HERKRY: A HERMITE–GAUSSIAN HIERARCHY WITH RECURRENT KRYLOV TRAJECTORY MODELING FOR ROBUST SPARSE-VIEW CT RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Sparse-view computed tomography (SVCT) relies heavily on learned priors, leaving it vulnerable to clinical distribution shifts. While a coarse-to-fine hierarchy can regularize early reconstruction by restricting weakly constrained degrees of freedom, directly extending data consistency (DC) to coarse scales is nontrivial. Image and sinogram downsampling do not generally commute with fan-beam projection, introducing geometry-dependent inconsistencies. To resolve this, we propose Hermite-Krylov CT (HerKry), which models images via a continuous Hermite–Gaussian (HG) hierarchy shifting from sparse, high-order primitives to dense, lower-order ones. This parameterization decouples representation resolution from measurement scale, enabling geometry-consistent DC directly against native sinograms. Crucially, higher-order HG modes can absorb complex spatial deformations via linear coefficient combination. This allows DC to freeze nonlinear parameters and optimize exclusively in linear coefficient space via Hermite-Coefficient CGLS (HC-CGLS), slashing DC latency by over 90% with better fidelity. A Krylov-Trajectory GRU (KT-GRU) further encodes intermediate solver iterates to propagate optimization dynamics across stages. Extensive evaluations across cross-organ transfer, varying noise, and extreme sparsity demonstrate superior robustness, outperforming competing methods by up to 4 dB. The code is available at https://anonymous.4open.science/r/Code-for-HerKry-DB40.

1 INTRODUCTION

Sparse-view computed tomography (SVCT) reconstructs CT images from substantially fewer angular projections than standard acquisition protocols, providing an effective means of reducing ionizing radiation exposure Goldman (2007); Wang et al. (2024). However, the resulting inverse problem is severely ill-conditioned, and modern reconstruction networks therefore rely heavily on learned image or sinogram priors Jin et al. (2017); Zhang et al. (2023a; 2025a); Chen et al. (2025b). When test-time anatomy or acquisition conditions differ from those observed during training, models frequently introduce amplified streaks or structural hallucinations Antun et al. (2020); Bhadra et al. (2021); Ma et al. (2026). As illustrated in Fig. 1(a), performing physical data consistency directly on a full-resolution pixel grid overfits weakly constrained directions, degrading fine structures.

One intuitive remedy is a coarse-to-fine hierarchy Hauptmann et al. (2020) that restricts spatial degrees of freedom during early inference. However, pairing low-resolution CT states with downsampled sinograms introduces severe physical inconsistency, for the image downsampling and divergent-beam integration do not commute. As quantified in Fig. 1(b), this geometric mismatch escalates rapidly at coarser scales, rendering conventional CT-sinogram pyramids unsuitable for hierarchical data consistency (DC).

To resolve this dilemma, we propose Hermite-Krylov CT (HerKry), a reconstruction framework that decouples the representation resolution and measurement scale. As shown in Fig. 1(a), HerKry represents each intermediate reconstruction using continuous Hermite-Gaussian (HG) primitives and projects them through the original scanning system. The resulting hierarchy therefore dynamically shifts the spatial density of the parameterization, while keeping the underlying physical measurement model strictly consistent.

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.

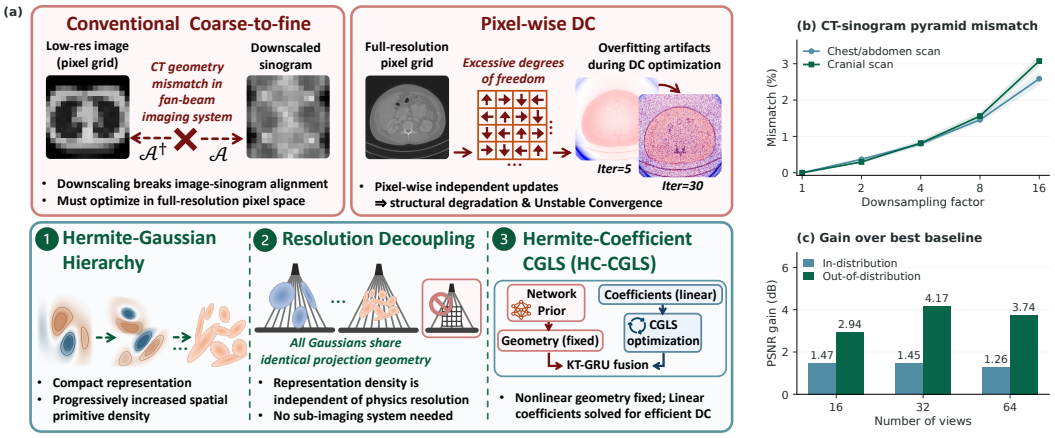

Figure 1: **Motivation and performance highlights of HerKry. (a)** Conventional hierarchical DC suffers from fan-beam geometric mismatches, while full-resolution DC overfits weakly constrained directions; HerKry decouples representation resolution from measurement scale via a continuous HG hierarchy and linear HC-CGLS. **(b)** geometric error in fan-beam CT-sinogram pyramids scales up with the downsampling factor. **(c)** PSNR gains over the strongest baseline under in-distribution and zero-shot OOD evaluations across different sparsity levels.

At coarse stages, HerKry allocates representational capacity to sparsely distributed primitives with higher-order HG modes, shifting toward denser, lower-order primitives at finer stages. Critically, higher-order HG modes are polynomial–Gaussian functions that span the same finite-order function space as Gaussian derivative bases. Their linear combinations can naturally approximate the local deformations through Taylor expansion Refregier (2003). Consequently, instead of relying on computationally expensive nonlinear optimizers (e.g., Adam Kingma & Ba (2014)) to adjust all Gaussian parameters, the DC refinements can be absorbed purely through efficient linear optimizers (e.g., conjugate gradient method Hestenes et al. (1952)) under a frozen spatial geometry. This bridges the gap between rich structural representation and efficient DC process of HG hierarchy.

Building on this insight, we decouple the reconstruction into geometric parameter synthesis and physical coefficient optimization. Specifically, the feed-forward networks predict the nonlinear Gaussian positions and shapes, while physical DC is executed exclusively in coefficient space via Hermite-Coefficient Conjugate Gradient Least Squares (HC-CGLS), slashing the DC latency by over 90%. Beyond rapid convergence, HC-CGLS natively exposes a structured sequence of intermediate Krylov iterates rather than merely a terminal output. We exploit this dynamic by introducing a Krylov-Trajectory GRU (KT-GRU) that encodes the ordered iterates, carrying a persistent solver history across stages. Together, the continuous HG hierarchy, decoupled coefficient solver, and recurrent trajectory model form a hierarchical, physics-consistent reconstruction pipeline.

Our main contributions are summarized as follows:

1) We identify the geometric mismatch caused by naive downsampling in fan-beam CT-sinogram pyramids, and propose HerKry to address it through continuous Hermite-Gaussian hierarchy. By decoupling representation resolution from measurement scale, HerKry enforces geometry-consistent DC directly against original sinograms across all stages.

2) We exploit the representational expressiveness of Hermite-Gaussian bases, decomposing costly full-parameter optimization into geometric parameter synthesis and physical coefficient optimization. The HC-CGLS can perform efficient DC in linear coefficient space, with latency reduced by over 90%. A KT-GRU is further developed to propagate Krylov solver dynamics across stages for implicit regularization.

3) Extensive evaluations across in-distribution benchmarks, cross-organ transfer, varying acquisition noise, and extreme angular sparsity demonstrate the superior fidelity and out-of-distribution (OOD) robustness of HerKry, achieving gains of up to 4 dB PSNR over competitive baselines (Fig. 1(c)).

2 PRELIMINARY

2.1 SPARSE-VIEW CT FORMULATION

In sparse-view CT, the discrete acquisition process is modeled as $\boldsymbol{y}_s = \mathcal{A}\boldsymbol{x} + \boldsymbol{\epsilon}$, where $\boldsymbol{x} \in \mathbb{R}^{N_{\text{pix}}}$ denotes the unknown attenuation image, $\boldsymbol{y}_s \in \mathbb{R}^M$ is the measured sparse-view sinogram, $\mathcal{A}$ is the discrete fan-beam projection operator, and $\boldsymbol{\epsilon}$ denotes measurement noise Kak & Slaney (2001). Sparse angular sampling leaves many image directions weakly constrained, resulting in a highly ill-conditioned inverse problem. For interpretation, let $\mathcal{A}^\dagger$ denote a pseudoinverse of the forward operator Hansen (2010). A candidate reconstruction can be written as: $\hat{\boldsymbol{x}} = \mathcal{A}^\dagger \boldsymbol{y}_s + \left(\mathcal{I} - \mathcal{A}^\dagger \mathcal{A}\right) \boldsymbol{z}$, where the first term is anchored by the measurements, whereas the second contains image components that are weakly or not directly constrained by the acquisition. In our implementation, filtered back-projection (FBP) Kak & Slaney (2001) serves as the practical approximation of $\mathcal{A}^\dagger$ used for reconstruction initialization. Under distribution shifts, learned reconstruction priors may introduce erroneous structures primarily through these weakly constrained directions Bhadra et al. (2021).

2.2 THE GEOMETRIC OBSTACLE OF EXTENDING DATA CONSISTENCY TO COARSE SCALES

A coarse-to-fine reconstruction pipeline would ideally perform physical DC at each representation scale. A naive CT-sinogram pyramid implicitly requires the commutation relation:

$$\underbrace{\mathcal{A}^{\text{coarse}}\left(\mathcal{D}_x \boldsymbol{x}\right)}_{\text{projection of downsampled image}} \overset{?}{=} \underbrace{\mathcal{D}_y\left(\mathcal{A}^{\text{fine}} \boldsymbol{x}\right)}_{\text{downsampled fine-scale measurement}}, \quad \forall \boldsymbol{x}, \tag{1}$$

where $\mathcal{D}_x$ and $\mathcal{D}_y$ denote 2D downsampling on image and 1D downsampling on the detector axis of sinogram, respectively. For divergent-beam CT, this equality generally does not hold: $\mathcal{D}_y \mathcal{A}^{\text{fine}} \neq \mathcal{A}^{\text{coarse}} \mathcal{D}_x$. Because adjacent detector rays follow divergent trajectories, their measurements cannot be physically aggregated into an equivalent coarse ray via sinogram downsampling. Likewise, image downsampling discards sub-pixel variations that govern native line integrals. Consequently, naively constructing coarse CT–sinogram pairs introduces a geometry-dependent modeling error. We further provide the theoretical derivations and empirical analyses in Appendix B.

3 METHOD

3.1 OVERVIEW OF HERKRY

As illustrated in Fig. 2, HerKry alternates between an outer continuous representation hierarchy and an inner physical coefficient solver across stages $l = 1, \ldots, L$:

Continuous HG Hierarchy (Sec. 3.2): As shown in Fig. 2(a), we parameterize CT images via Hermite-Gaussian primitives governed by an order-density schedule. Shifting from sparse, high-order modes to dense, zeroth-order Gaussians enables direct projection into native sinograms $\boldsymbol{y}_s$, bypassing discrete geometric mismatches.

HC-CGLS & Trajectory Modeling (Sec. 3.3, 3.4): As shown in Fig. 2(b), within each stage, we decouple the update into neural geometry synthesis and physical coefficient optimization. Under frozen geometries, HC-CGLS resolves linear data consistency efficiently. As shown in Fig. 2(c), a KT-GRU and spatial encoder then capture intermediate Krylov iterates into persistent memories $(\boldsymbol{h}_l^K, \boldsymbol{U}_l)$ to regularize subsequent stages.

3.2 RESOLUTION-DECOUPLED HERMITE-GAUSSIAN HIERARCHY

3.2.1 2D ANISOTROPIC HERMITE-GAUSSIAN PRIMITIVES

Standard 2D Gaussian primitives provide a continuous and spatially adaptive representation for images and visual features Zhang et al. (2025b); Huang et al. (2024); Zhang et al. (2026); Chen et al. (2025a). While effective for low-frequency regions, modeling directional edges with standard Gaussians requires dense spatial centers, which introduces excessive unconstrained degrees of freedom. To capture rich structural orientations within compact spatial support, we modulate Gaussian envelopes with orthogonal Hermite polynomials Refregier (2003).

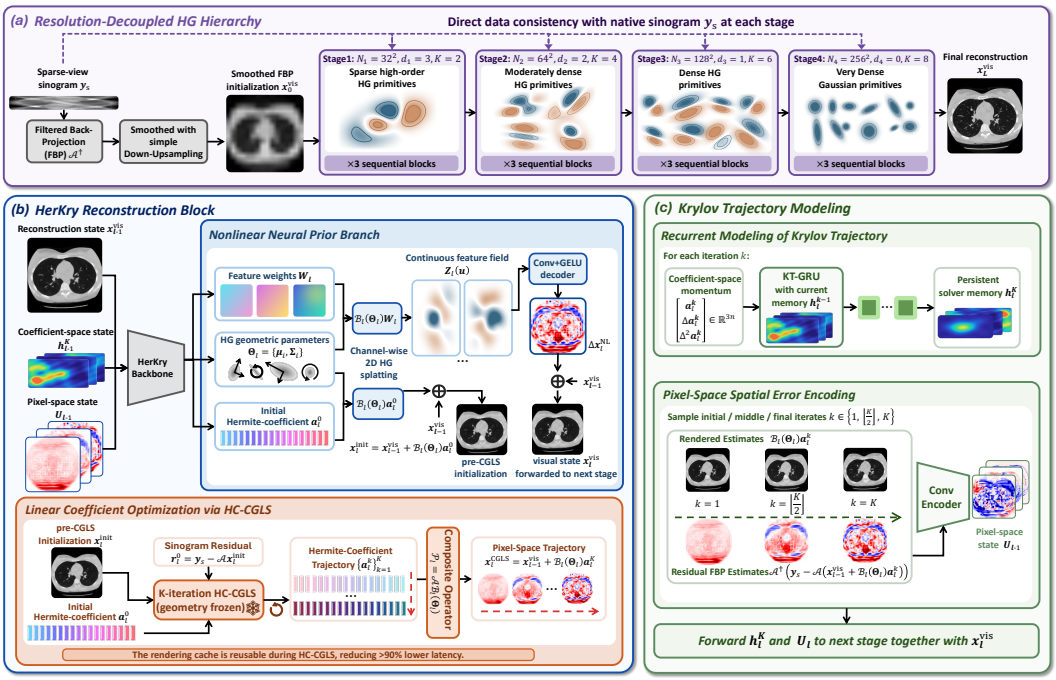

Figure 2: **Overview of HerKry. (a)** Continuous HG hierarchy transitioning from sparse high-order to dense low-order primitives for geometry-consistent DC on native sinograms $\boldsymbol{y}_s$. **(b)** HerKry block under frozen geometry $\boldsymbol{\Theta}_l$, combining nonlinear prior decoding with linear HC-CGLS coefficient solving. **(c)** Trajectory modeling via KT-GRU tracking solver momentum into memory $\boldsymbol{h}_l^K$, and spatial encoding into $\boldsymbol{U}_l$ across key iterates.

Each primitive $i \in \{1, \ldots, N_l\}$ at stage l is parameterized by its spatial geometry $\boldsymbol{\theta}_i = \{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$, where $\boldsymbol{\mu}_i \in \mathbb{R}^2$ is the center and $\boldsymbol{\Sigma}_i \in \mathbb{R}^{2\times 2}$ is a positive-definite covariance matrix. For a spatial coordinate $\boldsymbol{u} \in \mathbb{R}^2$, the normalized local coordinate is given by $\boldsymbol{v}_i(\boldsymbol{u}) = \boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{u} - \boldsymbol{\mu}_i) = [v_{i,x}, v_{i,y}]^\top$. The 2D tensor-product Hermite-Gaussian basis of order (p, q) is defined as:

$$\psi_{pq}(\boldsymbol{v}_i(\boldsymbol{u})) = \bar{H}_p(v_{i,x})\bar{H}_q(v_{i,y}) \cdot \frac{1}{2\pi} \exp\left(-\frac{1}{2}\|\boldsymbol{v}_i(\boldsymbol{u})\|^2\right), \tag{2}$$

where $\bar{H}_n(t) = H_n(t)/\sqrt{2^n n!}$ denotes the normalized Hermite polynomial of degree n, with orders bounded by $\mathcal{S}_{d_l} = \{(p, q) \mid 0 \le p, q \le d_l\}$.

Crucially, the spatial derivatives of a Gaussian are polynomial–Gaussian functions. Since Hermite polynomials span finite-degree polynomial spaces, these derivatives can be represented by finite-order Hermite–Gaussian modes Refregier (2003). Specifically, under a local displacement $\Delta\boldsymbol{\mu}_i$, the perturbed base Gaussian $g_i(\boldsymbol{u}) \triangleq \psi_{00}(\boldsymbol{v}_i(\boldsymbol{u}))$ can be approximated via Taylor expansion:

$$g_i(\boldsymbol{u} - \Delta\boldsymbol{\mu}_i) \approx g_i(\boldsymbol{u}) + \sum_{k=1}^{d_l} \frac{(-1)^k}{k!}(\Delta\boldsymbol{\mu}_i \cdot \nabla)^k g_i(\boldsymbol{u}) = \sum_{(p,q)\in\mathcal{S}_{d_l}} c_{pq}(\Delta\boldsymbol{\mu}_i)\, \psi_{pq}(\boldsymbol{v}_i(\boldsymbol{u})), \tag{3}$$

where the directional derivative operator $(\Delta\boldsymbol{\mu}_i \cdot \nabla)^k$ acting on g_i directly yields polynomial-Gaussian terms that reside in $\operatorname{span}\{\psi_{pq}\}_{(p,q)\in\mathcal{S}_{d_l}}$. Consequently, local translations and anisotropic stretch can be synthesized purely through linear coefficient modulation even under a fixed spatial geometry $\boldsymbol{\theta}_i$, bypassing costly full-parameter nonlinear optimization during DC process. Detailed derivations and empirical analyses are provided in Appendix C.

3.2.2 ORDER-DENSITY COMPLEMENTARY SCHEDULE

To regularize the ill-conditioned null space, HerKry progressively shifts representation capacity from per-primitive basis order to spatial density across four stages:

$$(N_l, d_l): \quad (32^2, 3) \longrightarrow (64^2, 2) \longrightarrow (128^2, 1) \longrightarrow (256^2, 0). \tag{4}$$

Early stages allocate fewer spatial centers 32^2 with high-order bases $d_1 = 3$ to suppress unconstrained spatial artifacts while preserving anisotropic expressiveness. Subsequent stages trade basis order for spatial density, eventually collapsing to 256^2 dense zeroth-order Gaussians that refine localized and sub-pixel details. Specifically, given fixed geometries $\boldsymbol{\Theta}_l = \{\boldsymbol{\theta}_i\}_{i=1}^{N_l}$, the image increment at stage l is synthesized as:

$$\Delta \boldsymbol{x}_l^{\mathrm{HG}}(\boldsymbol{u}) = \mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l = \sum_{i=1}^{N_l} \sum_{(p,q)\in\mathcal{S}_{d_l}} a_{i,pq}^{(l)}\, \psi_{pq}(\boldsymbol{v}_i(\boldsymbol{u})), \tag{5}$$

where $\boldsymbol{a}_l \in \mathbb{R}^{N_l(d_l+1)^2}$ gathers the HG coefficients and $\mathcal{B}_l(\boldsymbol{\Theta}_l)$ denotes the linear 2D Gaussian splatting operator. Because ψ_{pq} is analytically continuous over $\mathbb{R}^2$, it can be directly mapped into the native measurement space through the linear forward projector $\mathcal{A}$, yielding the composite linear forward operator:

$$\mathcal{P}_l \triangleq \mathcal{A}\mathcal{B}_l(\boldsymbol{\Theta}_l): \quad \mathbb{R}^{N_l(d_l+1)^2} \longrightarrow \mathbb{R}^M. \tag{6}$$

By evaluating $\mathcal{P}_l$ directly against the original sparse sinogram $\boldsymbol{y}_s$, HerKry enforces physical consistency without synthesizing downsampled sinograms, completely bypassing the non-commutative geometric mismatches of pixel-wise CT-sinogram pyramids.

3.3 HERMITE-GAUSSIAN PARAMETER UPDATE DECOMPOSITION

Jointly optimizing nonlinear primitive geometries $\boldsymbol{\Theta}_l$ and linear coefficients $\boldsymbol{a}_l$ via nonlinear optimizers (e.g., Adam) incurs severe GPU memory-bandwidth bottlenecks, as dynamic geometries require rebuilding spatial rendering caches at every iteration. Because HG bases can absorb local structural deformations through linear combinations alone, each reconstruction block decomposes the update into two parallel branches sharing the synthesized geometry $\boldsymbol{\Theta}_l$:

1) Nonlinear Neural Prior Branch: For each primitive $i \in \{1, \ldots, N_l\}$, the neural backbone operates on a persistent latent state to predict shared geometric parameters $\boldsymbol{\Theta}_l = \{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}_{i=1}^{N_l}$, and $(C+1)$ sets of HG coefficients across the active modes $\mathcal{S}_{d_l}$, totaling $(C+1)(d_l+1)^2$ coefficients per primitive. We partition these coefficients into: 1) a multi-channel feature coefficient tensor $\boldsymbol{W}_l \in \mathbb{R}^{N_l \times C \times (d_l+1)^2}$ dedicated to the nonlinear prior branch; 2) a physical initial coefficient vector $\boldsymbol{a}_l^0 \in \mathbb{R}^{N_l(d_l+1)^2}$ dedicated to the linear solver branch.

For the nonlinear branch, rendering each of the C feature channels via the shared HG bases yields a continuous multi-channel spatial feature field $\boldsymbol{Z}_l(\boldsymbol{u}) = [Z_l^{(1)}(\boldsymbol{u}), \ldots, Z_l^{(C)}(\boldsymbol{u})]^\top \in \mathbb{R}^C$:

$$Z_l^{(c)}(\boldsymbol{u}) = [\mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{W}_l^{(c)}](\boldsymbol{u}) = \sum_{i=1}^{N_l} \sum_{(p,q)\in\mathcal{S}_{d_l}} w_{i,c,pq}^{(l)}\, \psi_{pq}(\boldsymbol{v}_i(\boldsymbol{u})), \quad c \in \{1, \ldots, C\}. \tag{7}$$

A lightweight convolutional decoder then collapses $\boldsymbol{Z}_l$ into a single-channel spatial increment: $\Delta \boldsymbol{x}_l^{\mathrm{NL}} = \mathrm{Conv}(\mathrm{GELU}(\boldsymbol{Z}_l))$. Combined with the spatial render of the initial linear coefficients $\mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^0$, the network produces the direct reconstruction state: $\boldsymbol{x}_l^{\mathrm{vis}} = \boldsymbol{x}_{l-1}^{\mathrm{vis}} + \Delta \boldsymbol{x}_l^{\mathrm{NL}} + \mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^0$. The visual state $\boldsymbol{x}_l^{\mathrm{vis}}$ is then passed forward as the primary image representation to condition the subsequent reconstruction block.

2) Linear Coefficient Optimization Branch via HC-CGLS: Concurrently, the physical DC operates under the frozen geometry $\boldsymbol{\Theta}_l$. Anchored to the previous block's output $\boldsymbol{x}_{l-1}^{\mathrm{vis}}$, the linear branch defines its step-0 solver state without the current block's nonlinear increment $\Delta \boldsymbol{x}_l^{\mathrm{NL}}$: $\boldsymbol{x}_l^{\mathrm{init}} = \boldsymbol{x}_{l-1}^{\mathrm{vis}} + \mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^0$. Consequently, the step-0 measurement discrepancy is evaluated as: $\boldsymbol{r}_l^0 = \boldsymbol{y}_s - \mathcal{A}\boldsymbol{x}_l^{\mathrm{init}} = (\boldsymbol{y}_s - \mathcal{A}\boldsymbol{x}_{l-1}^{\mathrm{vis}}) - \mathcal{P}_l\boldsymbol{a}_l^0$. Starting from $\boldsymbol{a}_l^0$ and guided by $\boldsymbol{r}_l^0$, HC-CGLS solves for the physical refinement purely within the linear Hermite-coefficient space:

$$\min_{\boldsymbol{a}_l} \frac{1}{2} \left\| \mathcal{P}_l \boldsymbol{a}_l - (\boldsymbol{y}_s - \mathcal{A}\boldsymbol{x}_{l-1}^{\mathrm{vis}}) \right\|_2^2. \tag{8}$$

Because primitive geometries are static during this solve, rendering caches are built once and reused across all K iterations. This reformulates DC into an efficient, convex least-squares problem, reducing execution latency by over 90% compared to full-parameter nonlinear optimization. Therefore, the HC-CGLS output is formed as: $\boldsymbol{x}_l^{\mathrm{CGLS}} = \boldsymbol{x}_{l-1}^{\mathrm{vis}} + \mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^K$. Notably, to prevent semi-convergence artifacts of HC-CGLS Hansen (2010) from corrupting downstream predictions, we cascade the visual state $\boldsymbol{x}_l^{\mathrm{vis}}$ rather than $\boldsymbol{x}_l^{\mathrm{CGLS}}$ across stages and output $\boldsymbol{x}_L^{\mathrm{vis}}$ as the final reconstruction, distilling HC-CGLS solver dynamics purely via encoded states $\boldsymbol{h}_l^K$ and $\boldsymbol{U}_l$ in next subsection.

3.4 Recurrent Modeling of Krylov Trajectories

CGLS is a classic Krylov subspace solver that iteratively projects the solution onto nested subspaces $\mathcal{K}_k$ spanned by repeated operator applications $\mathcal{P}_l^\top \mathcal{P}_l$ Hestenes et al. (1952); Hansen (2010); Sabaté Landman et al. (2023). For ill-conditioned inverse problems, Krylov iterates exhibit intrinsic semi-convergence Hansen (2010); Gazzola & Sabaté Landman (2020). The early steps recover stable low-frequency structures, while later ones fit noise along ill-conditioned directions. Because this monotonic expansion forms a strictly ordered state trajectory, it naturally motivates recurrent modeling GRU Cho et al. (2014) to extract implicit regularization cues.

Moreover, this trajectory connects two complementary domains: the compact coefficient space where CGLS solver operates, and the spatial image grid where visual errors manifest. We thus decouple trajectory modeling into recurrent coefficient tracking and spatial residual encoding.

1) Coefficient-Space Dynamics with KT-GRU: Starting from $\boldsymbol{a}_l^0$, HC-CGLS generates an ordered sequence of iterates $\{\boldsymbol{a}_l^k\}_{k=1}^K$ spanning the affine Krylov subspace. To capture solver momentum, we compute the update velocity $\Delta\boldsymbol{a}_l^k = \boldsymbol{a}_l^k - \boldsymbol{a}_l^{k-1}$ and acceleration $\Delta^2\boldsymbol{a}_l^k = \Delta\boldsymbol{a}_l^k - \Delta\boldsymbol{a}_l^{k-1}$. A unidirectional KT-GRU recurrently tracks this trajectory:

$$\boldsymbol{h}_l^k = \mathrm{GRU}\left(\mathcal{E}_l\left(\left[\boldsymbol{a}_l^k,\, \Delta\boldsymbol{a}_l^k,\, \Delta^2\boldsymbol{a}_l^k\right]\right),\, \boldsymbol{h}_l^{k-1}\right), \tag{9}$$

where $\mathcal{E}_l$ is a linear layer projecting stage-dependent coefficient dimensions into a shared latent space. Notably, the coefficient-space state $\boldsymbol{h}_l^K$ propagates across all stages as persistent memory.

2) Pixel-Space Spatial Error Encoding: To supply spatial error context without recurrent overhead on full-resolution images, we subsample key trajectory iterates $k \in \mathcal{T}_{\text{sub}}$ from initial, midpoint, and terminal states and back-project their remaining DC residuals:

$$\boldsymbol{U}_l = \mathrm{Encode}\left(\mathrm{Concat}_{k\in\mathcal{T}_{\text{sub}}}\left(\left[\mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^k,\, \mathcal{A}^\dagger\left(\boldsymbol{y}_s - \mathcal{A}\left(\boldsymbol{x}_{l-1}^{\mathrm{vis}} + \mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^k\right)\right)\right]\right)\right). \tag{10}$$

Finally, a lightweight convolutional encoder processes the pixel-space state $\boldsymbol{U}_l$ and fuses it with the coefficient-space state $\boldsymbol{h}_l^K$ to condition the subsequent reconstruction stage.

3.5 Learning Objectives

HerKry is trained end-to-end using stage-wise ℓ_1 supervision against the ground-truth CT slice $\boldsymbol{x}_{\mathrm{gt}}$. The loss supervises both the visual state $\boldsymbol{x}_l^{\mathrm{vis}}$ and the pre-CGLS initialization $\boldsymbol{x}_l^{\mathrm{init}}$:$\mathcal{L} = \sum_{l=1}^L \lambda_l \left(\|\boldsymbol{x}_l^{\mathrm{vis}} - \boldsymbol{x}_{\mathrm{gt}}\|_1 + \rho_l\|\boldsymbol{x}_l^{\mathrm{init}} - \boldsymbol{x}_{\mathrm{gt}}\|_1\right)$, where $\boldsymbol{\lambda} = [1/8, 1/4, 1/2, 1]$ progressively emphasizes finer stages to account for their higher resolution, and $\boldsymbol{\rho} = [1, 1/2, 1/4, 0]$ encourages early network predictions $\boldsymbol{a}_l^0$ to be plausible before physical refinement and taper off at the final stage.

3.6 Experimental Setup

Datasets: We evaluate HerKry across two clinical CT benchmarks (details in Appendix E): **1) AAPM Benchmark Moen et al. (2021):** A 196-patient chest and abdominal CT benchmark, subsampled to 16, 32, and 64 views to probe reconstruction under severe ill-posedness, supplemented by a cohort of 1,136 cranial CT slices to test zero-shot cross-organ out-of-distribution (OOD) generalization. **2) LDCT Grand Challenge McCollough et al. (2017):** Authentic raw helical projections rebinned to fan-beam sinograms Wagner et al. (2023) of 1152 views $\times$ 368 detectors on abdominal scans to validate resilience against real scanning geometry and measurement noise, with 72, 144, and 288 views evaluated under mixed conditions coupled with 25% and 50% retained-dose levels, alongside unseen dose shifts (75% and 20%).

Table 1: Reconstruction on AAPM and LDCT Grand Challenge. Bold / underlined: best / second best. DC: observed-angle sinogram L_1 error against GT. LDCT percentages denote retained dose.

(a) AAPM PSNR (dB)↑ / SSIM↑ / LPIPS (10^{-2})↓ / DC (10^{-1})↓

Method	Chest / abdomen (In-Distribution)			Cranial (OOD)		
	16-view	32-view	64-view	16-view	32-view	64-view
FBP	18.10/0.413/40.94/57.82	21.44/0.510/32.88/31.80	26.23/0.646/22.59/12.69	22.05/0.521/39.40/48.16	25.58/0.602/31.32/29.72	31.38/0.723/19.86/11.96
FBPConvNet	29.74/0.838/7.35/6.48	32.49/0.877/5.18/3.57	35.86/0.916/3.11/1.67	29.95/0.863/7.88/8.64	34.46/0.919/3.75/4.80	40.26/0.960/1.26/2.31
LEARN++	29.71/0.830/7.79/2.57	35.23/0.910/3.32/0.74	38.45/0.939/1.85/0.45	31.66/0.892/5.91/2.76	40.88/0.974/0.96/0.54	46.75/0.991/0.29/0.30
DREAM-Net	31.59/0.864/6.39/2.94	34.76/0.905/3.95/1.34	36.90/0.927/2.59/1.50	33.19/0.910/4.11/3.24	38.08/0.954/1.77/1.71	42.41/0.978/0.87/1.35
FreeSeed	30.09/0.837/7.27/6.76	32.32/0.875/5.42/6.56	35.08/0.911/3.87/1.91	31.38/0.883/6.53/6.74	34.22/0.895/4.65/8.14	40.89/0.972/1.14/1.88
ProCT	29.21/0.839/7.54/6.72	32.39/0.879/5.20/3.78	35.47/0.913/3.41/2.10	29.23/0.840/10.31/14.89	33.57/0.889/6.49/9.95	38.83/0.928/3.41/6.54
CvG-Diff	30.87/0.856/6.78/5.37	33.04/0.888/4.87/4.82	35.17/0.920/2.60/3.99	31.56/0.888/6.70/6.70	34.42/0.922/4.62/5.72	38.72/0.959/2.40/3.77
PrideDiff	29.30/0.813/5.40/4.43	33.19/0.875/2.78/1.94	36.76/0.922/1.50/0.92	29.81/0.845/6.96/7.99	36.19/0.931/2.90/2.26	42.92/0.976/0.82/1.07
ReCo-Diff	30.14/0.842/7.77/4.32	34.04/0.895/4.68/2.16	38.00/0.935/1.97/0.92	30.78/0.867/8.54/9.32	35.67/0.922/5.02/5.93	42.67/0.966/2.16/2.33
PAIL	32.97/0.882/5.30/2.01	36.12/0.918/3.07/1.12	38.99/0.943/1.65/2.85	36.53/0.948/2.40/1.54	41.46/0.977/0.89/0.97	47.64/0.992/0.23/2.35
HerKry (ours)	**34.44/0.898/4.48/1.47**	**37.58/0.928/2.69/0.46**	**40.25/0.949/1.35/0.37**	**39.47/0.967/1.53/0.76**	**45.63/0.989/0.38/0.29**	**51.38/0.996/0.08/0.20**

(b) LDCT Grand Challenge PSNR (dB)↑ / SSIM↑

Method	In-Distribution: training-range dose				OOD: unseen dose			
	144-view		288-view		144-view		288-view	
Dose	50%	25%	50%	25%	75%	20%	75%	20%
FBP	24.70/0.485	23.94/0.452	29.27/0.695	28.20/0.653	24.96/0.496	23.57/0.437	29.65/0.710	27.71/0.634
FBPConvNet	33.76/0.825	33.60/0.822	34.55/0.838	34.41/0.836	33.81/0.826	33.52/0.820	34.59/0.838	34.34/0.835
LEARN++	33.66/0.821	33.44/0.819	34.81/0.857	34.55/0.854	33.72/0.822	33.33/0.818	34.89/0.858	34.40/0.853
DREAM-Net	33.63/0.826	33.44/0.822	34.85/0.855	34.57/0.850	33.70/0.827	33.34/0.820	34.95/0.857	34.43/0.848
FreeSeed	33.46/0.819	33.37/0.818	33.88/0.836	33.82/0.835	33.49/0.820	33.32/0.817	33.89/0.837	33.79/0.834
ProCT	33.99/0.832	33.82/0.829	35.26/0.867	34.98/0.862	34.05/0.833	33.73/0.828	35.36/0.868	34.84/0.860
CvG-Diff	34.24/0.836	34.06/0.832	35.28/0.864	35.02/0.858	34.30/0.838	33.98/0.830	35.38/0.867	34.90/0.855
PrideDiff	33.62/0.837	33.41/0.832	35.12/0.873	34.77/0.866	33.69/0.839	33.32/0.830	35.25/0.876	34.62/0.863
ReCo-Diff	34.26/0.840	34.08/0.835	35.36/0.871	35.07/0.863	34.33/0.841	34.00/0.833	35.47/0.873	34.95/0.860
PAIL	34.09/0.829	33.92/0.826	34.76/0.838	34.62/0.836	34.15/0.830	33.84/0.825	34.81/0.839	34.54/0.835
HerKry (ours)	**34.39/0.844**	**34.19/0.840**	**35.54/0.877**	**35.22/0.871**	**34.46/0.846**	**34.09/0.838**	**35.58/0.879**	**35.08/0.868**

Baselines and Evaluation Protocols: We compare against ten representative baselines spanning analytical reconstruction (**FBP** Kak & Slaney (2001)), image-domain post-processing (**FBPConvNet** Jin et al. (2017)), deep unfolding networks (**LEARN++** Zhang et al. (2023a), **DreamNet** Zhang et al. (2023b), **PAIL** Zhang et al. (2025a)), cross-domain/context transformers (**FreeSeed** Ma et al. (2023), **ProCT** Ma et al. (2026)), and conditional diffusion models (**CVGDiff** Chen et al. (2025b), **PRIDeDiff** Lu et al. (2024), **ReCoDiff** Choi et al. (2026)).

Reconstruction quality is assessed within $[-1000, 400]$ HU using **PSNR**, **SSIM** Wang et al. (2004), and **LPIPS** Zhang et al. (2018). Data fidelity is measured via the ℓ_1 sinogram error (DC) on observed angles against ground truth. All checkpoints are chosen based strictly on validation fidelity. Full specifications are detailed in Appendix E.

3.7 Reconstruction Results and Robustness Analysis

In-Distribution Evaluation: As reported in Table 1(a), HerKry sets the new state-of-the-art across all view settings on AAPM in-distribution chest and abdominal CT. Under extreme 16-view sampling, HerKry achieves 34.44 dB PSNR, surpassing the strongest baseline PAIL by $+1.47$ dB, with substantial reductions in LPIPS perceptual error and sinogram DC residuals. In the authentic clinical scenario of LDCT Grand Challenge (Table 1(b)), which combines authentic helical-to-fan projection geometries with Poisson–Gaussian photon starvation, HerKry steadily outperforms both unrolled solvers and diffusion baselines across all training-range dose levels (reaching 35.54 dB at 288-view, 50% dose). These results confirm that optimizing physical DC within linear Hermite-coefficient space maintains strict consistency against real scanner physics.

Out-of-Distribution (OOD) Robustness: The robustness advantages of HerKry are even more prominent under distribution shifts. In zero-shot cranial OOD evaluation on AAPM (Table 1(a)), where dense cranial bone induces severe streak artifacts in baseline models, HerKry achieves exceptional gains, outperforming PAIL by $+2.94$ dB (16-view) and $+4.17$ dB (32-view), while reducing sinogram DC error to 0.20×10^{-1} at 64 views. Under LDCT unseen dose conditions (75% and extreme 20% retained dose in Table 1(b)), HerKry similarly demonstrates superior stability, achieving 35.08 dB at 288 views. Decoupling representation scale from projection geometry enables HerKry to regularize weakly constrained null-space directions structurally.

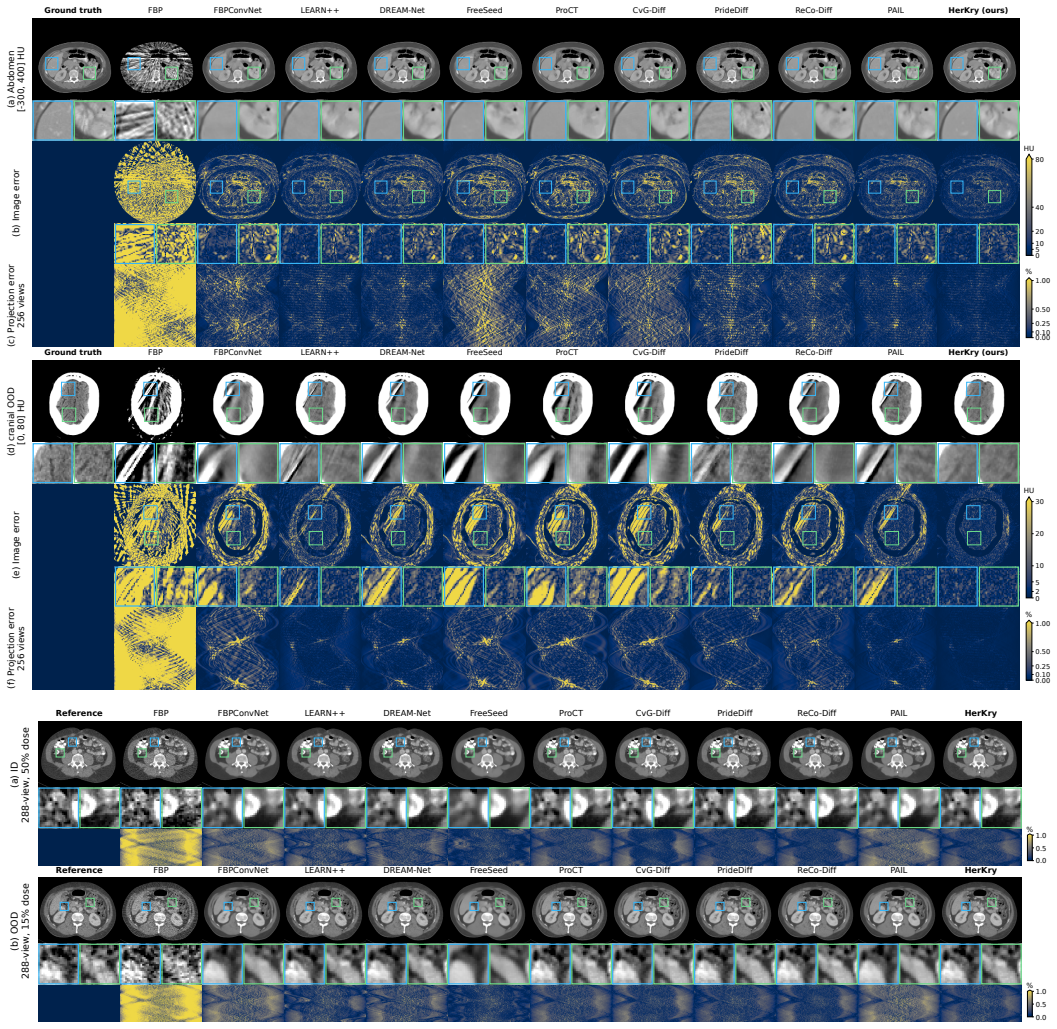

Figure 3: Qualitative reconstructions across AAPM (in-distribution abdomen and zero-shot cranial OOD) and LDCT Grand Challenge benchmarks. Panels display reconstructed slices with zoomed ROIs, absolute error maps, and projection error sinograms.

Qualitative and Projection Residual Analysis: Figure 3 visualizes the reconstructed slices, error maps, and full-view sinogram residuals across AAPM and LDCT Grand Challenge. For in-distribution scans, competing methods either leave noticeable radial streaks (FBP, FBPConvNet) or blur subtle vascular boundaries (CvG-Diff, ProCT). HerKry preserves sharp vessel interfaces and organ textures with errors predominantly below 10 HU. In cranial OOD transfers, baseline models suffer severe artifact flare-ups around high-attenuation bones, whereas HerKry reliably delineates fine ventricular margins. Furthermore, the 256-view projection error sinograms demonstrate that HerKry globally suppresses unobserved projection residuals across all geometries, verifying that our continuous HG hierarchy and KT-GRU momentum modeling strictly anchor the iterative optimization to physical CT acquisition geometry.

3.8 Ablation Study

We conduct systematic component diagnostics on 32-view AAPM CT to substantiate our key architectural and algorithmic choices. Table 2 summarizes the primary empirical comparisons.

Table 2: **Component diagnostics and solver comparisons on 32-view CT. (a)** Physical DC solver efficiency and fidelity across 12 intermediate unrolled blocks measured on RTX A6000. **(b)** Recurrent solver trajectory modeling across in-distribution and zero-shot cranial OOD regimes.

(a) Physical Data Consistency Formulation

Solver Variant	Local Block PSNR (dB) ↑	Latency (ms/12 blks) ↓
Adam-all (K steps)	32.13	264.18
Adam-all ($10K$ steps)	32.61	2313.44
HC-CGLS (K steps, ours)	**33.36**	**27.73**

(b) Recurrent Trajectory Modeling

Encoder Formulation	In-Dist. PSNR (dB) ↑	Cranial OOD PSNR ↑
No Coefficient Memory	36.81	44.93
GRU on Iterate $\boldsymbol{a}$	37.24	45.14
KT-GRU ($[\boldsymbol{a}, \Delta\boldsymbol{a}, \Delta^2\boldsymbol{a}]$, ours)	**37.58**	**45.63**

Linear HC-CGLS vs. Non-Convex Adam Optimization: As evaluated across all active unrolled solver blocks under identical initial states (Table 2(a)), optimizing linear coefficients via HC-CGLS achieves superior local fidelity (33.36 dB) while slashing execution latency by 89.5% (27.73 ms vs. 264.18 ms) compared to full-parameter Adam under matched step budgets K. Crucially, expanding Adam iterations to $10K$ steps improves CT image PSNR by 0.48 dB despite consuming $42\times$ more execution time (2313.44 ms). This demonstrates that non-convex coordinate updates exacerbate semi-convergence instabilities along ill-conditioned ray directions, whereas linear conjugate gradient updates in frozen-geometry HG space provide linear least-squares updates.

Continuous HG Hierarchy vs. Standard Gaussians: To isolate the benefit of polynomial basis modulation, we compare HerKry against a baseline configured strictly with zeroth-order Gaussians ($d = 0$) across identical stage structures (detailed in Appendix F). At the initial sparse stage ($N_1 = 32^2$), HerKry achieves a dramatic gain of $+2.48$ **dB** (31.90 dB vs. 29.42 dB, 95% CI $[2.35, 2.61]$). As primitive centers densify across stages, the margin naturally scales to $+1.23$ dB (Stage 2), $+0.50$ dB (Stage 3), and $+0.30$ dB (Stage 4). This monotonic progression empirically confirms our theoretical insight, that higher-order HG modes are indispensable for capturing anisotropic directional gradients when spatial centers are sparse, regularizing early degrees of freedom without sacrificing structural expressiveness.

Solver Dynamics Propagation via KT-GRU: Table 2(b) examines the impact of propagating internal solver dynamics across unrolled stages. Encoding the coefficient trajectory with velocity and acceleration cues $[\boldsymbol{a}_l^k, \Delta\boldsymbol{a}_l^k, \Delta^2\boldsymbol{a}_l^k]$ outperforms trajectory-agnostic models, consistently enhancing generalization under severe distribution shifts. This verifies that tracking solver momentum provides informative cues to regularize downstream blocks against high-frequency measurement noise.

Auxiliary Pre-Solve Supervision Annealing: We evaluate the stage-wise loss schedule $\boldsymbol{\rho}$. Relaxing intermediate constraints at the final stage via early-weighted supervision ($\boldsymbol{\rho} = [1, 0.5, 0.25, 0]$) achieves superior OOD generalization (45.63 dB) over uniform supervision (45.23 dB, where $\boldsymbol{\rho} = [0.5, 0.5, 0.5, 0.5]$). Under harsh simulated noise (0.5% sinogram RMS), early weighting retains $+2.14$ dB on validation and $+3.19$ dB on cranial CT over uniform supervision in Appendix F, corroborating that unconstraining the final dense primitive stage enables the model to absorb residual artifacts without overfitting early heuristics.

4 Conclusion

We propose HerKry to resolve the geometric mismatch in multiscale divergent-beam CT. By decoupling continuous Hermite–Gaussian representations from measurement scale, HerKry enforces geometry-consistent data consistency directly on native sinograms. Exploiting HG bases, HC-CGLS restricts physical DC optimization to linear coefficient space under frozen geometries, while KT-GRU recurrently propagates solver dynamics across stages. Extensive evaluations confirm superior fidelity and out-of-distribution robustness under clinical shifts.

AI USE STATEMENT

In this work, generative AI tools were used solely for English grammar editing, polishing academic writing clarity, and assisting with LaTeX formatting. All theoretical proofs and empirical findings were independently derived and implemented by the authors. The authors take full responsibility for the integrity, claims, and final content of this paper.

ETHICS STATEMENT

The primary objective of this work is to mitigate ionizing radiation risks in computed tomography by enabling high-fidelity reconstruction from sparse-view projections. All experiments were conducted using publicly available, fully anonymized clinical CT datasets that comply with established data governance protocols, requiring no direct involvement of human subjects or additional IRB approval.

REPRODUCIBILITY STATEMENT

To ensure reproducibility, all datasets used in our experiments are publicly available, and the complete source code—including model architectures, physical operators, and training pipelines—is provided in the anonymous repository at https://anonymous.4open.science/r/Code-for-HerKry-DB40.

REFERENCES

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.

Sayantan Bhadra, Varun A Kelkar, Frank J Brooks, and Mark A Anastasio. On hallucinations in tomographic image reconstruction. *IEEE transactions on medical imaging*, 40(11):3249–3260, 2021.

Du Chen, Liyi Chen, Zhengqiang Zhang, and Lei Zhang. Generalized and efficient 2d gaussian splatting for arbitrary-scale super-resolution. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26435–26445. IEEE, 2025a.

Jixiang Chen, Yiqun Lin, Yi Qin, Hualiang Wang, and Xiaomeng Li. Cross-view generalized diffusion model for sparse-view ct reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 140–150. Springer, 2025b.

Kyunghyun Cho, Bart Van Merriënboer, Çağlar Gulçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1724–1734, 2014.

Yong Eun Choi, Hyoung Suk Park, Kiwan Jeon, Hyun-Cheol Park, and Sung Ho Kang. Reco-diff: Residual-conditioned deterministic sampling for cold diffusion in sparse-view ct. *arXiv preprint arXiv:2603.02691*, 2026.

Qiaoqiao Ding, Hui Ji, Hao Gao, and Xiaoqun Zhang. Learnable multi-scale fourier interpolation for sparse view ct image reconstruction. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pp. 286–295, Cham, 2021. Springer International Publishing. ISBN 978-3-030-87231-1.

Reeves Fletcher and Colin M Reeves. Function minimization by conjugate gradients. *The computer journal*, 7(2):149–154, 1964.

Silvia Gazzola and Malena Sabaté Landman. Krylov methods for inverse problems: Surveying classical, and introducing new, algorithmic approaches. *GAMM-Mitteilungen*, 43(4):e202000017, 2020.

Lee W Goldman. Principles of ct: radiation dose and image quality. *Journal of nuclear medicine technology*, 35(4):213–225, 2007.

Per Christian Hansen. *Discrete inverse problems: insight and algorithms*. SIAM, 2010.

Andreas Hauptmann, Jonas Adler, Simon Arridge, and Ozan Öktem. Multi-scale learned iterative reconstruction. *IEEE transactions on computational imaging*, 6:843–856, 2020.

Magnus R Hestenes, Eduard Stiefel, et al. Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49(6):409–436, 1952.

Mu Hu and Guoqiang Xiao. Mascu-net: Memory-augmented adaptive shrinkage cascading unfolding network for interpretable sparse-view ct reconstruction. *Measurement Science and Technology*, 2026.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*, pp. 1–11, 2024.

Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing*, 26(9): 4509–4522, 2017.

Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis, et al. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Minjae Lee, Hyemi Kim, and Hee-Joung Kim. Sparse-view ct reconstruction based on multi-level wavelet convolution neural network. *Physica Medica*, 80:352–362, 2020. ISSN 1120-1797. doi: https://doi.org/10.1016/j.ejmp.2020.11.021.

Runrui Li, Qing Li, Hexi Wang, Saize Li, Juanjuan Zhao, Qiang Yan, and Long Wang. Ddptransformer: Dual-domain with parallel transformer network for sparse view ct image reconstruction. *IEEE Transactions on Computational Imaging*, 8:1101–1116, 2022. doi: 10.1109/TCI.2022.3207351.

Yingtai Li, Xueming Fu, Han Li, Shang Zhao, Ruiyang Jin, and S Kevin Zhou. 3dgr-ct: Sparse-view ct reconstruction with a 3d gaussian representation. *Medical Image Analysis*, 103:103585, 2025.

Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

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

Zexin Lu, Qi Gao, Tao Wang, Ziyuan Yang, Zhiwen Wang, Hui Yu, Hu Chen, Jiliu Zhou, Hongming Shan, and Yi Zhang. Pridediff: Physics-regularized generalized diffusion model for ct reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 9(2):157–168, 2024.

Chenglong Ma, Zilong Li, Junping Zhang, Yi Zhang, and Hongming Shan. Freeseed: Frequency-band-aware and self-guided network for sparse-view ct reconstruction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pp. 250–259, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-43999-5.

Chenglong Ma, Zilong Li, Junjun He, Junping Zhang, Yi Zhang, and Hongming Shan. Universal pre-training for generalizable incomplete-view ct reconstruction. *Pattern Recognition*, pp. 113513, 2026.

Cynthia H McCollough, Adam C Bartley, Rickey E Carter, Baiyu Chen, Tammy A Drees, Phillip Edwards, David R Holmes III, Alice E Huang, Farhana Khan, Shuai Leng, et al. Low-dose ct for the detection and classification of metastatic liver lesions: results of the 2016 low dose ct grand challenge. *Medical physics*, 44(10):e339–e352, 2017.

Taylor R Moen, Baiyu Chen, David R Holmes III, Xinhui Duan, Zhicong Yu, Lifeng Yu, Shuai Leng, Joel G Fletcher, and Cynthia H McCollough. Low-dose ct image and projection dataset. *Medical physics*, 48(2):902–911, 2021.

Long Peng, Anran Wu, Wenbo Li, Xinjie Zhang, Xueyuan Dai, Xin Di, Haoze Sun, Renjing Pei, Yang Wang, Yang Cao, et al. Pixel to gaussian: Ultra-fast continuous super-resolution with 2d gaussian modeling. In *International Conference on Learning Representations*, volume 2026, pp. 92172–92199, 2026.

Alexandre Refregier. Shapelets—i. a method for image analysis. *Monthly Notices of the Royal Astronomical Society*, 338(1):35–47, 2003.

Matteo Ronchetti. Torchradon: Fast differentiable routines for computed tomography. *arXiv preprint arXiv:2009.14788*, 2020.

Malena Sabaté Landman, Ander Biguri, Sepideh Hatamikia, Richard Boardman, John Aston, and Carola-Bibiane Schönlieb. On krylov methods for large-scale cbct reconstruction. *Physics in Medicine & Biology*, 68(15):155008, 2023.

Baoshun Shi, Ke Jiang, Qiusheng Lian, Xinran Yu, and Huazhu Fu. Prompting lipschitz-constrained network for multiple-in-one sparse-view ct reconstruction. *IEEE Transactions on Medical Imaging*, 2025.

Fabian Wagner, Mareike Thies, Laura Pfaff, Oliver Aust, Sabrina Pechmann, Daniela Weidner, Noah Maul, Maximilian Rohleder, Mingxuan Gu, Jonas Utz, et al. On the benefit of dual-domain denoising in a self-supervised low-dose ct setting. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5. IEEE, 2023.

Ce Wang, Kun Shang, Haimiao Zhang, Qian Li, and S. Kevin Zhou. Dudotrans: Dual-domain transformer for sparse-view ct reconstruction. In Nandinee Haq, Patricia Johnson, Andreas Maier, Chen Qin, Tobias Würfl, and Jaejun Yoo (eds.), *Machine Learning for Medical Image Reconstruction*, pp. 84–94, Cham, 2022. Springer International Publishing. ISBN 978-3-031-17247-2.

Tao Wang, Wenjun Xia, Jingfeng Lu, and Yi Zhang. A review of deep learning ct reconstruction from incomplete projection data. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 8(2):138–152, 2024. doi: 10.1109/TRPMS.2023.3316349.

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.

Jinxi Xiang, Yonggui Dong, and Yunjie Yang. Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging*, 40(5): 1329–1339, 2021.

Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. R^2-gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction. *arXiv preprint arXiv:2405.20693*, 2024.

Jianjia Zhang, Zirong Li, Jiayi Pan, Shaoyu Wang, and Weiwen Wu. Trustworthy limited data ct reconstruction using progressive artifact image learning. *IEEE Transactions on Image Processing*, 2025a.

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

Xinjie Zhang, Xingtong Ge, Tongda Xu, Dailan He, Yan Wang, Hongwei Qin, Guo Lu, Jing Geng, and Jun Zhang. Gaussianimage: 1000 fps image representation and compression by 2d gaussian splatting. In *European Conference on Computer Vision*, pp. 327–345. Springer, 2024.

Yi Zhang, Hu Chen, Wenjun Xia, Yang Chen, Baodong Liu, Yan Liu, Huaiqiang Sun, and Jiliu Zhou. Learn++: Recurrent dual-domain reconstruction network for compressed sensing ct. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 7(2):132–142, 2023a. doi: 10.1109/TRPMS.2022.3222213.

Yikun Zhang, Dianlin Hu, Shilei Hao, Jin Liu, Guotao Quan, Yi Zhang, Xu Ji, and Yang Chen. Dream-net: Deep residual error iterative minimization network for sparse-view ct reconstruction. *IEEE Journal of Biomedical and Health Informatics*, 27(1):480–491, 2023b. doi: 10.1109/JBHI.2022.3225697.

Yunxiang Zhang, Bingxuan Li, Alexandr Kuznetsov, Akshay Jindal, Stavros Diolatzis, Kenneth Chen, Anton Sochenov, Anton Kaplanyan, and Qi Sun. Image-gs: Content-adaptive image representation via 2d gaussians. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–11, 2025b.

Zhengqiang Zhang, Rongyuan Wu, Lingchen Sun, and Lei Zhang. Gpstoken: Gaussian parameterized spatially-adaptive tokenization for image representation and generation. *Advances in Neural Information Processing Systems*, 38:11218–11249, 2026.

Bo Zhou, S. Kevin Zhou, James S. Duncan, and Chi Liu. Limited view tomographic reconstruction using a cascaded residual dense spatial-channel attention network with projection data fidelity layer. *IEEE Transactions on Medical Imaging*, 40(7):1792–1804, 2021. doi: 10.1109/TMI.2021.3066318.

A Related Work

Sparse-view CT Reconstruction and OOD Problem: Deep reconstruction networks for sparse-view CT have evolved from image-domain artifact filtering Jin et al. (2017) and model-based deep unfolding Zhang et al. (2023a;b; 2025a) to dual-domain transformers Li et al. (2022); Wang et al. (2022) and generative diffusion priors Chen et al. (2025b); Lu et al. (2024); Choi et al. (2026). While these architectures achieve remarkable in-distribution fidelity, sparse angular sampling leaves substantial null-space directions unconstrained by measurements. Consequently, learned priors are susceptible to clinical distribution shifts, frequently fabricating hallucinations or amplifying streak artifacts under unseen anatomical structures or sparsity levels Antun et al. (2020); Bhadra et al. (2021). Existing robustness strategies predominantly modulate the learned prior via prompt conditioning or Lipschitz regularization Ma et al. (2026); Shi et al. (2025). HerKry addresses robustness structurally from the representation space. It regularizes early inference by progressively expanding representation degrees of freedom, while maintaining rigorous physical data consistency against native projections throughout the hierarchy.

Multiscale CT Reconstruction and Geometric Errors: Coarse-to-fine strategies spanning from wavelet transforms Lee et al. (2020), learnable Fourier pyramids Ding et al. (2021), and multi-resolution unrolled schemes Hauptmann et al. (2020) are widely adopted to regularize ill-posed inverse problems. However, existing frameworks operate on rigid, isotropic pixel grids, restricting their ability to represent localized directional structures at coarse scales. More fundamentally, discrete spatial decimation does not commute with divergent fan-beam projections ($\mathcal{D}_y\mathcal{A}^{\text{fine}} \neq \mathcal{A}^{\text{coarse}}\mathcal{D}_x$, Appendix B), injecting geometric errors that prevent enforcing physically valid DC in compact discrete spaces. HerKry circumvents this discrete mismatch by modeling the image via continuous and anisotropic Hermite–Gaussian primitives, decoupling representation capacity from projection geometry and enabling physical DC directly against native sinograms across all stages.

Gaussian Primitives and Continuous Representations: Originating from view synthesis, 3D Gaussian Splatting Kerbl et al. (2023) has been extended to tomographic reconstruction via radiative attenuation modeling and per-scan test-time volume optimization Zha et al. (2024); Li et al. (2025). Concurrently, 2D Gaussian primitives have emerged as compact continuous representations

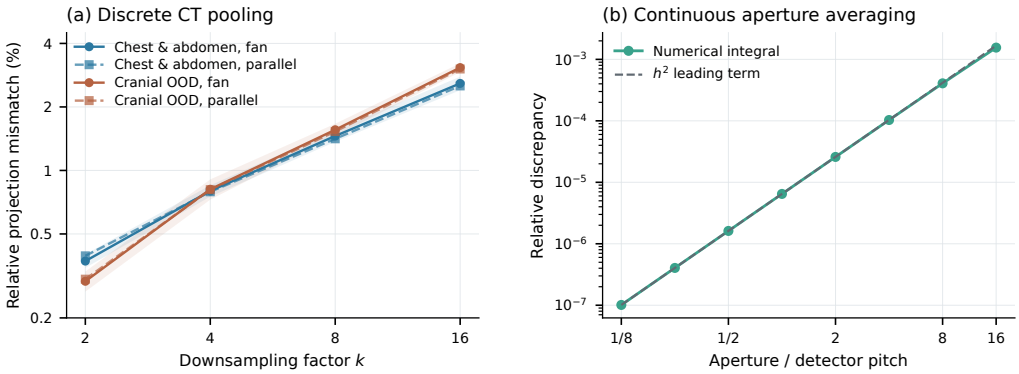

Figure 4: CT-sinogram pyramid mismatch: (a) discrete fan-beam or parallel-beam comparisons; (b) continuous aperture averaging and its h^2 leading term.

for image compression, arbitrary-scale super-resolution, and visual tokenization Zhang et al. (2024; 2025b); Peng et al. (2026); Zhang et al. (2026). However, both 2D and 3D frameworks rely strictly on standard zeroth-order Gaussian envelopes, necessitating computationally heavy, non-convex optimization over spatial coordinates and covariances to capture complex textures and deformations. Departing from per-scene 3D optimization, HerKry designs a feed-forward 2D continuous hierarchy modulated by orthogonal Hermite polynomials Refregier (2003). Because Hermite-Gaussian bases span the local spatial derivatives of Gaussians, HerKry absorbs local geometric deformations purely in linear coefficient space under a frozen geometry, transforming physical DC from non-convex parameter tuning to an efficient linear least-squares solve.

Physical Data Consistency and Krylov Trajectory Modeling: In deep unfolded CT reconstruction, physical data consistency (DC) is frequently simplified to heuristic measurement replacement or few-step gradient descent Zhang et al. (2023a); Zhou et al. (2021); Xiang et al. (2021); Hu & Xiao (2026); Zhang et al. (2023b; 2025a); Wang et al. (2022); Ma et al. (2023); Choi et al. (2026), bypassing least-squares optimization. While Krylov solvers like CGLS offer principled regularization through semi-convergence Fletcher & Reeves (1964); Hansen (2010); Gazzola & Sabaté Landman (2020), their existing multi-step architectures treat solvers as black boxes—discarding intermediate convergence trajectories Hu & Xiao (2026). HerKry circumvents this bottleneck by solving CGLS efficiently in reduced Hermite-coefficient space, and recurrently modeling the internal Krylov iterates to carry optimization dynamics across stages.

B Geometric Inconsistency of Coarse-Scale Fan-Beam Projections

In this section, we formalize why the non-commutation relation in equation 1 intensifies across downsampling levels. In clinical practice, raw helical CT acquisitions are routinely rebinned into sequential 2D fan-beam geometries Wagner et al. (2023) to facilitate slice-by-slice reconstruction. We therefore focus our theoretical analysis on the canonical fan-beam geometry Kak & Slaney (2001). Let $p(\theta, s)$ denote the continuous 2D Radon transform in parallel-beam coordinates, where $\theta \in [0, 2\pi)$ is the projection angle and $s \in \mathbb{R}$ is the radial detector coordinate. Under an equiangular circular fan-beam trajectory with source-to-isocenter distance D, the native fan-beam projection $g(\beta, \xi)$ is related to $p(\theta, s)$ through the coordinate transformation:

$$\theta(\beta, \xi) = \beta + \arctan\left(\frac{\xi}{D}\right), \qquad s(\xi) = \frac{D\xi}{\sqrt{D^2 + \xi^2}}, \tag{11}$$

such that $g(\beta, \xi) = p(\theta(\beta, \xi), s(\xi))$, where β denotes the source rotation angle.

Asymptotic Expansion of Detector Binning: Consider an idealized 1D detector-axis downsampling operator $\mathcal{D}_{y,h}$ with an effective averaging aperture $h > 0$:

$$(\mathcal{D}_{y,h}g)(\beta,\xi) = \frac{1}{h}\int_{-h/2}^{h/2} g(\beta,\xi+\eta)\,\mathrm{d}\eta, \tag{12}$$

where η serves as the local integration dummy variable along the detector array. Let $\theta_0 = \theta(\beta,\xi)$ and $s_0 = s(\xi)$ denote the reference coordinates, and define the local metric derivatives as:

$$a \triangleq \frac{\partial\theta}{\partial\xi}, \quad b \triangleq \frac{\partial s}{\partial\xi}, \quad a' \triangleq \frac{\partial^2\theta}{\partial\xi^2}, \quad b' \triangleq \frac{\partial^2 s}{\partial\xi^2}. \tag{13}$$

Expanding $g(\beta,\xi+\eta) = p(\theta(\beta,\xi+\eta), s(\xi+\eta))$ around (θ_0, s_0) via a bivariate Taylor series in η, and integrating over $[-h/2, h/2]$, yields:

$$(\mathcal{D}_{y,h}g)(\beta,\xi) = p(\theta_0,s_0) + \frac{h^2}{24}\left(b^2 p_{ss} + 2ab p_{\theta s} + a^2 p_{\theta\theta} + a' p_\theta + b' p_s\right) + \mathcal{O}(h^4), \tag{14}$$

where subscripts on p denote partial derivatives evaluated at (θ_0, s_0) (e.g., $p_{ss} \triangleq \frac{\partial^2 p}{\partial s^2}$, $p_{\theta s} \triangleq \frac{\partial^2 p}{\partial\theta\partial s}$, and $p_s \triangleq \frac{\partial p}{\partial s}$), and the odd powers of η vanish symmetrically.

In contrast, performing downsampling purely on CT image approximates to spatial low-pass filtering. In the projection domain, this corresponds to filtering solely along the radial coordinate s under fixed beam direction θ_0. To isolate the effect of detector-induced angular coupling, we introduce an idealized radial averaging reference at θ_0, with locally matched width $\delta s = |b|h$. Its expansion gives:

$$(\mathcal{D}_{s,|b|h}p)(\theta_0,s_0) = \frac{1}{|b|h}\int_{-|b|h/2}^{|b|h/2} p(\theta_0, s_0+\zeta)\,\mathrm{d}\zeta = p(\theta_0,s_0) + \frac{h^2}{24}b^2 p_{ss} + \mathcal{O}(h^4), \tag{15}$$

where ζ is the radial integration offset. Subtracting equation 15 from equation 14 reveals the intrinsic discrepancy specific to fan-beam geometry:

$$e_h(\beta,\xi) \triangleq (\mathcal{D}_{y,h}g)(\beta,\xi) - (\mathcal{D}_{s,|b|h}p)(\theta_0,s_0) = \frac{h^2}{24}C_{\mathrm{geo}}(\beta,\xi) + \mathcal{O}(h^4), \tag{16}$$

with the geometric coupling factor $C_{\mathrm{geo}}(\beta,\xi)$:

$$C_{\mathrm{geo}}(\beta,\xi) \triangleq 2ab\,p_{\theta s} + a^2\,p_{\theta\theta} + a'\,p_\theta + b'\,p_s. \tag{17}$$

The presence of non-radial derivative terms ($p_\theta, p_{\theta\theta}, p_{\theta s}$) directly stems from divergent ray trajectories. The traversing across adjacent detector channels simultaneously changes the projection angle θ (i.e., $a \neq 0$). These terms characterize the local discrepancy between detector averaging and the matched radial reference.

Scale-Dependent Error Scaling: Let h_0 denote the detector pitch in this rescaled coordinate. For a multiscale pyramid with downsampling factor $r > 1$, the effective aperture width at stage l scales as $h_l = r^l h_0$. Substituting h_l into equation 16 gives:

$$e_{h_l}(\beta,\xi) = r^{2l}\cdot\left[\frac{h_0^2}{24}C_{\mathrm{geo}}(\beta,\xi)\right] + \mathcal{O}(r^{4l}h_0^4). \tag{18}$$

For smooth projections with $C_{\mathrm{geo}} \neq 0$, the discrepancy has an h_l^2 leading term in the small-aperture regime.

Empirical Verification: To validate this scale-dependent mismatch without interference from neural networks or learned priors, we benchmark the normalized operator mismatch directly in projection space:

$$E_l = \frac{\left\|\mathcal{D}_y^{(l)}\mathcal{A}^{\mathrm{fine}}\boldsymbol{x} - \mathcal{A}^{(l)}\mathcal{D}_x^{(l)}\boldsymbol{x}\right\|_2}{\left\|\mathcal{D}_y^{(l)}\mathcal{A}^{\mathrm{fine}}\boldsymbol{x}\right\|_2}, \tag{19}$$

where $\mathcal{D}_x^{(l)}$ and $\mathcal{D}_y^{(l)}$ denote cumulative l-level 2D image pooling and 1D sinogram pooling respectively, and $\mathcal{A}^{(l)}$ is the discrete fan-beam projection matrix matched to the coarsened image grid. As plotted in Fig. 4, the discrete mismatch increases with coarsening, with log-log slopes of 0.93 and 1.10 for chest/abdomen and cranial scans over factors $2, 4, 8, 16$.

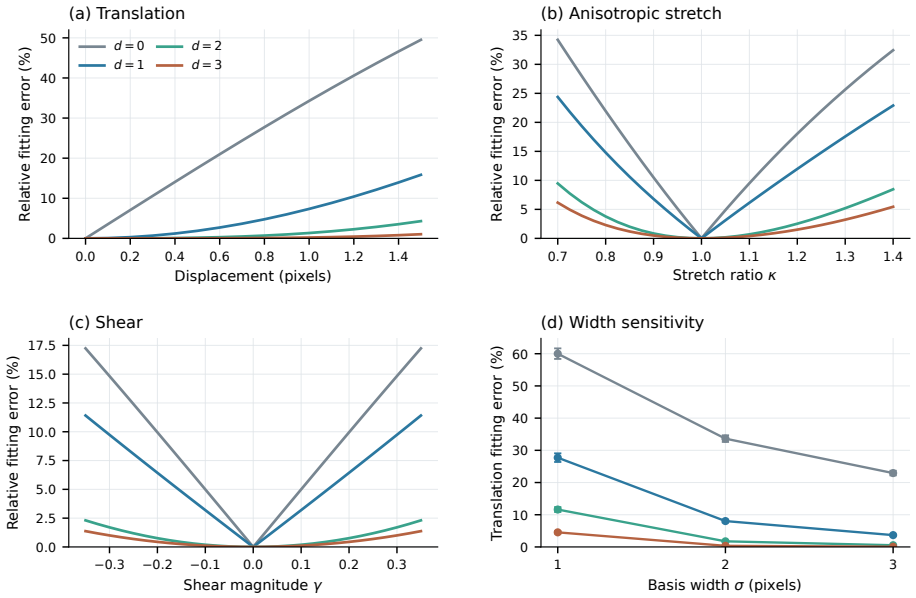

Figure 5: Fixed-geometry HG fitting under translation, stretch, shear, and varying basis width. Only coefficients are optimized. Error bars indicate 95% bootstrap intervals.

C Local Deformation Expressiveness of HG Modes

In HerKry, the linear physical optimization branch freezes Gaussian spatial parameters $\boldsymbol{\theta}_i = \{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$ and relies entirely on HC-CGLS to optimize linear coefficients $\boldsymbol{a}_l$.

Below, we show via Taylor analysis that the tangent space of local geometric perturbations, such as translations and anisotropic stretches, is algebraically subsumed by the linear span of Hermite-Gaussian bases Refregier (2003). Besides, we emphasize that HerKry does not explicitly compute Taylor coefficients at inference. Rather, this property guarantees that a convex linear solver (e.g., HC-CGLS) intrinsically possesses the expressive capacity to absorb local geometric misalignments.

Absorption of Local Translations: Let $G(\boldsymbol{v}) = (2\pi)^{-1}\exp(-\frac{1}{2}\|\boldsymbol{v}\|^2)$ denote the normalized 2D standard Gaussian envelope, where $\boldsymbol{v} = \boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{u} - \boldsymbol{\mu}_i)$. A local spatial translation $\Delta\boldsymbol{\mu}_i$ induces a normalized displacement $\boldsymbol{\delta}_i = \boldsymbol{\Sigma}_i^{-1/2}\Delta\boldsymbol{\mu}_i$. Expanding the shifted Gaussian around $\boldsymbol{v}$ yields:

$$G(\boldsymbol{v} - \boldsymbol{\delta}_i) = G(\boldsymbol{v}) - \boldsymbol{\delta}_i \cdot \nabla G(\boldsymbol{v}) + \frac{1}{2}\boldsymbol{\delta}_i^\top \nabla^2 G(\boldsymbol{v})\boldsymbol{\delta}_i + \dots \tag{20}$$

Under the normalization in equation 2, $\partial_{v_x} G = -\psi_{10}/\sqrt{2}$ and $\partial_{v_y} G = -\psi_{01}/\sqrt{2}$. Hence,

$$G(\boldsymbol{v} - \boldsymbol{\delta}_i) = \psi_{00}(\boldsymbol{v}) + \frac{\delta_{i,x}}{\sqrt{2}}\psi_{10}(\boldsymbol{v}) + \frac{\delta_{i,y}}{\sqrt{2}}\psi_{01}(\boldsymbol{v}) + \mathcal{O}(\|\boldsymbol{\delta}_i\|^2). \tag{21}$$

Substituting these identities into the first-order truncation of equation 20 gives:

$$G(\mathbf{v} - \boldsymbol{\delta}_i) \approx \psi_{00}(\mathbf{v}) + \frac{\delta_{i,x}}{\sqrt{2}}\psi_{10}(\mathbf{v}) + \frac{\delta_{i,y}}{\sqrt{2}}\psi_{01}(\mathbf{v}) \in \mathrm{span}\{\psi_{pq} \mid p + q \le 1\}. \tag{22}$$

Consequently, correcting a local displacement $\Delta\boldsymbol{\mu}_i$ to first-order accuracy does not require shifting the spatial center $\boldsymbol{\mu}_i$. It can be achieved purely by modulating the linear combination of the zeroth- and first-order HG coefficients.

Absorption of Anisotropic Shape Perturbations: Similarly, consider a small coordinate distortion $\boldsymbol{v} \mapsto (\boldsymbol{I} + \boldsymbol{E})\boldsymbol{v}$, where $\boldsymbol{E} \in \mathbb{R}^{2\times 2}$ captures local shear and anisotropic stretching on the premise

of $\|\boldsymbol{E}\|_2 \ll 1$. The first-order expansion of the perturbed envelope yields:

$$G((\boldsymbol{I}+\boldsymbol{E})\boldsymbol{v}) = G(\boldsymbol{v})\left[1 - \frac{1}{2}\boldsymbol{v}^\top(\boldsymbol{E}+\boldsymbol{E}^\top)\boldsymbol{v}\right] + \mathcal{O}(\|\boldsymbol{E}\|_2^2). \tag{23}$$

The perturbation term $\boldsymbol{v}^\top(\boldsymbol{E}+\boldsymbol{E}^\top)\boldsymbol{v}$ is an arbitrary homogeneous quadratic polynomial in (v_x, v_y). Since quadratic polynomial-Gaussian products are spanned by the second-order HG modes:

$$v_x^2 G(\boldsymbol{v}),\ v_y^2 G(\boldsymbol{v}),\ v_x v_y G(\boldsymbol{v}) \in \mathrm{span}\{\psi_{00}, \psi_{20}, \psi_{02}, \psi_{11}\}, \tag{24}$$

the second-order HG bases can locally compensate for anisotropic scaling and shearing without explicitly optimizing the covariance matrix $\boldsymbol{\Sigma}_i$.

Empirical Verification via Linear Least-Squares Solver: To verify that this theoretical property translates into practical utility for linear solvers, we conduct controlled fitting experiments under a frozen geometry $\boldsymbol{\theta}_i = \{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$. We synthesize target primitives subject to random translations $\|\Delta\boldsymbol{\mu}\|_2 \le 1.5$ pixels and anisotropic stretch ratios $\kappa \in [0.7, 1.4]$. Instead of optimizing spatial coordinates, we only solve a linear least-squares problem in coefficient space via conjugate gradient method Hestenes et al. (1952):

$$E_{\mathrm{fit}}(d) = \frac{\min_{\boldsymbol{a}} \|\boldsymbol{x}_{\mathrm{target}} - \mathcal{B}_d(\boldsymbol{\theta}_i)\boldsymbol{a}\|_2}{\|\boldsymbol{x}_{\mathrm{target}}\|_2}. \tag{25}$$

As shown in Fig. 5, while standard Gaussians of $d = 0$ fail to fit translated or stretched shapes under a fixed center, order $d = 1$ reduces shift-induced error by over 75%, and order $d = 2$ effectively absorbs anisotropic shear. This confirms that linear solvers like HC-CGLS can reliably absorb geometric misalignments under fixed primitive geometries.

D Algorithmic Pipeline of HerKry

Algorithm 1 details the stage-wise reconstruction pipeline of HerKry. Note that the within-stage loop is omitted for brevity. Each stage applies three blocks, carrying the updated image and memory forward. Block 12 skips the physical-solver and trajectory steps.

Algorithm 1 HerKry Reconstruction Pipeline

Require: Sparse-view sinogram $\boldsymbol{y}_s$, forward projector $\mathcal{A}$, stages L, HC-CGLS iterations K
Ensure: Final reconstructed CT slice $\boldsymbol{x}_L^{\mathrm{vis}}$

Initialize visual base $\boldsymbol{x}_0^{\mathrm{vis}} \leftarrow \mathrm{Smooth}(\mathrm{FBP}(\boldsymbol{y}_s))$, memory $\boldsymbol{h}_0^K \leftarrow \boldsymbol{0}$, residual feature $\boldsymbol{U}_0 \leftarrow \boldsymbol{0}$
for $l = 1, \ldots, L$ **do**
  ▷ **Nonlinear Neural Prior Branch**
  Predict $\{\boldsymbol{\Theta}_l, \boldsymbol{a}_l^0, \boldsymbol{W}_l\} \leftarrow \mathrm{Backbone}_l(\boldsymbol{x}_{l-1}^{\mathrm{vis}}, \boldsymbol{h}_{l-1}^K, \boldsymbol{U}_{l-1})$
  Direct update: $\boldsymbol{x}_l^{\mathrm{vis}} \leftarrow \boldsymbol{x}_{l-1}^{\mathrm{vis}} + \mathrm{Conv}(\mathrm{GELU}(\mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{W}_l)) + \mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^0$
  ▷ **Linear Physical Optimization Branch**
  Linear init: $\boldsymbol{x}_l^{\mathrm{init}} \leftarrow \boldsymbol{x}_{l-1}^{\mathrm{vis}} + \mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^0$
  Instantiate operator $\mathcal{P}_l \leftarrow \mathcal{A}\mathcal{B}_l(\boldsymbol{\Theta}_l)$ and residual $\boldsymbol{r}_l^0 \leftarrow \boldsymbol{y}_s - \mathcal{A}\boldsymbol{x}_l^{\mathrm{init}}$
  Solve $\min_{\boldsymbol{a}_l} \frac{1}{2}\|\mathcal{P}_l\boldsymbol{a}_l - (\boldsymbol{y}_s - \mathcal{A}\boldsymbol{x}_{l-1}^{\mathrm{vis}})\|_2^2$ via K CGLS steps from $\boldsymbol{a}_l^0 \implies \{\boldsymbol{a}_l^k\}_{k=1}^K$
  ▷ **Krylov Trajectory Modeling**
  Initialize solver state $\boldsymbol{h}_l^0 \leftarrow \boldsymbol{h}_{l-1}^K$, set $\Delta\boldsymbol{a}_l^0 \leftarrow \boldsymbol{0}$
  for $k = 1, \ldots, K$ **do**
    $\Delta\boldsymbol{a}_l^k \leftarrow \boldsymbol{a}_l^k - \boldsymbol{a}_l^{k-1}, \quad \Delta^2\boldsymbol{a}_l^k \leftarrow \Delta\boldsymbol{a}_l^k - \Delta\boldsymbol{a}_l^{k-1}$
    $\boldsymbol{h}_l^k \leftarrow \mathrm{GRU}\left(\mathcal{E}_l([\boldsymbol{a}_l^k, \Delta\boldsymbol{a}_l^k, \Delta^2\boldsymbol{a}_l^k]),\ \boldsymbol{h}_l^{k-1}\right)$
  end for
  Form spatial residual features $\boldsymbol{U}_l$ via equation 10
end for
return $\boldsymbol{x}_L^{\mathrm{vis}}$

E Detailed Experimental Settings

E.1 Dataset and Preprocessing

AAPM Benchmark: The AAPM Low Dose CT dataset Moen et al. (2021) consists of chest, abdominal, and cranial scans. Slices are normalized, truncated to $[-1000, 400]$ HU, and resized to 256×256. We partition the chest and abdominal cases into 136 training subjects (17,294 slices), 30 validation subjects (3,542 slices), and 30 testing subjects (4,021 slices). Projections are numerically generated via the forward operator $\mathcal{A}$ under an equiangular circular fan-beam trajectory (source-to-isocenter distance $D = 1075$ mm, source-to-detector distance $D_{\text{sd}} = 1536$ mm, with 256 detector bins over 360°). Subsampling to 16, 32, and 64 views corresponds to $\frac{1}{32}$, $\frac{1}{16}$, and $\frac{1}{8}$ of full-view sampling, respectively. To rigorously evaluate OOD cross-organ generalization, 1,136 cranial CT slices are isolated exclusively for zero-shot inference.

LDCT Grand Challenge: To assess performance under real sinogram and different measurement noise levels, we utilize raw projection data from the Low Dose CT Grand Challenge McCollough et al. (2017). The abdominal helical raw sinograms are rebinned into 2D fan-beam sinograms with 1152 views and 368 detector bins following standard protocols Wagner et al. (2023). The dataset is partitioned by subject into 8 training cases (4,185 slices), 1 validation case (503 slices), and 1 test case (463 slices). The sparse subsets are extracted at 72, 144, and 288 views. To systematically test model robustness under photon starvation, from 25% to 50% low-dose acquisition conditions are simulated by injecting calibrated Poisson–Gaussian noise into the sinograms according to McCollough et al. (2017), and the 75% and 20% are evaluated as unseen conditions. The reconstructions are generated at 368×368 and evaluated within the standard 256×256 central region against reference CT images.

E.2 Implementation Details

HerKry Architecture HerKry employs an $L = 4$ stage architecture governed by the order-density complementary schedule: $(N_l, d_l) \in \{(32^2, 3), (64^2, 2), (128^2, 1), (256^2, 0)\}$. Each stage comprises 3 sequential blocks, each executing a complete cycle of HG prediction and DC process. The network maintains a persistent 48-channel latent feature state defined over the Gaussian grid across blocks.

For each primitive, the backbone predicts the geometric parameters alongside $(C+1) = 8$ sets of HG mode coefficients (totaling $8(d_l + 1)^2$ values per primitive), partitioned into $C = 7$ feature channels $\boldsymbol{W}_l$ for nonlinear decoding and 1 physical initial coefficient set $\boldsymbol{a}_l^0$ for linear HC-CGLS solving. Across stages, the physical branch executes $K \in \{2, 4, 6, 8\}$ HC-CGLS iterations. The KT-GRU maintains a compact 16-channel hidden state to track solver velocity and acceleration trajectories, which is subsequently reprojected into the 48-channel persistent latent state. The stage-wise loss weights in the loss function are set to $\boldsymbol{\lambda} = [1/8, 1/4, 1/2, 1]$ and $\boldsymbol{\rho} = [1, 1/2, 1/4, 0]$.

Training Specifications: All models are optimized using AdamW Loshchilov & Hutter (2017) with an initial learning rate of 2×10^{-4} decayed to 10^{-6} via a cosine annealing schedule Loshchilov & Hutter (2016) over 50 epochs with a batch size of 8. CT projection and back-projection operations are implemented via the TorchRadon library Ronchetti (2020). Training is conducted on NVIDIA RTX A6000 GPUs. All competing baselines are trained using identical data splits and spatial augmentations, including random horizontal flips, vertical flips, and 90° rotations. The checkpoints are selected based on validation PSNR.

E.3 Evaluation Metrics and Protocols

Reconstruction quality is quantitatively assessed within the standard clinical soft-tissue window of $[-1000, 400]$ HU using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) Wang et al. (2004), and Learned Perceptual Image Patch Similarity (LPIPS) with an AlexNet backbone Zhang et al. (2018) to capture perceptual fidelity. The data fidelity is evaluated by measuring the ℓ_1 residual norm between the reconstruction's projections and the ground-truth sinograms exclusively across the observed sparse views.

E.4 Computational Complexity and Runtime Analysis

Table 3: Publication venues, model parameters, and empirical inference cost per slice for 32-view reconstruction (256×256) on an NVIDIA RTX A6000 GPU.

Method	Venue	Params (M)	Time (ms/slice)
FBP	–	0.00	0.03
FBPConvNet	TIP 2017	31.43	5.13
LEARN++	TRPMS 2023	2.17	64.36
DREAM-Net	JBHI 2023	10.02	22.16
FreeSeed	MICCAI 2023	32.48	13.69
ProCT	PR 2026	6.52	37.40
CvG-Diff	MICCAI 2025	34.42	221.27
PrideDiff	TRPMS 2025	5.08	54.17
ReCo-Diff	arXiv 2026	34.42	306.80
PAIL	TIP 2025	7.53	274.86
HerKry (ours)	–	8.52	56.29

Notes: The average inference time per slice is measured for 32-view, 256×256 reconstruction on an NVIDIA RTX A6000, batch size 8, over 50 distinct test batches after 10 warmup batches. The model parameters refer to the evaluated 32-view implementation. FLOPs are omitted because the available profiler does not count custom Hermite-Gaussian splatting kernels.

As reported in Table 3, HerKry maintains a compact footprint of 8.52M parameters and an inference latency of 56.29 ms/slice. While delivering over 4.5 dB higher PSNR on 32-view reconstruction, HerKry runs $4\times$ to $5.5\times$ faster than diffusion-based counterparts (CvG-Diff at 221.27 ms and ReCo-Diff at 306.80 ms). This validates that solving physical DC strictly within linear Hermite-coefficient space avoids repetitive, expensive coordinate re-rendering caches, achieving practical clinical reconstruction throughput.

E.5 Visualization of HerKry Reconstruction Process

Evolution of Continuous HG Primitive Atlas across Stages: Fig. 8 visualizes the multi-stage structural transition governed by our order-density complementary schedule. For the abdominal example, the active spatial centers progressively densify around organ contours and complex tissue interfaces. In Stage 1 ($N_1 = 32^2, d_1 = 3$), primitives are sparsely distributed yet capture directional gradients and boundary orientations via cubic Hermite polynomials. Across subsequent stages, the basis order is gradually traded for spatial sampling density, ultimately culminating in Stage 4 ($N_4 = 256^2, d_4 = 0$) where zeroth-order Gaussians resolve localized subtle parenchyma and fine vascular bifurcations. The 1D feature cross-sections in Fig. 8 demonstrate that the higher-order HG modes smoothly bound the target field through linear superposition, validating that the underlying parameterization can accurately express local variations under frozen geometric centers.

Block-Wise Unrolled Trajectories and Solver Rectification: Figs. 10 and 11 dissect the stage-by-stage iteration dynamics across all 12 blocks on in-distribution abdomen and zero-shot cranial scans under 32 views. At each intermediate block, the neural prior predicts an initial linear increment $\mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^0$, which is subsequently rectified by HC-CGLS against measurement sinograms $\boldsymbol{y}_s$ to enforce physical consistency. Concurrently, the nonlinear visual update $\Delta\boldsymbol{x}_l^{\mathrm{NL}}$ is integrated to form the next visual state $\boldsymbol{x}_l^{\mathrm{vis}}$.

Notably, in Stage 4 (Blocks 10–12), the pre-HC-CGLS visual state $\boldsymbol{x}_l^{\mathrm{init}}$ displays noticeable global contrast deviations and lower-frequency variations. This behavior directly aligns with our stage-wise loss formulation. The supervision weight on the pre-solve state is deliberately annealed to zero at the final stage ($\rho_4 = 0$). Eliminating pre-solve constraints at Stage 4 allows the dense primitive network (256^2) unconstrained flexibility to allocate its representational capacity toward high-frequency details, rather than penalizing intermediate pre-solver approximations. As depicted in the subsequent rows, these pre-solve discrepancies are immediately eliminated by the physical HC-CGLS solve and balanced by the parallel visual state branch, yielding clean, artifact-free reconstructions at the final output $\boldsymbol{x}_L^{\mathrm{vis}}$.

E.5.1 Robustness across Acquisition Conditions:

As shown in Fig. 9, we evaluate frozen HerKry checkpoints across 8-128 views on AAPM and 5-100% retained relative dose on LDCT, without test-time adaptation. The reconstruction quality generally improves with increasing view count or retained dose. On AAPM dataset, the performance deterioration becomes pronounced at 8-10 views below the training range. On LDCT Grand Challenge dataset, reducing the retained dose from 25% to 5% decreases PSNR by 1.13/1.46 dB at 144/288 views, respectively. These results characterize the sensitivity to acquisition conditions; the LDCT evaluation contains one test patient.

F Detailed Ablation Studies and Diagnostic Analyses

F.1 Stage-Wise Expressiveness of the Continuous HG Hierarchy

To rigorously validate the effect of Hermite-Gaussian bases across scales, Table 4 benchmarks stage-by-stage reconstructions on the full 4,021-slice AAPM test cohort of 32-view sparsity setting against a baseline constrained to standard zeroth-order Gaussians ($d = 0$).

Table 4: **Stage-wise reconstruction performance across the hierarchy on 32-view AAPM test.** ΔPSNR denotes slice-averaged gain with 95% patient-cluster bootstrap confidence intervals.

Hierarchy Stage	Gaussian-0 ($d = 0$) PSNR (dB) / SSIM	HerKry (HG Hierarchy) PSNR (dB) / SSIM	ΔPSNR [95% CI] (dB)
Stage 1 ($N = 32^2$)	29.419 / 0.8276	**31.903 / 0.8702**	**+2.484** [2.346, 2.608]
Stage 2 ($N = 64^2$)	34.004 / 0.8962	**35.234 / 0.9085**	**+1.229** [1.110, 1.339]
Stage 3 ($N = 128^2$)	36.384 / 0.9195	**36.883 / 0.9238**	**+0.499** [0.436, 0.554]
Stage 4 ($N = 256^2$)	37.274 / 0.9260	**37.576 / 0.9281**	**+0.303** [0.257, 0.345]

The empirical results exhibit a pronounced coarse-to-fine trend: at Stage 1 where spatial centers are sparse ($N_1 = 1024$), high-order HG polynomials ($d_1 = 3$) yield a substantial gain of $+2.484$ dB. This confirms that directional Hermite bases effectively regularize early unconstrained degrees of freedom while maintaining superior geometric expressiveness. At finer stages, as spatial sampling densifies, lower-order modes suffice, validating our order-density complementary schedule.

F.2 Comprehensive Comparison of Physical DC Solvers

Table 5 expands the comparison between HC-CGLS and various gradient-based Adam configurations evaluated across 11 intermediate unrolled solver blocks.

Table 5: **Comprehensive diagnostic comparison across physical DC solver variants.** Evaluated on 8 patient volumes across 12 active solver blocks under matched intermediate inputs on an NVIDIA RTX A6000 GPU.

Solver Formulation	Step Budget	Geo. Mean Sino. MSE ↓	Local Block PSNR (dB) ↑	Total Latency (ms) ↓
Adam-linear	K	0.80345	31.501	23.00
Adam-linear	$10K$	0.02805	32.690	512.05
Adam-all	K	0.04384	32.129	264.18
Adam-all	$10K$	0.00563	32.606	2313.44
Adam-all ($\eta = 10^{-3}$)	$10K$	0.00595	31.426	2314.36
HC-CGLS (ours)	K	0.02119	**33.355**	**27.73**

Semi-Convergence and Non-Convex Pitfalls: In severely ill-conditioned sparse-view tomography, aggressively minimizing measurement residual without convex spectral filtering fits measurement noise and amplifies streak artifacts along null-space directions. HC-CGLS strikes the optimal balance through Krylov semi-convergence, achieving top image fidelity in merely 27.73 ms.

F.3 Noise Sensitivity under Supervision Weight Schedules

Table 6 assesses model resilience across various sinogram noise levels ($\sigma_{\text{noise}}/\text{RMS}_{\text{sino}}$) under distinct pre-solve loss weighting schemes $\boldsymbol{\rho}$.

Table 6: **Reconstruction robustness under varying acquisition noise across supervision schemes (32-view).** Evaluated over test slices across in-distribution validation and cranial OOD cohorts.

Supervision	In-Dist. Validation PSNR (dB) ↑			Zero-Shot Cranial OOD PSNR (dB) ↑		
Scheme $\boldsymbol{\rho}$	Clean	Noise 0.2%	Noise 0.5%	Clean	Noise 0.2%	Noise 0.5%
Uniform $[.5, .5, .5, .5]$	37.38	36.32	32.84	45.43	42.15	38.67
Unsupervised Pre-Solve $[0, 0, 0, 0]$	**37.59**	36.85	34.61	45.55	43.18	41.22
Early-Weighted $[1, .5, .25, 0]$	37.54	**36.91**	**34.98**	**45.63**	**43.34**	**41.86**

Under higher noise contamination (0.5%), our early-weighted annealing outperforms uniform supervision by $+2.14$ dB on validation and $+3.19$ dB on cranial OOD. Enforcing strict pre-solver fidelity at the terminal stage ($\rho_4 > 0$) severely penalizes necessary fine-grained adjustments, forcing the network to underfit noisy sinograms, whereas annealing $\rho_4 \to 0$ grants the final HG representation maximum capacity to suppress structured noise.

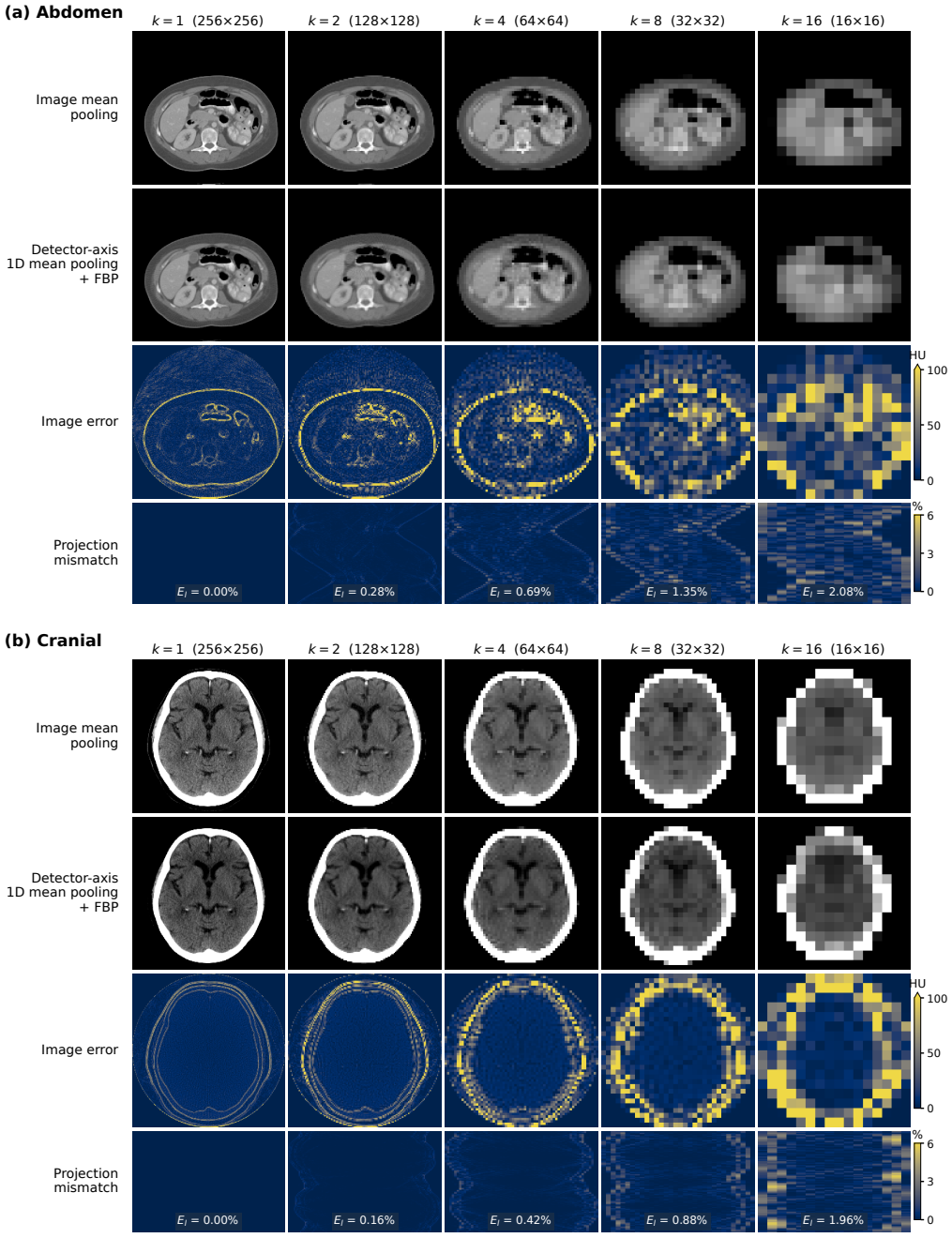

Figure 6: Qualitative CT-sinogram pyramid mismatch on real AAPM slices with 256-view projections. Each group shows image 2D mean pooling, detector-axis 1D mean pooling followed by FBP, their absolute image difference, and direct projection mismatch. CT windows are $[-300, 400]$ HU for abdomen and $[0, 80]$ HU for cranial images. Residuals use fixed linear scales: 0–100 HU and 0–6% of the full-resolution sinogram peak. Image differences additionally include FBP discretization error.

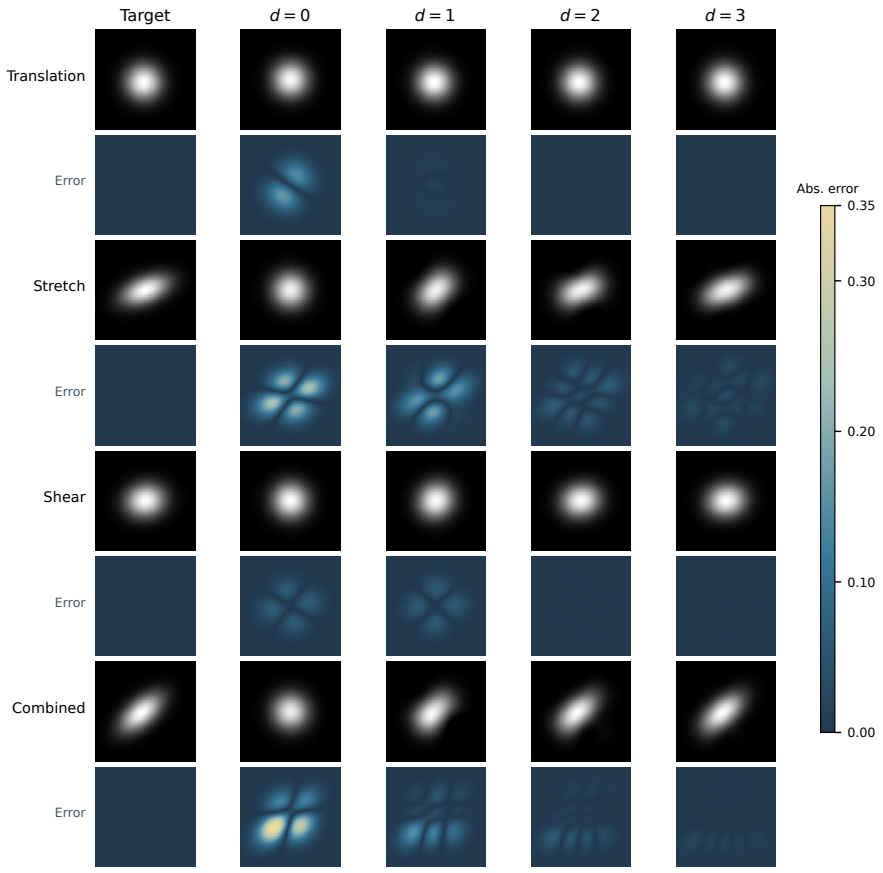

Figure 7: Frozen-geometry fitting examples and absolute errors for HG orders $d = 0, \ldots, 3$, using shared display scales.

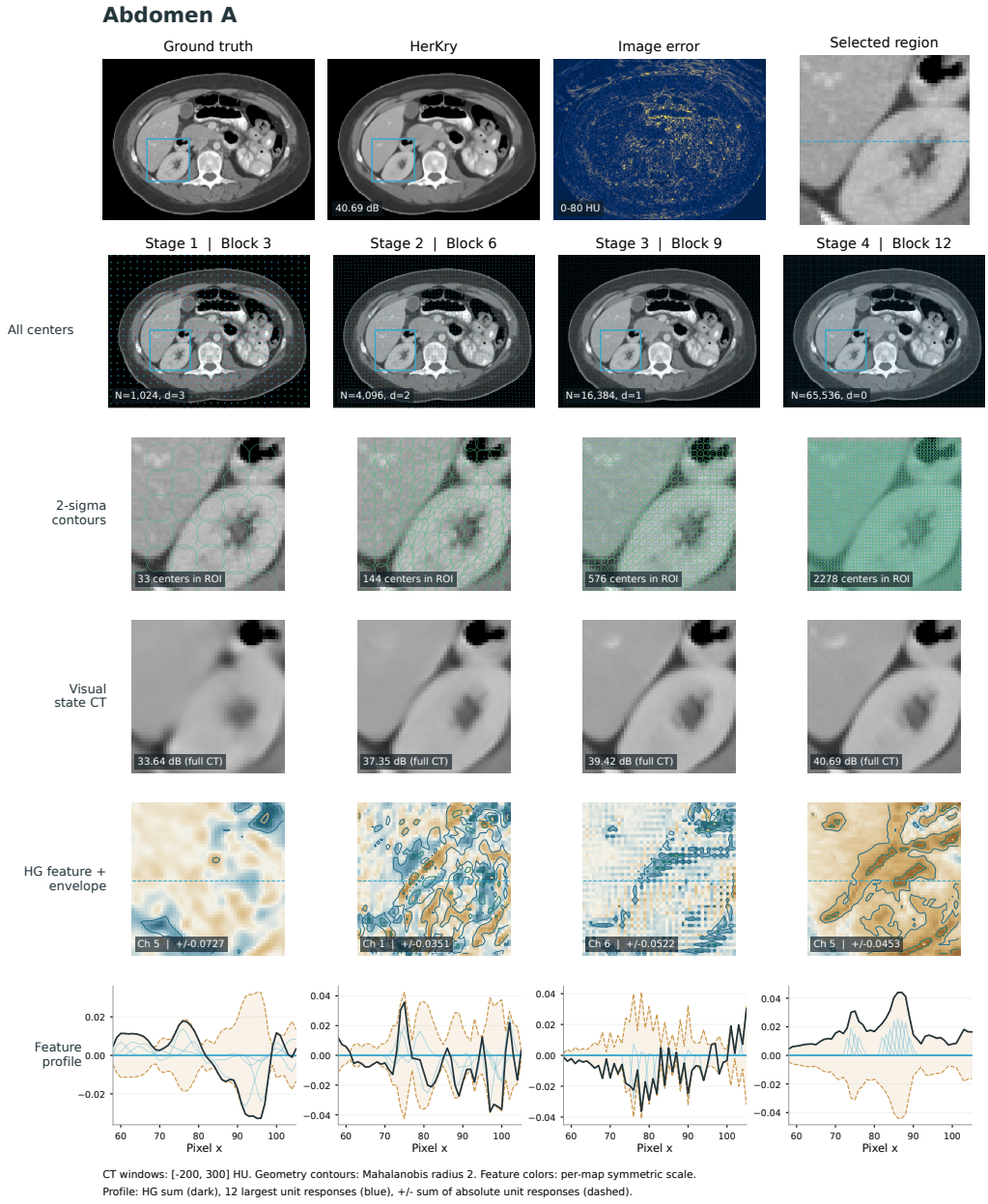

Figure 8: **Stage-wise spatial Gaussian distributions and Hermite–Gaussian response atlas. Top row:** Ground-truth reference, HerKry final visual reconstruction with corresponding PSNR and the selected ROI. **Rows 2-5:** Evolution across four stages ($l = 1, \ldots, 4$). (1) Center locations and 2σ Mahalanobis ellipses illustrating the progression from sparse, high-order primitives ($N = 1024, d = 3$) to dense, zeroth-order primitives ($N = 65536, d = 0$); (2) Visual state CT $\boldsymbol{x}_l^{\text{vis}}$ exhibiting steady PSNR gains; (3) Splatted spatial feature maps $\boldsymbol{Z}_l^{(c)}$ of the dominant channel correlated with reference image gradients; (4) 1D cross-sectional response profiles along the dashed transect, depicting the overall Hermite-Gaussian sum with dark line, the top 12 individual basis responses with blue line, and the upper/lower absolute envelope bounds with dashed lines.

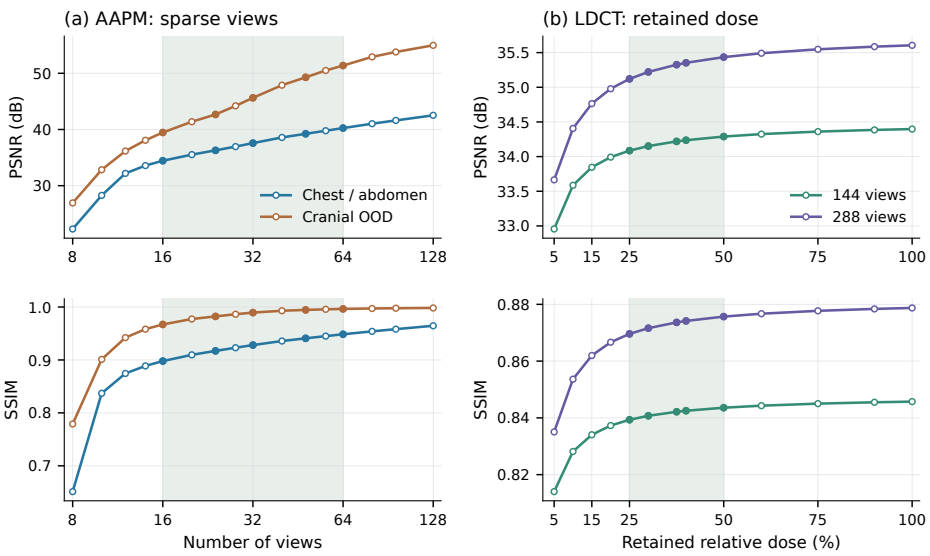

Figure 9: HerKry robustness to view sparsity and retained-dose variations. (a) AAPM: 16 view-count settings evaluated on chest / abdominal test slices and cranial OOD slices. (b) LDCT: 13 retained-dose levels at 144/288 views, evaluated on test slices. Shading indicates the training view range (16-64) and dose interval (25-50%). Filled markers denote actual training view counts in (a) and doses within the training interval in (b). Lower retained dose corresponds to stronger measurement noise. Each point reports slice-mean PSNR or SSIM.

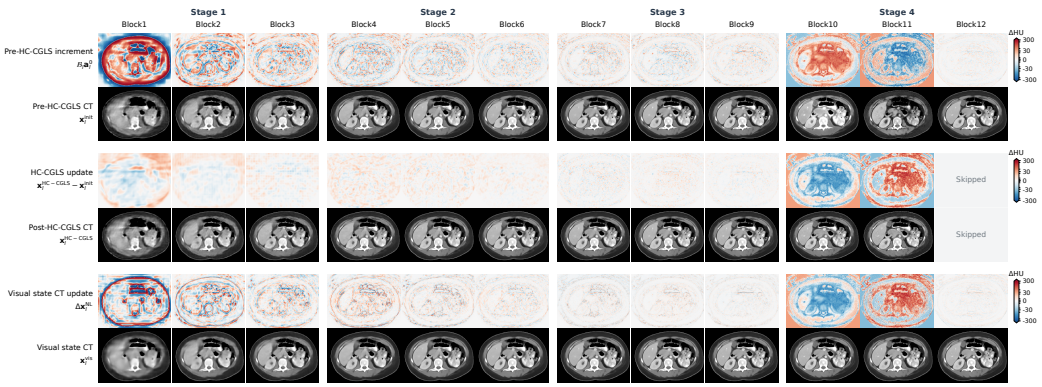

Figure 10: **Block-wise unrolled progression on an in-distribution abdominal slice with 32-view sparsity setting.** The reconstruction unfolds across 12 blocks. Rows from top to bottom: (1) Pre-HC-CGLS linear coefficient increment $\mathcal{B}_l(\boldsymbol{\Theta}_l)\boldsymbol{a}_l^0$ in ΔHU (range $[-300, 300]$); (2) Pre-HC-CGLS initial state $\boldsymbol{x}_l^{\text{init}}$; (3) Physical HC-CGLS correction update $\boldsymbol{x}_l^{\text{HC-CGLS}} - \boldsymbol{x}_l^{\text{init}}$ (skipped at Block 12); (4) Post-HC-CGLS CT slice $\boldsymbol{x}_l^{\text{HC-CGLS}}$; (5) Nonlinear visual update $\Delta\boldsymbol{x}_l^{\text{NL}}$; (6) Cumulative visual state $\boldsymbol{x}_l^{\text{vis}}$ cascading to the next block.

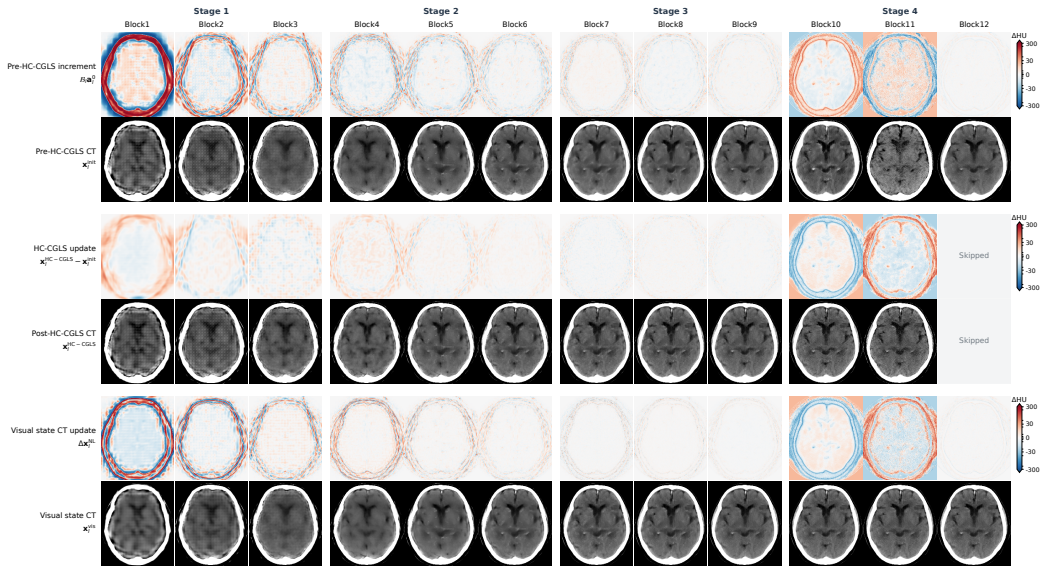

Figure 11: **Block-wise unrolled progression on a zero-shot cranial OOD slice (32-view).** Rows display the paired linear initializations, HC-CGLS physical projection rectifications, and nonlinear neural updates under a dedicated $[0, 80]$ HU clinical brain window. Block 12 bypasses HC-CGLS to directly produce the final visual CT output $\boldsymbol{x}_L^{\mathrm{vis}}$.