sum_{i=0}^{\kappa} \gamma_i \mathbf{x}_{m+i}, \quad (11)$$

where $\sum_{i=0}^{\kappa} \gamma_i = 1$. Substituting (9) in (11) and using (10) and $\sum_{i=0}^{\kappa} \gamma_i = 1$, we have

$$\begin{aligned} \mathbf{x}_{(m,\kappa)} &= \sum_{i=0}^{\kappa} \gamma_i (\mathbf{x}^* + \mathbf{e}_{m+i}) \\ &= \mathbf{x}^* + \sum_{i=0}^{\kappa} \gamma_i \mathbf{e}_{m+i} \\ &= \mathbf{x}^* + \sum_{i=0}^{\kappa} \gamma_i \mathbf{A}^i \mathbf{e}_m. \end{aligned} \quad (12)$$

Note that the optimal $\{\gamma_i\}$ and κ should be chosen so as to force $\sum_{i=0}^{\kappa} \gamma_i \mathbf{A}^i \mathbf{e}_m = \mathbf{0}$. This way, we attain the FP through only one extrapolation step.

More broadly speaking, given a nonzero matrix $\mathbf{B} \in \mathfrak{R}^{N \times N}$ and an arbitrary nonzero vector $\mathbf{u} \in \mathfrak{R}^N$, we can find a unique polynomial $P(z)$ with smallest degree to yield $P(\mathbf{B})\mathbf{u} = \mathbf{0}$. Such a $P(z)$ is called the minimal polynomial of \mathbf{B} with respect to the vector \mathbf{u} . Notice

that the zeros of $P(z)$ are the eigenvalues of \mathbf{B} . Thus, assume that the minimal polynomial of \mathbf{A} with respect to \mathbf{e}_m can be represented as

$$P(z) = \sum_{i=0}^{\kappa} c_i z^i, \quad c_{\kappa} = 1 \quad (13)$$

resulting in $P(\mathbf{A})\mathbf{e}_m = \mathbf{0}$. So, we have

$$\sum_{i=0}^{\kappa} c_i \mathbf{A}^i \mathbf{e}_m = \sum_{i=0}^{\kappa} c_i \mathbf{e}_{m+i} = \mathbf{0}. \quad (14)$$

Multiplying both sides of (14) by \mathbf{A} results in $\sum_{i=0}^{\kappa} c_i \mathbf{A} \mathbf{e}_{m+i} = \sum_{i=0}^{\kappa} c_i \mathbf{e}_{m+i+1} = \mathbf{0}$, and thus we receive

$$\sum_{i=0}^{\kappa} c_i \mathbf{e}_{m+i} = \sum_{i=0}^{\kappa} c_i \mathbf{e}_{m+i+1} = \mathbf{0}. \quad (15)$$

Subtracting these expressions gives

$$\begin{aligned} \sum_{i=0}^{\kappa} c_i (\mathbf{e}_{m+i+1} - \mathbf{e}_{m+i}) &= \sum_{i=0}^{\kappa} c_i (\mathbf{x}_{m+i+1} - \mathbf{x}_{m+i}) \\ &= \sum_{i=0}^{\kappa} c_i \mathbf{u}_{m+i} = \mathbf{0}. \end{aligned} \quad (16)$$

This suggests that $\{c_i\}$ could be determined by solving the linear equations posed in (16). Once obtaining $\{c_i\}$, $\{\gamma_i\}$ are calculated through $\gamma_i = \frac{c_i}{\sum_{j=0}^{\kappa} c_j}$. Note that $\sum_{j=0}^{\kappa} c_j \neq 0$ if $\mathbf{I} - \mathbf{A}$ is not singular yielding $\sum_{j=0}^{\kappa} c_j = P(1) \neq 0$. Assuming κ is the degree of the minimal polynomial of \mathbf{A} with respect to \mathbf{e}_m , we can find a set of $\{\gamma_i\}$ to satisfy $\sum_{i=0}^{\kappa} \gamma_i = 1$ resulting in $\sum_{i=0}^{\kappa} \gamma_i \mathbf{x}_{m+i} = \mathbf{x}^*$. However, the degree of the minimal polynomial of \mathbf{A} can be as large as N , which in our case is very high. Moreover, we also do not have a way to obtain this degree with an easy algorithm. Because of these two difficulties, some approximate methods are developed to extrapolate the next vector via the previous ones and we will discuss them in Section 3.2.

Turning to the nonlinear case, denote \mathbf{F} as the FP function to evaluate the next vector,

$$\mathbf{x}_{i+1} = \mathbf{F}(\mathbf{x}_i), i = 0, 1, \dots, \quad (17)$$

where \mathbf{F} is an N -dimensional vector-valued function, $\mathbf{F}: \mathfrak{R}^N \rightarrow \mathfrak{R}^N$. We say \mathbf{x}^* is a FP of \mathbf{F} if $\mathbf{x}^* = \mathbf{F}(\mathbf{x}^*)$. Expanding $\mathbf{F}(\mathbf{x})$ in its Taylor series yields

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{x}^*) + \mathbf{F}'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + O(\|\mathbf{x} - \mathbf{x}^*\|^2) \text{ as } \mathbf{x} \rightarrow \mathbf{x}^*,$$

where $\mathbf{F}'(\cdot)$ is the Jacobian matrix of $\mathbf{F}(\cdot)$. Recalling $\mathbf{F}(\mathbf{x}^*) = \mathbf{x}^*$, we have

$$\mathbf{F}(\mathbf{x}) = \mathbf{x}^* + \mathbf{F}'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + O(\|\mathbf{x} - \mathbf{x}^*\|^2) \text{ as } \mathbf{x} \rightarrow \mathbf{x}^*.$$

Assuming the sequence $\mathbf{x}_0, \mathbf{x}_1, \dots$ converges to \mathbf{x}^* (if $\rho(\mathbf{F}'(\mathbf{x})) < 1$), it follows that \mathbf{x}_i will be close enough to \mathbf{x}^* for all large i , and hence

$$\mathbf{x}_{i+1} = \mathbf{x}^* + \mathbf{F}'(\mathbf{x}^*)(\mathbf{x}_i - \mathbf{x}^*) + O(\|\mathbf{x}_i - \mathbf{x}^*\|^2), \text{ as } i \rightarrow \infty.$$

Then, we rewrite this in the form

$$\mathbf{x}_{i+1} - \mathbf{x}^* = \mathbf{F}'(\mathbf{x}^*)(\mathbf{x}_i - \mathbf{x}^*) + O(\|\mathbf{x}_i - \mathbf{x}^*\|^2), \text{ as } i \rightarrow \infty.$$

For large i , the vectors $\{\mathbf{x}_i\}$ behave as in the linear system of the form $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}$ through

$$\mathbf{x}_{i+1} = \mathbf{A}\mathbf{x}_i + \mathbf{b}, \quad i = 0, 1, \dots,$$

where $\mathbf{A} = \mathbf{F}'(\mathbf{x}^*)$, $\mathbf{b} = [\mathbf{I} - \mathbf{F}'(\mathbf{x}^*)]\mathbf{x}^*$. This implies that the nonlinear system yields the same formula as the linear one and motivates us to extrapolate the next vector by the previous ones as in linear systems. Indeed, such an extension has been shown to be successful

⁵ We refer the interesting readers to [20] and the references therein to explore further the VE technique.

in various areas of science and engineering, e.g., computational fluid dynamics, semiconductor research, tomography and geometrical image processing [19,24,20].

3.2. Derivations of MPE, RRE and SVD-MPE

We turn to discuss how to utilize an approximate way to obtain the next vector by extrapolating the previous ones. Due to the fact that the degree of the minimal polynomial can be as large as N and we cannot obtain it, an arbitrary positive number is set as the degree, being much smaller than the true one. With such a replacement, the linear equations in (16) become inconsistent and there does not exist a solution for $\{c_i\}, c_\kappa = 1$ in the ordinary sense. Alternatively, we solve instead

$$\min_{\mathbf{c}} \|\mathbf{U}_\kappa^m \mathbf{c}\|_2^2, \text{ s.t. } c_\kappa = 1 \quad (18)$$

where $\mathbf{c} = [c_0 \ \dots \ c_\kappa]^T$ and $\mathbf{U}_\kappa^m = [\mathbf{u}_m \ \dots \ \mathbf{u}_{m+\kappa}]$. Then evaluating γ_i through $c_i / (\sum_{i=0}^{\kappa} c_i)$ results in the next vector $\mathbf{x}_{(m,\kappa)} = \sum_{i=0}^{\kappa} \gamma_i \mathbf{x}_{m+i}$ as a new approximation. This method is known as Minimal Polynomial Extrapolation (MPE) [25].

The detailed steps for obtaining the next vector through MPE are shown in Algorithm 1. To solve the constrained problem in (18), we suggest utilizing QR decomposition with the modified Gram-Schmidt (MGS) [25,26]. The MGS procedure for the matrix \mathbf{U}_κ^m is shown in Algorithm 2.

Now, let us discuss the other two variants of VE, i.e., Reduced Rank Extrapolation (RRE) [25] and SVD-MPE [22]. The main difference among RRE, MPE and SVD-MPE is at Step 2 in Algorithm 1 regarding the evaluation of $\{\gamma_i\}$. In RRE and SVD-MPE, we utilize the following methods to obtain $\{\gamma_i\}$:

Algorithm 1. Minimal Polynomial Extrapolation (MPE)

Initialization:

A sequence of vectors $\{\mathbf{x}_m, \mathbf{x}_{m+1}, \mathbf{x}_{m+2}, \dots, \mathbf{x}_{m+\kappa+1}\}$ is produced by the baseline algorithm (FP in our case).

Output:

A new vector $\mathbf{x}_{(m,\kappa)}$.

1: Construct the matrix

$$\mathbf{U}_\kappa^m = [\mathbf{x}_{m+1} - \mathbf{x}_m, \dots, \mathbf{x}_{m+\kappa+1} - \mathbf{x}_{m+\kappa}] \in \mathbb{R}^{N \times (\kappa+1)}$$

and then compute its QR factorization via Algorithm 2,

$$\mathbf{U}_\kappa^m = \mathbf{Q}_\kappa \mathbf{R}_\kappa.$$

2: Denote $\mathbf{r}_{\kappa+1}$ as the $\kappa + 1$ th column of \mathbf{R}_κ without the last row and solve the following $\kappa \times \kappa$ upper triangular system

$$\mathbf{R}_{\kappa-1} \mathbf{c}' = -\mathbf{r}_{\kappa+1} \mathbf{c}' = [c_0, c_1, \dots, c_{\kappa-1}]^T$$

where $\mathbf{R}_{\kappa-1}$ is the previous κ columns of \mathbf{R}_κ without the last row. Finally, evaluate $\{\gamma_i\}$ through $\left\{ \frac{c_i}{\sum_{i=0}^{\kappa} c_i} \right\}$.

3: Compute $\xi = [\xi_0, \xi_1, \dots, \xi_{\kappa-1}]^T$ through

$$\xi_0 = 1 - \gamma_0; \xi_j = \xi_{j-1} - \gamma_j, j = 1, \dots, \kappa - 1$$

4: Compute $\boldsymbol{\eta} = [\eta_0, \eta_1, \dots, \eta_{\kappa-1}]^T = \mathbf{R}_{\kappa-1} \xi$. Then we attain $\mathbf{x}_{(m,\kappa)} = \mathbf{x}_m + \mathbf{Q}_{\kappa-1} \boldsymbol{\eta}$ as the new initial vector where $\mathbf{Q}_{\kappa-1}$ represents the previous κ columns of \mathbf{Q}_κ .

Algorithm 2. Modified Gram-Schmidt (MGS)

Output: \mathbf{Q}_κ and \mathbf{R}_κ (r_{ij} denotes the (i, j) th element of \mathbf{R}_κ and \mathbf{q}_i and \mathbf{u}_i represent the i th column of \mathbf{Q}_κ and \mathbf{U}_κ^m , respectively.).

1: Compute $r_{11} = \|\mathbf{u}_1\|_2$ and $\mathbf{q}_1 = \mathbf{u}_1 / r_{11}$.

2: **for** $i = 2, \dots, \kappa + 1$ **do**

3: Set $\mathbf{u}_i^{(1)} = \mathbf{u}_i$

4: **for** $j = 1, \dots, i - 1$ **do**

5: $r_{ji} = \mathbf{q}_j^T \mathbf{u}_i^{(j)}$ and $\mathbf{u}_i^{(j+1)} = \mathbf{u}_i^{(j)} - r_{ji} \mathbf{q}_j$

6: **end for**

7: Compute $r_{ii} = \|\mathbf{u}_i^{(i)}\|_2$ and $\mathbf{q}_i = \mathbf{u}_i^{(i)} / r_{ii}$

8: **end for**

RRE: Solving $\mathbf{R}_\kappa^T \mathbf{R}_\kappa \mathbf{d} = \mathbf{1}$ through forward and backward substitution, we obtain $\boldsymbol{\gamma}$ through $\frac{\mathbf{d}}{\sum_i d_i}$. Actually, such a formulation of $\boldsymbol{\gamma}$ is the solution of:

$$\min_{\boldsymbol{\gamma}} \|\mathbf{U}_\kappa^m \boldsymbol{\gamma}\|_2^2, \text{ s.t. } \sum_i \gamma_i = 1. \quad (19)$$

SVD-MPE: Computing the SVD decomposition of $\mathbf{R}_\kappa = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T$, we have $\boldsymbol{\gamma} = \frac{\mathbf{v}_{\kappa+1}}{\sum_i v_{i,\kappa+1}}$ where $\mathbf{v}_{\kappa+1}$ and $\mathbf{v}_{i,\kappa+1}$ represent the last column and the $(i, \kappa + 1)$ th element of matrix \mathbf{V} , respectively.

Here are two remarks regarding RRE, MPE and SVD-MPE:

- Observing the derivations of MPE, SVD-MPE and RRE, we notice that RRE's solution must exist unconditionally, while MPE and SVD-MPE may not exist because the sum of $\{c_i\}$ and $\{v_{i,\kappa+1}\}$ in MPE and SVD-MPE may become zero. Thus RRE may be more robust in practice [24]. However, MPE and RRE are related, as revealed in [27]. Specifically, if MPE does not exist, we have $\mathbf{x}_{(m,\kappa)}^{\text{RRE}} = \mathbf{x}_{(m,\kappa-1)}^{\text{RRE}}$. Otherwise, the following holds

$$\mu_\kappa \mathbf{x}_{(m,\kappa)}^{\text{RRE}} = \mu_{\kappa-1} \mathbf{x}_{(m,\kappa-1)}^{\text{RRE}} + \nu_\kappa \mathbf{x}_{(m,\kappa)}^{\text{MPE}}, \mu_\kappa = \mu_{\kappa-1} + \nu_\kappa$$

where $\mu_\kappa, \mu_{\kappa-1}$ and ν_κ are positive scalars depending only on $\mathbf{x}_{(m,\kappa)}^{\text{RRE}}, \mathbf{x}_{(m,\kappa-1)}^{\text{RRE}}$ and $\mathbf{x}_{(m,\kappa)}^{\text{MPE}}$, respectively. Furthermore, the performance of MPE and RRE is similar – both of the methods either perform well or work poorly [20].

- Observe that we only need to store $\kappa + 2$ vectors in memory at all steps in Algorithm 2. Formulating the matrix \mathbf{U}_κ^m , we overwrite the vector \mathbf{x}_{m+i} with $\mathbf{u}_{m+i} = \mathbf{x}_{m+i} - \mathbf{x}_{m+i-1}$ when the latter is computed and only \mathbf{x}_m is always in the memory. Next, \mathbf{u}_{m+i} is overwritten by $\mathbf{q}_i, i = 1, \dots, \kappa + 1$ in computing the matrix \mathbf{Q}_κ . Thus, we do not need to save the vectors $\mathbf{x}_{m+1}, \dots, \mathbf{x}_{m+\kappa+1}$, which implies that no additional memory is required in running Algorithm 2.

3.3. Embedding VE in the baseline algorithm

We introduce VE in its cycling formulation for practical usage. One cycling means we activate the baseline algorithm to produce $\{\mathbf{x}_i\}$ and then utilize VE once to evaluate the new vector as a novel initial point. Naturally, we repeat such a cycling many times. The steps of utilizing VE in its cycling mode are shown in Algorithm 3. Few comments are in order:

- In practice, we utilize VE in its cycling formulation. Specially, the iterative form shown in Algorithm 3 is named as full cycling [20]. To save the complexity of computing $\{\gamma_i\}$, one may reuse the previous $\{\gamma_i\}$, a method known as cycling with frozen γ_i . Parallel VE can be also developed if more machines are available. Explaining details of the last two strategies is out of the scope of this paper. We refer the reader to [20] for more information.
- Numerical experience also indicates that cycling with even moderately large $m > 0$ will avoid stalling from happening [28]. Moreover, we also recommend setting $m > 0$ when the problem becomes challenging to solve.
- In our case, the termination criterion in Algorithm 3 can be the number of total iterations (the number of calling of the baseline algorithm) or the difference between consecutive two vectors. Furthermore, we also recommend giving additional iterations to activate the baseline algorithm after terminating the VE, which can stabilize the accelerated algorithm in practice.

Algorithm 3. Baseline Algorithm + Vector Extrapolation

Initialization:

Choose nonnegative integers m and κ and an initial vector \mathbf{x}_0 . The baseline algorithm is the FP method, as given in (7).

Output:

Final Solution \mathbf{x}^* .

1: while 1 do

2: Obtain the series of \mathbf{x}_i through the baseline algorithm where $1 \leq i \leq m + \kappa + 1$, and save \mathbf{x}_{m+i} for $0 \leq i \leq \kappa + 1$ to formulate \mathbf{U}_m^κ .

3: Call Algorithm 1 to obtain $\mathbf{x}_{(m,\kappa)}$.

4: If the termination of the algorithm is satisfied, set $\mathbf{x}^* = \mathbf{x}_{(m,\kappa)}$ and break, otherwise, set $\mathbf{x}_{(m,\kappa)}$ as the new initial point \mathbf{x}_0 and go to Step 2.

5: end while

3.4. Convergence and stability properties

We mention existing results regarding the convergence and stability properties of VE for understanding this technique better. A rich literature has examined the convergence and stability properties of RRE, MPE and SVD-MPE in linear systems [29,30]. Assuming the matrix \mathbf{A} is diagonalizable, then in the k th iteration \mathbf{x}_k should have the form $\mathbf{x}_k = \mathbf{x}^* + \sum_{i=1}^{\kappa} \mathbf{v}_i \lambda_i^k$ where $(\lambda_i, \mathbf{v}_i)$ are some or all of the eigenvalues and corresponding eigenvectors of \mathbf{A} , with distinct nonzero eigenvalues. By ordering λ_i as $|\lambda_1| \geq |\lambda_2| \geq \dots$, the following asymptotic performance holds for all of the three variants of VE when $|\lambda_k| > |\lambda_{k+1}|$:

$$\mathbf{x}_{(m,\kappa)} - \mathbf{x}^* = O(\lambda_{\kappa+1}^m) \text{ as } m \rightarrow \infty. \quad (20)$$

This implies that the sequence $\{\mathbf{x}_{(m,\kappa)}\}_{m=0}^{\infty}$ converges to \mathbf{x}^* faster than the original sequence $\{\mathbf{x}_k\}$.

As shown in (20), for a large m , (8) reduces the contributions of the smaller λ_i to the error $\mathbf{x}_{(m,\kappa)} - \mathbf{x}^*$, while VE eliminates the contributions of the κ largest λ_i . This indicates that $\mathbf{x}_{(m,\kappa)} - \mathbf{x}^*$ is smaller than each of the errors $\mathbf{x}_{m+i} - \mathbf{x}^*$, $i = 0, 1, \dots, \kappa$, when m is large enough. We mention another observation that an increasing κ generally results in a faster convergence of VE. However, a large κ has to increase the storage requirements and also requires a much higher computational cost. Numerical experiments indicate that a moderate κ can already works well in practice.

If the following condition is held, we say VE is stable:

$$\sup_m \sum_{i=0}^{\kappa} |\gamma_i^{(m,\kappa)}| < \infty. \quad (21)$$

Here, we denote $\{\gamma_i\}$ by $\{\gamma_i^{(m,\kappa)}\}$ to show their dependence on m and κ . If (21) holds true, the error in \mathbf{x}_i will not magnify severely. As shown in [29–31], MPE and RRE obey such a stability property.

For nonlinear systems, the analysis of convergence and stability becomes extremely challenging. One of the main results is the quadratic convergence theorem [23,32,33]. This theorem is built on one special assumption that κ is set to be the degree of the minimal polynomial of $\mathbf{F}'(\mathbf{x}^*)$. The proof of the quadratic convergence was shown in [32]. In a following work, Smith, Ford and Sidi noticed that there exists a gap in the previous proof [23]. Jbilou et al. suggested two more conditions in order to close the gap [33]:

- The matrix $\mathbf{F}'(\mathbf{x}^*) - \mathbf{I}$ is nonsingular
- $\mathbf{F}'(\cdot)$ satisfies the following Lipschitz condition:

$$\|\mathbf{F}'(\mathbf{x}) - \mathbf{F}'(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|L > 0.$$

Surprisingly, these two conditions are met by the RED scheme. The first condition is satisfied by the fact

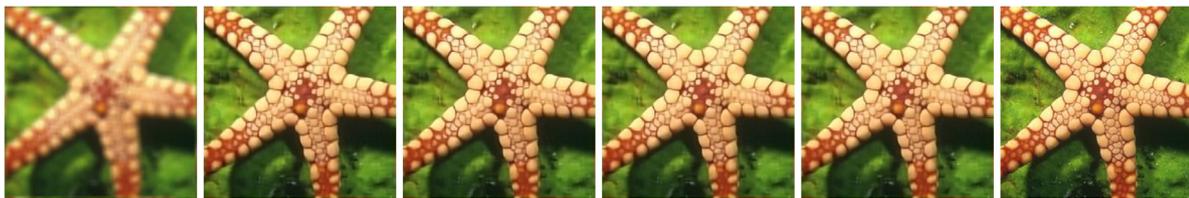
$$\rho\left(\left[\frac{1}{\sigma^2}\mathbf{H}^T\mathbf{H} + \alpha\mathbf{I}\right]^{-1} \alpha\nabla_{\mathbf{x}}f(\mathbf{x})\right) < 1.$$

The second one is also true, due to the assumption in RED that the denoiser $f(\mathbf{x})$ is differentiable. So we claim that it is possible for VE to solve RED with quadratic convergence rate.

Although VE can lead to a quadratic convergence rate, trying to achieve such a rate may not be realistic because κ can be as large as N . However, we may obtain a linear but fast convergence in practice with even moderate values of m and κ , which is also demonstrated in the following numerical experiments.

4. Experimental results

We follow the same experiments of image deblurring and super-resolution as presented in [17] to investigate the performance of VE in acceleration. The trainable nonlinear reaction diffusion (TNRD) method [6] is chosen as the denoising engine. Mainly, we choose the FP method as our baseline algorithm. For a fair comparison, the same parameters suggested in [17] for different image processing tasks are set in our experiments. In [17], the authors compared RED with other popular algorithms in image deblurring and super-resolution tasks, showing its superiority. As the main purpose in this paper is to present the acceleration of our method for solving RED, we omit the comparisons with other popular algorithms. In the following, we mainly show the acceleration of applying MPE with FP for solving RED first and then discuss the choice of parameters in VE, i.e., m and κ . In addition, we compare our method with three other methods, steepest descent (SD), Nesterov's acceleration [34] and Limited-memory BFGS (L-BFGS) [35]. Note that we need to determine a proper step-size for the above methods [35]. However, evaluating the objective value or gradient in RED is expensive implying that any line-search method becomes prohibitive. Note that, in contrast, in our framework as described in Algorithm 3 does not suffer from such a problem. In the following, we manually choose a fixed step-size for getting a good convergence behavior. Finally, we compare the difference among RRE, MPE and SVD-MPE. All of the experiments are conducted on a workstation with Intel(R) Xeon(R) CPU E5-2699 @2.20 GHz.

(a) SD: 22.56dB (input) \rightarrow 25.65dB \rightarrow 26.47dB \rightarrow 26.97dB \rightarrow 27.35dB \rightarrow original.(b) SD-MPE: 22.56dB (input) \rightarrow 27.46dB \rightarrow 28.62dB \rightarrow 29.29dB \rightarrow 29.69dB \rightarrow original.(c) Nesterov: 22.56dB (input) \rightarrow 27.44dB \rightarrow 28.78dB \rightarrow 29.43dB \rightarrow 29.79dB \rightarrow original.(d) LBFGS: 22.56dB (input) \rightarrow 27.80dB \rightarrow 29.52dB \rightarrow 30.14dB \rightarrow 30.40dB \rightarrow original.(e) FP: 22.56dB (input) \rightarrow 28.91dB \rightarrow 29.76dB \rightarrow 30.13dB \rightarrow 30.31dB \rightarrow original.(f) FP-MPE: 22.56dB (input) \rightarrow 30.07dB \rightarrow 30.55dB \rightarrow 30.60dB \rightarrow 30.60dB \rightarrow original.**Fig. 1.** Reconstruction results of different algorithms in various iterations for the uniform kernel. From left to right: Blurred one \rightarrow #20 \rightarrow #40 \rightarrow #60 \rightarrow #80 \rightarrow Ground truth.

4.1. Image deblurring

In this experiment, we degrade the test images by convolving with two different point spread functions (PSFs), i.e., 9×9 uniform blur and a Gaussian blur with a standard derivation of 1.6. In both of these cases, we add an additive Gaussian noise with $\sigma = \sqrt{2}$ to the blurred images. The parameters m and κ in Algorithm 3 are set to 0 and 5 for the image deblurring task. Additionally, we apply VE to the case where the baseline algorithm is SD, called SD-MPE,

with the parameters m and κ are chosen as 0 and 8. The value of the cost function and peak signal to noise ratio (PSNR) versus iteration or CPU time are given in Figs. 2 and 3⁶. These correspond to both a uniform and a Gaussian blur kernels, all tested on the

⁶ The goal of this paper is to investigate the performance of solving RED with VE rather than the restoration results. Therefore, we present the recovered PSNR versus iteration or running time. One can utilize some no-reference quality metrics like NFERM [36] and ARISM_c [37] to further examine the restoration results.

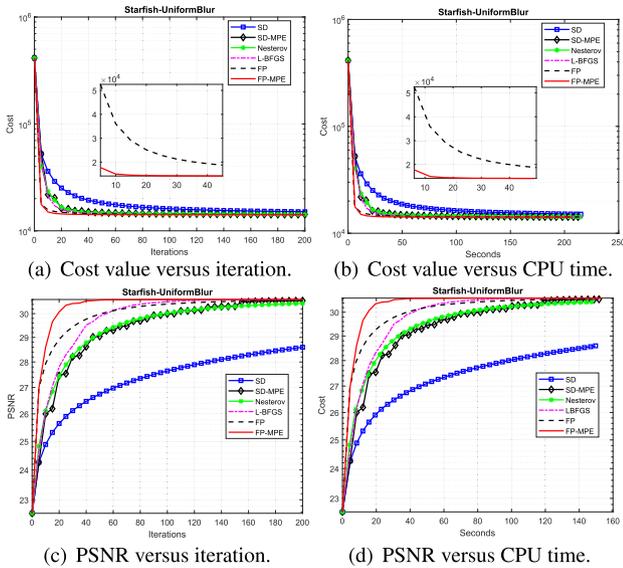


Fig. 2. Image Deblurring - Uniform Kernel, "Starfish" Image.

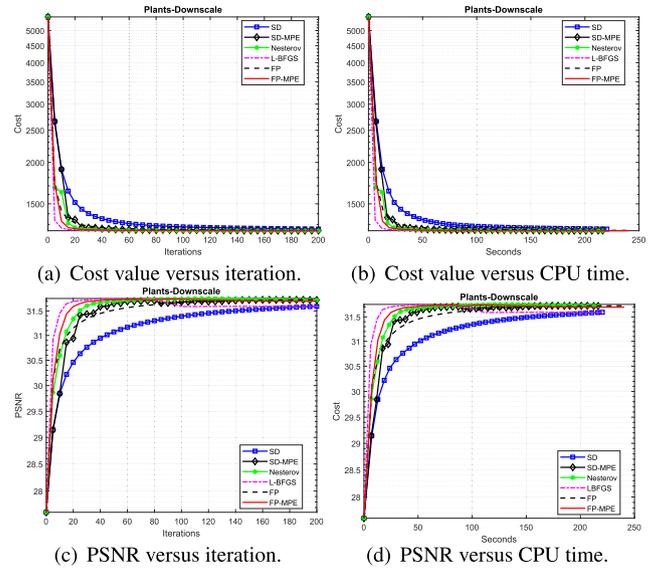


Fig. 4. Image Super-Resolution, "Plant" Image.

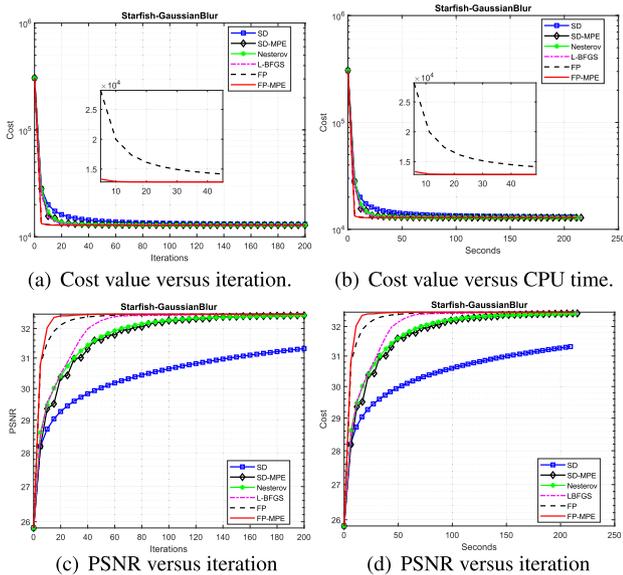


Fig. 3. Image Deblurring - Gaussian Kernel, "Starfish" Image.

"starfish" image. Clearly, we observe that SD is the slowest algorithm. Surprisingly, SD-MPE and Nesterov's method yield almost the same convergence speed, despite their totally different scheme. Moreover, We note that FP is faster than L-BFGS, Nesterov's method and SD-MPE. Undoubtedly, FP-MPE is the fastest one, both in terms of iterations and CPU time, which indicates the effectiveness of MPE's acceleration. To provide a visual effect, we show the change in recon-

structed quality of different algorithms in Fig. 1. Clearly, the third column of FP-MPE achieves the best reconstruction faster, while other methods need more iterations to obtain a comparable result.

Additional nine test images suggested in [17] are also included in our experiments, in order to investigate the performance of VE further. In this experiment we focus on the comparison between FP-MPE and FP for the additional images. We run the native FP method 200 iterations first and denote the final image by \mathbf{x}^* . Clearly, the corresponding cost-value is $E(\mathbf{x}^*)$. We activate Algorithm 3 with the same initial value as used in the FP method to examine how many iterations are needed to attain the same or lower objective value than $E(\mathbf{x}^*)$. The final number of iterations with different images are given in Table 1. Clearly, an acceleration is observed in all the test images in the image deblurring task.

4.2. Image super-resolution

We generate a low resolution image by blurring the ground truth one with a 7×7 Gaussian kernel with standard derivation 1.6 and then downsample by a factor of 3. Afterwards, an additive Gaussian noise with $\sigma = 5$ is added to the resulting image. The same parameters m and κ used in the deblurring task for FP-MPE are adopted here. For SD-MPE, the parameters m and κ are set to 1 and 10, respectively. We choose "Plants" as our test image because it needs more iterations for FP-MPE to converge. As observed from Fig. 4, while L-BFGS and the Nesterov's method are faster than the FP method, our acceleration method (FP-MPE) is quite competitive with both. Furthermore, we investigate all of the test images as shown in [17] to see how many iterations are needed for MPE to achieve the same or lower cost compared with the FP method. The results are shown in Table 2. As can be seen,

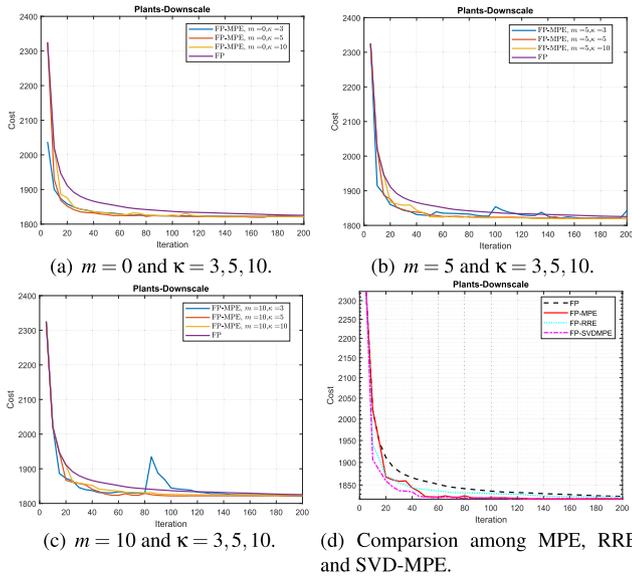
Table 1 The number of iterations with different images for FP and FP-MPE in image deblurring task to attain the same cost.

Image	Butterfly	Boats	C. Man	House	Parrot	Lena	Barbara	Starfish	Peppers	Leaves
Deblurring: Uniform kernel, $\sigma = \sqrt{2}$										
RED: FP-TNRD	200	200	200	200	200	200	200	200	200	200
RED: FP-MPE-TNRD	60	55	55	80	50	50	55	50	55	55
Deblurring: Gaussian kernel, $\sigma = \sqrt{2}$										
RED: FP-TNRD	200	200	200	200	200	200	200	200	200	200
RED: FP-MPE-TNRD	70	65	55	65	55	80	45	55	95	80

Table 2

The number of iterations with different images for FP and FP-MPE in image super-resolution task to attain the same cost.

Super-Resolution, scaling =3, $\sigma = 5$									
Image	Butterfly	Flower	Girl	Parth.	Parrot	Raccoon	Bike	Hat	Plants
RED: FP-TNRD	200	200	200	200	200	200	200	200	200
RED: FP-MPE-TNRD	60	65	50	70	55	60	60	50	70

**Fig. 5.** Exploring the robustness to the choice of the parameters in MPE, and the difference among the three VE schemes. All these graphs correspond to the test image “Plants” for the single image super-resolution task.

MPE works better than the FP method indicating an effective acceleration for solving RED.

4.3. The choice of the parameters and the difference among RRE, MPE and SVD-MPE

Conclude by discussing the robustness to the choice of parameters m and κ for the MPE algorithm. To this end, the single image super-resolution task is chosen as our study. Furthermore, we choose to demonstrate this robustness on the “Plants” image since it required the largest number of iterations in the MPE recovery process. As seen from Figs. 5(a)–(c), MPE always converges faster than the regular FP method with different m and κ . Moreover, we also observe that a lower objective value is attained through MPE. Notice that MPE has some oscillations because it is not a monotonically accelerated technique. However, we still see a lower cost is achieved if additional iterations are given.

In part (d) of Fig. 5, an optimal pair of m and κ is chosen for MPE, RRE and SVD-MPE for the single image super-resolution task with the “Plants” image.⁷ We see that all three methods yield an acceleration and a lower cost, demonstrating the effectiveness of the various variants of VE. Moreover, we see that SVD-MPE converges faster at the beginning, but MPE yields a lowest eventual cost.

5. Conclusion

The work reported in [17] introduced RED – a flexible framework for using arbitrary image denoising algorithms as priors for

general inverse problems. This scheme amounts to iterative algorithms in which the denoiser is called repeatedly. While appealing and quite practical, there is one major weakness to the RED scheme – the complexity of denoising algorithms is typically high which implies that the use of RED is likely to be costly in run-time. This work aims at deploying RED efficiently, alleviating the above described shortcoming. An accelerated technique is proposed in this paper, based on the Vector Extrapolation (VE) methodology. The proposed algorithms are demonstrated to substantially reduce the number of overall iterations required for the overall recovery process. We also observe that the choice of the parameters in the VE scheme is robust.

Declaration of Competing Interest

There is no conflict of interest.

Acknowledgment

Y. R. thanks the Zuckerman Institute, ISEF Foundation, and the Viterbi Fellowship, Technion, for providing additional research support.

References

- [1] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, vol. 2, 2005, pp. 60–65.
- [2] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Trans. Image Process.* 15 (12) (2006) 3736–3745.
- [3] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, *IEEE Trans. Image Process.* 16 (8) (2007) 2080–2095.
- [4] D. Zoran, Y. Weiss, From learning models of natural image patches to whole image restoration, in: 2011 IEEE International Conference on Computer Vision (ICCV), IEEE, 2011, pp. 479–486.
- [5] W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally centralized sparse representation for image restoration, *IEEE Trans. Image Process.* 22 (4) (2013) 1620–1630.
- [6] Y. Chen, T. Pock, Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6) (2017) 1256–1272.
- [7] P. Chatterjee, P. Milanfar, Is denoising dead?, *IEEE Trans Image Process.* 19 (4) (2010) 895–911.
- [8] P. Milanfar, A tour of modern image filtering: New insights and methods, both practical and theoretical, *IEEE Signal Process. Mag.* 30 (1) (2013) 106–128.
- [9] A. Levin, B. Nadler, Natural image denoising: Optimality and inherent bounds, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 2833–2840.
- [10] M. Protter, M. Elad, H. Takeda, P. Milanfar, Generalizing the nonlocal-means to super-resolution reconstruction, *IEEE Trans. Image Process.* 18 (1) (2009) 36–51.
- [11] A. Danielyan, V. Katkovnik, K. Egiazarian, Bm3d frames and variational image deblurring, *IEEE Trans. Image Process.* 21 (4) (2012) 1715–1728.
- [12] C.A. Metzler, A. Maleki, R.G. Baraniuk, Optimal recovery from compressive measurements via denoising-based approximate message passing, in: 2015 International Conference on Sampling Theory and Applications (SampTA), IEEE, 2015, pp. 508–512.
- [13] S.V. Venkatakrisnan, C.A. Bouman, B. Wohlberg, Plug-and-play priors for model based reconstruction, in: 2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2013, pp. 945–948.
- [14] S. Sreehari, S.V. Venkatakrisnan, B. Wohlberg, G.T. Buzzard, L.F. Drummy, J.P. Simmons, C.A. Bouman, Plug-and-play priors for bright field electron tomography and sparse interpolation, *IEEE Trans. Comput. Imag.* 2 (4) (2016) 408–423.

⁷ The optimal m and κ are obtained by searching in the range [0, 10] with $\kappa \geq 2$, seeking the fastest convergence for these three methods.

- [15] S.H. Chan, X. Wang, O.A. Elgendy, Plug-and-play admm for image restoration: Fixed-point convergence and applications, *IEEE Trans. Comput. Imag.* 3 (1) (2017) 84–98.
- [16] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Machine Learn.* 3 (1) (2011) 1–122.
- [17] Y. Romano, M. Elad, P. Milanfar, The little engine that could: Regularization by denoising (red), *SIAM J. Imag. Sci.* 10 (4) (2017) 1804–1844.
- [18] S. Cabay, L. Jackson, A polynomial extrapolation method for finding limits and antilimits of vector sequences, *SIAM J. Num. Anal.* 13 (5) (1976) 734–752.
- [19] N. Rajeevan, K. Rajgopal, G. Krishna, Vector-extrapolated fast maximum likelihood estimation algorithms for emission tomography, *IEEE Trans. Med. Imaging* 11 (1) (1992) 9–20.
- [20] A. Sidi, *Vector Extrapolation Methods with Applications*, vol. 17, SIAM, 2017.
- [21] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, vol. 120, Springer Science & Business Media, 2011.
- [22] A. Sidi, "Svd-mpe: An svd-based vector extrapolation method of polynomial type," arXiv preprint arXiv:1505.00674, 2015.
- [23] D.A. Smith, W.F. Ford, A. Sidi, Extrapolation methods for vector sequences, *SIAM Rev.* 29 (2) (1987) 199–233.
- [24] G. Rosman, L. Dascal, A. Sidi, R. Kimmel, Efficient beltrami image filtering via vector extrapolation methods, *SIAM J. Imag. Sci.* 2 (3) (2009) 858–878.
- [25] A. Sidi, Efficient implementation of minimal polynomial and reduced rank extrapolation methods, *J. Comput. Appl. Math.* 36 (3) (1991) 305–337.
- [26] G.H. Golub, C.F. Van Loan, *Matrix Computations*, vol. 3, JHU Press, 2012.
- [27] A. Sidi, Minimal polynomial and reduced rank extrapolation methods are related, *Adv. Comput. Math.* 43 (1) (2017) 151–170.
- [28] A. Sidi, Y. Shapira, Upper bounds for convergence rates of acceleration methods with initial iterations, *Num. Algorithms* 18 (2) (1998) 113–132.
- [29] A. Sidi, W.F. Ford, D.A. Smith, Acceleration of convergence of vector sequences, *SIAM J. Num. Anal.* 23 (1) (1986) 178–196.
- [30] A. Sidi, J. Bridger, Convergence and stability analyses for some vector extrapolation methods in the presence of defective iteration matrices, *J. Comput. Appl. Math.* 22 (1) (1988) 35–61.
- [31] A. Sidi, Convergence and stability properties of minimal polynomial and reduced rank extrapolation algorithms, *SIAM J. Num. Anal.* 23 (1) (1986) 197–209.
- [32] S. Skelboe, Computation of the periodic steady-state response of nonlinear networks by extrapolation methods, *IEEE Trans. Circ. Syst.* 27 (3) (1980) 161–175.
- [33] K. Jbilou, H. Sadok, Some results about vector extrapolation methods and related fixed-point iterations, *J. Comput. Appl. Math.* 36 (3) (1991) 385–398.
- [34] Y. Nesterov, *Lectures on Convex Optimization*, Springer, 2018.
- [35] J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer, 2006.
- [36] K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment, *IEEE Trans. Multimedia* 17 (1) (2015) 50–63.
- [37] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, No-reference image sharpness assessment in autoregressive parameter space, *IEEE Trans. Image Process.* 24 (10) (2015) 3218–3231.