RELAY: CROSS-STEP DATA-CONSISTENCY REUSE FOR FEW-STEP MEDICAL IMAGE RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Sparse-view computed tomography (CT) and accelerated magnetic resonance imaging (MRI) reduce radiation dose and scan time but yield severely ill-posed inverse problems. Plug-and-play reconstruction avoids protocol-specific retraining by combining a pretrained image prior with operator-specific data consistency (DC), while consistency models reduce prior evaluation to only a few neural function evaluations (NFEs). For sparse-view Radon and multi-coil encoding, however, DC typically requires iterative conjugate gradients (CG), making its truncation a critical source of error under tight computational budgets. We show that this inner solve becomes the primary bottleneck as operator conditioning worsens, while noise injection and momentum provide little benefit. To address this challenge, we introduce **RELAY**, a training-free solver that transfers the DC solution across outer steps and restarts standard CG using the exact residual of the current system. RELAY improves DC accuracy under unchanged inner-iteration and NFE budgets while preserving the exact DC solution. We further establish non-asymptotic guarantees that characterize its cross-step error propagation and conditioning-dependent iteration savings. Across sparse-view CT and accelerated multi-coil low-field MRI, RELAY achieves the best PSNR among learned baselines across all tested settings, uses substantially fewer CG iterations than a cold start, and on CT surpasses compressed sensing run to convergence.

1 INTRODUCTION

Sparse-view computed tomography uses fewer projection angles to lower radiation dose, while accelerated magnetic resonance imaging undersamples k-space to shorten scan time and mitigate motion artifacts (Pruessmann et al., 1999; Lustig et al., 2007). Both improve acquisition efficiency but leave severely ill-posed inverse problems to be solved from incomplete measurements (Fig. 1a). Classical compressed sensing (Donoho, 2006) relies on fixed hand-designed priors, whereas unrolled networks (Hammernik et al., 2018; Zhang & Ghanem, 2018) typically couple learned reconstruction to the operators and sampling patterns seen during training, limiting transfer across clinical protocols. Plug-and-play (PnP) reconstruction (Venkatakrishnan et al., 2013) instead combines a reusable pretrained image prior with operator-specific data consistency (DC). With consistency models (CMs) (Song et al., 2023b) reducing the prior update to as few as one or two neural function evaluations (NFEs), the operator-dependent computational burden increasingly shifts to the DC step.

The computational cost of DC is determined by the acquisition operator. Closed-form updates require a tractable pseudo-inverse, singular value decomposition, or diagonalizing transform (Wang et al., 2022; Zhu et al., 2023). Such structure is available for common natural-image operators such as blur, masking, and downsampling, but generally not for sparse-view Radon or multi-coil encoding with scan-specific coil sensitivities: their normal operators cannot be reduced by a fixed transform to element-wise inversion (Fig. 1b). DC is therefore typically implemented by solving a positive-definite linear system, commonly with conjugate gradients (CG) (Block et al., 2007; Chung et al., 2024). With only a few inner iterations, CG may stop far from convergence, especially for poorly conditioned systems, leaving substantial error in the data-consistency update.

Few-step reconstruction is governed by two computational budgets: outer prior evaluations and inner DC iterations. Existing methods have focused primarily on the outer loop, improving measurement guidance (Chung et al., 2022; Song et al., 2023a), noise scheduling, and sampling stabilizers (Gulle

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

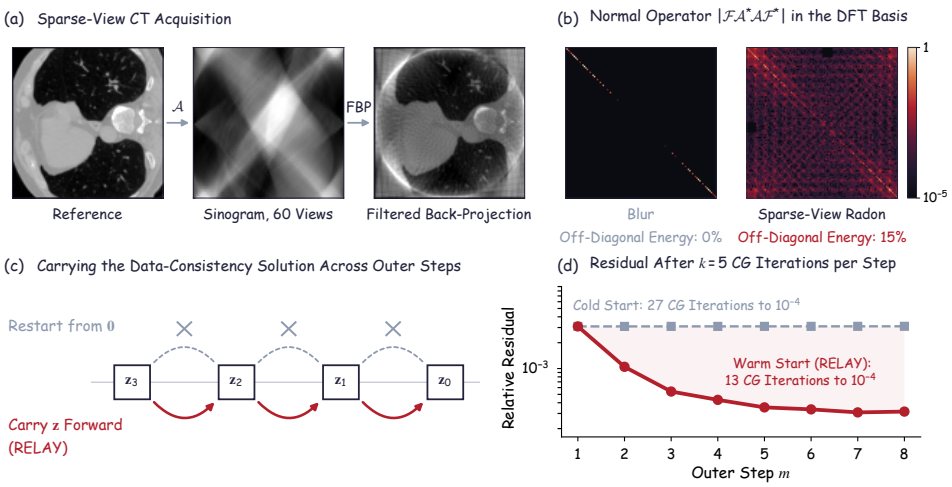


Figure 1: Sparse-view CT and truncated data consistency. (a) Reference, sinogram, and filtered back-projection. (b) DFT-domain normal operators for blur and sparse-view Radon. (c) Cold start versus RELAY across outer steps. (d) DC residuals with five CG iterations per step.

et al., 2026). The inner DC solve, by contrast, is typically assigned a fixed iteration budget. This is often benign when DC is inexpensive or solved to sufficient accuracy, but breaks down for medical operators with poorly conditioned normal systems. Under a fixed reconstruction budget, computation devoted to the prior and to DC must therefore be balanced, yet which side limits reconstruction quality and how that balance depends on the acquisition operator remain unclear.

We identify operator conditioning as the key factor governing this trade-off. For poorly conditioned medical operators, reconstruction quality is limited primarily by the accuracy of the truncated DC solve, while noise injection and momentum offer little benefit and can even be detrimental. We propose **RELAY**, a training-free scheme that reuses the DC solution across outer steps instead of repeatedly solving each system from a cold start. Consecutive systems share the same acquisition operator and change only gradually across outer steps, allowing RELAY to carry the previous solution forward, recompute the residual for the current system, and restart standard CG from a substantially better initialization (Fig. 1c,d). This improves DC accuracy under the same inner-iteration and NFE budgets while preserving the exact DC solution. We evaluate RELAY under a zero-shot protocol across sparse-view CT and accelerated multi-coil MRI on three datasets. Our contributions:

- We identify truncated DC accuracy as a key bottleneck in few-step reconstruction for ill-conditioned medical operators, and show that operator conditioning governs the trade-off between outer prior evaluations and inner solver iterations.
- We propose **RELAY**, a training-free solver that carries the DC solution across outer steps and restarts standard CG with the exact residual of the current system, improving inner-solve accuracy without changing the underlying DC problem.
- We establish non-asymptotic guarantees for RELAY under the varying penalty schedule, including an error recursion across outer steps and a bound on the iteration savings with respect to operator conditioning.
- Across sparse-view CT and accelerated multi-coil MRI on three datasets, RELAY consistently improves reconstruction under matched computational budgets and substantially reduces the CG iterations required to reach comparable accuracy.

2 Related Work

Learning-Based Medical Reconstruction Physics-driven unrolled networks interleave learned regularization with DC and are widely used for accelerated MRI (Aggarwal et al., 2018; Hammernik et al., 2023). Their learned components are typically optimized for a particular acquisition distribution, and performance can degrade under changes in sampling pattern or protocol (Knoll et al., 2019), with distribution-shift instabilities (Antun et al., 2020) studied more broadly in learned

reconstruction. Self-supervised and scan-specific methods reduce dependence on fully sampled targets (Yaman et al., 2020; 2021), but generally require acquisition-specific optimization or adaptation. Generative priors instead decouple image modeling from the measurement operator and have been applied to accelerated MRI and tomographic reconstruction (Chung & Ye, 2022; Song et al., 2021; Jalal et al., 2021; Barbano et al., 2025). Their original diffusion implementations, however, require many prior evaluations, motivating recent few-step alternatives based on consistency models.

Generative Priors for Inverse Problems Diffusion and score-based priors (Ho et al., 2020; Song et al., 2020b) originally required tens to hundreds of neural evaluations (Song et al., 2020a; Karras et al., 2022), motivating step reduction through distillation (Salimans & Ho, 2022; Yin et al., 2024) and consistency models (Song et al., 2023b). Existing inverse solvers differ mainly in how they incorporate the measurements. DPS (Chung et al., 2022) and ΠGDM (Song et al., 2023a) guide the sampling trajectory without an inner linear solve. DDNM (Wang et al., 2022), DiffPIR (Zhu et al., 2023), and CM4IR (Garber & Tirer, 2025) instead use closed-form or pseudo-inverse-based data updates that require additional structure from the forward operator. CoSIGN (Zhao et al., 2024) learns operator conditioning through a separately trained ControlNet, while DDS (Chung et al., 2024) embeds a truncated conjugate-gradient solve within the generative sampler. These approaches therefore span measurement guidance, direct operator-dependent updates, and iterative DC.

Table 1: Representative zero-shot generative solvers for medical inverse problems. *Matrix-free*: forward and adjoint operator applications only. DM: diffusion model. CM: consistency model.

			Efficiency	Generality		
Method	**Domain**	**Prior**	**Prior evaluations**	**Operator-agnostic**	**Matrix-free**	**Data-fidelity mechanism**
No inner solve: the measurements guide the sampling trajectory						
DPS (Chung et al., 2022)	Natural	DM	$\sim 10^3$	✓	✓	Likelihood gradient
ΠGDM (Song et al., 2023a)	Natural	DM	$\sim 10^2$	✓	✗	Pseudoinverse guidance
No inner solve: the data step is available in closed form						
DDNM (Wang et al., 2022)	Natural	DM	$\sim 10^2$	✓	✗	Null-space projection
DiffPIR (Zhu et al., 2023)	Natural	DM	$\sim 10^2$	✓	✗	Proximal update
CM4IR (Garber & Tirer, 2025)	Natural	CM	≤ 10	✓	✗	Back-projection
CoSIGN (Zhao et al., 2024)	Natural, CT	CM	≤ 10	✗	✗	ControlNet conditioning
Inner linear solve without reuse across outer steps						
DDS (Chung et al., 2024)	CT, MRI	DM	$\sim 10^1$	✓	✓	CG in the sampler
Inner linear solve reusing the previous solution						
RELAY (ours)	CT, MRI	CM	≤ 10	✓	✓	CG on the correction system

Iterative Data Consistency in Generative Reconstruction Plug-and-play (PnP) reconstruction decouples priors from data fidelity through variable splitting (Venkatakrishnan et al., 2013; Afonso et al., 2010; Chan et al., 2016), with convergence and fixed-point properties studied extensively (Ryu et al., 2019; Nair et al., 2021). PnP has been applied to MRI (Ahmad et al., 2020) and more recently combined with diffusion and consistency priors (Xu & Chi, 2024; Wu et al., 2024; Park et al., 2025). PnP-CM (Gulle et al., 2026), for example, combines a consistency-model prior with noise injection and iterate momentum at low NFE. When no closed-form data update is available, however, each outer step still requires an iterative linear solve, typically restarted from scratch. These solves form a sequence of related systems, for which numerical linear algebra has long used warm starts and Krylov-subspace recycling (Parks et al., 2006). RELAY instead carries forward only the previous DC solution, recomputes the current residual, and restarts standard CG without storing or recycling a Krylov basis. Tab. 1 compares RELAY with representative zero-shot generative solvers.

3 METHODOLOGY

Problem Formulation Sparse-view CT and accelerated MRI admit the linear measurement model $\mathbf{y} = \mathcal{A}\mathbf{x} + \mathbf{n}$, where $\mathcal{A}$ is the known forward operator, $\mathcal{A}^*$ its adjoint, $\mathbf{y}$ the acquired measurements, and $\mathbf{n}$ measurement noise. We consider the regularized inverse problem as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \ \tfrac{1}{2}\|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\, \Phi(\mathbf{x}), \tag{1}$$

where $\lambda > 0$ balances data fidelity and the image prior. PnP replaces the proximal map of Φ by a pretrained consistency model while retaining the physics-based data-fidelity step. We study two fixed computational budgets: N prior evaluations and k iterations per inner DC solve. Our goal is to improve reconstruction within the same (N, k) budget rather than increasing either.

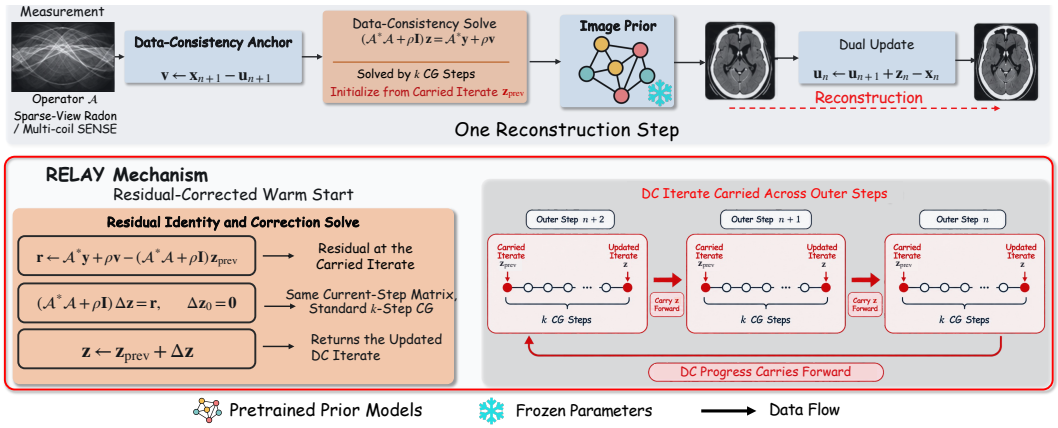


Figure 2: Overview of RELAY. Top: a standard PnP outer step with a DC solve of k CG steps and one frozen-prior evaluation. Bottom: RELAY carries the previous DC solution forward and forms its exact residual for the new system, so standard CG solves only the correction at unchanged NFE.

Overview Fig. 2 summarizes RELAY. Each outer step alternates three updates: a physics-based DC solve, one evaluation of a frozen consistency-model prior, and a dual update, following the variable-splitting formulation of PnP-ADMM (Chan et al., 2016; Boyd et al., 2011). For sparse-view CT and multi-coil MRI, the DC update is a positive-definite linear system solved approximately by k CG iterations. The key design of RELAY lies in how these truncated solves are connected across outer steps: rather than restarting each solve from zero, RELAY carries the preceding DC solution forward and uses it to initialize the next system (Fig. 2, bottom). In this way, progress made within one inner budget is retained by the next, while the outer reconstruction updates remain unchanged.

Outer Reconstruction Step The reconstruction follows a reverse schedule of N outer steps, $n = N-1, \ldots, 0$. Splitting $\mathbf{x} = \mathbf{z}$ in Eq. 1 (Afonso et al., 2010) gives the scaled augmented Lagrangian

$$\mathcal{L}_\rho(\mathbf{z}, \mathbf{x}, \mathbf{u}) = \tfrac{1}{2}\|\mathcal{A}\mathbf{z} - \mathbf{y}\|_2^2 + \lambda\Phi(\mathbf{x}) + \tfrac{\rho}{2}\|\mathbf{z} - \mathbf{x} + \mathbf{u}\|_2^2 - \tfrac{\rho}{2}\|\mathbf{u}\|_2^2, \tag{2}$$

whose block updates, with the proximal map of Φ replaced by f_θ, form one outer step. Let $\mathbf{x}_{n+1}$ and $\mathbf{u}_{n+1}$ denote the primal and dual states entering the step, and let $\rho = \rho_{n+1} > 0$ be the corresponding DC penalty. Defining the DC anchor $\mathbf{v} = \mathbf{x}_{n+1} - \mathbf{u}_{n+1}$, the three block updates are

$$\begin{aligned} &\textsc{Step 1:} \quad \mathbf{z}_n = \arg\min_{\mathbf{z}} \ \tfrac{1}{2}\|\mathcal{A}\mathbf{z} - \mathbf{y}\|_2^2 + \tfrac{\rho}{2}\|\mathbf{z} - \mathbf{v}\|_2^2, \\ &\textsc{Step 2:} \quad \mathbf{x}_n = f_\theta(\mathbf{z}_n + \mathbf{u}_{n+1}, \sigma), \\ &\textsc{Step 3:} \quad \mathbf{u}_n = \mathbf{u}_{n+1} + \mathbf{z}_n - \mathbf{x}_n, \end{aligned} \tag{3}$$

where $\sigma = \sigma(t_{n+1})$ is the scheduled noise level and f_θ is the frozen consistency-model prior. Each outer step evaluates f_θ once, giving $\mathrm{NFE} = N$. The DC update in Step 1 is quadratic and satisfies the normal system

$$(\mathcal{A}^*\mathcal{A} + \rho\mathbf{I})\mathbf{z} = \mathcal{A}^*\mathbf{y} + \rho\mathbf{v}, \tag{4}$$

which is positive definite and is solved approximately with k CG iterations. RELAY uses the scheduled prior directly and requires none of the additional noise injection or iterate momentum that some few-step solvers add to improve low-NFE quality (Gulle et al., 2026).

RELAY Data-Consistency Update With only k CG iterations, each DC system is generally solved only approximately. Yet consecutive outer steps solve closely related systems: the acquisition term $\mathcal{A}^*\mathcal{A}$ is shared, while the anchor $\mathbf{v}$ and penalty ρ evolve across the reconstruction trajectory. RELAY carries the truncated DC solution from the preceding outer step, denoted by $\mathbf{z}_{\text{prev}}$, into the current system. Let $\mathbf{M} = \mathcal{A}^*\mathcal{A} + \rho\mathbf{I}$ and $\mathbf{b} = \mathcal{A}^*\mathbf{y} + \rho\mathbf{v}$. Rather than solving $\mathbf{M}\mathbf{z} = \mathbf{b}$ from zero, RELAY first evaluates the exact residual of the current system at the carried iterate,

$$\mathbf{r} = \mathbf{b} - \mathbf{M}\mathbf{z}_{\text{prev}} = \mathcal{A}^*\mathbf{y} + \rho\mathbf{v} - (\mathcal{A}^*\mathcal{A} + \rho\mathbf{I})\mathbf{z}_{\text{prev}}. \tag{5}$$

Algorithm 1 RELAY

Input: f_θ, $\mathcal{A}$, $\mathbf{y}$, $\{t_n, \rho_n\}_{n=1}^N$, CG budget k

1: $\mathbf{x}_N, \mathbf{u}_N, \mathbf{z}_{\text{prev}} \leftarrow \mathbf{0}$
2: **for** $n = N{-}1, \dots, 0$ **do**
3: $\quad \mathbf{M} \leftarrow \mathcal{A}^*\mathcal{A} + \rho_{n+1}\mathbf{I}, \quad \mathbf{b} \leftarrow \mathcal{A}^*\mathbf{y} + \rho_{n+1}(\mathbf{x}_{n+1} - \mathbf{u}_{n+1})$
4: $\quad \mathbf{r} \leftarrow \mathbf{b} - \mathbf{M}\mathbf{z}_{\text{prev}}$ ▷ exact residual
5: $\quad \Delta\mathbf{z} \leftarrow \mathrm{CG}_k(\mathbf{M}, \mathbf{r})$ ▷ k CG steps from $\mathbf{0}$
6: $\quad \mathbf{z}_n \leftarrow \mathbf{z}_{\text{prev}} + \Delta\mathbf{z}, \quad \mathbf{z}_{\text{prev}} \leftarrow \mathbf{z}_n$ ▷ carry forward
7: $\quad \mathbf{x}_n \leftarrow f_\theta\big(\mathbf{z}_n + \mathbf{u}_{n+1},\, \sigma(t_{n+1})\big)$
8: $\quad \mathbf{u}_n \leftarrow \mathbf{u}_{n+1} + \mathbf{z}_n - \mathbf{x}_n$
9: **end for**
10: **return** $\mathbf{x}_0$

It then solves for the correction $\Delta\mathbf{z} = \mathbf{z} - \mathbf{z}_{\text{prev}}$,

$$(\mathcal{A}^*\mathcal{A} + \rho\mathbf{I})\Delta\mathbf{z} = \mathbf{r}, \tag{6}$$

and recovers the current DC iterate as

$$\mathbf{z} = \mathbf{z}_{\text{prev}} + \Delta\mathbf{z}. \tag{7}$$

The correction system uses exactly the same matrix as Eq. 4, so standard CG applies without modification from $\Delta\mathbf{z}_0 = \mathbf{0}$. This is algebraically equivalent to applying CG directly to Eq. 4 with initialization $\mathbf{z}_0 = \mathbf{z}_{\text{prev}}$ (Saad, 2003). Thus, RELAY changes only the finite-budget trajectory of the DC solver: the underlying system and its exact solution remain unchanged, with the conventional cold start recovered when $\mathbf{z}_{\text{prev}} = \mathbf{0}$. Alg. 1 states RELAY in full. The only state it adds to a standard PnP-ADMM loop is the carried iterate $\mathbf{z}_{\text{prev}}$.

Finite-Budget Guarantee The correction form is an exact change of variables: an exactly solved correction Eq. 6 returns the solution of Eq. 4, and $\mathbf{z}_{\text{prev}} = \mathbf{0}$ recovers the cold start (Appendix A.2). RELAY therefore changes only the error left after k CG iterations, which we bound next. Index the executed outer steps forward by $m = N - n \in \{1, \dots, N\}$, with penalty $\beta_m = \rho_{n+1}$, system matrix $\mathbf{M}_m = \mathcal{A}^*\mathcal{A} + \beta_m\mathbf{I}$, right-hand side $\mathbf{b}_m$, exact solution $\mathbf{z}_m^\star = \mathbf{M}_m^{-1}\mathbf{b}_m$, $\kappa_m = \mathrm{cond}(\mathbf{M}_m)$ and $c_m = (\sqrt{\kappa_m} - 1)/(\sqrt{\kappa_m} + 1)$. Let ζ_m be the DC solution stored at step m (line 6 of Alg. 1) and $\varepsilon_m = \|\mathbf{z}_m^\star - \zeta_m\|_{\mathbf{M}_m}$ its error in the energy norm $\|\mathbf{w}\|_{\mathbf{M}} = (\mathbf{w}^*\mathbf{M}\mathbf{w})^{1/2}$.

Theorem 1 (Cross-step error recursion)**.** *Assume $\beta_m > 0$ for every m and CG in exact arithmetic with $k \geq 1$. Then $\varepsilon_1 \leq 2c_1^k\|\mathbf{z}_1^\star\|_{\mathbf{M}_1}$ and*

$$\varepsilon_m \;\leq\; 2c_m^k\big(\theta_m\,\varepsilon_{m-1} + D_m\big), \qquad 2 \leq m \leq N, \tag{8}$$

where $\theta_m = \max\{1, \sqrt{\beta_m/\beta_{m-1}}\}$ accounts for the energy-norm change and D_m bounds the drift of the exact solution between consecutive steps. A cold-started solve instead satisfies $\varepsilon_m^{\text{cold}} \leq 2c_m^k\|\mathbf{z}_m^\star\|_{\mathbf{M}_m}$ and resets its initialization error at every step. If $N \geq 2$ and $q_N = \max_{2\leq j\leq N} 2c_j^k\theta_j < 1$, the inherited error contracts geometrically to a drift-level floor.

The explicit drift term, the geometric bound and all proofs are given in Appendix A. The recursion also translates into an iteration saving. For a starting error X in $\|\cdot\|_{\mathbf{M}_m}$, let $k^\star(X)$ be the smallest $k \geq 1$ with $2c_m^k X \leq \varepsilon$ for a target error ε, and let $X_m^{\text{c}} = \|\mathbf{z}_m^\star\|_{\mathbf{M}_m}$ and $X_m^{\text{w}} = \theta_m\varepsilon_{m-1} + D_m$ be the cold starting error and a bound on the warm one of Theorem 1.

Corollary 2 (Conditioning-dependent iteration savings)**.** *If $\kappa_m > 1$ and $\varepsilon/2 \leq X_m^{\text{w}} < X_m^{\text{c}}$, then*

$$k^\star\big(X_m^{\text{c}}\big) - k^\star\big(X_m^{\text{w}}\big) \;\geq\; \frac{\ln\big(X_m^{\text{c}}/X_m^{\text{w}}\big)}{\ln(1/c_m)} - 1 \;\geq\; \frac{\sqrt{\kappa_m} - 1}{2}\,\ln\frac{X_m^{\text{c}}}{X_m^{\text{w}}} - 1. \tag{9}$$

For a fixed warm-start advantage $X_m^{\text{c}}/X_m^{\text{w}}$, the certified saving grows with κ_m, so RELAY is predicted to help most when the acquisition operator makes the DC system poorly conditioned.

4 Experiments

Datasets and Acquisition Settings We evaluate on two sparse-view CT benchmarks and one accelerated multi-coil MRI benchmark. **OrganAMNIST** (Yang et al., 2023) contains 64×64 organ

CT slices, from which we simulate parallel-beam measurements with Radon operators at varying projection counts, and we evaluate on 50 test slices. **LoDoPaB-CT** (Leuschner et al., 2021) provides clinical CT images at 256×256, evaluated on 100 test slices under the same acquisition model. Both use additive Gaussian noise with $\sigma_y = 0.05$ unless stated otherwise. **M4Raw** (Lyu et al., 2023) provides acquired 0.3 T low-field brain k-space with four coils and T1, T2 and FLAIR contrasts at 256×256. We use 40 test slices per contrast at accelerations $R \in \{4, 8, 12\}$ with a 1D Gaussian mask. The forward model is SENSE (Pruessmann et al., 1999), $\mathcal{A}\mathbf{x} = \mathcal{M} \odot \mathcal{F}(\mathbf{S}\mathbf{x})$, with mask $\mathcal{M}$, Fourier transform $\mathcal{F}$ and ESPIRiT coil maps $\mathbf{S}$ (Uecker et al., 2014).

Baselines and Metrics We compare RELAY with three classes of methods. Conventional baselines are FBP for CT, zero-filled reconstruction for MRI, and compressed sensing with total variation (TV-CS) (Sidky & Pan, 2008) or an ℓ_1-wavelet prior (Lustig et al., 2007). Generative baselines are the diffusion-based DPS (Chung et al., 2022), ΠGDM (Song et al., 2023a), DDS (Chung et al., 2024) and DiffPIR (Zhu et al., 2023), and the consistency-based CM (Song et al., 2023b), CM4IR (Garber & Tirer, 2025) and PnP-CM (Gulle et al., 2026). We retain the published sampling and guidance settings of each learned baseline, apart from the MRI sampling schedule. Since Radon and multi-coil encoding admit no operator-specific closed form, every method that needs an iterative data-fidelity solve uses the same CG implementation. We report PSNR and SSIM (Wang et al., 2004) on every benchmark and LPIPS (Zhang et al., 2018) on CT.

Implementation Details For each benchmark, we train one unconditional image prior and reuse it across all acquisition settings without protocol-specific fine-tuning. An ADM U-Net (Dhariwal & Nichol, 2021) is trained as an EDM teacher (Karras et al., 2022) and distilled into a consistency model (Song et al., 2023b). For M4Raw, the teacher is initialized from a public fastMRI (Zbontar et al., 2018) brain model (Gülle et al., 2026). Distillation takes 6k–38k steps across benchmarks and is performed on two NVIDIA RTX PRO 6000 GPUs (96 GB each). Diffusion baselines use the corresponding EDM teacher, while consistency-based methods use the distilled model. All iterative PnP methods share the same CG implementation and ADMM penalty schedule (Gulle et al., 2026). Unless stated otherwise, CT runs use N=6 outer steps with k=20 CG iterations per step and MRI runs N=4 with k=10. Full training and reconstruction settings are provided in Appendix B.

4.1 Main Results

Tab. 2 shows that RELAY consistently improves reconstruction across both acquisition modalities. On OrganAMNIST and LoDoPaB-CT, it achieves the best performance at every reported view count, outperforming the strongest learned baseline, PnP-CM, by 2.58–8.11 dB and 4.55–8.12 dB, respectively. RELAY also surpasses the tuned TV and ℓ_1-wavelet baselines across all CT settings, including TV-CS run to convergence at 12–23× its operator budget. This advantage persists beyond the operating points in the main table: Fig. 3 shows that RELAY remains the strongest few-step method throughout the full 8–120-view sweep and continues to benefit from additional projections after the competing methods begin to saturate. The same behavior transfers to multi-coil MRI. Across the nine contrast–acceleration combinations in Tab. 3, RELAY attains the highest PSNR in every case and the highest SSIM at the higher accelerations. At R=12, DPS, DDS, DiffPIR and CM4IR trail it by 2.0 to 5.4 dB, ΠGDM diverges at every rate, and at R=8 RELAY also exceeds TV-CS run to convergence on all three contrasts. Its advantage over classical CS also becomes more pronounced at higher acceleration, supporting the effectiveness of the same cross-step DC reuse under a distinct acquisition operator.

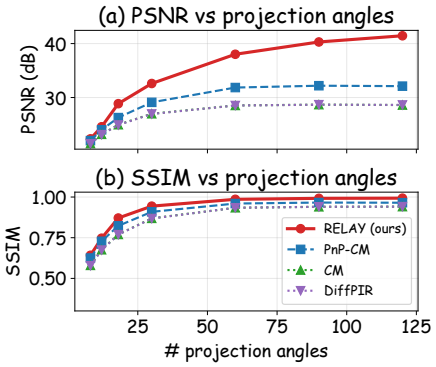


Figure 3: OrganAMNIST robustness across projection counts. PSNR/SSIM of four few-step learned solvers from 8 to 120 views.

Qualitative Results Fig. 4 compares representative reconstructions and absolute-error maps. On OrganAMNIST (Fig. 4a), CM over-smooths organ boundaries and PnP-CM retains structured errors,

Table 2: Main results, PSNR in dB. On CT, TV-CS *(converged)* uses 12–23× RELAY's operator budget. '/': not applicable to the modality. **Bold** and underline: the best and second-best.

	OrganAMNIST								**LoDoPaB-CT**								**M4Raw** MRI, R=8					
	18		30		60		90		18		30		60		90		T1		T2		FLAIR	
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FBP	19.08	.631	19.84	.707	20.15	.739	20.17	.740	21.62	.523	22.00	.581	22.32	.661	22.41	.692	/		/		/	
Zero-filled	/		/		/		/		/		/		/		/		22.46	.515	23.63	.475	22.30	.464
TV-CS	27.11	.820	30.67	.918	35.21	.977	36.40	.983	30.40	.751	32.37	.803	35.28	.884	36.84	.922	23.44	.505	24.39	.515	23.27	.497
TV-CS *(converged)*	28.74	.870	32.48	.943	37.87	.985	40.13	.991	34.36	.867	36.96	.910	40.27	.951	42.01	.966	26.00	.537	21.97	.348	21.44	.344
ℓ_1-wavelet CS	26.52	.795	30.09	.898	34.76	.969	36.06	.978	30.27	.743	32.21	.795	35.10	.878	36.69	.918	24.57	.540	24.72	.492	24.05	.480
DPS	15.32	.164	15.31	.164	15.30	.163	15.30	.163	17.36	.521	17.29	.520	17.25	.520	17.25	.520	22.18	.454	23.00	.366	21.86	.376
ΠGDM	24.42	.682	26.96	.793	29.79	.885	30.54	.903	25.68	.578	28.20	.680	29.78	.771	29.93	.788			*diverges*			
DDS	23.46	.741	24.33	.818	24.51	.842	24.49	.844	27.04	.680	26.92	.722	26.50	.764	26.33	.773	23.44	.483	23.74	.438	22.98	.435
DiffPIR	24.96	.770	26.99	.869	28.53	.934	28.67	.940	29.17	.716	30.17	.762	31.01	.843	31.18	.876	20.18	.286	18.60	.226	18.60	.243
CM	24.96	.770	27.00	.869	28.55	.934	28.68	.940	29.17	.716	30.17	.762	31.02	.843	31.18	.876	24.23	.566	24.77	.484	23.74	.491
CM4IR	15.54	.375	15.59	.390	15.62	.396	15.63	.397	18.35	.249	18.29	.258	18.27	.266	18.27	.269	23.96	.456	21.98	.376	21.67	.373
PnP-CM	26.28	.823	29.11	.909	31.85	.960	32.19	.966	29.99	.728	31.55	.774	33.36	.849	34.08	.887	27.04	.599	25.98	.542	25.23	.522
RELAY (ours)	**28.86**	**.871**	**32.60**	**.944**	**38.02**	**.986**	**40.30**	**.992**	**34.54**	**.868**	**37.12**	**.911**	**40.36**	**.952**	**42.20**	**.967**	**27.48**	**.618**	**26.38**	**.562**	**25.62**	**.540**

Table 3: M4Raw reconstruction across acceleration rates and MRI contrasts. PSNR is reported in dB. Higher is better for both metrics. **Bold** and underline denote the best and second-best results.

	T1						T2						FLAIR					
	R=4		R=8		R=12		R=4		R=8		R=12		R=4		R=8		R=12	
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Zero-filled	25.94	.644	22.46	.515	21.99	.483	26.32	.619	23.63	.475	23.18	.435	25.21	.599	22.30	.464	21.77	.425
TV-CS	27.21	.590	23.44	.505	22.66	.466	26.68	.549	24.39	.515	23.57	.420	26.13	.553	23.27	.497	22.33	.417
ℓ_1-wavelet CS	28.49	.669	24.57	.540	23.20	.490	27.24	.610	24.72	.492	23.88	.450	26.90	.599	24.05	.480	22.79	.434
DPS	25.41	.567	22.18	.454	21.76	.449	25.42	.508	23.00	.366	22.54	.362	24.50	.494	21.86	.376	21.52	.367
ΠGDM			*diverges*						*diverges*						*diverges*			
DDS	25.51	.520	23.44	.483	23.12	.494	24.13	.455	23.74	.438	23.74	.452	24.02	.466	22.98	.435	22.99	.445
DiffPIR	20.16	.299	20.18	.286	21.37	.337	18.12	.244	18.60	.226	20.57	.277	18.29	.261	18.60	.243	20.44	.293
CM	27.98	**.679**	24.23	.566	23.38	.530	27.22	**.637**	24.77	.484	24.04	.457	26.76	**.614**	23.74	.491	23.00	.451
CM4IR	21.18	.331	23.96	.456	24.21	.490	19.32	.281	21.98	.376	23.32	.410	19.23	.289	21.67	.373	22.61	.397
PnP-CM	27.43	.610	27.04	.599	26.04	.565	26.38	.559	25.98	.542	25.59	.510	25.74	.543	25.23	.522	24.69	.492
RELAY (ours)	**28.54**	.671	**27.48**	**.618**	**26.35**	**.580**	**27.26**	.612	**26.38**	**.562**	**25.92**	**.525**	**26.93**	.601	**25.62**	**.540**	**25.02**	**.507**

whereas RELAY preserves sharper boundaries with markedly lower residual error. On M4Raw at R=8 (Fig. 4b), RELAY suppresses the pronounced artifacts of zero-filling and better preserves cortical and ventricular structure across both slices and all three contrasts.

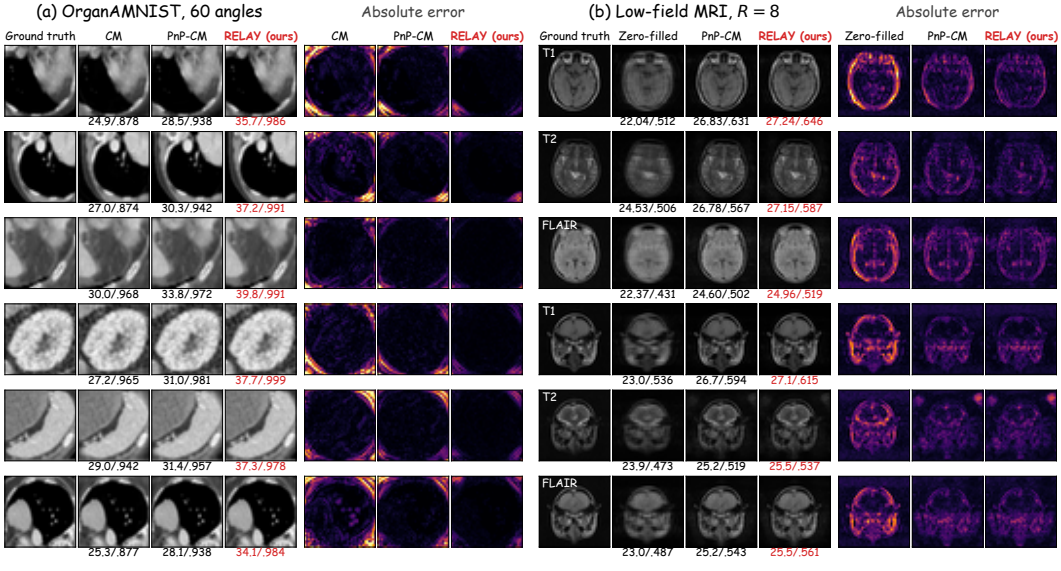


Figure 4: Reconstructions (PSNR/SSIM) and absolute errors. (a) OrganAMNIST, 60 angles. (b) M4Raw, R=8, two slices. Full results in Fig. 11 and Fig. 14.

Computational Cost Fig. 5 compares reconstruction accuracy against operator applications, which account for over 97% of the CT runtime. At the same 6 NFE, RELAY requires only 7% more operator applications than PnP-CM (276 vs. 258) while improving PSNR by 6.17 dB. RELAY also outperforms TV-CS under the matched 276-application budget, and remains better even when TV-CS is run to convergence at 12–23× the operator cost.

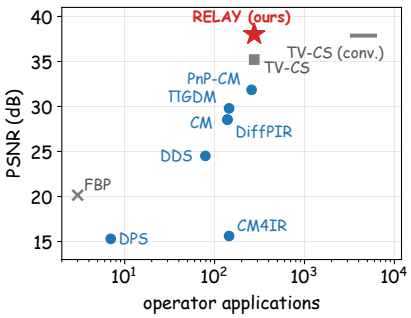


Figure 5: PSNR versus operator applications on OrganAMNIST at 60 angles.

Perceptual Quality Tab. 4 reports LPIPS on the CT benchmarks. RELAY achieves the lowest perceptual error at every view count, reaching 2.6× lower LPIPS than PnP-CM at 90 views on LoDoPaB-CT and 8.5× lower on OrganAMNIST. The advantage also grows with additional measurements. On LoDoPaB-CT, RELAY reduces LPIPS by a factor of 5.6 from 18 to 90 views, compared with 3.1 for PnP-CM and at most 2.4 for the remaining baselines. This trend shows that RELAY's gains extend beyond pixel-wise fidelity to perceptual reconstruction quality.

Table 4: Perceptual quality on sparse-view CT (LPIPS↓), for the runs in Tab. 2.

	OrganAMNIST				LoDoPaB-CT			
# angles	18	30	60	90	18	30	60	90
FBP	.1797	.1383	.1344	.1380	.4786	.4266	.3387	.2930
DPS	.7036	.7038	.7038	.7037	.5984	.5990	.5995	.5996
ΠGDM	.1226	.0517	.0251	.0279	.3738	.2999	.2360	.2200
DDS	.1395	.1213	.1370	.1408	.3916	.3511	.3102	.3071
DiffPIR	.0944	.0458	.0273	.0283	.3674	.3050	.2043	.1505
CM	.0940	.0460	.0269	.0280	.3670	.3046	.2041	.1504
CM4IR	.3855	.3825	.3804	.3800	.8564	.8510	.8478	.8467
PnP-CM	.0732	.0295	.0071	.0068	.3578	.2922	.1816	.1139
RELAY (ours)	**.0517**	**.0162**	**.0013**	**.0008**	**.2438**	**.1921**	**.0967**	**.0436**

Table 5: DC efficiency on M4Raw at R=8 (PSNR in dB). Δ: warm-start gain.

	CG iterations k							
Solver	2	3	4	6	8	12	16	24
T1								
Cold start	23.86	24.09	25.18	26.71	26.97	27.07	27.09	27.11
RELAY (ours)	**24.71**	**25.83**	**26.63**	**26.88**	**27.28**	**27.35**	**27.41**	**27.45**
Δ	+0.85	+1.74	+1.45	+0.17	+0.31	+0.28	+0.32	+0.34
T2								
Cold start	24.92	24.90	25.42	26.01	26.03	25.96	25.98	25.99
RELAY (ours)	**25.49**	**25.79**	**26.23**	**26.11**	**26.32**	**26.35**	**26.36**	**26.38**
Δ	+0.57	+0.89	+0.81	+0.10	+0.29	+0.39	+0.38	+0.39
FLAIR								
Cold start	23.59	23.68	24.19	25.14	25.25	25.22	25.25	25.26
RELAY (ours)	**24.11**	**24.77**	**25.31**	**25.33**	**25.52**	**25.55**	**25.58**	**25.60**
Δ	+0.52	+1.09	+1.12	+0.19	+0.27	+0.33	+0.33	+0.34

4.2 Framework Analysis

Data-Consistency Budget Tab. 5 varies the per-step CG budget k on M4Raw at R=8. RELAY provides its largest gains under tight inner budgets, improving the cold start by 0.52–1.74 dB for $k \leq 4$. More importantly, RELAY with only k=8 already surpasses the cold start at k=24 on all three contrasts, reducing the required inner-iteration budget by 3×. This confirms that cross-step reuse converts the carried DC solution into substantially faster finite-budget convergence.

Hyperparameter Sensitivity Analysis Fig. 6 varies the inner CG budget, NFE, and measurement noise. In Fig. 6a, RELAY with k=10 already surpasses the cold start at k=30, consistent with the iteration savings in Corollary 2. In Fig. 6b, RELAY at only N=2 exceeds the best PnP-CM result over the full NFE sweep. Without re-tuning, RELAY also remains best at every noise level in Fig. 6c.

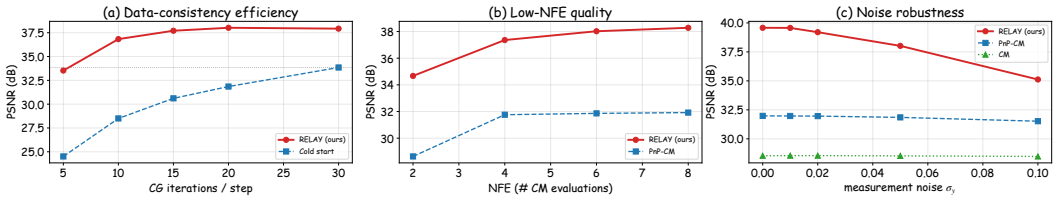


Figure 6: Sensitivity to the CG budget, NFE and measurement noise on OrganAMNIST at 60 angles.

Ablation Study Tab. 6 shows that carrying the previous DC solution is the key factor: at k=20, it improves the cold start by 2.55–7.77 dB across all view counts and consistently outperforms initialization from the prior output or DC anchor. Tab. 7 shows that noise injection and momentum provide little benefit, confirming that the gain comes primarily from cross-step DC reuse. Applying the

same initialization to DDS, with its sampler and budgets unchanged, yields a further 2.37–7.37 dB improvement, indicating that the mechanism transfers beyond RELAY's reconstruction scheme.

Table 6: Warm-start ablation on OrganAMNIST at k=20, within RELAY and in DDS.

	PSNR (dB) ↑			
# angles	18	30	60	90
DC initialization				
Zero (cold start)	26.31	29.20	32.13	32.53
Prior output $\mathbf{x}_{n+1}$	28.10	31.82	36.82	38.44
DC anchor $\mathbf{v}$	28.50	32.49	37.34	38.77
Carried $\mathbf{z}_{\text{prev}}$ (RELAY)	**28.86**	**32.60**	**38.02**	**40.30**
Δ (carried vs. zero)	+2.55	+3.40	+5.89	+7.77
Transfer to DDS				
DDS	23.46	24.33	24.51	24.49
DDS + carried init.	**25.83**	**28.53**	**31.37**	**31.86**

Table 7: Noise injection and momentum on LoDoPaB-CT at 60 angles.

	Noise injection		Momentum	
PSNR (dB)	on	off	on	off
Cold start	33.36	33.59	33.36	33.36
RELAY (ours)	40.36	40.36	40.35	40.35

Table 8: Matrix–vector products↓ to reach a fixed residual over six steps at 60 angles.

Anchor drift	0.005	0.02	0.05	0.15
Cold start	168	168	168	184
Subspace recycling	252	264	265	282
RELAY (ours)	**33**	**93**	**111**	**143**

Comparison with Subspace Recycling Tab. 8 compares RELAY with classical Krylov-subspace recycling. An 8-dimensional recycled subspace (Parks et al., 2006) requires 2–8× more matrix–vector products than RELAY and does not outperform the cold start. RELAY remains most efficient across all drift levels, with its gain over cold start decreasing from 5.1× to 1.3× as drift increases, consistent with Theorem 1. It adds only one image-sized vector of memory.

Residual Trace and Training Budget Fig. 7a shows where the DC saving comes from: at every budget the cold residual stays flat across outer steps, whereas the carried solve is 6 to 10× lower by outer step m=8, as Theorem 1 predicts. Fig. 7b varies the CM distillation budget on LoDoPaB-CT with the solver fixed, and across 2k–38k steps PSNR varies by at most 0.71 dB.

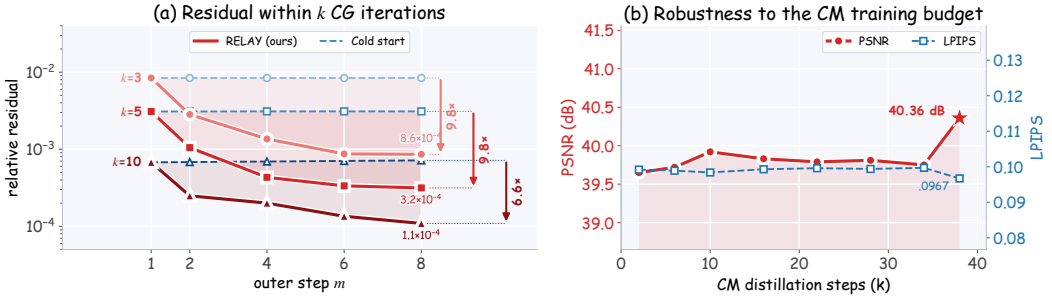


Figure 7: (a) Relative DC residual per outer step at 60 angles, with the anchor drifting by 0.02 per step. (b) PSNR and LPIPS on LoDoPaB-CT at 60 angles against the CM distillation budget.

5 Conclusion

This paper proposes **RELAY**, a training-free reconstruction solver that preserves data-consistency progress across outer steps without increasing the prior-evaluation budget. RELAY carries the preceding DC solution forward, recomputes the exact residual of the current system, and restarts standard CG while leaving the underlying DC problem and its exact solution unchanged. Our analysis shows that the inherited finite-budget error contracts across steps when $q_N < 1$ and that the resulting iteration savings become more pronounced as operator conditioning worsens. Across sparse-view CT and accelerated multi-coil low-field MRI, RELAY achieves the best PSNR among learned baselines across all tested settings, with substantially fewer CG iterations than a cold start.

AI USE STATEMENT

We used LLMs for polishing the language of the manuscript, for writing the plotting code that turns the recorded results into the figures of this paper, and for checking the consistency of the mathematical notation across the text, the equations and the algorithm. All experiments, results and claims are the authors' own, and the authors verified every figure and every statement produced with this assistance.

REPRODUCIBILITY STATEMENT

All three benchmarks are public: OrganAMNIST from MedMNIST v2 (Yang et al., 2023), LoDoPaB-CT (Leuschner et al., 2021) and M4Raw (Lyu et al., 2023). Section 4 and Appendix B give the prior architecture and training schedule, the acquisition and noise settings of both modalities, the evaluation sets and reconstruction budgets, the definition of every reported metric, and the configuration of each baseline, with the ADMM penalty schedule taken from the public PnP-CM implementation (Gulle et al., 2026). The assumptions of the theory are stated in Appendix A.1, and complete proofs of every claim are given in Appendix A.3.

REFERENCES

Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE transactions on image processing*, 19(9): 2345–2356, 2010.

Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging*, 38(2):394–405, 2018.

Rizwan Ahmad, Charles A Bouman, Gregery T Buzzard, Stanley Chan, Sizhuo Liu, Edward T Reehorst, and Philip Schniter. Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery. *IEEE signal processing magazine*, 37(1):105–116, 2020.

Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, 2020.

Riccardo Barbano, Alexander Denker, Hyungjin Chung, Tae Hoon Roh, Simon Arridge, Peter Maass, Bangti Jin, and Jong Chul Ye. Steerable conditional diffusion for out-of-distribution adaptation in medical image reconstruction. *IEEE Transactions on Medical Imaging*, 44(5):2093–2104, 2025.

Amir Beck and Marc Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE transactions on image processing*, 18(11):2419–2434, 2009.

Kai Tobias Block, Martin Uecker, and Jens Frahm. Undersampled radial mri with multiple coils. iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 57(6):1086–1098, 2007.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play admm for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1): 84–98, 2016.

Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *International conference on learning representations*, volume 2024, pp. 38922–38949, 2024.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

Tomer Garber and Tom Tirer. Zero-shot image restoration using few-step guidance of consistency models (and beyond). In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2398–2407. IEEE, 2025.

Merve Gülle, Junno Yun, Yaşar Utku Alçalar, and Mehmet Akçakaya. Consistency models for fast MRI using regularization by denoising. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2026.

Merve Gulle, Junno Yun, Yasar Utku Alçalar, and Mehmet Akçakaya. Pnp-cm: Consistency models as plug-and-play priors for inverse problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 38464–38474, 2026.

Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine*, 79(6):3055–3071, 2018.

Kerstin Hammernik, Thomas Küstner, Burhaneddin Yaman, Zhengnan Huang, Daniel Rueckert, Florian Knoll, and Mehmet Akçakaya. Physics-driven deep learning for computational magnetic resonance imaging: Combining physics and machine learning for improved medical imaging. *IEEE signal processing magazine*, 40(1):98–114, 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. Robust compressed sensing mri with deep generative priors. *Advances in neural information processing systems*, 34:14938–14954, 2021.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Florian Knoll, Kerstin Hammernik, Erich Kobler, Thomas Pock, Michael P Recht, and Daniel K Sodickson. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magnetic resonance in medicine*, 81(1):116–128, 2019.

Johannes Leuschner, Maximilian Schmidt, Daniel Otero Baguer, and Peter Maass. Lodopab-ct, a benchmark dataset for low-dose computed tomography reconstruction. *Scientific Data*, 8(1):109, 2021.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

Mengye Lyu, Lifeng Mei, Shoujin Huang, Sixing Liu, Yi Li, Kexin Yang, Yilong Liu, Yu Dong, Linzheng Dong, and Ed X Wu. M4raw: A multi-contrast, multi-repetition, multi-channel mri k-space dataset for low-field mri research. *Scientific Data*, 10(1):264, 2023.

Pravin Nair, Ruturaj G Gavaskar, and Kunal Narayan Chaudhury. Fixed-point and objective convergence of plug-and-play algorithms. *IEEE Transactions on Computational Imaging*, 7:337–348, 2021.

Frank Ong and Michael Lustig. SigPy: A python package for high performance iterative reconstruction. In *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*, pp. 4819, 2019.

Chicago Y Park, Yuyang Hu, Michael T McCann, Cristina Garcia-Cardona, Brendt Wohlberg, and Ulugbek S Kamilov. Plug-and-play priors as a score-based method. In *2025 IEEE International Conference on Image Processing (ICIP)*, pp. 49–54. IEEE, 2025.

Michael L Parks, Eric De Sturler, Greg Mackey, Duane D Johnson, and Spandan Maiti. Recycling krylov subspaces for sequences of linear systems. *SIAM Journal on Scientific Computing*, 28(5): 1651–1674, 2006.

Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, and Peter Boesiger. Sense: sensitivity encoding for fast mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

Ernest Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Plug-and-play methods provably converge with properly trained denoisers. In *International Conference on Machine Learning*, pp. 5546–5557. PMLR, 2019.

Yousef Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022.

Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17):4777–4807, 2008.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International conference on learning representations*, 2023a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. *arXiv preprint arXiv:2111.08005*, 2021.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. *arXiv preprint arXiv:2303.01469*, 2023b.

Martin Uecker, Peng Lai, Mark J Murphy, Patrick Virtue, Michael Elad, John M Pauly, Shreyas S Vasanawala, and Michael Lustig. Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa. *Magnetic resonance in medicine*, 71(3):990–1001, 2014.

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pp. 945–948. IEEE, 2013.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. *Advances in Neural Information Processing Systems*, 37:118389–118427, 2024.

Xingyu Xu and Yuejie Chi. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. *Advances in Neural Information Processing Systems*, 37:36148–36184, 2024.

Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic resonance in medicine*, 84(6):3172–3191, 2020.

Burhaneddin Yaman, Seyed Amir Hossein Hosseini, and Mehmet Akçakaya. Zero-shot self-supervised learning for mri reconstruction. *arXiv preprint arXiv:2102.07737*, 2021.

Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific data*, 10(1):41, 2023.

Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6613–6623. IEEE, 2024.

Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018.

Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1828–1837. IEEE, 2018.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 586–595. IEEE, 2018.

Jiankun Zhao, Bowen Song, and Liyue Shen. Cosign: Few-step guidance of consistency model to solve general inverse problems. In *European Conference on Computer Vision*, pp. 108–126. Springer, 2024.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1219–1229. IEEE, 2023.

Roadmap Appendix A develops the theory of Section 3: the setup and assumptions (Appendix A.1), the full statements behind Theorem 1 and Corollary 2 together with a reading of the bound (Appendix A.2), and all proofs (Appendix A.3). Appendix B details the prior pretraining, the acquisition, noise and budget settings of both modalities, the diagnostics behind the solver analyses, the evaluation metrics and the baselines. Appendix C reports additional results: the DC analysis behind Section 4.2 (Appendix C.1), classical compressed sensing and operator cost (Appendix C.2), robustness to the source domain of the prior (Appendix C.3), and qualitative comparisons on sparse-view CT and low-field MRI (Appendix C.4).

A Theory

A.1 Setup and Assumptions

Let $m = N - n$ index the executed outer steps, so the loop of Alg. 1 runs $m = 1, \dots, N$ with penalty $\beta_m := \rho_{n+1} > 0$ and DC anchor $\mathbf{v}_m := \mathbf{x}_{n+1} - \mathbf{u}_{n+1}$, and works with

$$\mathbf{M}_m := \mathcal{A}^*\mathcal{A} + \beta_m \mathbf{I}, \qquad \mathbf{b}_m := \mathcal{A}^*\mathbf{y} + \beta_m \mathbf{v}_m, \qquad \mathbf{z}_m^\star := \mathbf{M}_m^{-1}\mathbf{b}_m. \tag{10}$$

Write ζ_m for the iterate stored by line 6, with $\zeta_0 = \mathbf{0}$, and $\|\mathbf{w}\|_{\mathbf{M}}^2 := \mathbf{w}^*\mathbf{M}\mathbf{w}$ for the energy norm, so that

$$\varepsilon_m := \|\mathbf{z}_m^\star - \zeta_m\|_{\mathbf{M}_m} \tag{11}$$

is the error the truncated solve leaves at outer step m. Both $\mathbf{M}_m$ and its norm change with m whenever the schedule $\{\rho_n\}$ is not constant, and the analysis tracks that dependence explicitly. We assume **(H1)** $\mathcal{A}$ is linear with $\beta_m > 0$ for every m, so $\mathbf{M}_m \succ \mathbf{0}$ and $\|\cdot\|_{\mathbf{M}_m}$ is a norm, and **(H2)** CG is run in exact arithmetic with $k \geq 1$ iterations, so for $\mathbf{M} \succ \mathbf{0}$ the classical energy-norm bound

$$\|\mathbf{w}^\star - \mathbf{w}_k\|_{\mathbf{M}} \;\leq\; 2c^{\,k}\,\|\mathbf{w}^\star - \mathbf{w}_0\|_{\mathbf{M}}, \qquad c = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}, \quad \kappa = \mathrm{cond}(\mathbf{M}), \tag{12}$$

applies (Saad, 2003, §6.11). Since $\kappa_m \leq (\|\mathcal{A}\|_2^2 + \beta_m)/\beta_m < \infty$ for every $\beta_m > 0$, we have $c_m < 1$ strictly.

For the 60-angle operator of Section 4.2, $\lambda_{\max}(\mathcal{A}^*\mathcal{A}) = 3.43 \times 10^3$, so every penalty $\beta_m \geq 5$ gives $\kappa_m \leq 687$ and $2c_m^{20} \leq 0.44$. The condition $q_N < 1$ therefore holds at k=20 for any such schedule whose penalty rises by less than a factor of 5 per step.

A.2 Full Statements

Proofs of all statements are collected in Appendix A.3.

Proposition 3 (Exactness)**.** *Let $\mathbf{M} \succ \mathbf{0}$, let $\mathbf{b}$ and $\mathbf{z}_{\mathrm{prev}}$ be arbitrary, and set $\mathbf{r} = \mathbf{b} - \mathbf{M}\mathbf{z}_{\mathrm{prev}}$. If $\Delta\mathbf{z}$ satisfies $\mathbf{M}\,\Delta\mathbf{z} = \mathbf{r}$, then $\mathbf{z} = \mathbf{z}_{\mathrm{prev}} + \Delta\mathbf{z}$ satisfies $\mathbf{M}\mathbf{z} = \mathbf{b}$. If $\mathbf{z}_{\mathrm{prev}} = \mathbf{0}$, then $\mathbf{r} = \mathbf{b}$ and $\mathbf{z} = \Delta\mathbf{z}$.*

Lemma 4 (Consecutive energy norms)**.** *For every $\mathbf{w}$ and every $m \geq 2$,*

$$\begin{aligned} \|\mathbf{w}\|_{\mathbf{M}_m} \;&\leq\; \theta_m\,\|\mathbf{w}\|_{\mathbf{M}_{m-1}}, \\ \theta_m &:= \max\Big\{1,\ \sqrt{\beta_m/\beta_{m-1}}\Big\}. \end{aligned} \tag{13}$$

Lemma 5 (Drift of the exact solution)**.** *Let $\Delta\mathbf{v}_m := \mathbf{v}_m - \mathbf{v}_{m-1}$ and $\Delta\beta_m := \beta_m - \beta_{m-1}$. For every $m \geq 2$,*

$$\begin{aligned} \|\mathbf{z}_m^\star - \mathbf{z}_{m-1}^\star\|_{\mathbf{M}_m} \;&\leq \sqrt{\beta_m}\,\|\Delta\mathbf{v}_m\|_2 \\ &\quad + \frac{|\Delta\beta_m|}{\sqrt{\beta_m}}\,\|\mathbf{v}_{m-1} - \mathbf{z}_{m-1}^\star\|_2. \end{aligned} \tag{14}$$

We write D_m for the right-hand side of Eq. 14.

Proposition 6 (Geometric envelope)**.** *Under (H1)–(H2), suppose further that $N \geq 2$ and set*

$$q_N := \max_{2 \leq j \leq N} 2c_j^k\theta_j, \quad \bar{D}_N := \max_{2 \leq j \leq N} 2c_j^k D_j. \tag{15}$$

If $q_N < 1$, then Alg. 1 satisfies, for every $2 \leq m \leq N$,

$$\varepsilon_m \;\leq\; q_N^{m-1}\varepsilon_1 + \frac{\bar{D}_N}{1 - q_N}. \tag{16}$$

For a solve whose starting error in $\|\cdot\|_{\mathbf{M}_m}$ is $X > 0$, the budget $k^\star(X)$ of Corollary 2 is

$$k^\star(X) \;=\; \max\Big\{1,\; \Big\lceil \frac{\ln(2X/\varepsilon)}{\ln(1/c_m)} \Big\rceil \Big\}. \tag{17}$$

Interpreting the Bound Comparing the warm bound Eq. 8 with the cold bound of Theorem 1, the warm guarantee improves on the cold one at outer step m exactly when

$$\theta_m\, \varepsilon_{m-1} + D_m \;<\; \|\mathbf{z}_m^\star\|_{\mathbf{M}_m}, \tag{18}$$

that is, when the carried iterate lies closer to the new optimum than the origin does, the formal content of the statement that the ADMM iterates change slowly. The cold bound resets to $\|\mathbf{z}_m^\star\|_{\mathbf{M}_m}$ at every outer step, whereas Eq. 16 settles the warm error at a level governed by the drift term $\bar{D}_N$ rather than by the size of the solution, the sense in which solver progress carries across outer steps. The per-step gain requires only Eq. 18, so it extends beyond the uniform envelope $q_N < 1$, and the certified savings are conservative: Eq. 12 depends on $\mathbf{M}_m$ only through κ_m whereas CG convergence depends on the full spectrum, so the measured savings exceed the bound. Fig. 8 visualizes the bound on a run with a fixed penalty at k=5: the cold iterate stays on an outer level set of the shared objective, whereas the carried iterate moves inward and reaches a 27× smaller objective gap by outer step m=8.

A.3 Proofs

Proof of Proposition 3. Since $\mathbf{M}$ acts linearly,

$$\begin{aligned} \mathbf{M}\mathbf{z} &= \mathbf{M}\big(\mathbf{z}_{\text{prev}} + \Delta\mathbf{z}\big) = \mathbf{M}\mathbf{z}_{\text{prev}} + \mathbf{M}\,\Delta\mathbf{z} \\ &= \mathbf{M}\mathbf{z}_{\text{prev}} + \mathbf{r} \\ &= \mathbf{M}\mathbf{z}_{\text{prev}} + \mathbf{b} - \mathbf{M}\mathbf{z}_{\text{prev}} = \mathbf{b}, \end{aligned} \tag{19}$$

using the hypothesis $\mathbf{M}\,\Delta\mathbf{z} = \mathbf{r}$ and then the definition of $\mathbf{r}$. For the second claim, setting $\mathbf{z}_{\text{prev}} = \mathbf{0}$ in that definition gives $\mathbf{r} = \mathbf{b} - \mathbf{M}\mathbf{0} = \mathbf{b}$, so the system solved for $\Delta\mathbf{z}$ is $\mathbf{M}\,\Delta\mathbf{z} = \mathbf{b}$, and $\mathbf{z} = \mathbf{0} + \Delta\mathbf{z} = \Delta\mathbf{z}$. □

Proof of Lemma 4. Fix $\mathbf{w}$ and $m \geq 2$. Expanding the quadratic form for an arbitrary index j,

$$\begin{aligned} \|\mathbf{w}\|_{\mathbf{M}_j}^2 &= \mathbf{w}^*\big(\mathcal{A}^*\mathcal{A} + \beta_j \mathbf{I}\big)\mathbf{w} \\ &= \mathbf{w}^*\mathcal{A}^*\mathcal{A}\mathbf{w} + \beta_j\, \mathbf{w}^*\mathbf{w} \\ &= \|\mathcal{A}\mathbf{w}\|_2^2 + \beta_j \|\mathbf{w}\|_2^2. \end{aligned} \tag{20}$$

Write Eq. 20 at $j = m$ and at $j = m-1$ and subtract the second from the first. The term $\|\mathcal{A}\mathbf{w}\|_2^2$ is common to both and cancels, leaving

$$\|\mathbf{w}\|_{\mathbf{M}_m}^2 = \|\mathbf{w}\|_{\mathbf{M}_{m-1}}^2 + (\beta_m - \beta_{m-1})\|\mathbf{w}\|_2^2. \tag{21}$$

Suppose first $\beta_m \leq \beta_{m-1}$. The second term of Eq. 21 is then non-positive, so $\|\mathbf{w}\|_{\mathbf{M}_m}^2 \leq \|\mathbf{w}\|_{\mathbf{M}_{m-1}}^2$, hence $\|\mathbf{w}\|_{\mathbf{M}_m} \leq \|\mathbf{w}\|_{\mathbf{M}_{m-1}}$, which is Eq. 13 with $\theta_m = 1$.

Suppose instead $\beta_m > \beta_{m-1}$. Evaluating Eq. 20 at $j = m-1$ and discarding the non-negative term $\|\mathcal{A}\mathbf{w}\|_2^2$ gives $\|\mathbf{w}\|_{\mathbf{M}_{m-1}}^2 \geq \beta_{m-1}\|\mathbf{w}\|_2^2$. As $\beta_{m-1} > 0$ by (H1), this rearranges to

$$\|\mathbf{w}\|_2^2 \;\leq\; \frac{1}{\beta_{m-1}}\, \|\mathbf{w}\|_{\mathbf{M}_{m-1}}^2. \tag{22}$$

The coefficient $\beta_m - \beta_{m-1}$ multiplying $\|\mathbf{w}\|_2^2$ in Eq. 21 is positive in this case, so substituting Eq. 22 gives

$$\begin{aligned} \|\mathbf{w}\|_{\mathbf{M}_m}^2 \;&\leq\; \|\mathbf{w}\|_{\mathbf{M}_{m-1}}^2 + \frac{\beta_m - \beta_{m-1}}{\beta_{m-1}} \|\mathbf{w}\|_{\mathbf{M}_{m-1}}^2 \\ &=\; \frac{\beta_m}{\beta_{m-1}}\, \|\mathbf{w}\|_{\mathbf{M}_{m-1}}^2. \end{aligned} \tag{23}$$

Taking square roots gives $\|\mathbf{w}\|_{\mathbf{M}_m} \leq \sqrt{\beta_m/\beta_{m-1}}\, \|\mathbf{w}\|_{\mathbf{M}_{m-1}}$.

The two cases are exhaustive, and the constant obtained in each is the corresponding branch of the maximum defining θ_m. □

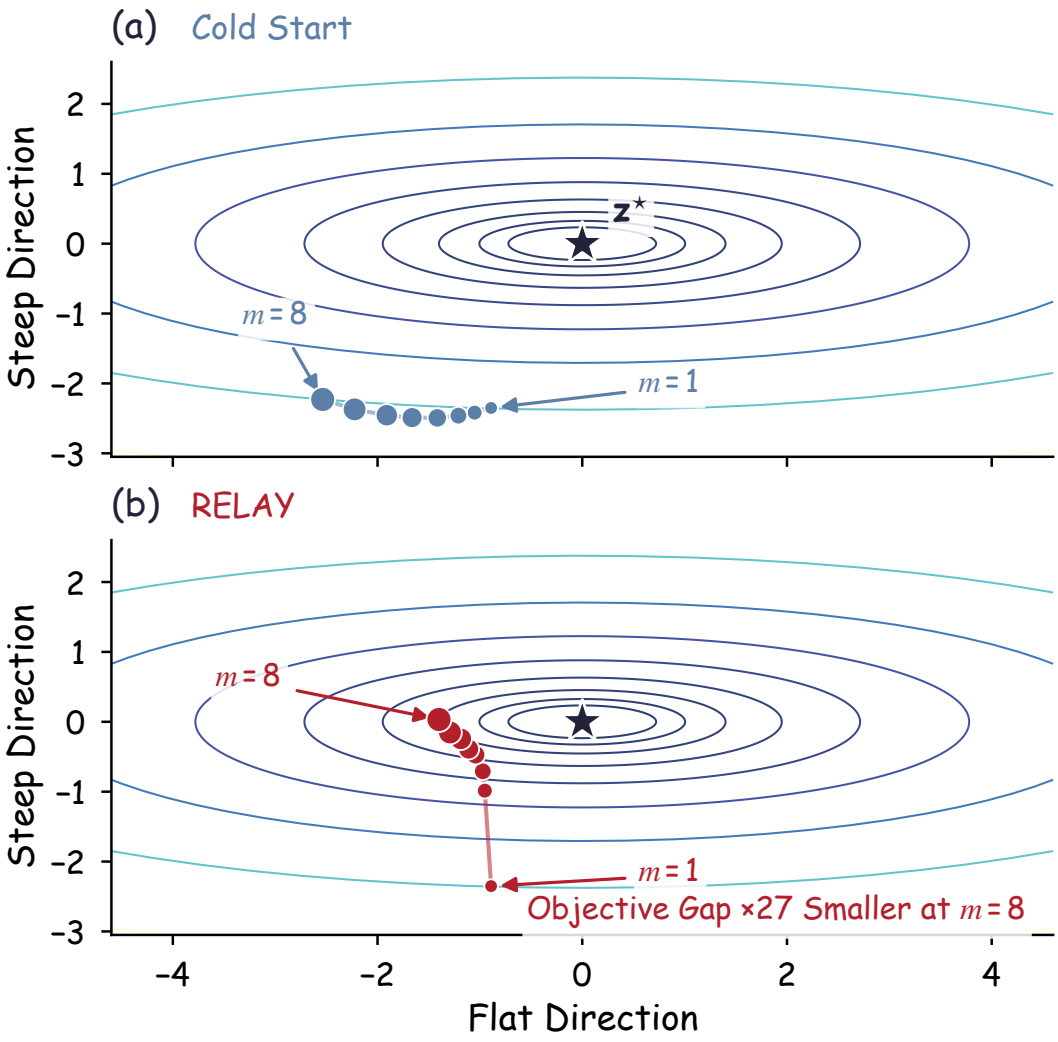


Figure 8: DC error over eight outer steps of a 60-angle reconstruction, in the plane spanned by the two principal directions of $\mathbf{M} = \mathcal{A}^*\mathcal{A} + \rho\mathbf{I}$ with the penalty held fixed. (a) Every step restarts from zero. (b) Every step starts from the carried iterate, annotated with the final objective-gap ratio. Marker size grows with m.

Proof of Lemma 5. Fix $m \geq 2$. The two exact solutions solve systems that differ in both the matrix and the right-hand side, so the first task is to write their difference as a single system in $\mathbf{M}_m$. By Eq. 10 at index $m-1$ we have $\mathbf{M}_{m-1}\mathbf{z}^\star_{m-1} = \mathbf{b}_{m-1}$, and the definition of $\mathbf{M}_m$ gives $\mathbf{M}_m = \mathbf{M}_{m-1} + \Delta\beta_m\mathbf{I}$. Applying the latter to $\mathbf{z}^\star_{m-1}$,

$$\mathbf{M}_m\mathbf{z}^\star_{m-1} = \mathbf{M}_{m-1}\mathbf{z}^\star_{m-1} + \Delta\beta_m\mathbf{z}^\star_{m-1} = \mathbf{b}_{m-1} + \Delta\beta_m\mathbf{z}^\star_{m-1}. \tag{24}$$

Subtracting Eq. 24 from $\mathbf{M}_m\mathbf{z}^\star_m = \mathbf{b}_m$,

$$\mathbf{M}_m\big(\mathbf{z}^\star_m - \mathbf{z}^\star_{m-1}\big) = \mathbf{b}_m - \mathbf{b}_{m-1} - \Delta\beta_m\mathbf{z}^\star_{m-1}. \tag{25}$$

Substitute $\mathbf{b}_j = \mathcal{A}^*\mathbf{y} + \beta_j\mathbf{v}_j$ for $j = m$ and $j = m-1$. The term $\mathcal{A}^*\mathbf{y}$ is common to both and cancels. Adding and subtracting $\beta_m\mathbf{v}_{m-1}$ then separates the change in the anchor from the change in the penalty,

$$\begin{aligned}\mathbf{b}_m - \mathbf{b}_{m-1} &= \beta_m\mathbf{v}_m - \beta_{m-1}\mathbf{v}_{m-1} \\ &= \beta_m(\mathbf{v}_m - \mathbf{v}_{m-1}) + (\beta_m - \beta_{m-1})\mathbf{v}_{m-1} \\ &= \beta_m\Delta\mathbf{v}_m + \Delta\beta_m\mathbf{v}_{m-1}.\end{aligned} \tag{26}$$

Inserting Eq. 26 into Eq. 25 and collecting the two $\Delta\beta_m$ terms,

$$\begin{aligned}\mathbf{M}_m\big(\mathbf{z}^\star_m - \mathbf{z}^\star_{m-1}\big) &= \mathbf{g}_m,\\ \mathbf{g}_m &:= \beta_m\Delta\mathbf{v}_m + \Delta\beta_m\big(\mathbf{v}_{m-1} - \mathbf{z}^\star_{m-1}\big).\end{aligned} \tag{27}$$

By (H1) $\mathbf{M}_m$ is invertible, so $\mathbf{z}^\star_m - \mathbf{z}^\star_{m-1} = \mathbf{M}_m^{-1}\mathbf{g}_m$.

It remains to bound this in $\|\cdot\|_{\mathbf{M}_m}$. Since $\mathbf{M}_m$ is Hermitian,

$$\begin{aligned}\big\|\mathbf{M}_m^{-1}\mathbf{g}_m\big\|^2_{\mathbf{M}_m} &= \big(\mathbf{M}_m^{-1}\mathbf{g}_m\big)^*\mathbf{M}_m\big(\mathbf{M}_m^{-1}\mathbf{g}_m\big)\\ &= \mathbf{g}_m^*\mathbf{M}_m^{-1}\mathbf{M}_m\mathbf{M}_m^{-1}\mathbf{g}_m = \mathbf{g}_m^*\mathbf{M}_m^{-1}\mathbf{g}_m.\end{aligned} \tag{28}$$

Since $\mathbf{M}_m \succ \mathbf{0}$, so is $\mathbf{M}_m^{-1}$, and the Rayleigh quotient bound gives $\mathbf{g}_m^*\mathbf{M}_m^{-1}\mathbf{g}_m \le \lambda_{\max}(\mathbf{M}_m^{-1})\|\mathbf{g}_m\|_2^2$ with $\lambda_{\max}(\mathbf{M}_m^{-1}) = 1/\lambda_{\min}(\mathbf{M}_m)$. The eigenvalues of $\mathbf{M}_m = \mathcal{A}^*\mathcal{A} + \beta_m\mathbf{I}$ are those of $\mathcal{A}^*\mathcal{A}$ each shifted by β_m, and $\mathcal{A}^*\mathcal{A} \succeq \mathbf{0}$ gives $\lambda_{\min}(\mathcal{A}^*\mathcal{A}) \ge 0$, so

$$\lambda_{\min}(\mathbf{M}_m) = \lambda_{\min}(\mathcal{A}^*\mathcal{A}) + \beta_m \;\ge\; \beta_m. \tag{29}$$

Combining Eq. 28, the Rayleigh bound and Eq. 29, then taking square roots,

$$\|\mathbf{z}^\star_m - \mathbf{z}^\star_{m-1}\|_{\mathbf{M}_m} \;\le\; \frac{\|\mathbf{g}_m\|_2}{\sqrt{\beta_m}}. \tag{30}$$

By the triangle inequality applied to Eq. 27,

$$\begin{aligned}\|\mathbf{g}_m\|_2 \;&\le \beta_m\|\Delta\mathbf{v}_m\|_2\\ &\quad + |\Delta\beta_m|\,\|\mathbf{v}_{m-1} - \mathbf{z}^\star_{m-1}\|_2.\end{aligned} \tag{31}$$

Dividing by $\sqrt{\beta_m} > 0$ and substituting the result into Eq. 30 gives the bound Eq. 14. □

Proof of Theorem 1 and Proposition 6. Fix m with $2 \le m \le N$. Line 4 of Alg. 1 forms $\mathbf{r}_m = \mathbf{b}_m - \mathbf{M}_m\zeta_{m-1}$ and line 5 runs k CG steps on $\mathbf{M}_m\,\Delta\mathbf{z} = \mathbf{r}_m$ from $\Delta\mathbf{z}_0 = \mathbf{0}$. As $\mathbf{M}_m$ is invertible by (H1), that system has the unique solution

$$\begin{aligned}\Delta\mathbf{z}^\star_m &= \mathbf{M}_m^{-1}\mathbf{r}_m = \mathbf{M}_m^{-1}\big(\mathbf{b}_m - \mathbf{M}_m\zeta_{m-1}\big)\\ &= \mathbf{M}_m^{-1}\mathbf{b}_m - \zeta_{m-1} = \mathbf{z}^\star_m - \zeta_{m-1},\end{aligned} \tag{32}$$

the last equality by the definition of $\mathbf{z}^\star_m$ in Eq. 10. Line 6 sets $\zeta_m = \zeta_{m-1} + \Delta\mathbf{z}_k$, where $\Delta\mathbf{z}_k$ denotes the k-th CG iterate, so subtracting,

$$\mathbf{z}^\star_m - \zeta_m = \big(\mathbf{z}^\star_m - \zeta_{m-1}\big) - \Delta\mathbf{z}_k = \Delta\mathbf{z}^\star_m - \Delta\mathbf{z}_k. \tag{33}$$

The error the outer step leaves therefore coincides with the CG error of the inner solve. Applying Eq. 12 to the inner solve with $\mathbf{M} = \mathbf{M}_m$ and $\mathbf{w}_0 = \mathbf{0}$,

$$\begin{aligned}\varepsilon_m &\overset{(33)}{=} \|\Delta\mathbf{z}^\star_m - \Delta\mathbf{z}_k\|_{\mathbf{M}_m} \;\le\; 2c_m^k\|\Delta\mathbf{z}^\star_m - \mathbf{0}\|_{\mathbf{M}_m}\\ &\overset{(32)}{=} 2c_m^k\|\mathbf{z}^\star_m - \zeta_{m-1}\|_{\mathbf{M}_m}.\end{aligned} \tag{34}$$

Inserting $\mathbf{z}^\star_{m-1}$ and applying the triangle inequality in $\|\cdot\|_{\mathbf{M}_m}$,

$$\|\mathbf{z}^\star_m - \zeta_{m-1}\|_{\mathbf{M}_m} \;\le\; \|\mathbf{z}^\star_m - \mathbf{z}^\star_{m-1}\|_{\mathbf{M}_m} + \|\mathbf{z}^\star_{m-1} - \zeta_{m-1}\|_{\mathbf{M}_m}. \tag{35}$$

Lemma 5 bounds the first summand by D_m. The second is measured in $\|\cdot\|_{\mathbf{M}_m}$, whereas ε_{m-1} is defined in $\|\cdot\|_{\mathbf{M}_{m-1}}$ by Eq. 11. These differ whenever $\beta_m \ne \beta_{m-1}$. Lemma 4 applied to $\mathbf{w} = \mathbf{z}^\star_{m-1} - \zeta_{m-1}$ converts one into the other,

$$\|\mathbf{z}^\star_{m-1} - \zeta_{m-1}\|_{\mathbf{M}_m} \;\le\; \theta_m\,\|\mathbf{z}^\star_{m-1} - \zeta_{m-1}\|_{\mathbf{M}_{m-1}} = \theta_m\,\varepsilon_{m-1}. \tag{36}$$

Substituting Eq. 35, Lemma 5 and Eq. 36 into Eq. 34 gives the recursion in Eq. 8.

For $m = 1$, the initialization of Alg. 1 sets $\zeta_0 = \mathbf{0}$, so Eq. 32 reads $\Delta\mathbf{z}^\star_1 = \mathbf{z}^\star_1$ and Eq. 34 becomes $\varepsilon_1 \le 2c_1^k\|\mathbf{z}^\star_1\|_{\mathbf{M}_1}$. A cold start is by definition the variant in which ζ_{m-1} is replaced by $\mathbf{0}$ at every m. Repeating the computation Eq. 32–Eq. 34 with that substitution gives $\varepsilon_m^{\text{cold}} \le 2c_m^k\|\mathbf{z}^\star_m\|_{\mathbf{M}_m}$ for every m, with no term inherited from outer step $m-1$.

For the closed form, let $N \geq 2$ and take $q_N, \bar{D}_N$ as in Eq. 15. Assume $q_N < 1$. Since $2c_j^k \theta_j \leq q_N$ and $2c_j^k D_j \leq \bar{D}_N$ for every such j, the recursion gives $\varepsilon_j \leq q_N \varepsilon_{j-1} + \bar{D}_N$ for $2 \leq j \leq N$. Unrolling this $m-1$ times, since $q_N \geq 0$,

$$\begin{aligned} \varepsilon_m &\leq q_N^{m-1}\varepsilon_1 + \bar{D}_N \sum_{j=0}^{m-2} q_N^j \leq q_N^{m-1}\varepsilon_1 + \bar{D}_N \sum_{j=0}^{\infty} q_N^j \\ &= q_N^{m-1}\varepsilon_1 + \frac{\bar{D}_N}{1-q_N}, \end{aligned} \tag{37}$$

which is Eq. 16. □

Proof of Corollary 2. Since $\kappa_m > 1$ we have $c_m \in (0,1)$, so $\ln c_m = -\ln(1/c_m) < 0$. For $X, \varepsilon > 0$,

$$2c_m^k X \leq \varepsilon \iff k \ln c_m \leq \ln \frac{\varepsilon}{2X} \iff k \geq \frac{\ln(2X/\varepsilon)}{\ln(1/c_m)}, \tag{38}$$

and the smallest integer satisfying this together with $k \geq 1$ from (H2) is Eq. 17.

For Eq. 9, write $A(X) = \ln(2X/\varepsilon)/\ln(1/c_m)$, which is increasing in X, and set $a = A(X_m^{\mathrm{c}})$, $b = A(X_m^{\mathrm{w}})$, so $a \geq b$ and $a - b = \ln(X_m^{\mathrm{c}}/X_m^{\mathrm{w}})/\ln(1/c_m)$. The hypothesis $2X_m^{\mathrm{w}} \geq \varepsilon$ says $b \geq 0$. Three cases cover $b \geq 0$. If $b \geq 1$, neither maximum in Eq. 17 binds and $k^\star(X_m^{\mathrm{c}}) - k^\star(X_m^{\mathrm{w}}) = \lceil a \rceil - \lceil b \rceil \geq a - (b+1)$, using $\lceil a \rceil \geq a$ and $\lceil b \rceil < b+1$. If $a \leq 1$, both sides of Eq. 17 clamp to 1, the left side of Eq. 9 is 0, and $a - b - 1 \leq a - 1 \leq 0$ because $b \geq 0$. If $b < 1 < a$, then $k^\star(X_m^{\mathrm{w}}) = 1$ and $k^\star(X_m^{\mathrm{c}}) = \lceil a \rceil \geq a$, so the difference is at least $a - 1 \geq a - b - 1$, again because $b \geq 0$. This proves the first inequality of Eq. 9.

For the second, $1/c_m = (\sqrt{\kappa_m}+1)/(\sqrt{\kappa_m}-1) = 1 + 2/(\sqrt{\kappa_m}-1)$, and $\ln(1+x) \leq x$ for $x > -1$ applied with $x = 2/(\sqrt{\kappa_m}-1) > 0$ gives $\ln(1/c_m) \leq 2/(\sqrt{\kappa_m}-1)$, hence $1/\ln(1/c_m) \geq (\sqrt{\kappa_m}-1)/2$. Substituting into the first inequality of Eq. 9 completes the proof. □

B Experimental Details

Prior Pretraining Each benchmark has a single unconditional generative model trained in advance on its own training split. The backbone is the ADM U-Net (Dhariwal & Nichol, 2021), and training proceeds in two stages, EDM teacher training (Karras et al., 2022) followed by consistency distillation (Song et al., 2023b) into the model that serves as the prior update. Tab. 9 lists the settings that differ across the three benchmarks, and the rest are shared: AdamW (Loshchilov & Hutter, 2017) with weight decay set to zero and a constant learning rate, mixed-precision training in bfloat16, EDM preconditioning with σ_{data}=0.5, $\sigma \in [0.002, 80]$ and ρ_{EDM}=7, and a lognormal noise sampler with P_{mean}=−1.2, P_{std}=1.2. Distillation minimizes the squared distance with the weighting of Karras et al. (2022) between the student and its target network on a Heun step of the teacher ODE, with the target updated by an exponential moving average with decay 0.95 and the number of discretization steps rising from 2 to 150 over a 100k-step schedule. Training ran on two RTX PRO 6000 (96 GB) GPUs.

Acquisition and Noise The CT operator uses equally spaced angles over $[0°, 180°)$ and one detector bin per image column. Gaussian noise is added to the sinogram, $\mathbf{y} \leftarrow \mathbf{y} + \sigma_y \boldsymbol{\epsilon}$, with σ_y=0.05 throughout except in Tab. 15, which uses σ_y=0.2 on OrganAMNIST and σ_y=0.8 on LoDoPaB-CT, both close to 1% of the sinogram RMS of the respective benchmark. The MRI mask draws phase encodes from a Gaussian profile of width $N_{\text{pe}}/4$ centred on the k-space origin and keeps a fully sampled centre of 24, 12 and 8 lines at R=4, 8 and 12.

Reconstruction Budgets We evaluate on 50 OrganAMNIST slices, 100 LoDoPaB-CT slices and 40 M4Raw slices per contrast, and the ablations reuse the same sets. CT runs use N=6 outer steps with k=20 CG iterations, MRI runs N=4 with k=10, and the sweep of Tab. 5 varies k at N=4. The penalty ρ_n rises across the outer steps following the public implementation (Gulle et al., 2026), with its task-specific endpoints for CT and for MRI at R=4 and $R \in \{8, 12\}$.

Table 9: Prior pretraining for the three benchmarks. Batch sizes are per GPU, and every run uses two GPUs with data parallelism.

	Organ{A,C,S}	LoDoPaB-CT	M4Raw
Architecture			
Base width	128	256	256
Width multipliers	$1, 2, 3, 4$	$1, 1, 2, 2, 4, 4$	
Residual blocks	2	2	2
Attention (downsampling)	$16, 8$	$32, 16, 8$	
Image channels	1	1	2
Teacher			
Steps	12k	200k	20k[†]
Batch size, LR	64, 10^{-4}	16, 10^{-4}	12, $2{\cdot}10^{-5}$
Distillation			
Steps	6k	38k	12k
Batch size, LR	64, 10^{-4}	12, 10^{-4}	12, 10^{-5}
Training images	61,468	35,840	12,288

[†]Fine-tuning of a public fastMRI brain model (Gülle et al., 2026).

Classical Compressed Sensing On CT both solvers run FISTA (Beck & Teboulle, 2009), the total-variation step using Chambolle's dual projection and the wavelet step an orthogonal Daubechies-4 transform with soft thresholding. The regularization weight is swept over $\{3{\cdot}10^{-4}, 10^{-3}, 3{\cdot}10^{-3}, 10^{-2}, 3{\cdot}10^{-2}\}$ and selected per setting. Budget-matched runs stop after 138 iterations and converged runs after 3200. On MRI both solvers are the SigPy reconstruction apps (Ong & Lustig, 2019), with the wavelet weight 0.008 and the total-variation weight tuned per contrast.

Inner Budgets of the Baselines Each baseline keeps its own data step, so the operator counts of Tab. 14 do not follow RELAY's k: DDS runs five CG iterations per sampling step, whereas DiffPIR and CM apply a single closed-form or regularized update.

Solver Diagnostics Fig. 7a, Fig. 8 and Tab. 8 run the DC solve alone on an ellipse phantom of 64×64 pixels at the fixed penalty ρ=5. The anchor follows $\mathbf{v}_m = \mathbf{v}_{m-1} + s\,\boldsymbol{\epsilon}_m$ with $\boldsymbol{\epsilon}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where s=0.02, a relative drift of 5.5% per outer step, in the trace and is swept in Tab. 8. The DFT panel of Fig. 1 compares the 60-angle operator with a Gaussian blur on a 16×16 grid and reports the off-diagonal energy $1 - \sum_i G_{ii}^2 / \sum_{i,j} G_{ij}^2$ of $G = |\mathcal{F}\mathcal{A}^*\mathcal{A}\mathcal{F}^*|$.

Evaluation Metrics Let $\mathbf{x} \in [0,1]^d$ be the reference image and $\hat{\mathbf{x}}$ the reconstruction, which for RELAY is the output $\mathbf{x}_0$ of Alg. 1, both multiplied by the same support mask before evaluation, which for MRI is the support of the coil sensitivities applied to the magnitude images. Aggregate results are means over the test set, whereas the qualitative figures report per-image values. Reconstruction quality is measured by PSNR, SSIM (Wang et al., 2004) and LPIPS (Zhang et al., 2018),

$$\mathrm{PSNR}(\hat{\mathbf{x}}, \mathbf{x}) = 10 \log_{10} \frac{L^2}{\frac{1}{d}\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2}, \tag{39}$$

$$\mathrm{SSIM}(\hat{\mathbf{x}}, \mathbf{x}) = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \frac{\big(2\mu_{\hat{\mathbf{x}},p}\,\mu_{\mathbf{x},p} + C_1\big)\big(2s_{\hat{\mathbf{x}}\mathbf{x},p} + C_2\big)}{\big(\mu_{\hat{\mathbf{x}},p}^2 + \mu_{\mathbf{x},p}^2 + C_1\big)\big(s_{\hat{\mathbf{x}},p}^2 + s_{\mathbf{x},p}^2 + C_2\big)}, \tag{40}$$

$$\mathrm{LPIPS}(\hat{\mathbf{x}}, \mathbf{x}) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \big\| \boldsymbol{\omega}_l \odot \big(\hat{\mathbf{f}}_{hw}^l - \mathbf{f}_{hw}^l\big) \big\|_2^2, \tag{41}$$

with data range $L = 1$. In Eq. 40, $\mu_{\cdot,p}$, $s_{\cdot,p}^2$ and $s_{\hat{\mathbf{x}}\mathbf{x},p}$ are the local means, variances and covariance over the 7×7 uniform window centred at pixel $p \in \mathcal{P}$, and $C_1 = (0.01L)^2$, $C_2 = (0.03L)^2$. In Eq. 41, $\mathbf{f}_{hw}^l, \hat{\mathbf{f}}_{hw}^l \in \mathbb{R}^{C_l}$ are the channel-normalized activations of layer l of a pretrained AlexNet

at position (h, w) on an $H_l \times W_l$ grid, computed on the images scaled to $[-1, 1]$ and replicated across three channels, and $\boldsymbol{\omega}_l$ are the learned channel weights of Zhang et al. (2018). Higher PSNR and SSIM and lower LPIPS are better. The warm-start gain Δ in the DC-efficiency and ablation tables is the PSNR of RELAY minus that of the cold-started solve at the same budget.

The solver analyses assess the DC solve of $\mathbf{Mz} = \mathbf{b}$ directly. For an iterate $\mathbf{z}$ they report the relative residual and, for CG started from a given point, the iterations needed to reach a tolerance,

$$\mathrm{relres}(\mathbf{z}) = \frac{\|\mathbf{b} - \mathbf{Mz}\|_2}{\|\mathbf{b}\|_2}, \qquad k_\tau = \min\left\{k \geq 0 : \ \mathrm{relres}(\mathbf{z}_k) \leq \tau\right\}, \quad \tau = 10^{-4}, \tag{42}$$

where $\mathbf{z}_k$ is the k-th CG iterate (Fig. 7a, Tab. 11). The objective gap of Fig. 8 is $\mathcal{J}(\mathbf{z}) - \mathcal{J}(\mathbf{z}^\star) = \frac{1}{2}\|\mathbf{z} - \mathbf{z}^\star\|_{\mathbf{M}}^2$ for $\mathcal{J}(\mathbf{z}) = \frac{1}{2}\mathbf{z}^*\mathbf{Mz} - \mathrm{Re}(\mathbf{b}^*\mathbf{z})$ and $\mathbf{z}^\star = \mathbf{M}^{-1}\mathbf{b}$. Cost is counted as matrix–vector products with $\mathbf{M}$ (Tab. 8) or as operator applications, the total number of calls to $\mathcal{A}$ and $\mathcal{A}^*$ (Fig. 5, Tab. 14).

Baselines The learned baselines use the benchmark's EDM teacher for the diffusion samplers and its distilled consistency model for the consistency-based solvers, without task-specific training. For the diffusion baselines, $\mathbf{x}_t$ denotes the noisy sample at diffusion time t and $\hat{\mathbf{x}}_0(\mathbf{x}_t)$ its Tweedie estimate of the clean image $\mathbf{x}_0$, not the outer iterates of Alg. 1.

- **FBP** reconstructs sparse-view CT analytically by filtering each projection and back-projecting it.
- **Zero-filled** reconstruction applies the adjoint SENSE operator $\mathcal{A}^*$ to the undersampled k-space, combining the coil images with the conjugate sensitivity maps.
- **TV-CS** (Sidky & Pan, 2008) minimizes $\frac{1}{2}\|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\,\mathrm{TV}(\mathbf{x})$, solved with FISTA (Beck & Teboulle, 2009) on CT and with the SigPy reconstruction app (Ong & Lustig, 2019) on MRI.
- **ℓ_1-wavelet CS** (Lustig et al., 2007) replaces the total variation with the ℓ_1 norm of the wavelet coefficients, with the same solvers as TV-CS.
- **DPS** (Chung et al., 2022) guides a diffusion sampler with the gradient of $\|\mathbf{y} - \mathcal{A}\hat{\mathbf{x}}_0(\mathbf{x}_t)\|_2^2$, back-propagated through the network from its Tweedie estimate $\hat{\mathbf{x}}_0$.
- **ΠGDM** (Song et al., 2023a) approximates $p(\mathbf{x}_0 \mid \mathbf{x}_t)$ by a Gaussian of variance r_t^2 around the Tweedie estimate and adds the resulting pseudoinverse-weighted vector–Jacobian product as guidance.
- **DDS** (Chung et al., 2024) runs a few CG steps initialized at the Tweedie estimate inside DDIM sampling, which keeps the data-consistency update in the tangent space of the denoised sample without backpropagating through the network.
- **DiffPIR** (Zhu et al., 2023) alternates, by half-quadratic splitting, a diffusion denoising step with the data subproblem $\min_{\mathbf{x}} \|\mathbf{y} - \mathcal{A}\mathbf{x}\|_2^2 + \gamma_t\|\mathbf{x} - \hat{\mathbf{x}}_0(\mathbf{x}_t)\|_2^2$ with anchor weight γ_t, which it solves in closed form for deblurring, inpainting and super-resolution.
- **CM** (Song et al., 2023b) maps any point of the diffusion ODE trajectory directly to an image, and we alternate it with the shared regularized DC solve.
- **CM4IR** (Garber & Tirer, 2025) guides a few-step consistency-model sampler with back-projection guidance through the pseudoinverse $\mathcal{A}^\dagger$, combined with an improved initialization and a dedicated noise-injection mechanism.
- **PnP-CM** (Gulle et al., 2026) uses the consistency model as the proximal operator of the prior inside PnP-ADMM and adds noise perturbations and momentum-based updates for the low-NFE regime.

On MRI each learned baseline uses the better of two sampling schedules, the schedule for each R in the public implementation (Gulle et al., 2026) and the default CT schedule. ΠGDM diverges on M4Raw under both, as its guidance scales as $1/r_t^2$, which its fixed noise schedule drives towards 10^{-4}.

C Additional Results

C.1 Data-Consistency Analysis

DC Efficiency Tab. 10 lists the values plotted in Fig. 6a. The warm-start gain is largest where the budget is tightest, $+9.02$ dB at $k{=}5$, and RELAY at $k{=}10$ already exceeds the cold start at $k{=}30$.

Table 10: DC efficiency on sparse-view CT at 60 angles for k CG iterations per step. Δ: warm-start gain.

	PSNR (dB) ↑				
CG iterations k	5	10	15	20	30
Cold start	24.51	28.50	30.61	31.85	33.85
RELAY (ours)	**33.53**	**36.83**	**37.71**	**38.02**	**37.93**
Δ	$+9.02$	$+8.33$	$+7.10$	$+6.17$	$+4.08$

Per-Step Residual Trace Tab. 11 follows one run step by step at three per-step budgets and lists the values plotted in Fig. 7a. The cold relative residual $\mathrm{relres}(\mathbf{z})$ of Eq. 42 stays flat across the outer steps at every budget, whereas the carried solve lowers it by up to an order of magnitude and cuts the CG iterations needed to reach 10^{-4} from 27 to 12–16 by the final outer step $m{=}8$. This is the behaviour that Eq. 16 describes: the inherited error contracts towards a level set by the drift rather than resetting to the size of the solution.

Table 11: DC trace across outer steps at 60 angles, with the anchor drifting by 0.02 per step.

Outer step m	1	2	4	6	8
budget $k = 3$					
Cold-start residual	8.41e−3	8.41e−3	8.41e−3	8.41e−3	8.42e−3
RELAY residual	8.41e−3	**2.83e−3**	**1.36e−3**	**8.70e−4**	**8.60e−4**
Cold-start iterations	27	27	27	27	27
RELAY iterations	27	**24**	**20**	**17**	**16**
budget $k = 5$					
Cold-start residual	3.09e−3	3.09e−3	3.10e−3	3.10e−3	3.10e−3
RELAY residual	3.09e−3	**1.05e−3**	**4.32e−4**	**3.35e−4**	**3.15e−4**
Cold-start iterations	27	27	27	27	27
RELAY iterations	27	**23**	**15**	**14**	**13**
budget $k = 10$					
Cold-start residual	6.76e−4	6.82e−4	6.93e−4	7.05e−4	7.17e−4
RELAY residual	6.76e−4	**2.49e−4**	**2.00e−4**	**1.35e−4**	**1.09e−4**
Cold-start iterations	27	27	27	27	27
RELAY iterations	27	**18**	**9**	**12**	**12**

Ablations on LoDoPaB-CT Tab. 12 sweeps each hyperparameter for RELAY and the cold start. The warm start improves PSNR at every CG budget, by 11.8 dB at $k{=}2$ and still by 7.0 dB at the default budget $k{=}20$, so RELAY at $k{=}5$ already exceeds the cold start at $k{=}20$, a $4\times$ reduction. As the NFE grows the cold start saturates by $N{=}4$, whereas RELAY keeps improving through $N{=}8$: for the cold start the truncated data step, not the prior, limits the quality. The stabilizer columns match Tab. 7.

Table 12: Ablations on LoDoPaB-CT at 60 angles over 100 slices, with Δ the warm-start gain.

PSNR (dB) ↑	CG iterations k				NFE N				Noise injection		Momentum	
Solver	2	5	10	20	2	4	6	8	on	off	on	off
Cold start	20.33	26.51	30.91	33.36	31.46	33.34	33.36	33.38	33.36	33.59	33.36	33.36
RELAY (ours)	32.14	37.63	39.46	40.36	36.71	39.82	40.36	40.94	40.36	40.36	40.35	40.35
Δ	+11.81	+11.12	+8.55	+7.00	+5.25	+6.48	+7.00	+7.56	+7.00	+6.77	+6.99	+6.99

C.2 Classical Compressed Sensing and Cost

Classical CS at Convergence Tab. 13 runs TV-CS and ℓ_1-wavelet CS to convergence. On OrganAMNIST at 60 angles this is 12–23× RELAY's operator budget. RELAY stays ahead of both in PSNR and SSIM at every angle count on both CT benchmarks.

Table 13: Classical CS run to convergence (12–23× RELAY's operator budget) at 18–90 projection angles. PSNR in dB, higher is better for both metrics.

	18		30		60		90	
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
OrganAMNIST								
TV-CS *(converged)*	28.74	.870	32.48	.943	37.87	.985	40.13	.991
ℓ_1-wavelet *(converged)*	26.63	.791	29.47	.874	35.42	.970	38.37	.982
RELAY (ours)	**28.86**	**.871**	**32.60**	**.944**	**38.02**	**.986**	**40.30**	**.992**
LoDoPaB-CT								
TV-CS *(converged)*	34.36	.867	36.96	.910	40.27	.951	42.01	.966
ℓ_1-wavelet *(converged)*	30.80	.748	33.26	.815	37.27	.905	39.41	.940
RELAY (ours)	**34.54**	**.868**	**37.12**	**.911**	**40.36**	**.952**	**42.20**	**.967**

Computational Cost Tab. 14 counts operator applications, which account for more than 97% of the CT runtime. For 7% more applications than PnP-CM at the same 6 NFE (276 against 258), RELAY gains +6.17 dB, at 2.88 s per reconstruction on a single RTX PRO 6000. Budget-matched TV-CS runs 138 FISTA iterations at two applications each, and converged TV-CS needs 12 to 23 times that budget.

Table 14: Cost per reconstruction on OrganAMNIST at 60 angles. Operator applications sum calls to $\mathcal{A}$ and $\mathcal{A}^*$.

Method	Operator applications	NFE
FBP	3	0
DPS	7	6
DDS	79	6
CM	139	6
DiffPIR	139	6
CM4IR	145	6
ΠGDM	145	6
PnP-CM	258	6
RELAY (ours)	276	6
TV-CS *(budget-matched)*	276	0
TV-CS *(converged)*	3200–6400	0

Fixed Relative Noise Level Lowering the tube current to reduce dose raises the noise floor at every angle count. Tab. 15 raises σ_y to $\approx$1% of the sinogram RMS and re-evaluates the methods of this table with unchanged hyperparameters. RELAY keeps the best PSNR at every setting without

re-tuning, whereas converged TV-CS loses up to 23 dB relative to Tab. 13 and falls below the learned solvers reported here from 60 angles upward.

Table 15: Fixed relative noise level ($\approx$1% of the sinogram RMS) at 18–90 projection angles.

	PSNR (dB) ↑			
# angles	18	30	60	90
OrganAMNIST, σ_y=0.2				
ℓ_1-wavelet	23.21	22.65	18.34	20.53
TV-CS *(converged)*	<u>26.73</u>	27.10	24.55	25.33
CM	24.84	26.78	28.27	28.48
PnP-CM	26.00	<u>28.44</u>	<u>30.43</u>	<u>30.81</u>
RELAY (ours)	**27.18**	**28.87**	**30.88**	**31.24**
LoDoPaB-CT, σ_y=0.8				
ℓ_1-wavelet	27.30	25.88	21.06	17.54
TV-CS *(converged)*	29.03	26.90	22.23	19.02
CM	28.91	29.74	30.46	30.68
PnP-CM	<u>29.58</u>	<u>30.65</u>	<u>31.22</u>	<u>31.20</u>
RELAY (ours)	**29.95**	**31.05**	**31.64**	**31.66**

C.3 Robustness to the Prior

Tab. 16 runs the same solver on M4Raw with a fastMRI knee CM (Gulle et al., 2026), a fastMRI 3 T brain CM (Gülle et al., 2026) and the M4Raw 0.3 T brain CM. RELAY keeps its advantage with a prior from another scanner, within 1.9 dB of the matched prior, so the gain comes from the DC warm start rather than from a prior matched to the scanner.

Table 16: Effect of the prior's source domain (R=8, PSNR (dB) ↑).

	PSNR (dB) ↑		
Prior (source domain)	T1	T2	FLAIR
Knee CM (fastMRI, 3 T, OOD)	<u>26.32</u>	<u>25.48</u>	<u>25.01</u>
Brain CM (fastMRI, 3 T, OOD)	25.64	24.72	24.01
M4Raw CM (0.3 T)	**27.48**	**26.38**	**25.62**

C.4 Qualitative Comparisons

Sparse-View CT Fig. 9 shows one slice at 18 to 90 angles, and RELAY degrades most gracefully as views are removed. In Fig. 10 the baselines leave *structured* residuals, angular streaks for FBP and over-smoothed organ boundaries for CM, DiffPIR and PnP-CM. RELAY produces the most uniform, lowest-magnitude error and the best per-image PSNR/SSIM on every slice, both here and across the seven evenly spaced slices of Fig. 11. Its margin over PnP-CM is largest where the anatomical detail is finest, where a truncated DC solve leaves its residual along boundaries.

Classical CS Fig. 12 places budget-matched TV-CS beside the learned solvers on four LoDoPaB slices. TV-CS recovers the large-scale anatomy but smooths the vessel walls and the airway rim, whereas RELAY keeps those edges sharp, and FBP leaves the streaks that both suppress. At the fixed relative noise level of Tab. 15 (Fig. 13) the converged TV-CS reconstruction develops high-frequency mottling across the parenchyma, whereas RELAY adds no structure the measurements do not support.

Low-Field MRI On M4Raw at R=8 (Fig. 14), zero-filling and TV-CS retain the coarse aliasing that the wavelet solver only partly removes, whereas RELAY recovers the cortical boundary and the ventricular margin. RELAY leads ℓ_1-wavelet CS in PSNR on at least 37 of the 40 test slices for every contrast.

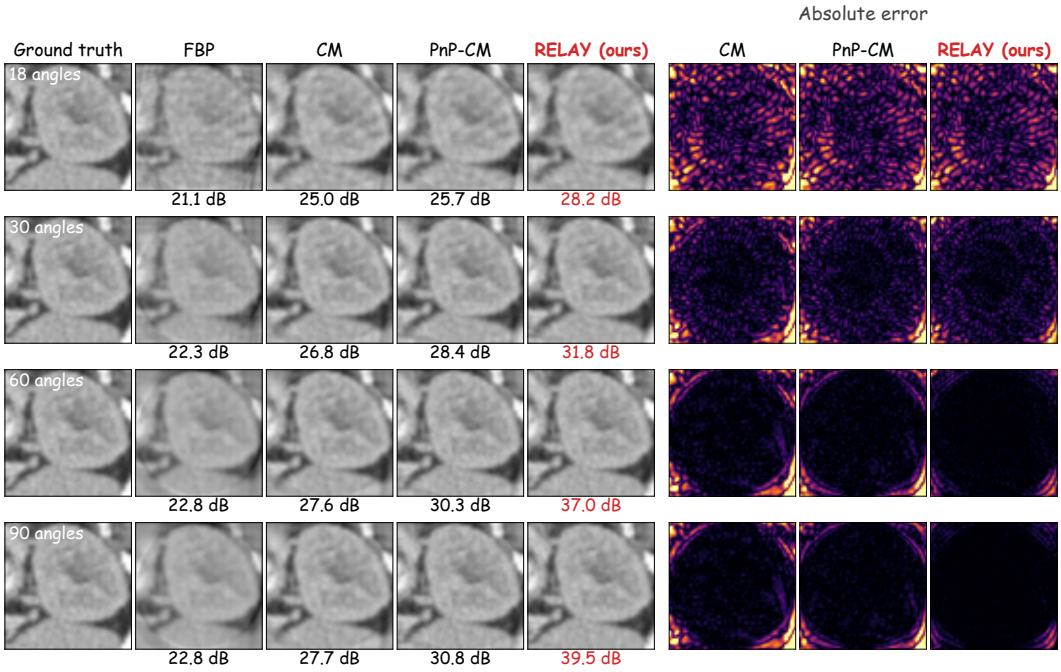


Figure 9: Effect of view sparsity on one OrganAMNIST test slice at 18, 30, 60 and 90 projection angles (rows), with per-image PSNR and absolute errors.

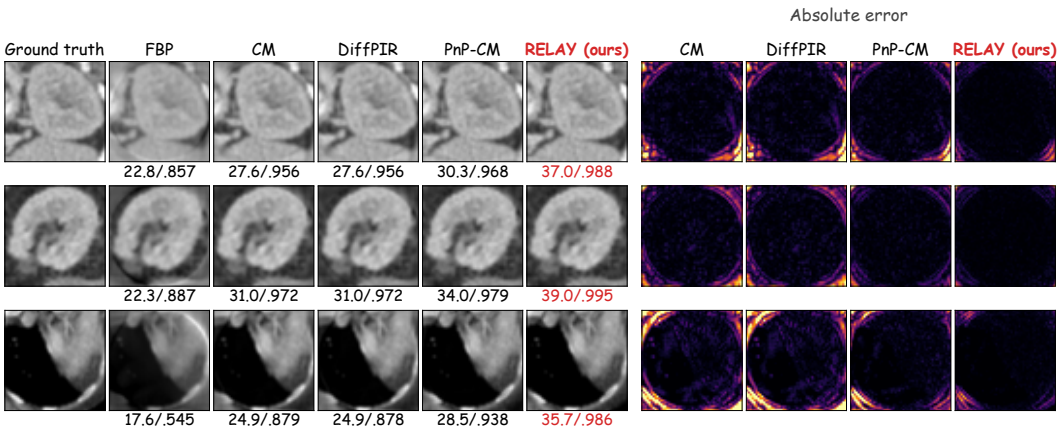


Figure 10: OrganAMNIST at 60 angles on three test slices, with per-image PSNR/SSIM and absolute errors $|\hat{\mathbf{x}} - \mathbf{x}|$ on one shared scale.

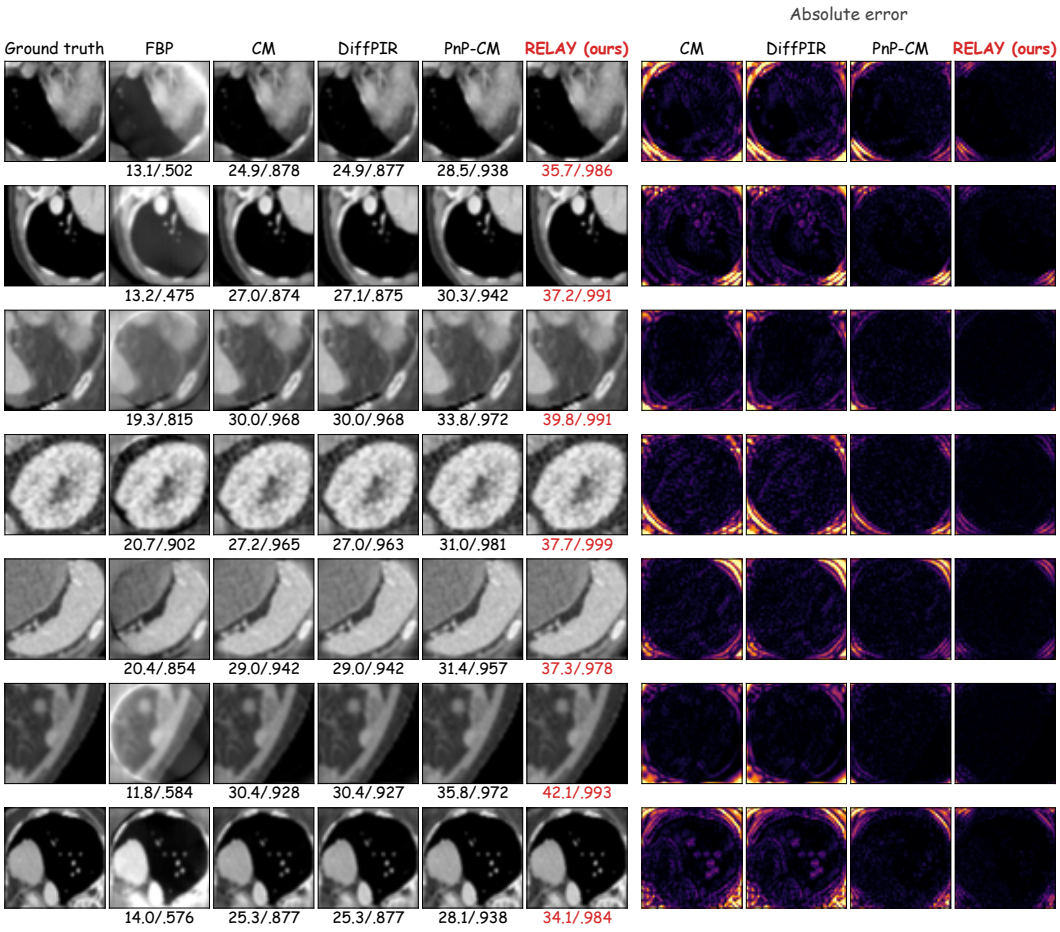


Figure 11: OrganAMNIST at 60 angles. Seven test slices at evenly spaced indices, with per-image PSNR/SSIM and absolute errors.

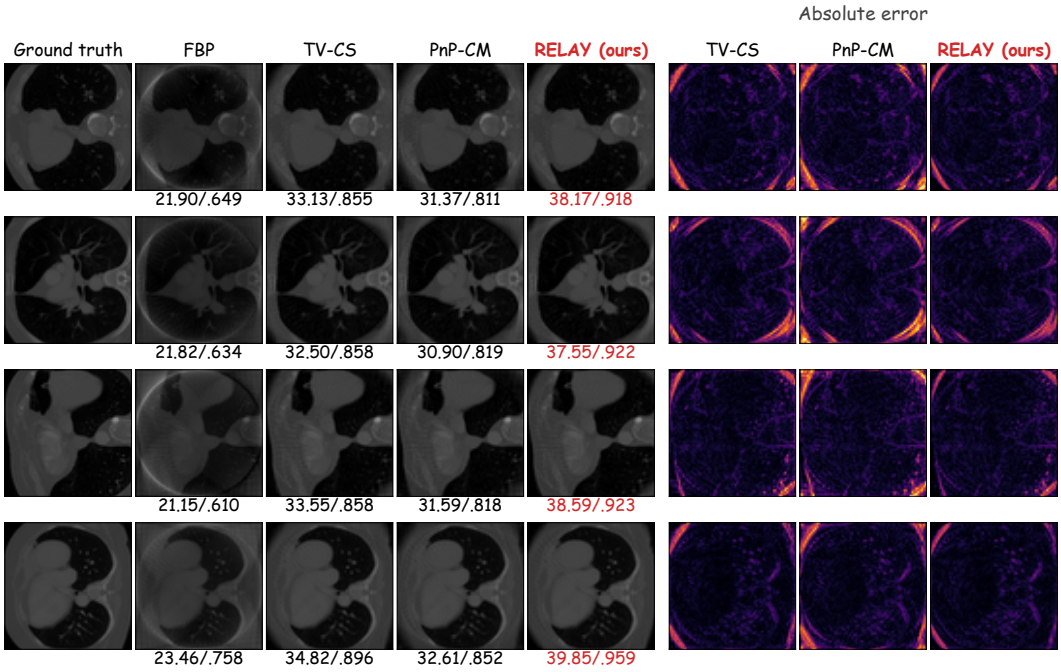


Figure 12: Classical CS on LoDoPaB-CT at 60 angles under an equal operator budget, with per-image PSNR/SSIM and absolute errors. Four slices at the median gain of RELAY over TV-CS.

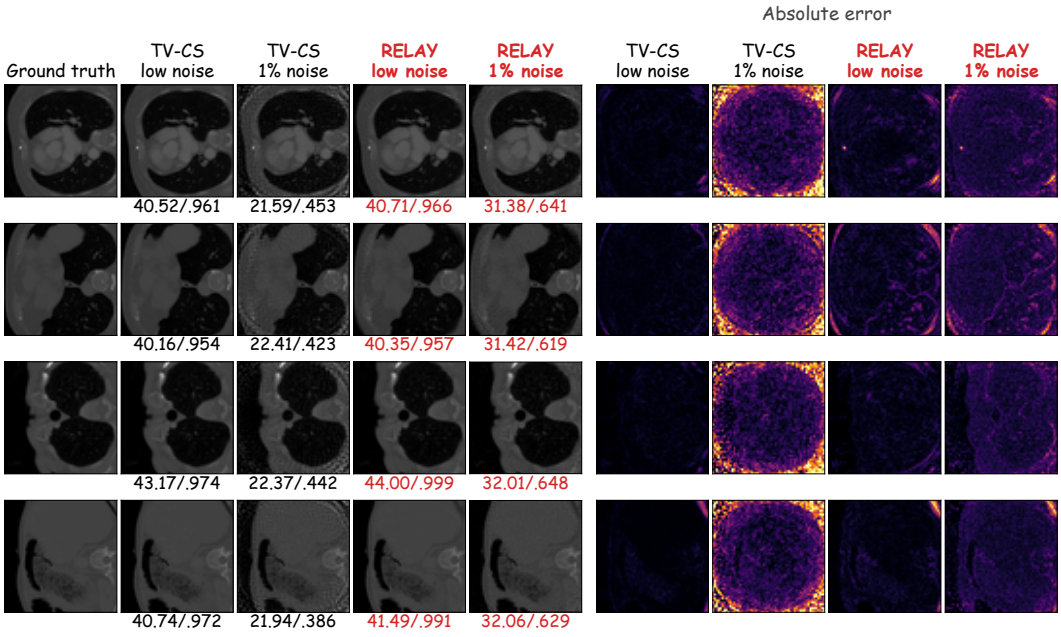


Figure 13: Classical CS on LoDoPaB-CT at 60 angles, default versus $\approx 1\%$ relative noise, one slice per row, with per-image PSNR/SSIM and absolute errors.

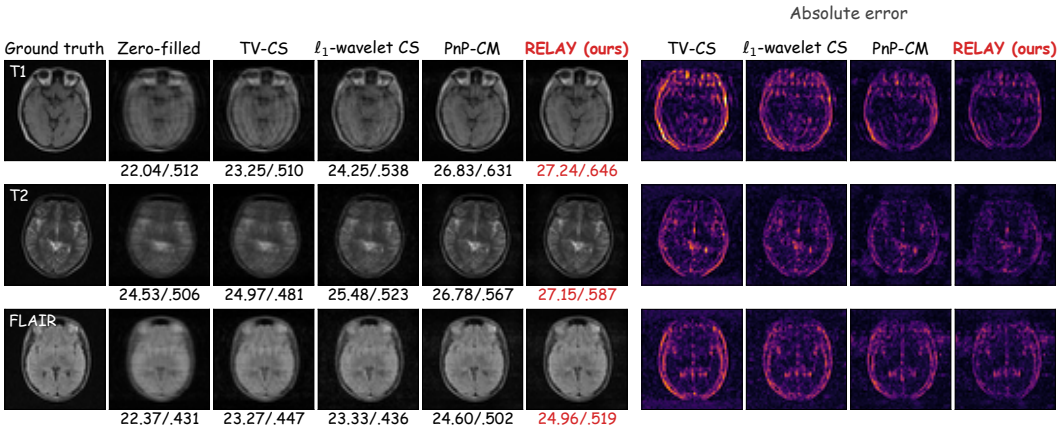


Figure 14: M4Raw at R=8, with per-image PSNR/SSIM and absolute errors. Each row is the slice closest to the median gap between RELAY and ℓ_1-wavelet CS.