Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Adaptive and Accelerated Golden-Ratio Primal-Dual Algorithms for Total Variation Image Denoising

**Faiz Ur Rahman**
King Fahd University of Petroleum and Minerals
**Izhar Ahmad**
drizhar@kfupm.edu.sa

King Fahd University of Petroleum and Minerals

**Research Article**

**Keywords:**

**Posted Date:** November 2nd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11336696/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Adaptive and Accelerated Golden-Ratio Primal-Dual Algorithms for Total Variation Image Denoising

Faiz Ur Rahman[1] and Izhar Ahmad[1,2*]

[1]Department of Mathematics, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia.
[2]Center for Intelligent Secure Systems, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia.

*Corresponding author(s). E-mail(s): drizhar@kfupm.edu.sa;

**Abstract**

Total variation (TV) denoising is a widely used variational approach for recovering images from noisy observations while preserving sharp edges. The resulting Rudin–Osher–Fatemi (ROF) model is convex but nonsmooth and can be efficiently treated by primal–dual first-order methods. This paper specializes adaptive and accelerated golden-ratio primal–dual algorithms to the ROF TV-denoising problem. First, an extended golden-ratio primal–dual algorithm (E-GRPDA) with a non-increasing adaptive stepsize rule is considered. The resulting scheme avoids explicit estimation of the discrete gradient operator norm and backtracking while retaining simple matrix-free proximal updates. We then exploit the strong convexity of the quadratic fidelity term to specialize a parameter-free accelerated golden-ratio primal–dual algorithm (PF-AGRPDA), for which an $\boldsymbol{O(1/N^2)}$ convergence rate is obtained for the corresponding weighted ergodic quantities. Numerical experiments on the Cameraman, House, and Peppers images at two Gaussian-noise levels compare the primal–dual hybrid gradient (PDHG) method, E-GRPDA, and PF-AGRPDA. PDHG gives the fastest decrease of the reported nonergodic objective residual, whereas PF-AGRPDA provides competitive reconstruction quality and achieves the highest final PSNR and the lowest relative $\boldsymbol{L_2}$ error in the reported tests. The results demonstrate the practical viability of the adaptive and accelerated golden-ratio frameworks while highlighting the distinction between theoretical ergodic acceleration and finite-iteration computational performance.

**Keywords:** Total variation denoising; ROF model; primal–dual algorithms; golden-ratio algorithm; adaptive stepsizes; convex optimization

## 1 Introduction

Variational image denoising is commonly formulated as an optimization problem that balances fidelity to a noisy observation with a regularity prior that encodes desirable image structure Rudin et al. (1992). In this setting, quadratic $L_2$ regularization is effective at suppressing noise but tends to oversmooth discontinuities, because it penalizes large gradients uniformly and therefore blurs edges Rudin et al. (1992). The Rudin–Osher–Fatemi (ROF) model addressed this limitation by introducing total variation regularization, which promotes piecewise-smooth reconstructions and preserves sharp transitions by penalizing the magnitude of the image gradient rather than its

squared magnitude Rudin et al. (1992). Because many imaging tasks depend on accurate boundary localization, the ROF model established total variation denoising as a principled edge-preserving variational approach and motivated the development of numerous efficient numerical methods for TV-based models Rudin et al. (1992); Chan et al. (1999); Chambolle (2004); Zhu and Chan (2008).

From an algorithmic perspective, the ROF objective is convex but nonsmooth because of the total variation term. This nonsmoothness has motivated the development of specialized numerical methods for TV-based optimization problems Rudin et al. (1992); Chan et al. (1999). Early approaches relied on variational and dual formulations, while subsequent developments introduced efficient splitting and primal–dual strategies tailored to the nonsmooth structure of TV regularization Chan et al. (1999); Chambolle (2004); Goldstein and Osher (2009); Zhu and Chan (2008). These developments motivated efficient first-order methods for large-scale TV models, including accelerated gradient, splitting, and primal–dual approaches Beck and Teboulle (2009); Goldstein and Osher (2009); Zhu and Chan (2008); Chambolle and Pock (2011). More recently, Mahdi et al. Mahdi et al. (2026) investigated derivative-free conjugate-gradient-type methods for nonlinear equations arising in image restoration. In particular, Chan et al. Chan et al. (1999) developed a nonlinear primal–dual method for TV-based image restoration by introducing an additional dual variable associated with the TV term. This primal–dual viewpoint is even more significant when looking for scalable first order methods that do not require the solution of large linear systems, a distinguishing benefit in high dimensional imaging applications Zhu and Chan (2008).

With the advancement of convex optimization tools, proximal splitting frameworks emerged as a systematic approach for constructing convergent algorithms for structured convex problems Combettes and Wajs (2005); Combettes and Pesquet (2012); Condat (2013). Combettes and Wajs Combettes and Wajs (2005); Combettes and Pesquet (2012) formalized proximal forward–backward splitting for signal recovery, highlighting the proximity operator as a fundamental computational building block for non-smooth convex terms Combettes and Wajs (2005).However, TV denoising involves a nonsmooth function composed with a linear operator, namely the discrete gradient, and this linear-composite structure is naturally handled by primal–dual splitting methods that update the primal and dual variables separately Chambolle and Pock (2011); Combettes and Pesquet (2012); Condat (2013); Bertsekas (2009). Alternative splitting and preconditioning strategies have also been developed for convex–concave saddle-point formulations, including Douglas–Rachford-type methods with applications to discrete TV denoising Bredies and Sun (2015). From a more general optimization viewpoint, convex–concave saddle-point problems can also be approached by subgradient-type schemes, and Nedić and Ozdaglar developed subgradient methods for saddle-point problems that provide convergence guarantees under standard assumptions Nedić and Ozdaglar (2009). Although subgradient methods are broadly applicable, TV restoration typically benefits from exploiting additional structure (separable proximals, linear operators, and strong convexity of the data term), which motivates specialized primal–dual algorithms for imaging Chambolle and Pock (2011); Zhu and Chan (2008).

These considerations led to the widespread adoption of first-order primal–dual algorithms in imaging, where the per-iteration cost can be made matrix-free and dominated by applications of the operator and its adjoint Pock et al. (2009); Chambolle and Pock (2011). Pock, Cremers, Bischof, and Chambolle illustrated the strength of primal–dual schemes in challenging imaging energies such as formulations related to the Mumford–Shah functional, showing how saddle-point updates can be used in segmentation and related problems Pock et al. (2009). Chambolle and Pock then presented a general primal–dual hybrid gradient (PDHG) method for composite convex problems, which directly covers ROF TV denoising and many other imaging models Chambolle and Pock (2011). In parallel, Zhu and Chan developed an efficient PDHG-type framework for TV image restoration and emphasized the practical benefits of explicit, low-memory primal–dual iterations for large-scale imaging Zhu and Chan (2008). A persistent practical issue in these methods is that convergence conditions impose restrictions on the

primal and dual stepsizes through the linear operator. This has motivated several strategies for reducing the burden of global parameter selection, including diagonal preconditioning and adaptive stepsize procedures Pock and Chambolle (2011); Goldstein et al. (2015).

The dependence of primal–dual splitting schemes on global problem constants has motivated several adaptive and linesearch strategies. Goldstein, Li, and Yuan proposed adaptive primal–dual splitting methods for statistical learning and image processing Goldstein et al. (2015), while Malitsky and Pock developed a primal–dual linesearch framework with convergence and ergodic-rate guarantees Malitsky and Pock (2018). At the same time, the linear-composite regularization structure that appears in TV denoising also arises in other structured convex models beyond imaging, such as the elastic net and the fused lasso in statistics and signal processing Zou and Hastie (2005); Tibshirani et al. (2005). The fused lasso, in particular, introduces an $\ell_1$ penalty on successive differences, which can be viewed as a one-dimensional analogue of total variation regularization Tibshirani et al. (2005). In compressive sensing, related non-smooth composite formulations such as $\ell_1$–$\ell_1$ minimization also motivate primal–dual and splitting algorithms that combine proximal operations with linear operators Xiao et al. (2013). Together, these developments underline that parameter-robust primal–dual algorithms are valuable not only for TV denoising, but for a broad class of structured convex optimization problems Condat (2013); Goldstein et al. (2015).

Golden-ratio ideas have also led to a distinct family of first-order methods for variational inequalities and structured convex optimization Malitsky (2020); Chang and Yang (2021). Chang and Yang Chang and Yang (2021) developed a golden-ratio primal–dual algorithm for structured convex optimization, in which a trajectory-based convex combination governed by the parameter yields a convergent primal–dual iteration with favorable theoretical properties Chang and Yang (2021). Building on this line, Soe, Tam, and Vetrivel developed an extended golden ratio primal–dual algorithm for convex–concave saddle-point problems and introduced new stepsize rules that remove the need to know global problem constants such as operator norm and avoid backtracking procedures, while preserving convergence guarantees Soe et al. (2026). Motivated by the operator-norm dependence of classical PDHG in TV denoising, the goal of this paper is to specialize the extended golden ratio framework (with the non-increasing adaptive stepsize Rule 1) to the ROF TV denoising problem, thereby obtaining a norm-free TV denoising solver that remains fully proximal and matrix-free Chambolle and Pock (2011); Soe et al. (2026).

The main contribution of this paper is a systematic specialization and computational assessment of adaptive and accelerated golden-ratio primal–dual frameworks for the discrete ROF TV-denoising problem. In particular, we derive explicit matrix-free ROF implementations of the corresponding primal–dual iterations, including the proximal mappings and the equivalence between the auxiliary-variable dual update and the $F^*$-proximal formulation. We then specialize the norm-free Rule 1 E-GRPDA framework of Soe, Tam, and Vetrivel Soe et al. (2026) to ROF denoising, yielding an implementation whose adaptive stepsize selection does not require prior estimation of the discrete gradient operator norm. Exploiting the 1-strong convexity of the quadratic ROF fidelity term, we further specialize the accelerated golden-ratio primal–dual framework of Soe and Vetrivel Soe and Vetrivel (2026) to the ROF model. The resulting specialization inherits the $O(1/N^2)$ ergodic convergence rate established for the weighted quantities covered by the corresponding convergence theorem. Finally, we provide a controlled numerical comparison of PDHG, norm-free Rule 1 E-GRPDA, and the accelerated specialization in terms of objective convergence, computational cost, and reconstruction quality.

The remainder of the paper is organized as follows. Section 2 introduces the discrete ROF TV-denoising model and establishes the notation. Section 3 summarizes the convex-analytic ingredients, derives the saddle-point formulation, and gives the proximal mappings used by the algorithms. Section 4 presents PDHG, the norm-free E-GRPDA Rule 1 specialization, and the accelerated ROF specialization, together with their relevant convergence properties. Section 5 presents the numerical experiments and results.

## 2 Problem Formulation

We consider a discrete grayscale image $u \in \mathcal{X}$, where $\mathcal{X} \cong \mathbb{R}^N$ represents pixel values on a 2D grid with $N$ pixels, and let $f \in \mathcal{X}$ denote the observed noisy image. The discrete ROF TV denoising problem is written as

$$\min_{u\in\mathcal{X}} \lambda\|Ku\|_{2,1} + \frac{1}{2}\|u-f\|_2^2, \tag{1}$$

where $\lambda > 0$ is the regularization parameter and $K : \mathcal{X} \to \mathcal{Y}$ is the discrete gradient operator Rudin et al. (1992); Zhu and Chan (2008); Chambolle and Pock (2011). Here $\mathcal{Y} \cong \mathbb{R}^{2N}$ stores horizontal and vertical components of the discrete gradient Zhu and Chan (2008). For an image indexed by $(i, j)$, a typical choice is a forward-difference gradient

$$(Ku)_{i,j} = \big((D_x u)_{i,j}, (D_y u)_{i,j}\big), \qquad (D_x u)_{i,j} = u_{i+1,j} - u_{i,j}, \tag{2}$$

with boundary conventions (e.g., Neumann or periodic) chosen to define $D_x, D_y$ on the boundary Zhu and Chan (2008); Chambolle and Pock (2011).

The isotropic discrete TV seminorm $\|Ku\|_{2,1}$ is defined as

$$\|Ku\|_{2,1} = \sum_{p=1}^{N} \|(Ku)_p\|_2 = \sum_{i,j} \sqrt{(D_x u)_{i,j}^2 + (D_y u)_{i,j}^2}, \tag{3}$$

where $(Ku)_p \in \mathbb{R}^2$ denotes the local gradient vector at pixel $p$ Rudin et al. (1992); Zhu and Chan (2008). For compactness, we define $G(u) := \frac{1}{2}\|u-f\|_2^2, F(z) := \lambda\|z\|_{2,1}, z \in \mathcal{Y}$, so that (1) becomes the structured convex minimization problem

$$\min_{u\in\mathcal{X}} G(u) + F(Ku), \tag{4}$$

which is the canonical form targeted by primal–dual splitting methods for imaging Chambolle and Pock (2011); Condat (2013).

## 3 Preliminaries

In this section, we introduce the standard convex analysis tools required to derive saddle-point formulation and the primal-dual algorithm. For the basic concepts of these tools, we refer to Bauschke and Combettes (2017); Peypouquet (2015).

**Definition 3.1** Bauschke and Combettes (2017); Peypouquet (2015) For a proper convex function $F : \mathcal{Y} \to (-\infty, +\infty]$, its conjugate $F^* : \mathcal{Y} \to (-\infty, +\infty]$ is $F^*(y) := \sup_{z\in\mathcal{Y}} \{\langle y, z\rangle - F(z)\}$.

**Definition 3.2** Bauschke and Combettes (2017); Peypouquet (2015) For a closed convex set $C \subseteq \mathcal{Y}$, the indicator function $\iota_C$ is

$$\iota_C(y) := \begin{cases} 0, & y \in C, \\ +\infty, & y \notin C. \end{cases}$$

Indicator functions represent constraints as convex penalties and arise naturally as conjugates of norms Bertsekas (2009); Mordukhovich and Nam (2014).

**Definition 3.3** Bauschke and Combettes (2017); Peypouquet (2015) For a proper convex function $H$ and stepsize $\gamma > 0$, the proximal map is

$$\operatorname{prox}_{\gamma H}(v) := \operatorname*{argmin}_x \left\{ H(x) + \frac{1}{2\gamma}\|x-v\|_2^2 \right\}.$$

Proximal splitting and primal–dual algorithms are typically designed so that each update reduces to evaluating proximal maps of constituent functions Combettes and Wajs (2005); Condat (2013).

### 3.1 Fenchel–Rockafellar saddle-point reformulation of TV denoising

A standard saddle-point representation of (4) is obtained by inserting the conjugate representation $F(Ku) = \sup_{y\in\mathcal{Y}} \{\langle Ku, y\rangle - F^*(y)\}$, which yields the convex–concave problem

$$\min_{u\in\mathcal{X}} \max_{y\in\mathcal{Y}} \langle Ku, y\rangle + G(u) - F^*(y). \tag{5}$$

This is the canonical form used in PDHG and related primal–dual schemes Chambolle and Pock (2011); Condat (2013).

To specialize (5) to TV denoising, we compute the conjugate of $F(z) = \lambda\|z\|_{2,1}$. The dual norm of $\|\cdot\|_{2,1}$ is $\|\cdot\|_{2,\infty}$, defined by

$$\|y\|_{2,\infty} := \max_{p=1,\dots,N} \|y_p\|_2, \qquad y_p \in \mathbb{R}^2. \tag{6}$$

Using Hölder's inequality and the standard conjugacy relation between a scaled norm and the indicator function of the corresponding dual-norm ball, we obtain

$$F^*(y) = \iota_{\{y:\ \|y\|_{2,\infty}\le\lambda\}}(y), \tag{7}$$

meaning the dual variable is constrained pixelwise to lie in an $\ell_2$ ball of radius $\lambda$ Bertsekas (2009); Mordukhovich and Nam (2014).

Substituting (7) into (5) yields the ROF saddle-point formulation

$$\min_{u\in\mathcal{X}} \max_{\|y\|_{2,\infty}\le\lambda} \langle Ku, y\rangle + \frac{1}{2}\|u-f\|_2^2, \qquad (8)$$

which makes the TV term tractable through simple proximal/projection operations on the dual variable Chambolle and Pock (2011); Zhu and Chan (2008).

### 3.2 Proximal mappings for ROF TV denoising

Both PDHG and the E-GRPDA specialization require evaluating proximal maps of $G$ and either $F^*$ or $F$ Chambolle and Pock (2011); Soe et al. (2026). The proximal map of the quadratic fidelity term is derived as follows: For $G(u) = \frac{1}{2}\|u-f\|_2^2$, the proximal map is $\mathrm{prox}_{\tau G}(v) = \mathrm{argmin}_u \left\{\frac{1}{2}\|u-f\|_2^2 + \frac{1}{2\tau}\|u-v\|_2^2\right\}$. Setting the gradient to zero gives $(u-f) + \frac{1}{\tau}(u-v) = 0 \implies u = \frac{v+\tau f}{1+\tau}$, so that $\mathrm{prox}_{\tau G}(v) = \frac{v+\tau f}{1+\tau}$. Closed-form proximal updates of this type are a key reason first-order primal–dual methods are attractive for ROF denoising Chambolle and Pock (2011); Zhu and Chan (2008).

For the dual TV term $F^*$, it follows from (7) that $F^*(y) = \iota_C(y)$, where $C = \{y : \|y\|_{2,\infty} \le \lambda\}$. Hence, $\mathrm{prox}_{\sigma F^*}(q) = \mathrm{Proj}_C(q)$, the Euclidean projection of $q$ onto $C$ Chambolle and Pock (2011). Since $C$ is separable across pixels, the projection acts pointwise, $(\mathrm{Proj}_C(q))_p = \frac{q_p}{\max\{1, \|q_p\|_2/\lambda\}}$ for each pixel $p$. In Chambolle and Pock (2011); Zhu and Chan (2008) this projection form is standard in PDHG implementations of isotropic TV denoising.

The proximal map of the primal TV term $F$ is: For $F(z) = \lambda\|z\|_{2,1}$, the proximal operator $\mathrm{prox}_{\alpha F}(v)$ is separable across pixels and corresponds to vector shrinkage: $(\mathrm{prox}_{\alpha F}(v))_p = \max\left\{1 - \frac{\alpha\lambda}{\|v_p\|_2}, 0\right\} v_p$. This block soft-thresholding rule is standard for $\ell_2$-block norms and is used widely in proximal splitting and primal–dual schemes for TV and related regularizers Combettes and Wajs (2005); Condat (2013).

The saddle-point formulation (8) and the explicit proximal mappings derived above provide the ingredients needed to construct implementable primal–dual algorithms for ROF denoising; see Chambolle and Pock (2011); Soe et al. (2026) and the references therein. These algorithms are presented next.

## 4 Primal–Dual Algorithms for TV Denoising

This section presents the baseline PDHG/Chambolle–Pock algorithm, the norm-free E-GRPDA Rule 1 specialization, and an accelerated golden-ratio primal–dual specialization for ROF TV denoising Chambolle and Pock (2011); Soe et al. (2026); Soe and Vetrivel (2026). All three schemes arise from closely related primal–dual formulations, but they differ in their stepsize requirements and theoretical convergence properties. In particular, E-GRPDA Rule 1 avoids prior estimation of $\|K\|$, whereas the accelerated scheme exploits the strong convexity of the ROF fidelity term to obtain an improved ergodic convergence rate Chambolle and Pock (2011); Soe et al. (2026); Soe and Vetrivel (2026). We emphasize that the rate statements below concern the specific ergodic quantities appearing in the cited convergence theorems. They do not imply that the raw objective value, PSNR, or pointwise iterate error of PF-AGRPDA must decrease faster than those of PDHG at every iteration or for every problem instance.

### 4.1 Baseline: Chambolle–Pock PDHG for ROF TV denoising

Starting from (8), PDHG alternates a proximal (projection) step on the dual variable with a proximal step on the primal variable, plus an extrapolation step Chambolle and Pock (2011). A standard PDHG iteration is

$$\begin{aligned} y^{k+1} &= \mathrm{prox}_{\sigma F^*}\left(y^k + \sigma K\bar{u}^k\right), \\ u^{k+1} &= \mathrm{prox}_{\tau G}\left(u^k - \tau K^* y^{k+1}\right), \\ \bar{u}^{k+1} &= u^{k+1} + \theta(u^{k+1} - u^k), \end{aligned} \qquad (9)$$

where $\tau, \sigma > 0$ and $\theta \in [0,1]$. Convergence requires a stepsize restriction involving the operator norm

$\|K\|$, commonly stated as

$$\tau\sigma\|K\|^2 < 1, \tag{10}$$

which motivates norm-free alternatives when $\|K\|$ is inconvenient to compute or estimate Chambolle and Pock (2011); Goldstein et al. (2015).

Specializing (9) using (3.2), we obtain explicit TV-denoising updates:

$$\begin{aligned} y_p^{k+1} &= \frac{(y^k + \sigma K\bar{u}^k)_p}{\max\{1, \|(y^k + \sigma K\bar{u}^k)_p\|_2/\lambda\}}, \\ u^{k+1} &= \frac{u^k - \tau K^* y^{k+1} + \tau f}{1+\tau}, \\ \bar{u}^{k+1} &= u^{k+1} + \theta(u^{k+1} - u^k). \end{aligned} \tag{11}$$

These iterations are matrix-free and require only applications of $K$ and $K^*$ plus pointwise operations.

## 4.2 Norm-free E-GRPDA specialization with Rule 1

Chang and Yang Chang and Yang (2021) introduced a golden ratio primal–dual algorithm for structured convex optimization, where a trajectory-based convex combination governed by a parameter $\psi$ yields a convergent primal–dual iteration. Soe, Tam, and Vetrivel Soe et al. (2026) extended this approach to convex–concave saddle-point problems and proposed stepsize rules that avoid explicit dependence on global problem constants such as $\|K\|$ . In particular, Rule 1 defines a non-increasing adaptive primal stepsize based on local iterate differences and corresponding operator differences, thereby eliminating the need to compute $\|K\|$ and avoiding backtracking Soe et al. (2026).

To apply this framework to TV denoising, we rewrite (4) by introducing an auxiliary variable $w = Ku$:

$$\min_{u\in\mathcal{X},\, w\in\mathcal{Y}} \; G(u) + \lambda\|w\|_{2,1} \quad \text{s.t.} \quad Ku - w = 0, \tag{12}$$

which matches the extended golden ratio primal–dual structure used in Soe et al. (2026). The resulting iteration alternates a proximal step for $u$, a proximal step for $w$ (block shrinkage), and a dual ascent step on the constraint residual, together with a golden-ratio extrapolation Soe et al. (2026); Condat (2013).

**Proposition 4.1** (Equivalence of the dual updates)**.** *Let $F(w) = \lambda\|w\|_{2,1}$, and let $\sigma_n > 0$. Define $w^n = \operatorname{prox}_{F/\sigma_n}\left(\frac{y^{n-1}}{\sigma_n} + Ku^n\right)$. Then the update $y^n = y^{n-1} + \sigma_n(Ku^n - w^n)$ is equivalent to $y^n = \operatorname{prox}_{\sigma_n F^*}\left(y^{n-1} + \sigma_n Ku^n\right)$.*

*Proof* By Moreau's decomposition Combettes and Wajs (2005), $\operatorname{prox}_{\sigma_n F^*}(q) = q - \sigma_n \operatorname{prox}_{F/\sigma_n}\left(\frac{q}{\sigma_n}\right)$. Taking $q = y^{n-1} + \sigma_n Ku^n$ gives $\frac{q}{\sigma_n} = \frac{y^{n-1}}{\sigma_n} + Ku^n$. Hence,

$$\operatorname{prox}_{\sigma_n F^*}\left(y^{n-1} + \sigma_n Ku^n\right) = y^{n-1} + \sigma_n Ku^n - \sigma_n \operatorname{prox}_{F/\sigma_n}\left(\frac{y^{n-1}}{\sigma_n} + Ku^n\right).$$

Using the definition of $w^n$, we obtain $\operatorname{prox}_{\sigma_n F^*}\left(y^{n-1} + \sigma_n Ku^n\right) = y^{n-1} + \sigma_n(Ku^n - w^n)$. Therefore, the two dual updates are identical. □

**Algorithm 1** E-GRPDA (Rule 1) specialized to ROF TV denoising

**Input:** Noisy image $f$, regularization parameter $\lambda > 0$, $\psi \in (1, \phi]$, where $\phi = \frac{1+\sqrt{5}}{2}$, $\beta > 0$, parameters $0 < 2\mu' < \mu < \frac{\psi}{2}$, initial points $u^0 \in X$, $y^0 \in Y$, and $\tau_0 > 0$.

1: Set $z^0 = u^0$.
2: **for** $n = 1, 2, \ldots$ **do**
3: $\quad z^n = \frac{\psi-1}{\psi}u^{n-1} + \frac{1}{\psi}z^{n-1}$.
4: $\quad u^n = \operatorname{prox}_{\tau_{n-1}G}\left(z^n - \tau_{n-1}K^*y^{n-1}\right)$.
5: $\quad$ Using $G(u) = \frac{1}{2}\|u - f\|_2^2$, compute $u^n = \frac{z^n - \tau_{n-1}K^*y^{n-1} + \tau_{n-1}f}{1+\tau_{n-1}}$.
6: $\quad$ Update the Rule 1 primal stepsize: $\tau_n = \min\left\{\tau_{n-1}, \frac{\mu\|u^n - u^{n-1}\|_2}{\sqrt{\beta}\|Ku^n - Ku^{n-1}\|_2}\right\}$.
7: $\quad$ Set $\sigma_n = \beta\tau_n$.
8: $\quad$ Compute $w^n = \operatorname{prox}_{F/\sigma_n}\left(\frac{y^{n-1}}{\sigma_n} + Ku^n\right)$, $F(w) = \lambda\|w\|_{2,1}$.
9: $\quad$ Update $y^n = y^{n-1} + \sigma_n(Ku^n - w^n)$.
10: **end for**

*Remark 4.2* If $\|Ku^n - Ku^{n-1}\|_2 = 0$, the quotient in the Rule 1 update is interpreted as $+\infty$, including

the case $u^n = u^{n-1}$. Hence $\tau_n = \tau_{n-1}$ whenever the denominator vanishes.

Algorithm 1 retains the matrix-free structure of primal–dual methods: each iteration requires only applications of $K$ and $K^*$ together with pointwise proximal operations. Its main distinction from the PDHG scheme considered above lies in the stepsize selection. While PDHG uses the norm-dependent condition, the Rule 1 update determines the stepsizes from local iterate and operator differences and therefore does not require prior estimation of $\|K\|$ Soe et al. (2026). This norm-free feature does not imply that E-GRPDA is uniformly faster than a well-tuned PDHG method; when $\|K\|$ is readily available, PDHG may admit more effective stepsize choices. The practical motivation for the adaptive rule is instead to remove prior operator-norm information, which is particularly relevant when the underlying operator is nonstandard or problem dependent.

## 4.3 Convergence guarantees and supporting results

Because both PDHG (9) and E-GRPDA (Algorithm 1) arise from primal–dual formulations, their convergence can be characterized through saddle-point optimality, function-value estimates, and feasibility or primal–dual residuals Chambolle and Pock (2011); Soe et al. (2026); Nedić and Ozdaglar (2009). This perspective also clarifies why explicit stepsize conditions (such as (10)) appear in classical methods and why norm-free update rules (such as Rule 1) are practically valuable see Goldstein et al. (2015); Soe et al. (2026) and the references therein for further details.

**Theorem 4.3** (Saddle-point optimality principle)**.** *(Nedić and Ozdaglar 2009, Theorem 1). Let $\mathcal{L}(u,y) = G(u) + \langle Ku, y\rangle - F^*(y)$ be the Lagrangian associated with the primal ROF problem $\min_{u\in X}\{G(u)+F(Ku)\}$, where $G(u) = \frac{1}{2}\|u-f\|_2^2$, $F(z) = \lambda\|z\|_{2,1}$. If $(u^\star, y^\star)$ is a saddle point of $\mathcal{L}$, then $u^\star$ solves the primal ROF problem, $y^\star$ solves the corresponding dual problem, and the primal and dual optimal values coincide.*

*Proof* Since $(u^\star, y^\star)$ is a saddle point, $\mathcal{L}(u^\star, y) \le \mathcal{L}(u^\star, y^\star) \le \mathcal{L}(u, y^\star)$, $\forall\, u \in X$, $y \in Y$. By Fenchel conjugacy, $G(u) + F(Ku) = \sup_{y\in Y} \mathcal{L}(u,y)$. Hence, $G(u^\star) + F(Ku^\star) = \mathcal{L}(u^\star, y^\star) \le G(u) + F(Ku)$, $\forall\, u \in X$, which proves that $u^\star$ is primal optimal. Similarly, defining the dual objective by $d(y) := \inf_{u\in X} \mathcal{L}(u,y)$, the saddle inequalities give $d(y) \le \mathcal{L}(u^\star, y^\star) = d(y^\star)$, $\forall\, y \in Y$. Thus $y^\star$ is dual optimal. Consequently, $\min_{u\in X}\{G(u)+F(Ku)\} = \mathcal{L}(u^\star, y^\star) = \max_{y\in Y} d(y)$, so the primal and dual optimal values are equal. □

Theorem 4.3 motivates measuring progress via a saddle-point gap of the form $\mathcal{L}(\bar{u}, y^\star) - \mathcal{L}(u^\star, \bar{y})$, since vanishing gap certifies approach to a saddle point Nedić and Ozdaglar (2009); Chambolle and Pock (2011). In ROF TV denoising, strong convexity of the quadratic fidelity term further implies uniqueness of the primal minimizer, which helps interpret convergence of iterates in terms of convergence of restored images Bertsekas (2009).

**Proposition 4.4** (Uniqueness of the ROF minimizer)**.** *Let $X \simeq \mathbb{R}^N$ be the discrete image space, let $f \in X$ be the observed image, let $\lambda > 0$, and let $K : X \to Y$ denote the discrete gradient operator. Define $J(u) = \lambda\|Ku\|_{2,1} + \frac{1}{2}\|u - f\|_2^2$, $u \in X$. Then $J$ is 1-strongly convex and coercive. Consequently, the ROF problem $\min_{u\in X}\left\{\lambda\|Ku\|_{2,1} + \frac{1}{2}\|u-f\|_2^2\right\}$ admits a unique minimizer $u^\star \in X$.*

*Proof* Write $J(u) = R(u) + G(u)$, where $R(u) = \lambda\|Ku\|_{2,1}$, $G(u) = \frac{1}{2}\|u-f\|_2^2$. Since $K$ is linear and $\|\cdot\|_{2,1}$ is convex, $R$ is convex. Moreover, for any $u, v \in X$, $G(v) = G(u) + \langle u - f, v - u\rangle + \frac{1}{2}\|v-u\|_2^2$, which shows that $G$ is 1-strongly convex. Hence the sum $J = R + G$ is also 1-strongly convex. Furthermore, $J(u) \ge \frac{1}{2}\|u-f\|_2^2$, and therefore $J(u) \longrightarrow +\infty$ as $\|u\|_2 \longrightarrow +\infty$. Thus $J$ is coercive. Since $J$ is continuous on the finite-dimensional space $X$, it attains its minimum. Finally, strong convexity implies that this minimizer is unique. Hence there exists a unique $u^\star \in X$ satisfying $u^\star = \arg\min_{u\in X}\left\{\lambda\|Ku\|_{2,1} + \frac{1}{2}\|u-f\|_2^2\right\}$. □

With this saddle-point viewpoint fixed, we first summarize the convergence guarantees supporting PDHG and E-GRPDA; the accelerated PF-AGRPDA specialization is treated separately in Section 4.4.

The classical PDHG / Chambolle–Pock algorithm admits a standard convergence theorem

under a norm-based stepsize restriction Chambolle and Pock (2011). This result also provides an explicit sublinear bound on the ergodic saddle-point gap, which supplies a standard theoretical measure of convergence for primal–dual methods Chambolle and Pock (2011); Zhu and Chan (2008).

**Corollary 4.5** (ROF specialization of the Chambolle–Pock convergence result)**.** *Let* $\mathcal{L}(u,y) = G(u) + \langle Ku, y\rangle - F^*(y)$, *where* $G(u) = \frac{1}{2}\|u - f\|_2^2$, $F(z) = \lambda\|z\|_{2,1}$, *and suppose that* $(u^\star, y^\star)$ *is a saddle point of the ROF saddle-point problem* $\min_{u\in X}\max_{y\in Y}\mathcal{L}(u,y)$. *Let* $\{(u^k, y^k)\}_{k\geq 0}$ *be generated by the PDHG iteration*

$$y^{k+1} = \operatorname{prox}_{\sigma F^*}\left(y^k + \sigma K\bar{u}^k\right),$$

$$u^{k+1} = \operatorname{prox}_{\tau G}\left(u^k - \tau K^* y^{k+1}\right),$$
$$\bar{u}^{k+1} = u^{k+1} + (u^{k+1} - u^k),$$

*corresponding to the standard choice* $\theta = 1$. *If* $\tau\sigma\|K\|^2 < 1$, *then the sequence* $(u^k, y^k)$ *converges to a saddle point of* $\mathcal{L}$. *Moreover, defining the ergodic averages* $U^N := \frac{1}{N}\sum_{k=1}^N u^k$, $Y^N := \frac{1}{N}\sum_{k=1}^N y^k$, *the primal–dual gap satisfies* $0 \leq \mathcal{L}(U^N, y^\star) - \mathcal{L}(u^\star, Y^N) \leq \frac{1}{N}\left(\frac{\|u^0-u^\star\|_2^2}{2\tau} + \frac{\|y^0-y^\star\|_2^2}{2\sigma}\right)$. *In particular,* $\mathcal{L}(U^N, y^\star) - \mathcal{L}(u^\star, Y^N) = O\left(\frac{1}{N}\right)$.

*Proof* For the ROF model, $G(u) = \frac{1}{2}\|u - f\|_2^2$ is proper, continuous, convex, and strongly convex, while $F(z) = \lambda\|z\|_{2,1}$ is proper, convex, and lower semicontinuous. Since the discrete gradient $K : X \to Y$ is linear on finite-dimensional spaces, it is bounded. Furthermore, the proximal mappings required by PDHG are well defined and are given by $\operatorname{prox}_{\tau G}(v) = \frac{v+\tau f}{1+\tau}$, and $\operatorname{prox}_{\sigma F^*}(q) = \operatorname{Proj}_{\{\|y\|_{2,\infty}\leq\lambda\}}(q)$. Hence, the ROF saddle-point problem satisfies the assumptions of the Chambolle–Pock convergence theorem (Chambolle and Pock 2011, Theorem 1). Under $\tau\sigma\|K\|^2 < 1$ and the standard choice $\theta = 1$, that result gives $(u^k, y^k) \longrightarrow (u^\star, y^\star)$, where $(u^\star, y^\star)$ is a saddle point of $\mathcal{L}$. It also yields $\mathcal{L}(U^N, y^\star) - \mathcal{L}(u^\star, Y^N) \leq \frac{1}{N}\left(\frac{\|u^0-u^\star\|_2^2}{2\tau} + \frac{\|y^0-y^\star\|_2^2}{2\sigma}\right)$. Since $(u^\star, y^\star)$ is a saddle point, $\mathcal{L}(u^\star, Y^N) \leq \mathcal{L}(u^\star, y^\star) \leq \mathcal{L}(U^N, y^\star)$, and therefore $0 \leq \mathcal{L}(U^N, y^\star) - \mathcal{L}(u^\star, Y^N)$. Thus, $0 \leq \mathcal{L}(U^N, y^\star) - \mathcal{L}(u^\star, Y^N) \leq \frac{C}{N}$, where $C = \frac{\|u^0-u^\star\|_2^2}{2\tau} + \frac{\|y^0-y^\star\|_2^2}{2\sigma}$. Hence, $\mathcal{L}(U^N, y^\star) - \mathcal{L}(u^\star, Y^N) = O\left(\frac{1}{N}\right)$. Finally, by Proposition 4.4, the primal ROF minimizer is unique, so the primal component converges to the unique ROF solution $u^\star$. □

Corollary 4.5 formalizes the practical observation that PDHG can be highly efficient for TV denoising when $\|K\|$ is known or can be estimated reliably, since the bound (10) directly determines admissible stepsizes Chambolle and Pock (2011); Zhu and Chan (2008). At the same time, the explicit appearance of $\|K\|$ is the main obstacle in settings where the operator norm is inconvenient to compute, which motivates norm-free stepsize strategies Goldstein et al. (2015); Soe et al. (2026).

*Remark 4.6* A closely related primal–dual splitting method for problems involving proximable terms and linear compositions also requires a stepsize condition based on the operator norm of the linear map, and in finite-dimensional settings the condition can be relaxed to allow equality in a norm-type bound. For further details, we refer the reader to Condat (2013) and the references therein. This justifies the above general statement, which also holds true for many primal–dual schemes, although for a variety of different reasons depending on the algorithmic details (see, for example, Condat (2013); Chambolle and Pock (2011)).

To derive convergent primal–dual iterations, golden ratio primal–dual methods Chang and Yang (2021) were proposed for trajectory-based extrapolation for solving the problem using $\psi$. Under the paper's assumptions Chang and Yang (2021), convergence to a saddle point follows in the original GRPDA. Collection of stepsize rules such as Rule 1 is then added to the extended framework of Soe, Tam, and Vetrivel Soe et al. (2026) the rules still ensure convergence without being aware of $\|K\|$ explicitly.

**Corollary 4.7** (ROF specialization of the GRPDA convergence result)**.** *Chang and Yang (2021) Let* $\mathcal{L}(u,y) = G(u) + \langle Ku, y\rangle - F^*(y)$, *where* $G(u) = \frac{1}{2}\|u - f\|_2^2$, $F(z) = \lambda\|z\|_{2,1}$, *and suppose that the ROF saddle-point problem* $\min_{u\in X}\max_{y\in Y}\mathcal{L}(u,y)$ *admits a saddle point* $(u^\star, y^\star)$. *Let* $\phi = \frac{1+\sqrt{5}}{2}$, $1 < \psi \leq \phi$, *and choose* $\tau, \sigma > 0$ *such that* $\tau\sigma\|K\|^2 < \psi$. *Starting from* $u^0 \in X$, $y^0 \in Y$, *and* $z^0 = u^0$, *consider the GRPDA iteration* $z^n = \frac{\psi-1}{\psi}u^{n-1} + \frac{1}{\psi}z^{n-1}$, $u^n = \operatorname{prox}_{\tau G}\left(z^n - \tau K^* y^{n-1}\right)$, $y^n =$

$\mathrm{prox}_{\sigma F^*}\left(y^{n-1}+\sigma K u^n\right)$. *Then* $(u^n, y^n)$ *converges to a saddle point of* $\mathcal{L}$. *In particular, since the ROF primal minimizer is unique,* $u^n \longrightarrow u^\star$, *where* $u^\star$ *is the unique solution of the ROF denoising problem.*

*Proof* For the ROF problem, $G(u) = \frac{1}{2}\|u-f\|_2^2$ is proper, convex, and continuous, and $F(z) = \lambda\|z\|_{2,1}$ is proper, convex, and lower semicontinuous. Moreover, the discrete gradient $K : X \to Y$ is a bounded linear operator. The required proximal mappings are well defined and satisfy $\mathrm{prox}_{\tau G}(v) = \frac{v+\tau f}{1+\tau}$, and $\mathrm{prox}_{\sigma F^*}(q) = \mathrm{Proj}_{\{\|y\|_{2,\infty}\le\lambda\}}(q)$. Thus, the ROF saddle-point formulation satisfies the assumptions of the GRPDA convergence result in Chang and Yang (2021). Under $1 < \psi \le \phi$, $\tau\sigma\|K\|^2 < \psi$, the GRPDA sequence therefore satisfies $(u^n, y^n) \longrightarrow (u^\star, y^\star)$, for some saddle point $(u^\star, y^\star)$ of $\mathcal{L}$. The saddle-point optimality conditions give $0 \in \partial G(u^\star) + K^* y^\star$, $0 \in \partial F^*(y^\star) - Ku^\star$, so $u^\star$ solves the primal ROF problem. Finally, by Proposition 4.4, the ROF objective is a strongly convex and hence has a unique minimizer. Therefore, $u^n \longrightarrow u^\star$, where $u^\star$ is the unique ROF denoised solution. □

In the extended framework, the first key technical ingredient is that the Rule 1 stepsize sequence is non-increasing but remains bounded away from zero, which prevents the algorithm from stalling due to vanishing stepsizes see Soe et al. (2026).

**Proposition 4.8** (Stepsize bound for the ROF specialization of Rule 1)**.** *Let* $\{\tau_n\}$ *be generated by the ROF specialization of Rule 1,* $\tau_n = \min\left\{\tau_{n-1}, \frac{\mu\|u^n-u^{n-1}\|_2}{\sqrt{\beta}\,\|Ku^n-Ku^{n-1}\|_2}\right\}$, *where* $\tau_0 > 0$, $\mu > 0$, *and* $\beta > 0$. *The quotient is interpreted as* $+\infty$ *whenever its denominator vanishes. Then* $\{\tau_n\}$ *is non-increasing and satisfies* $\tau_n \ge \eta := \min\left\{\tau_0, \frac{\mu}{\sqrt{\beta}\,\|K\|}\right\} > 0$. *Consequently,* $\{\tau_n\}$ *converges to a limit* $\tau_\infty > 0$ *satisfying* $\tau_\infty = \lim_{n\to\infty}\tau_n \ge \eta$.

*Proof* Set $d_n := u^n - u^{n-1}$. For $Kd_n \neq 0$, define $r_n := \frac{\mu\|d_n\|_2}{\sqrt{\beta}\,\|Kd_n\|_2}$. Then $\tau_n = \min\{\tau_{n-1}, r_n\}$, and therefore $\tau_n \le \tau_{n-1}$. Hence $\{\tau_n\}$ is non-increasing. Since $K$ is bounded, $\|Kd_n\|_2 \le \|K\|\,\|d_n\|_2$. Thus, whenever $Kd_n \neq 0$, $r_n = \frac{\mu\|d_n\|_2}{\sqrt{\beta}\,\|Kd_n\|_2} \ge \frac{\mu}{\sqrt{\beta}\,\|K\|}$. If $Kd_n = 0$, the quotient is interpreted as $+\infty$, so the same lower estimate is valid. Hence $\tau_n \ge \min\left\{\tau_{n-1}, \frac{\mu}{\sqrt{\beta}\,\|K\|}\right\}$. Iterating this inequality yields $\tau_n \ge \min\left\{\tau_0, \frac{\mu}{\sqrt{\beta}\,\|K\|}\right\} = \eta > 0$. Therefore, $\eta \le \tau_n \le \tau_{n-1}$. Thus $\{\tau_n\}$ is monotone and bounded below, and consequently there exists $\tau_\infty$ such that $\tau_n \longrightarrow \tau_\infty$, $\tau_\infty \ge \eta > 0$. □

We emphasize that the appearance of $\|K\|$ in this theoretical lower bound does not require its computation by the algorithm; the Rule 1 stepsize update itself uses only local iterate and operator differences.Proposition 4.8 verifies the positive lower-bound property required by Rule 1 in the ROF setting. The general convergence result of Soe, Tam, and Vetrivel (Soe et al. 2026, Theorem 3.1) can therefore be specialized to the present model. In this specialization, the smooth term $h$ is absent, $h \equiv 0$, $\nabla h \equiv 0$, while the nonsmooth quadratic fidelity term is represented by $G$. The resulting ROF convergence statement is given in the following corollary.

**Corollary 4.9** (Convergence of the ROF specialization of Rule 1)**.** *Let* $G(u) = \frac{1}{2}\|u-f\|_2^2$, $F(z) = \lambda\|z\|_{2,1}$, *and consider the ROF saddle-point problem* $\min_{u\in X}\max_{y\in Y}\mathcal{L}(u,y)$, $\mathcal{L}(u,y) = G(u) + \langle Ku, y\rangle - F^*(y)$. *Assume that the saddle-point set is nonempty, and let the iterates be generated by Algorithm 1. Suppose that* $1 < \psi \le \phi = \frac{1+\sqrt{5}}{2}$, $\beta > 0$, $0 < 2\mu' < \mu < \frac{\psi}{2}$. *Then the primal–dual sequence converges to a saddle point* $(u^\star, y^\star)$ *of* $\mathcal{L}$. *In particular,* $u^n \longrightarrow u^\star$, *where* $u^\star$ *is the unique minimizer of the ROF problem.*

*Proof* For the ROF specialization, $G(u) = \frac{1}{2}\|u-f\|_2^2$ is proper, convex, continuous, and proximable, while $F(z) = \lambda\|z\|_{2,1}$ is proper, convex, lower semicontinuous, and proximable. Moreover, the discrete gradient $K : X \to Y$ is a bounded linear operator.

The additional smooth term in the general E-GRPDA model is absent here; hence $h \equiv 0$, $\nabla h \equiv 0$. Accordingly, Rule 1 reduces to $\tau_n = \min\left\{\tau_{n-1}, \frac{\mu\|u^n-u^{n-1}\|_2}{\sqrt{\beta}\,\|Ku^n-Ku^{n-1}\|_2}\right\}$. By Proposition 4.8, $\tau_n \ge \eta > 0$, $\tau_n \longrightarrow \tau_\infty > 0$. Since $\sigma_n = \beta\tau_n$, we also have $\sigma_n \longrightarrow \beta\tau_\infty > 0$. Furthermore, by Proposition 4.1, the auxiliary-variable implementation in Algorithm 1 is equivalent to $y^n = \mathrm{prox}_{\sigma_n F^*}\left(y^{n-1}+\sigma_n K u^n\right)$. Thus the ROF iteration is an instance of the Rule 1 E-GRPDA framework. Therefore, under $1 < \psi \le \phi$, $0 < 2\mu' < \mu < \frac{\psi}{2}$, the assumptions of (Soe et al. 2026, Theorem 3.1) are

satisfied, and the generated sequence converges to a saddle point $(u^\star, y^\star)$ of $\mathcal{L}$.

Finally, Proposition 4.4 gives uniqueness of the primal ROF minimizer. Hence the primal component satisfies $u^n \longrightarrow u^\star$. □

Beyond the parameter range used in Corollary 4.9, Soe, Tam, and Vetrivel Soe et al. (2026) establish a further extension of the admissible golden-ratio parameter under stronger conditions on the Rule 1 parameters (Soe et al. 2026, Theorem 3.2). Specializing that result to the ROF model gives the following corollary.

**Corollary 4.10** (Extended parameter range for the ROF specialization of Rule 1)**.** *Let* $G(u) = \frac{1}{2}\|u - f\|_2^2$, $F(z) = \lambda\|z\|_{2,1}$, *and consider the ROF saddle-point problem* $\min_{u\in X}\max_{y\in Y}\mathcal{L}(u,y)$, $\mathcal{L}(u,y) = G(u) + \langle Ku, y\rangle - F^*(y)$. *Assume that the saddle-point set is nonempty and let the iterates be generated by Algorithm 1. Suppose that* $1 < \psi < 1 + \sqrt{3}$, *and choose* $\mu, \mu' > 0$ *such that* $0 < 3\mu' < \mu < \frac{\psi}{2} + \frac{\psi(1+\psi-\psi^2)}{2(\psi+1)}$. *Then the E-GRPDA sequence converges to a saddle point* $(u^\star, y^\star)$ *of* $\mathcal{L}$. *Consequently,* $u^n \longrightarrow u^\star$, *where* $u^\star$ *is the unique minimizer of the ROF problem.*

*Proof* For the ROF specialization, $G(u) = \frac{1}{2}\|u - f\|_2^2$ is proper, convex, continuous, and proximable, while $F(z) = \lambda\|z\|_{2,1}$ is proper, convex, lower semicontinuous, and proximable. Moreover, the discrete gradient $K : X \to Y$ is a bounded linear operator. The smooth term in the general E-GRPDA model is absent in the present problem, so $h \equiv 0$, $\nabla h \equiv 0$. Hence Rule 1 reduces to $\tau_n = \min\left\{\tau_{n-1}, \frac{\mu\|u^n - u^{n-1}\|_2}{\sqrt{\beta}\,\|Ku^n - Ku^{n-1}\|_2}\right\}$. By Proposition 4.8, $\tau_n \geq \eta > 0$, $\tau_n \longrightarrow \tau_\infty > 0$. Therefore, with $\theta_n := \frac{\tau_n}{\tau_{n-1}}$, we have $\theta_n \longrightarrow 1$. Furthermore, the auxiliary-variable dual update in Algorithm 1 is equivalent, by Moreau's identity, to $y^n = \mathrm{prox}_{\sigma_n F^*}\left(y^{n-1} + \sigma_n K u^n\right)$. Thus Algorithm 1 is the ROF specialization of the Rule 1 E-GRPDA scheme. Under $1 < \psi < 1 + \sqrt{3}$, $0 < 3\mu' < \mu < \frac{\psi}{2} + \frac{\psi(1+\psi-\psi^2)}{2(\psi+1)}$, the assumptions of the extended convergence result (Soe et al. 2026, Theorem 3.2) are satisfied. Hence $(u^n, y^n) \longrightarrow (u^\star, y^\star)$, where $(u^\star, y^\star)$ is a saddle point of $\mathcal{L}$.

Finally, Proposition 4.4 gives uniqueness of the primal ROF minimizer. Therefore, $u^n \longrightarrow u^\star$. □

Corollary 4.10 shows that the extended parameter range established in (Soe et al. 2026, Theorem 3.2) remains applicable after specialization to the ROF model. Thus, under the stated conditions on $\mu$ and $\mu'$, the admissible range is $1 < \psi < 1 + \sqrt{3}$.

In addition to convergence of the iterates, the E-GRPDA analysis establishes an ergodic $O(1/N)$ rate for the function-value residual and the constraint violation in the associated constrained optimization formulation (Soe et al. 2026, Theorem 4.1). We next specialize this result to the constrained ROF model.

**Corollary 4.11** (Ergodic convergence rate for the ROF specialization of Rule 1)**.** *The following result specializes (Soe et al. 2026, Theorem 4.1) to the constrained ROF model. Consider the constrained ROF problem* $\min_{u\in X,\, w\in Y}\Phi(u,w) := G(u) + \lambda\|w\|_{2,1}$ *subject to* $Ku - w = 0$, *where* $G(u) = \frac{1}{2}\|u - f\|_2^2$. *Let* $(u^\star, w^\star, y^\star)$ *be a primal–dual solution, so that* $w^\star = Ku^\star$. *Let* $\{(u^n, w^n, y^n)\}$ *be generated by Algorithm 1, and define the ergodic averages* $\widetilde{u}^N = \frac{1}{N}\sum_{n=1}^N u^n$, $\widetilde{w}^N = \frac{1}{N}\sum_{n=1}^N w^n$. *Then there exist constants* $C_1, C_2 > 0$, *independent of* $N$, *such that* $\left|\Phi(\widetilde{u}^N, \widetilde{w}^N) - \Phi(u^\star, w^\star)\right| \leq \frac{C_1}{N}$, *and* $\|K\widetilde{u}^N - \widetilde{w}^N\|_2 \leq \frac{C_2}{N}$. *Hence the ergodic function-value residual and constraint violation decay at order* $O\left(\frac{1}{N}\right)$.

*Proof* The ROF problem can be written in the constrained form $\min_{u,w}\left\{G(u) + \lambda\|w\|_{2,1}\right\}$ subject to $Ku - w = 0$. Here $G(u) = \frac{1}{2}\|u - f\|_2^2$ and $w \mapsto \lambda\|w\|_{2,1}$ are proper, convex, lower semicontinuous, and proximable, while $K$ is a bounded linear operator. Moreover, the smooth term in the general E-GRPDA formulation is absent in the ROF model, so that $h \equiv 0$, $\nabla h \equiv 0$. By the Moreau equivalence established above, the auxiliary-variable updates in Algorithm 1 agree with the corresponding E-GRPDA primal–dual formulation. Hence the constrained ROF problem satisfies the assumptions of the ergodic convergence result (Soe et al. 2026, Theorem 4.1). Applying that theorem to the present specialization yields constants $C_1, C_2 > 0$, independent of $N$, such that $\left|\Phi(\widetilde{u}^N, \widetilde{w}^N) - \Phi(u^\star, w^\star)\right| \leq \frac{C_1}{N}$, and $\|K\widetilde{u}^N - \widetilde{w}^N\|_2 \leq \frac{C_2}{N}$. Therefore, both the function-value residual and the feasibility residual converge with ergodic order $O\left(\frac{1}{N}\right)$. □

*Remark 4.12* Adaptive primal–dual splitting schemes can also attain sublinear ergodic convergence under suitable adaptive stepsize conditions; see, for example, (Goldstein et al. 2015, Theorem 1). In the E-GRPDA framework considered here, Rule 1 updates the stepsizes using local iterate and operator differences and does not require prior knowledge of the operator norm $\|K\|$ or a backtracking procedure Soe et al. (2026). This norm-free stepsize selection is the main practical distinction emphasized in the present ROF specialization.

Together, Corollaries 4.5 and 4.9, supported by the stepsize property in Proposition 4.8, provide the theoretical basis for the PDHG and E-GRPDA implementations considered in this work Chambolle and Pock (2011); Soe et al. (2026). The two algorithms differ primarily in their stepsize requirements: PDHG uses a stability condition involving $\|K\|$, whereas Rule 1 of E-GRPDA determines its stepsizes from local iterate information without requiring an operator-norm estimate. Section 4.4 introduces the third method, PF-AGRPDA, which exploits strong convexity to obtain an accelerated ergodic rate under the assumptions of the corresponding convergence theorem. The numerical experiments therefore compare these three distinct algorithmic regimes under a common ROF denoising protocol.

## 4.4 Accelerated ROF specialization under strong convexity

The quadratic fidelity term of the ROF model, $G(u) = \frac{1}{2}\|u - f\|_2^2$, is 1-strongly convex, as established in Proposition 4.4. This additional structure can be exploited to obtain improved theoretical convergence guarantees for suitable golden-ratio primal–dual schemes. Chang and Yang Chang and Yang (2021) established an accelerated version of GRPDA under strong convexity, while the recent adaptive accelerated framework of Soe and Vetrivel Soe and Vetrivel (2026) provides an ergodic $O(1/N^2)$ convergence rate under the corresponding assumptions. For the ROF model, the accelerated framework can be specialized by taking $G(u) = \frac{1}{2}\|u - f\|_2^2$, $F(w) = \lambda\|w\|_{2,1}$, $h \equiv 0$. Hence the strong-convexity modulus of the primal fidelity term is equal to one, while the local smoothness quantity associated with $h$ vanishes. We therefore consider the following ROF specialization of the accelerated method of Soe and Vetrivel Soe and Vetrivel (2026).

**Algorithm 2** PF-AGRPDA specialized to ROF TV denoising

**Input:** Noisy image $f$, regularization parameter $\lambda > 0$, $\beta_0 > 0$, $\psi \in (\psi_0, \phi)$, and $0 < \alpha \le 1/3$, where $\psi_0 \approx 1.32472$, $\psi_0^3 - \psi_0 - 1 = 0$, $\phi = \frac{1+\sqrt{5}}{2}$. Choose $u^0 \in \mathcal{X}$, $y^0 \in \mathcal{Y}$, and set $z^0 = u^0$. Define $\rho = \psi^{-1} + \psi^{-2}$, $\theta_0 = 1$, $C = \psi\|K\|^2$, $\Theta = \max\{\theta_0, \psi\rho\}$, $U = \sqrt{\frac{\rho\psi\alpha^2\Theta}{C\beta_0}}$, and choose $0 < \tau_0 \le U$.

1: **for** $n = 1, 2, \ldots$ **do**
2: $z^n = \frac{\psi-1}{\psi}u^{n-1} + \frac{1}{\psi}z^{n-1}$.
3: $u^n = \mathrm{prox}_{\tau_{n-1}G}\left(z^n - \tau_{n-1}K^*y^{n-1}\right)$.
4: Using $G(u) = \frac{1}{2}\|u - f\|_2^2$, compute $u^n = \frac{z^n - \tau_{n-1}K^*y^{n-1} + \tau_{n-1}f}{1+\tau_{n-1}}$.
5: Compute $\zeta_n = \frac{\psi-\rho}{\psi+\rho\tau_{n-1}}$.
6: Update $\beta_n = \beta_{n-1}\left(1 + \zeta_n\tau_{n-1}\right)$.
7: Update $\tau_n = \min\left\{\rho\tau_{n-1}, \frac{\alpha^2\theta_{n-1}}{\beta_n\|K\|^2\tau_{n-1}}\right\}$.
8: Set $\sigma_n = \beta_n\tau_n$.
9: Compute $w^n = \mathrm{prox}_{F/\sigma_n}\left(\frac{y^{n-1}}{\sigma_n} + Ku^n\right)$, $F(w) = \lambda\|w\|_{2,1}$.
10: Update $y^n = y^{n-1} + \sigma_n(Ku^n - w^n)$.
11: Update $\theta_n = \frac{\psi\tau_n}{\tau_{n-1}}$.
12: **end for**

*Remark 4.13* Unlike E-GRPDA with Rule 1, the PF-AGRPDA specialization above is not norm-free, since both its initialization bound and its stepsize update involve $\|K\|$. Accordingly, the two golden-ratio schemes address different objectives: E-GRPDA removes prior operator-norm information, whereas PF-AGRPDA exploits strong convexity to obtain the accelerated ergodic rate established in Soe and Vetrivel (2026).

**Corollary 4.14** (Accelerated convergence rate for the ROF specialization)**.** *Let the iterates be generated by Algorithm 2, and assume that the ROF saddle-point problem admits a solution. Define* $Q_N := \sum_{n=1}^{N}\beta_n\tau_n$, *and the weighted ergodic averages* $\widetilde{u}^N = \frac{1}{Q_N}\sum_{n=1}^{N}\beta_n\tau_n u^n$, $\widetilde{w}^N = \frac{1}{Q_N}\sum_{n=1}^{N}\beta_n\tau_n w^n$. *Since the ROF fidelity term*

*$G(u) = \frac{1}{2}\|u - f\|_2^2$ is 1-strongly convex, the accelerated convergence result (Soe and Vetrivel 2026, Theorem 6.1) applies. Consequently, there exist constants $C_1, C_2 > 0$, independent of $N$, such that $\left|\Phi(\widetilde{u}^N, \widetilde{w}^N) - \Phi(u^\star, w^\star)\right| \le \frac{C_1}{N^2+N}$, and $\|K\widetilde{u}^N - \widetilde{w}^N\|_2 \le \frac{C_2}{N^2+N}$. Hence both the ergodic function-value residual and the constraint violation converge at the accelerated rate $O\left(\frac{1}{N^2}\right)$.*

*Remark 4.15* The $O(1/N^2)$ statement in Corollary 4.14 concerns the weighted ergodic quantities defined in the theorem. It should not be interpreted as a claim that PF-AGRPDA necessarily produces a smaller nonergodic ROF objective error, a larger PSNR, or a shorter CPU time than PDHG at every finite iteration. These quantities measure different aspects of numerical behavior and are therefore reported separately in the experiments.

The three primal–dual schemes considered in this work therefore represent different algorithmic trade-offs. The standard PDHG method uses a norm-dependent stepsize condition and admits the classical $O(1/N)$ ergodic guarantee. E-GRPDA with Rule 1 removes the need for prior knowledge of $\|K\|$ while retaining an $O(1/N)$ ergodic rate. In contrast, the accelerated PF-AGRPDA specialization exploits the strong convexity of the ROF fidelity term to attain an $O(1/N^2)$ ergodic rate, but its stepsize construction still involves $\|K\|$. These distinctions motivate the numerical study in the next section. Rather than assuming that the accelerated ergodic guarantee implies uniform superiority in every numerical metric, we compare the three methods separately in terms of objective convergence, CPU time, reconstruction error, and PSNR. This allows the theoretical and practical trade-offs among norm-dependent PDHG, norm-free E-GRPDA, and strong-convexity-based PF-AGRPDA to be assessed directly.

# 5 Numerical Experiments

The numerical experiments compare PDHG, E-GRPDA, and PF-AGRPDA on the ROF total-variation denoising problem. Three standard grayscale test images are considered: Cameraman, House, and Peppers. All images are resized to $256 \times 256$ pixels. Additive Gaussian noise with standard deviations $\sigma = 0.08$ and $\sigma = 0.10$ is used. The same noisy realization is supplied to all three methods in each test in order to ensure a direct comparison. All algorithms were implemented in Julia and executed on an 11th-generation Intel Core i5 processor with 16 GB of RAM. The same initialization was used for all methods.

Unless otherwise stated, the regularization parameter is fixed at $\lambda = 0.12$, and each method is run for 300 iterations. The comparison is based on the relative objective residual, CPU time, peak signal-to-noise ratio (PSNR), and relative $L_2$ reconstruction error. In addition, a discrete sweep over $\lambda$ is performed to examine the influence of the regularization parameter on the reconstruction quality.

Table 1 summarizes the reconstruction accuracy after 300 iterations. The three methods produce closely comparable final reconstructions for all images and both noise levels. The differences in PSNR are small, but PF-AGRPDA gives the highest final PSNR in each of the reported tests. The relative $L_2$ errors show the same general behavior. Thus, although the convergence histories below show that PDHG decreases the measured nonergodic objective residual more rapidly, PF-AGRPDA remains competitive in terms of the quality of the reconstructed image.

**Table 1** Final reconstruction results after 300 iterations for PDHG, E-GRPDA, and PF-AGRPDA with $\lambda = 0.12$. Higher PSNR and lower relative $L_2$ error indicate better performance.

| Image | Noise | Method | PSNR (dB) | Rel. $L_2$ error |
|---|---|---|---|---|
| Cameraman | 0.08 | PDHG | 27.3372 | 0.0822 |
| | | E-GRPDA | 27.3390 | 0.0822 |
| | | PF-AGRPDA | **27.3723** | **0.0819** |
| | 0.10 | PDHG | 27.0029 | 0.0854 |
| | | E-GRPDA | 27.0050 | 0.0854 |
| | | PF-AGRPDA | **27.0378** | **0.0851** |
| House | 0.08 | PDHG | 30.2921 | 0.0530 |
| | | E-GRPDA | 30.2940 | 0.0530 |
| | | PF-AGRPDA | **30.3212** | **0.0528** |
| | 0.10 | PDHG | 29.8628 | 0.0560 |
| | | E-GRPDA | 29.8648 | 0.0560 |
| | | PF-AGRPDA | **29.8930** | **0.0558** |
| Peppers | 0.08 | PDHG | 28.2586 | 0.0752 |
| | | E-GRPDA | 28.2604 | 0.0752 |
| | | PF-AGRPDA | **28.2969** | **0.0749** |
| | 0.10 | PDHG | 27.8907 | 0.0784 |
| | | E-GRPDA | 27.8927 | 0.0784 |
| | | PF-AGRPDA | **27.9280** | **0.0781** |

The numerical differences in Table 1 are modest, which is consistent with the reconstruction panels presented below. For example, at $\sigma = 0.08$, PF-AGRPDA attains PSNR values of 27.3723,

30.3212, and 28.2969 dB for Cameraman, House, and Peppers, respectively. The corresponding PDHG values are 27.3372, 30.2921, and 28.2586 dB. The same pattern is observed at $\sigma = 0.10$. Hence, the principal numerical advantage of PF-AGRPDA in these tests is not a faster decrease of the measured nonergodic objective residual, but a slightly improved final reconstruction accuracy.

Table 2 provides a direct comparison of the computational effort required to reach a relative objective residual of $10^{-2}$. The results show a consistent ordering over all six test cases. PDHG reaches the prescribed tolerance first, followed by E-GRPDA, whereas PF-AGRPDA requires more iterations. The CPU-time measurements exhibit the same general behavior. This observation is consistent with the convergence curves presented below and shows that, for the present implementation, PDHG is the fastest method with respect to the reported nonergodic objective criterion.

**Table 2** Iterations and CPU time required to reach a relative objective residual of $10^{-2}$ for $\lambda = 0.12$. Smaller values indicate faster convergence.

| Image | Noise | Method | Iterations | CPU time (s) |
|---|---|---|---|---|
| Cameraman | 0.08 | PDHG | **40** | **0.214** |
| | | E-GRPDA | 58 | 0.418 |
| | | PF-AGRPDA | 158 | 0.801 |
| | 0.10 | PDHG | **33** | **0.134** |
| | | E-GRPDA | 53 | 0.317 |
| | | PF-AGRPDA | 142 | 0.768 |
| House | 0.08 | PDHG | **46** | **0.202** |
| | | E-GRPDA | 63 | 0.433 |
| | | PF-AGRPDA | 169 | 0.800 |
| | 0.10 | PDHG | **36** | **0.134** |
| | | E-GRPDA | 56 | 0.291 |
| | | PF-AGRPDA | 146 | 0.685 |
| Peppers | 0.08 | PDHG | **37** | **0.117** |
| | | E-GRPDA | 53 | 0.266 |
| | | PF-AGRPDA | 129 | 0.616 |
| | 0.10 | PDHG | **31** | **0.133** |
| | | E-GRPDA | 48 | 0.233 |
| | | PF-AGRPDA | 119 | 0.551 |

The results in Table 2 are remarkably consistent across the three images and the two noise levels. For Cameraman at $\sigma = 0.08$, for example, PDHG reaches the $10^{-2}$ tolerance in 40 iterations, compared with 58 iterations for E-GRPDA and 158 iterations for PF-AGRPDA. At $\sigma = 0.10$, the corresponding iteration counts are 33, 53, and 142, respectively.

A similar pattern is observed for House and Peppers. The smallest iteration count is obtained for Peppers at $\sigma = 0.10$, where PDHG, E-GRPDA, and PF-AGRPDA require 31, 48, and 119 iterations, respectively. The CPU-time results support the same conclusion. Thus, with respect to the nonergodic relative objective residual used in these experiments, PDHG provides the fastest finite-iteration convergence among the three methods.

The influence of the regularization parameter is shown first. Figure 1 reports PSNR as a function of $\lambda$ for both noise levels. The curves exhibit the expected trade-off: PSNR initially increases as the regularization becomes sufficiently strong to suppress the noise, reaches a maximum, and then decreases when stronger TV regularization produces excessive smoothing.

For Cameraman and Peppers at $\sigma = 0.08$, the largest PSNR among the tested values is obtained around $\lambda = 0.06$. For House, the best values lie approximately in the range 0.06–0.08. The three algorithms produce very similar $\lambda$-PSNR profiles, indicating that the choice of the regularization parameter has a substantially larger influence on the final reconstruction quality than the small differences among the three optimization schemes.

We first consider the Cameraman image. Figure 2 shows the noisy images and the reconstructions obtained by the three methods. All methods substantially suppress the Gaussian noise while retaining the dominant edges and structures of the image. The reconstructed images obtained by PDHG, E-GRPDA, and PF-AGRPDA are visually close at both noise levels.

The objective-residual curves in Figure 3 show a consistent ordering. PDHG decreases the nonergodic objective residual most rapidly, followed by E-GRPDA, whereas PF-AGRPDA has a slower initial decrease. This behavior is observed both against the iteration count and against CPU time. Nevertheless, all three methods show a sustained reduction of the objective residual over the computational budget.

The PSNR curves provide a complementary view of the reconstruction process. As shown in Figure 4, all three methods rapidly improve the noisy observation during the first iterations and subsequently approach a stable reconstruction level. PF-AGRPDA approaches this level more gradually and exhibits a larger early transient, whereas the three methods become close after sufficient iterations.

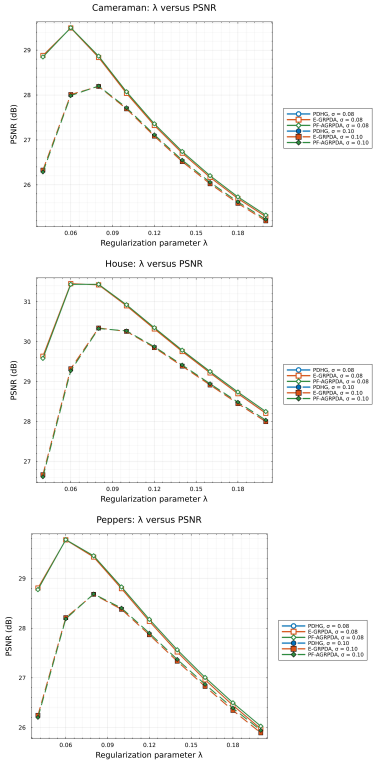



**Fig. 1** PSNR as a function of the regularization parameter $\lambda$ for the Cameraman, House, and Peppers images at noise levels $\sigma = 0.08$ and 0.10. The curves show a clear intermediate range of $\lambda$ that gives the best reconstruction quality, while larger values lead to a gradual reduction in PSNR.

Figure 5 shows the corresponding relative $L_2$ reconstruction error as a function of CPU time. The error decreases rapidly during the early iterations and then approaches a common neighborhood. PF-AGRPDA exhibits a larger initial error but later attains a reconstruction-error level comparable with the other two methods.

The House image provides a second test containing several pronounced edges and nearly piecewise-constant regions. The reconstruction results are displayed in Figure 6. All three methods remove a substantial part of the noise while preserving the roof boundaries, windows, and other dominant geometric features. Increasing the noise level from 0.08 to 0.10 produces the expected reduction in reconstruction quality, but the principal structures remain well preserved.

The objective-residual results in Figure 7 are consistent with those obtained for Cameraman. PDHG reaches smaller nonergodic objective residuals in fewer iterations and less CPU time, while E-GRPDA follows closely. PF-AGRPDA converges more gradually according to this particular nonergodic criterion.

The PSNR histories in Figure 8 show that useful reconstructions are obtained after relatively few iterations. PF-AGRPDA requires a longer transient but reaches a PSNR comparable to, and in part of the iteration history slightly above, the values produced by PDHG and E-GRPDA. The differences become small as the iterations progress.

The relative $L_2$ results in Figure 9 support the same observation. PDHG and E-GRPDA reach the neighborhood of their final reconstruction error quickly. PF-AGRPDA decreases more gradually but ultimately approaches essentially the same error range.

The Peppers image contains smoother regions together with several curved and high-contrast boundaries. Figure 10 shows that the three algorithms effectively suppress the noise and retain the principal shapes of the peppers. As in the previous experiments, the visual differences among the final reconstructions are relatively small.

Figure 11 again shows that PDHG has the fastest decrease of the nonergodic objective residual, with E-GRPDA giving the second-fastest decrease. PF-AGRPDA decreases the same residual more slowly, although the residual continues to decrease throughout the experiment. The same ordering is observed when CPU time replaces iteration number on the horizontal axis.

The PSNR histories are reported in Figure 12. All three methods rapidly improve the noisy image and eventually produce similar reconstruction-quality levels. PF-AGRPDA displays a more pronounced initial transient and reaches its high-PSNR region later than PDHG and E-GRPDA.

The relative $L_2$ error versus CPU time is shown in Figure 13. The three methods eventually approach a narrow error range. PF-AGRPDA initially decreases the reconstruction error more slowly but subsequently becomes competitive with the other two schemes.

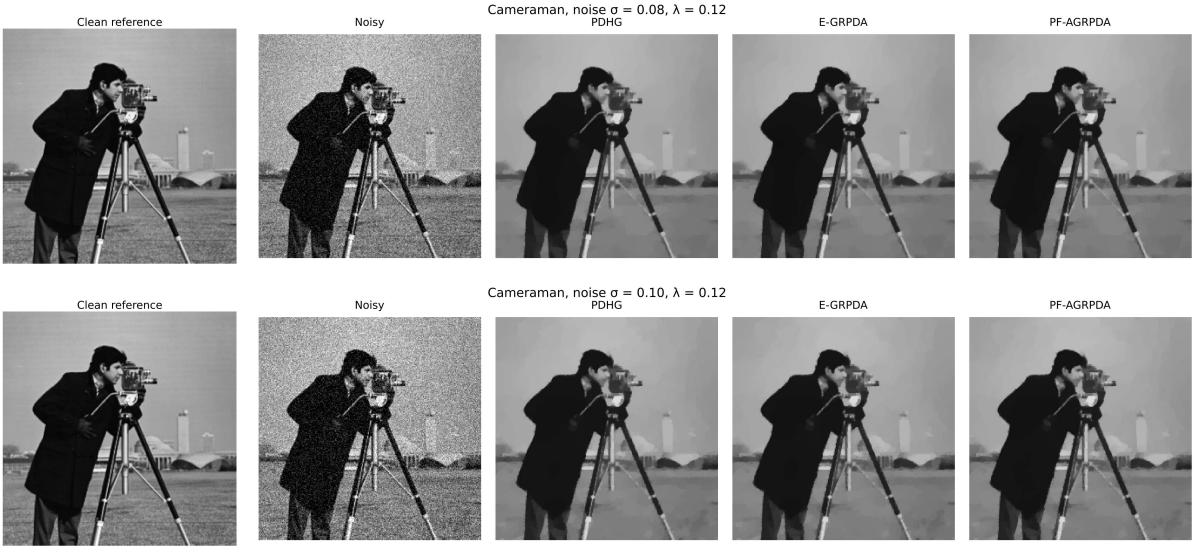



**Fig. 2** Denoising results for the Cameraman image at $\sigma = 0.08$ (top) and $\sigma = 0.10$ (bottom), with $\lambda = 0.12$. Each row shows the clean reference image, noisy observation, and reconstructions produced by PDHG, E-GRPDA, and PF-AGRPDA.

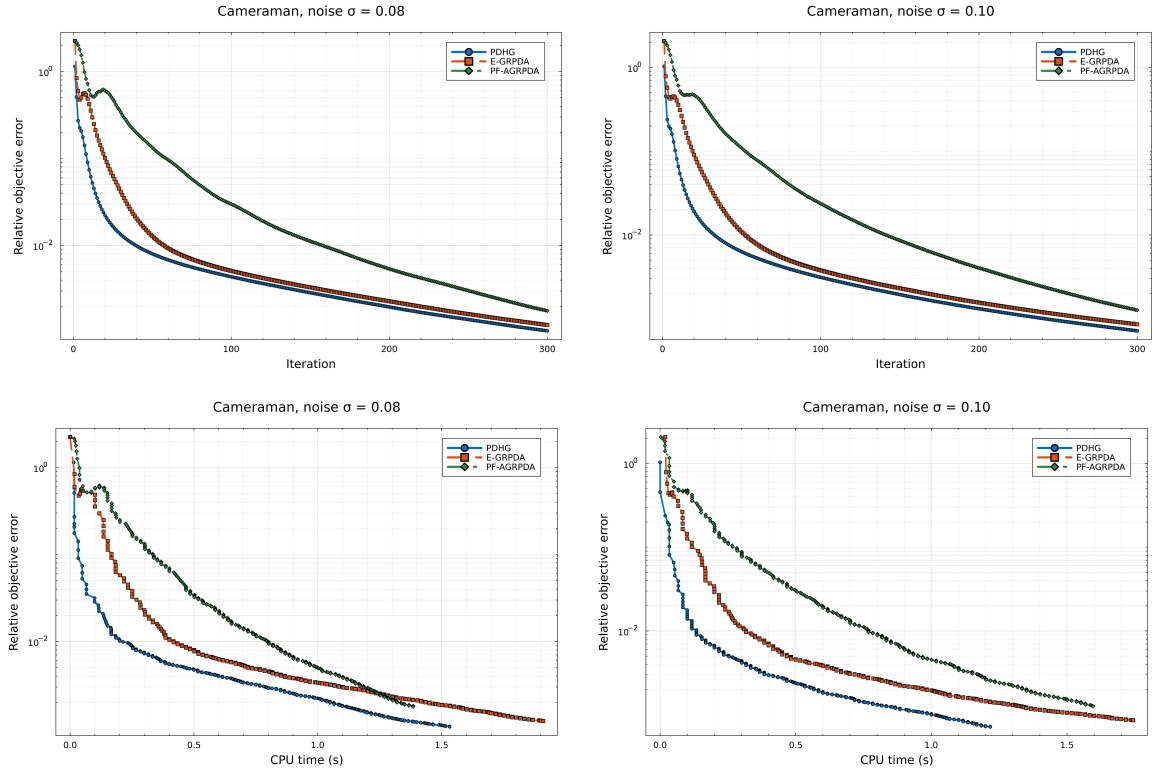



**Fig. 3** Relative objective residual for the Cameraman image. The top row shows convergence with respect to iteration number, and the bottom row shows convergence with respect to CPU time. The left and right columns correspond to $\sigma = 0.08$ and $\sigma = 0.10$, respectively.

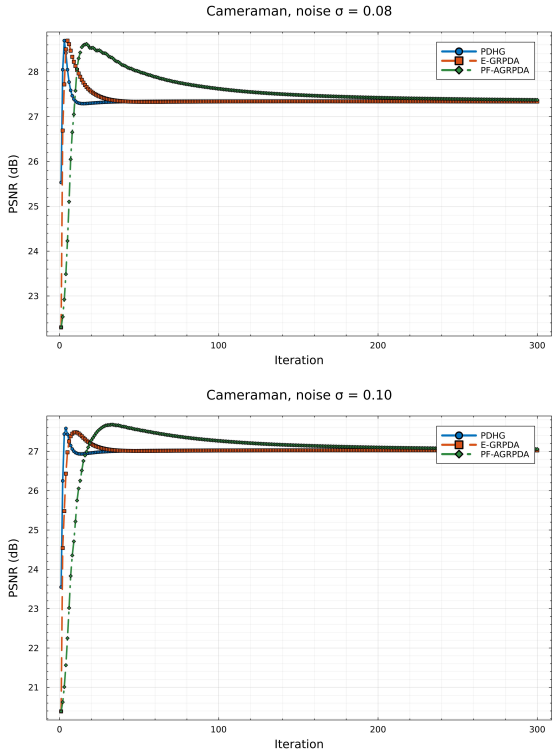



**Fig. 4** PSNR versus iteration for the Cameraman image at $\sigma = 0.08$ (top) and $\sigma = 0.10$ (bottom). All three algorithms approach comparable reconstruction-quality levels after the initial transient.

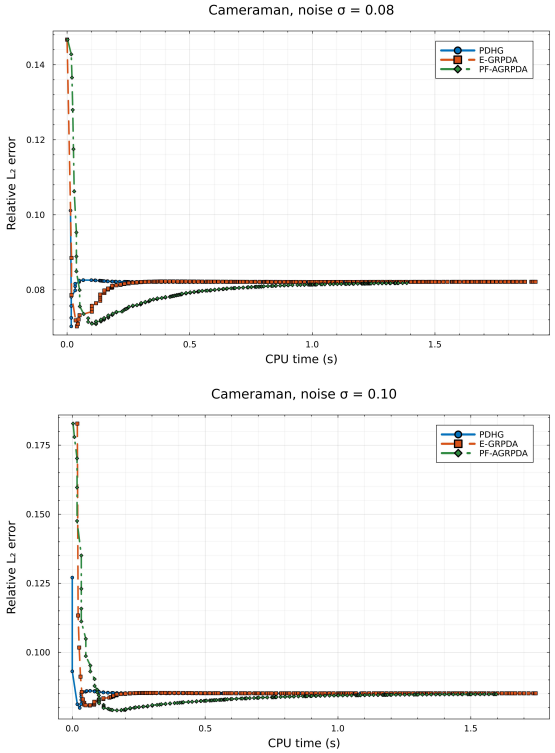



**Fig. 5** Relative $L_2$ reconstruction error versus CPU time for the Cameraman image at $\sigma = 0.08$ (top) and $\sigma = 0.10$ (bottom).

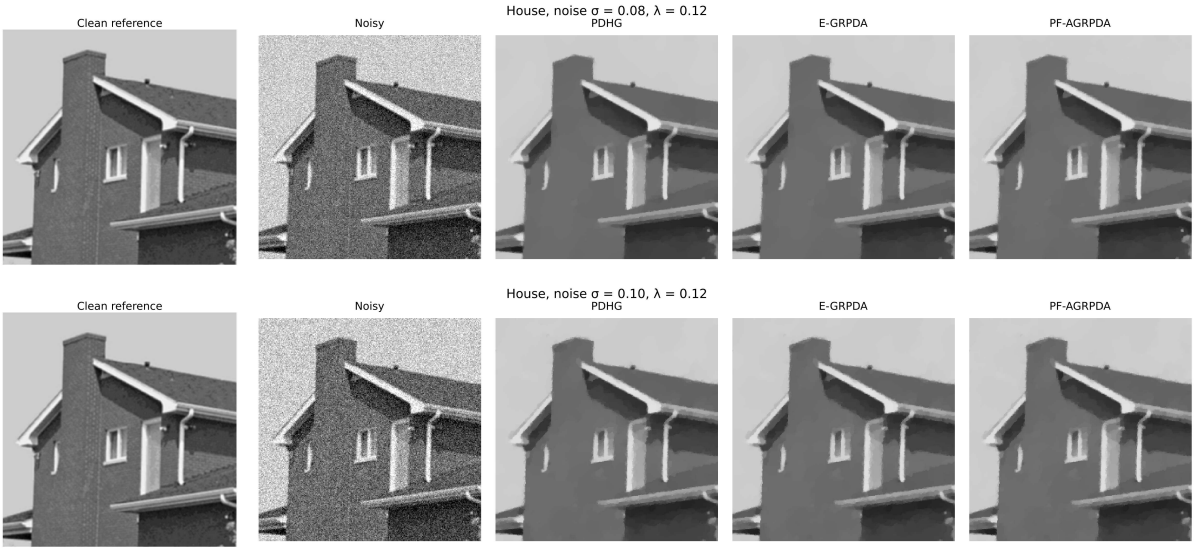



**Fig. 6** Denoising results for the House image at $\sigma = 0.08$ (top) and $\sigma = 0.10$ (bottom), with $\lambda = 0.12$. The columns show the clean reference image, noisy observation, PDHG reconstruction, E-GRPDA reconstruction, and PF-AGRPDA reconstruction.

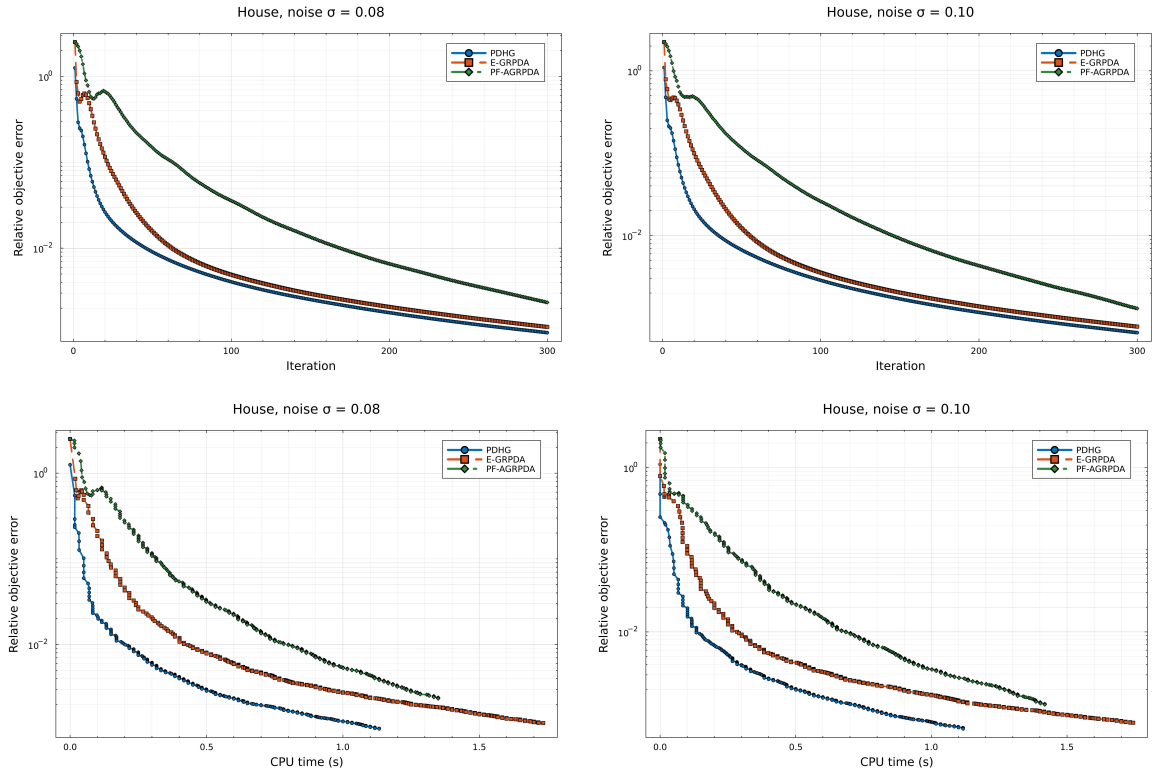



**Fig. 7** Relative objective residual for the House image. The top row reports residual versus iteration and the bottom row reports residual versus CPU time. The left and right columns correspond to $\sigma = 0.08$ and $\sigma = 0.10$, respectively.

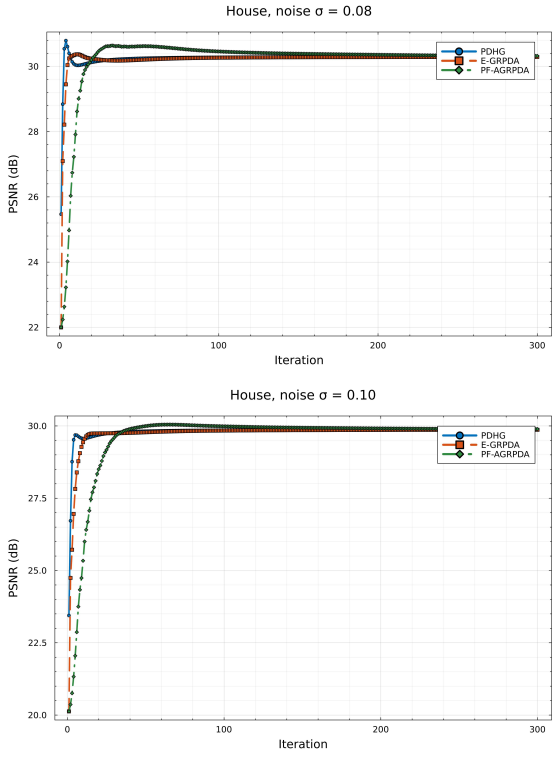



**Fig. 8** PSNR versus iteration for the House image at $\sigma = 0.08$ (top) and $\sigma = 0.10$ (bottom). The methods approach closely comparable PSNR values after their initial transients.

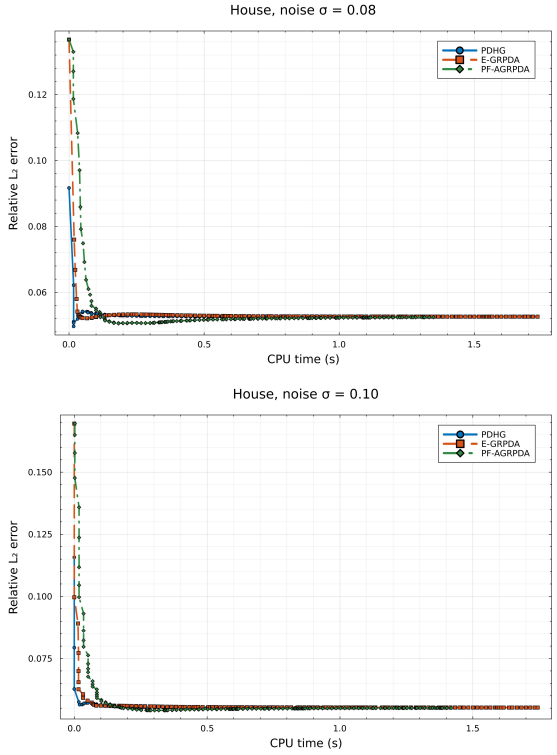



**Fig. 9** Relative $L_2$ reconstruction error versus CPU time for the House image at $\sigma = 0.08$ (top) and $\sigma = 0.10$ (bottom).

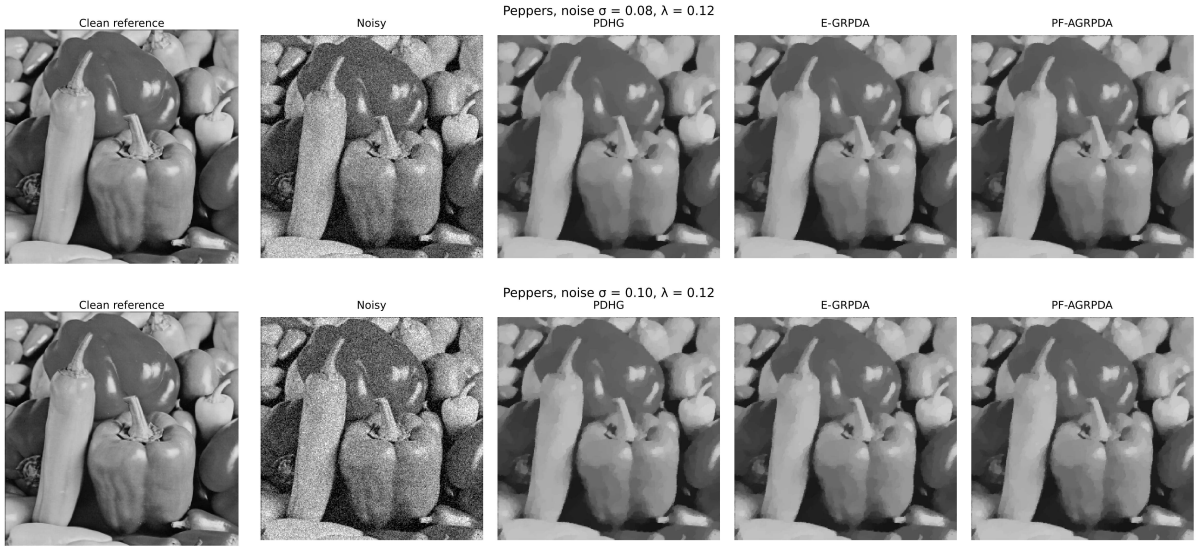



**Fig. 10** Denoising results for the Peppers image at $\sigma = 0.08$ (top) and $\sigma = 0.10$ (bottom), with $\lambda = 0.12$. The columns show the clean reference image, noisy observation, and reconstructions obtained by PDHG, E-GRPDA, and PF-AGRPDA.

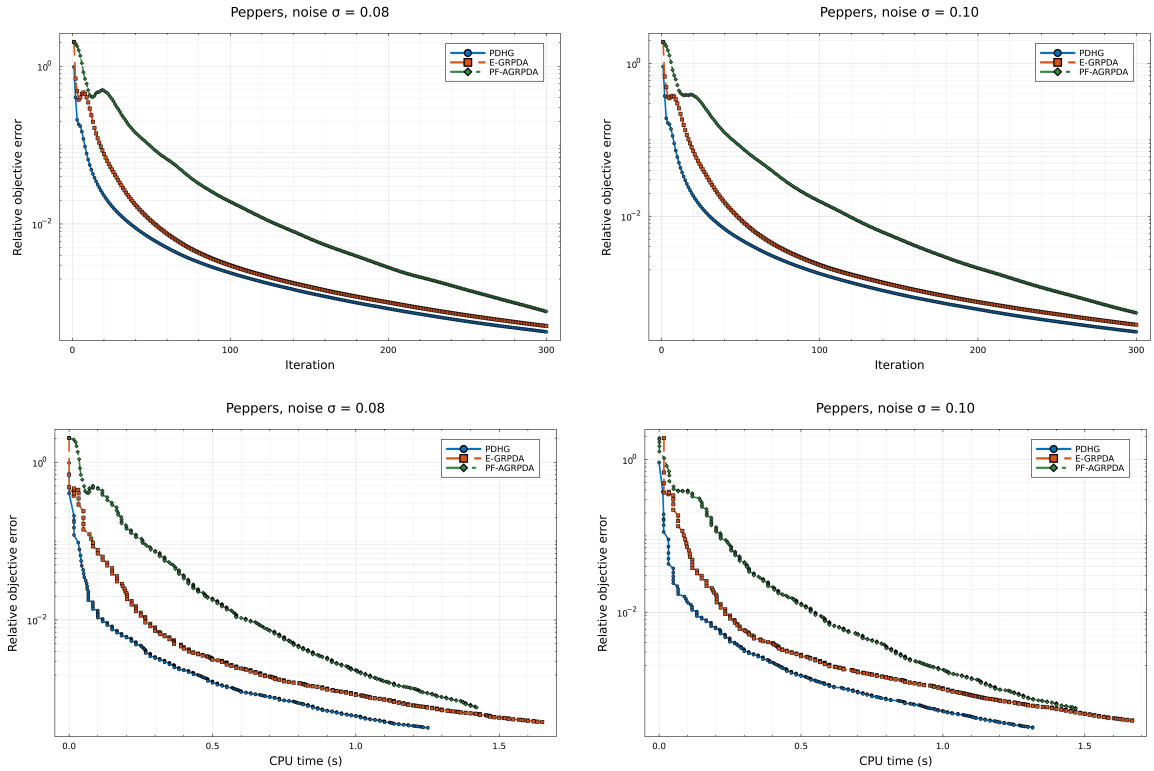



**Fig. 11** Relative objective residual for the Peppers image. The top row shows the residual versus iteration and the bottom row shows the residual versus CPU time. The left and right columns correspond to $\sigma = 0.08$ and $\sigma = 0.10$, respectively.

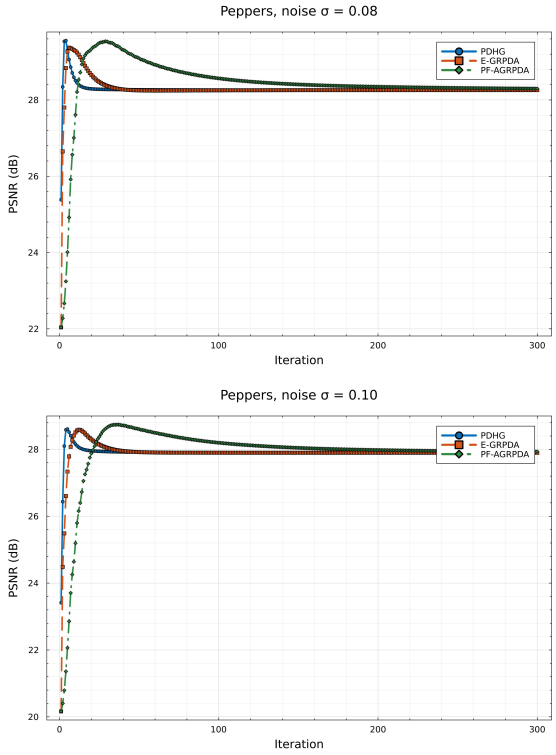



**Fig. 12** PSNR versus iteration for the Peppers image at $\sigma = 0.08$ (top) and $\sigma = 0.10$ (bottom). The three methods approach similar PSNR levels after the transient phase.

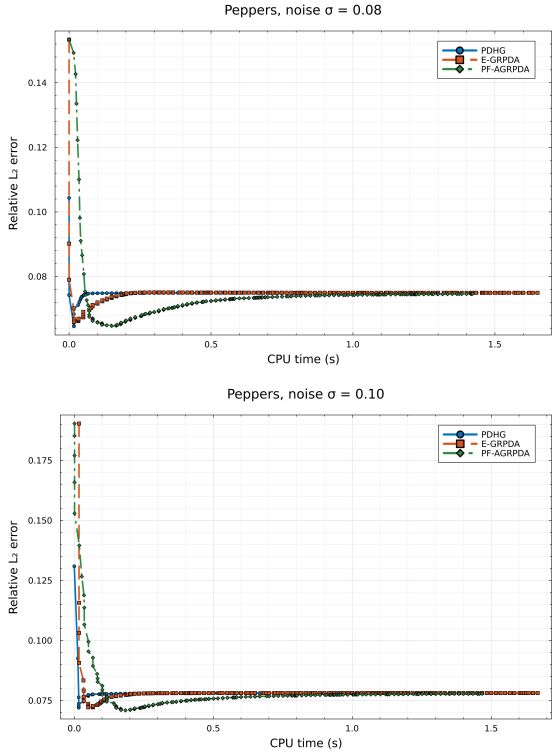



**Fig. 13** Relative $L_2$ reconstruction error versus CPU time for the Peppers image at $\sigma = 0.08$ (top) and $\sigma = 0.10$ (bottom).

# 6 Conclusion

The numerical experiments demonstrate that PDHG, E-GRPDA, and PF-AGRPDA are effective for ROF total-variation image denoising. PDHG gives the fastest decrease of the nonergodic objective residual and requires the least computational effort to reach the prescribed tolerance, while E-GRPDA generally exhibits intermediate behavior. PF-AGRPDA converges more gradually with respect to this criterion, but consistently provides competitive reconstruction quality and attains the highest final PSNR and the lowest relative $L_2$ error in the reported experiments.

These results highlight an important distinction between optimization speed and reconstruction accuracy. In particular, the accelerated $O(1/N^2)$ convergence result established for PF-AGRPDA concerns the weighted ergodic quantities and should not be interpreted as requiring faster finite-iteration nonergodic objective convergence or lower CPU time than PDHG.

Overall, the numerical results support the effectiveness of the three methods for ROF image denoising, while highlighting their different theoretical and computational advantages.

**Author contributions.** Faiz Ur Rahman and Izhar Ahmad contributed to the development of the manuscript, analysis of the results, and preparation of the final manuscript. Both authors read and approved the final manuscript.

**Funding.** The authors received no funding for this work.

**Competing interests.** The authors declare no competing interests.

**Data availability.** The images used in the numerical experiments are standard test images. The data supporting the findings of this study are available from the corresponding author upon reasonable request.

# References

Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. Springer, Cham (2017)

Bertsekas, D.P.: Convex Optimization Theory. Athena Scientific, Belmont, MA (2009)

Bredies, K., Sun, H.: Preconditioned Douglas–Rachford splitting methods for convex-concave saddle-point problems. SIAM Journal on Numerical Analysis **53**(1), 421–444 (2015)

Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Transactions on Image Processing **18**(11), 2419–2434 (2009)

Chan, T.F., Golub, G.H., Mulet, P.: A nonlinear primal-dual method for total variation-based image restoration. SIAM Journal on Scientific Computing **20**(6), 1964–1977 (1999)

Chambolle, A.: An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision **20**(1–2), 89–97 (2004)

Condat, L.: A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. Journal of Optimization Theory and Applications **158**(2), 460–479 (2013)

Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision **40**(1), 120–145 (2011)

Combettes, P.L., Pesquet, J.-C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators. Set-Valued and Variational Analysis **20**(2), 307–330 (2012)

Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Modeling & Simulation **4**(4), 1168–1200 (2005)

Chang, X., Yang, J.: A golden ratio primal-dual algorithm for structured convex optimization. Journal of Scientific Computing **87**(2), 47 (2021)

Goldstein, T., Li, M., Yuan, X.: Adaptive primal-dual splitting methods for statistical learning and image processing. In: Advances in Neural Information Processing Systems, vol. 28 (2015)

Goldstein, T., Osher, S.: The split bregman method for $L_1$-regularized problems. SIAM Journal on Imaging Sciences **2**(2), 323–343 (2009)

Malitsky, Y.: Golden ratio algorithms for variational inequalities. Mathematical Programming **184**, 383–410 (2020)

Mahdi, S., Elbatal, I., Abdelwahab, M.M., Tawhid, M., Mousavirad, S.J., Wagdy, A.: Derivative-free methods (conjugate gradient methods) for dealing with nonlinear equations and image restoration. Numerical Algorithms (2026) https://doi.org/10.1007/s11075-026-02316-9

Mordukhovich, B.S., Nam, N.M.: An Easy Path to Convex Analysis and Applications. Morgan & Claypool Publishers, San Rafael, CA (2014)

Malitsky, Y., Pock, T.: A first-order primal-dual algorithm with linesearch. SIAM Journal on Optimization **28**(1), 411–432 (2018)

Nedić, A., Ozdaglar, A.: Subgradient methods for saddle-point problems. Journal of Optimization Theory and Applications **142**(1), 205–228 (2009)

Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: IEEE International Conference on Computer Vision, pp. 1762–1769 (2011)

Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the Mumford–Shah functional. In: IEEE International Conference on Computer Vision, pp. 1133–1140 (2009)

Peypouquet, J.: Convex Optimization in Normed Spaces: Theory, Methods and Examples. SpringerBriefs in Optimization. Springer, Cham (2015)

Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**(1–4), 259–268 (1992)

Soe, S., Tam, M.K., Vetrivel, V.: The golden ratio primal-dual algorithm with two new stepsize rules for convex-concave saddle point problems. Computational Optimization and Applications **93**, 1355–1399 (2026)

Soe, S., Vetrivel, V.: Two adaptive accelerated golden ratio primal–dual algorithms with an application to Poisson imaging problem. arXiv:2607.08174v2 (2026)

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K.: Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**(1), 91–108 (2005)

Xiao, Y., Zhu, H., Wu, S.-Y.: Primal and dual alternating direction algorithms for $\ell_1$-$\ell_1$-norm minimization problems in compressive sensing. Numerical Linear Algebra with Applications **20**(4), 585–604 (2013)

Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08–34, UCLA (2008)

Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**(2), 301–320 (2005)