SPECTRAL STEERING: LOCAL INVERSE GEOMETRY FOR FROZEN DIFFUSION GENERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

A frozen diffusion generator may need to change several output properties while preserving others. We study when guidance faithfully realizes such a structured request and whether local inverse correction improves its direction. Our account links this benefit to the interaction between the request and the objective-specific remaining-sampler Jacobian: conditioning alone is insufficient. Spectral steering combines a request-specific cosine-gain predictor with damped inverse correction and fixed-time relinearization for finite edits. Scalar nulls and matched-movement controls test the predicted mechanism. Across 8,923 instances in a retrospectively identified local regime, predicted and realized gains correlate at Spearman 0.79, versus -0.11 for condition number. Experiments span images, audio, robotics, and control and planning, including complex character motion generation. On 50 paired inpainting cases, four corrections improve unobserved-pixel PSNR over Adam by 3.25 dB at equal total derivative traversals. Together, the analysis and controls connect request-dependent geometry to directional gains and establish where these gains translate into downstream performance.

1 INTRODUCTION

Frozen diffusion generators must often satisfy new tasks without retraining (Fig. 1). Differentiable guidance propagates output objectives to sampling states (Bansal et al., 2023; Chung et al., 2023; Yu et al., 2023), while differentiation through generation enables direct latent optimization (Wallace et al., 2023; Ben-Hamu et al., 2024; Karunratanakul et al., 2024). We ask when gradients faithfully realize a *specified pattern of downstream changes*, and whether directional correction gains can be predicted.

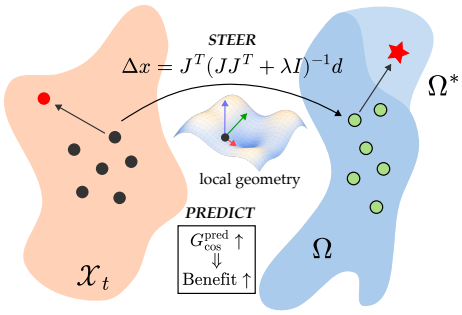


Figure 1: Spectral steering of a frozen generator. Spectral steering uses a damped local inverse of the requested downstream properties.

Requests combine change and preservation: brighten one region while holding another fixed, alter frequency bands while preserving context, or move a character's wrist while keeping its feet planted. Loss decrease alone does not describe which components move. We evaluate *request fidelity* alongside task success.

The relevant local input–output map is the Jacobian J_h of the selected properties through the entire remaining sampler. Pulling a request d back as $\eta J_h^\top d$ predicts $\eta J_h J_h^\top d$, so unequal singular gains can alter its relative components. Spectral steering uses classical damped least squares (Levenberg, 1944; Marquardt, 1963) to approximately compensate these gains. For finite edits, it refreshes the Jacobian and residual toward the original target at fixed diffusion time (Fig. 2).

Correction gains depend on the request. Even an ill-conditioned Jacobian preserves a supported single-mode request's direction. Distortion depends jointly on the spectrum and the request's modal composition. We express this dependence as a prediction of cosine improvement. An exact scalar null and matched-movement controls make the account experimentally testable beyond aggregate method rankings.

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.

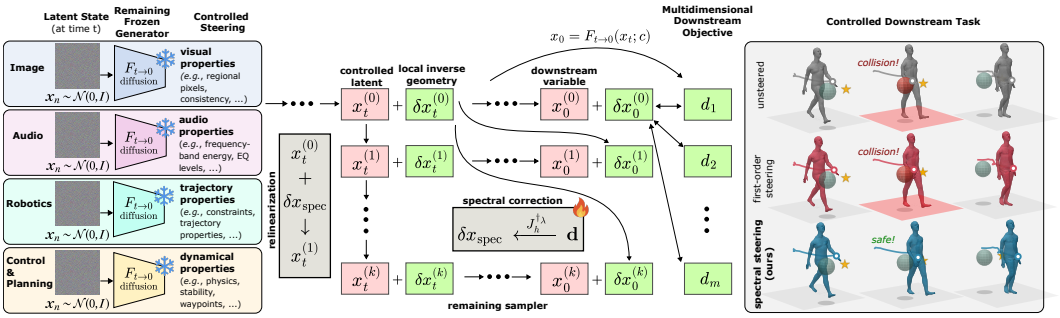


Figure 2: From gradient pullback to local inversion. The frozen remaining sampler maps an intervention to downstream properties. FO reweights a request by $J_h J_h^\top$; the damped inverse compensates unequal gains. K2/K4 refresh the Jacobian and residual at the same diffusion time. Right: unsteered, matched FO, and K4 character motions at two common instants; red spheres mark skeleton–obstacle intersection.

Downstream systems require selected quantities to change by specified amounts while others remain fixed. Character motion requires coordinated joint changes and preservation; robotics adds execution dynamics, and content editing requires preserved context. Our experiments connect predicted directional gains to realized edits and task outcomes, including failures of local approximation or downstream transfer.

Our contributions are:

(1) Request-specific geometry of faithful steering. We characterize how the objective-specific remaining-sampler geometry distorts structured requests, identifying the role of modal composition and an exact scalar null after magnitude matching (Sections 3 and 5.1).

(2) Predicting directional gains and refining finite edits. Our score $G_{\cos}^{\text{pred}}$ predicts local cosine improvement over first-order guidance. Spectral correction implements the damped inverse, and fixed-time relinearization refreshes the geometry for finite requests (Sections 3 and 5.2).

(3) Controlled evidence across generative domains. Movement controls and optimizer comparisons test directional gains, task success, and executed outcomes across images, audio, robotics, and control and planning—beginning with complex character motion generation (Section 5).

2 Related Work

Gradient guidance. Classifier guidance, Universal Guidance, Loss-Guided Diffusion and DPS condition sampling through gradients (Dhariwal & Nichol, 2021; Bansal et al., 2023; Song et al., 2023b; Chung et al., 2023); FreeDoM uses energy-based guidance (Yu et al., 2023). These methods establish general differentiable conditioning; our question is how faithfully a gradient transmits a specified vector displacement.

End-to-end optimization. Earlier text-guided face editing uses GAN inversion (Xia et al., 2021). Full-process differentiation is established: FlowGrad, DOODL and D-Flow optimize generative states through the resulting output (Liu et al., 2023; Wallace et al., 2023; Ben-Hamu et al., 2024); DNO applies this principle to motion and DITTO to music (Karunratanakul et al., 2024; Novack et al., 2024). D-Flow also analyzes the generator Jacobian's action on output gradients. OC-Flow formulates optimal control (Wang et al., 2025), and posterior and end-to-end guidance have been connected theoretically (Blasingame & Liu, 2025). We examine how a prescribed vector request interacts with this geometry, predicting directional correction gains and testing them under movement controls.

Geometry and inverse guidance. DDRM and DDNM exploit measurement structure (Kawar et al., 2022; Wang et al., 2023); ΠGDM uses measurement-model pseudoinverses for conditional-score estimation (Song et al., 2023a). Spherical Gaussian and manifold-preserving guidance constrain intermediate or data-manifold geometry (Yang et al., 2024; He et al., 2024); Park et al. analyze latent singular directions (Park et al., 2023). JAPS studies denoiser Jacobians, while CLAMP applies

denoiser-pulled curvature and manifold-aligned damping to posterior correction (Hen et al., 2026; Shin et al., 2026). We study the selected downstream representation composed with the complete remaining sampler, asking which requests benefit from inverse correction. The contribution combines a request-dependent prediction with scalar nulls, movement controls, and finite-edit evaluation. The damped least-squares update supplies the correction; it does not define a posterior sampler.

Planning and policy steering. Offline demonstration learning highlights the role of data and evaluation in manipulation (Mandlekar et al., 2021). Diffuser and Diffusion Policy generate trajectories and actions (Janner et al., 2022; Chi et al., 2023); SafeDiffuser and JM2D address constrained planning (Xiao et al., 2025; Jung et al., 2025). DynaGuide uses a dynamics model to steer policies, whereas DSRL learns a latent-noise controller from environment interaction (Du & Song, 2025; Wagenmaker et al., 2025). Spectral steering requires a differentiable specification without training another controller. Execution and safety remain separate evaluation requirements. Appendix A summarizes the control interpretation.

3 Objective-Specific Local Inverse Geometry

3.1 Structured requests and local geometry

At fixed diffusion time t, let $x_t \in \mathbb{R}^D$ be the intervention state and $z(x_t) = h(F_{t\to 0}(x_t; c)) \in \mathbb{R}^Z$ the downstream properties. The map includes the complete remaining sampler, decoding, and task-specific coordinate scaling; conditioning and remaining randomness are fixed. Its Jacobian is

$$J_{h,t} = \frac{\partial (h \circ F_{t\to 0})(x_t; c)}{\partial x_t}. \tag{1}$$

For target $z_\star = z(x_t) + d$, the request $d = [d_{\text{change}}^\top, 0_{\text{preserve}}^\top]^\top$ specifies changes and held properties. We seek a finite displacement $z(x_t + \Delta x) - z(x_t) \approx d$, using the local model

$$z(x_t + \Delta x) - z(x_t) \approx J_{h,t}\Delta x. \tag{2}$$

Preservation is a soft least-squares target; omitted properties are uncontrolled. Gradient descent on $\frac{1}{2}\|z(x) - z_\star\|_2^2$ gives

$$\Delta x_{\text{FO}} = \eta J_{h,t}^\top d, \qquad \Delta z_{\text{FO}} \approx \eta J_{h,t} J_{h,t}^\top d, \quad \eta > 0. \tag{3}$$

This is a descent direction when nonzero, but its predicted response need not align with d. A scalar step size changes the response magnitude without correcting the relative components. The relevant geometry depends on both the generator and the downstream representation: changing the measurements or their scaling changes the operator being inverted.

3.2 Damped inversion and request-specific prediction

For $\lambda > 0$, spectral steering solves the regularized local problem

$$\Delta x_{\text{spec}} = \arg\min_\delta \ \tfrac{1}{2}\|J_{h,t}\delta - d\|_2^2 + \tfrac{\lambda}{2}\|\delta\|_2^2, \tag{4}$$

$$= (J_{h,t}^\top J_{h,t} + \lambda I_D)^{-1} J_{h,t}^\top d = J_{h,t}^\top (J_{h,t} J_{h,t}^\top + \lambda I_Z)^{-1} d. \tag{5}$$

One correction is **K1**. The equivalent primal and dual forms solve in state and property space, respectively. For compact SVD $J_{h,t} = U\Sigma V^\top$ and request $d = \sum_i a_i u_i + d_\perp$, with $U^\top d_\perp = 0$, the predicted responses are

$$J_{h,t}\Delta x_{\text{FO}} = \eta \sum_i \sigma_i^2 a_i u_i, \qquad J_{h,t}\Delta x_{\text{spec}} = \sum_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda} a_i u_i. \tag{6}$$

For $\sigma_i^2 \gg \lambda$, inverse gains approach one; weak modes remain attenuated and $d_\perp$ is locally unreachable. A supported single-mode request has collinear responses regardless of the condition number. Directional benefit therefore depends on which modes the request occupies, not just the spectrum. For $Z = 1$, magnitude matching makes the latent updates identical (Appendix A).

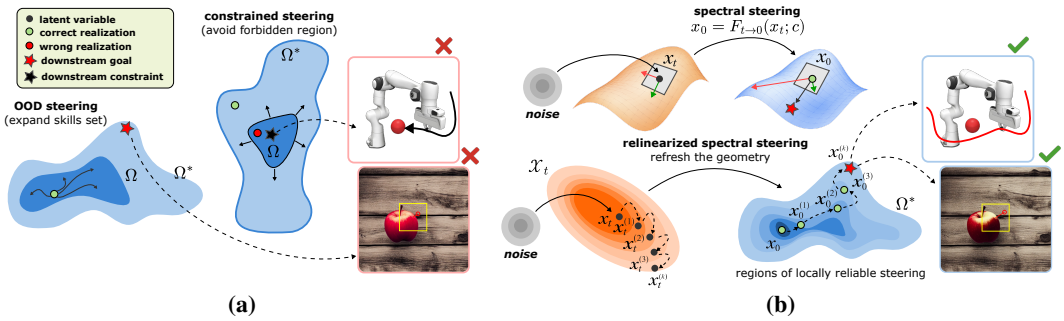


Figure 3: Structured steering and fixed-time relinearization. (a) Changes and preservation must be realized jointly through the frozen generator. (b) Successive corrections refresh the Jacobian and residual toward the original target at the same diffusion time. $\Omega^\star$ denotes the target action set that becomes reachable after steering.

Writing $G_h = J_{h,t}J_{h,t}^\top$, we predict the directional gain by

$$G_{\cos}^{\text{pred}} = \cos\big(G_h(G_h + \lambda I)^{-1}d, d\big) - \cos(G_h d, d). \tag{7}$$

For nonzero requests and responses, the score predicts local cosine gain before editing. Different requests can have different predicted gains at the same Jacobian. Agreement with nonlinear edits is tested empirically; magnitude and task utility are evaluated separately.

3.3 Fixed-time relinearization and computation

K2/K4 repeat the correction toward the original target. At $x_t^{(k)}$, recompute $J_k = \partial z/\partial x_t^{(k)}$, set $r_k = z_\star - z(x_t^{(k)})$, and update

$$x_t^{(k+1)} = x_t^{(k)} + J_k^\top(J_k J_k^\top + \lambda_k I)^{-1} r_k, \qquad \lambda_k = 0.01\sigma_1(J_k)^2. \tag{8}$$

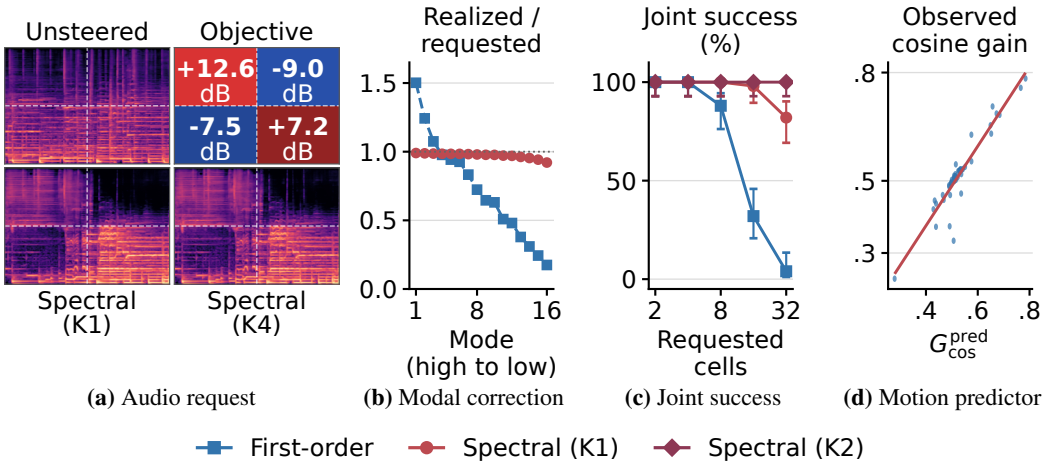
The index k counts corrections, not denoising steps; the final output is $F_{t\to 0}(x_t^{(K)}; c)$. Each candidate completes the same remaining sampler, measures its residual relative to the original target, and updates the intervention state. Keeping that target fixed avoids requesting the full edit again at every correction. Refreshing the local model addresses nonlinear error without guaranteeing convergence (Fig. 3). The coefficient 0.01 is fixed; "adaptive" damping means refreshing σ_1^2 at each correction, rather than a line search or a latent-norm bound.

Explicit solvers use reduced-SVD filtering without hard rank truncation and obtain σ_1 from the Jacobian. Matrix-free solvers use up to 16 request-started Golub–Kahan iterations, estimate σ_1 from the projected matrix, and solve the damped problem there. Each iteration uses one VJP and one JVP through the remaining sampler: one correction costs up to 16 of each, and K corrections up to $16K$ of each. "Direct" inversion refers to the local inverse relation, not iteration-free computation. Numerical-support thresholds are diagnostic, separate from damping; setting-specific thresholds, stopping rules, and solver variants are in Appendix I.1.

4 Experimental Protocol

Vision uses a CelebA-HQ DDPM (Ho et al., 2020) for restoration and Stable Diffusion 1.5 (Rombach et al., 2022) for structured edits. Audio covers MAESTRO waveform excerpts (Hawthorne et al., 2018) and generated mel spectrograms; robotics measures MuJoCo execution (Todorov et al., 2012). Control and planning begins with 60-frame MDM character motions (Tevet et al., 2023), followed by trajectory and dynamical-system tasks. Table 1 identifies models; Appendix L defines objectives and endpoints.

Matching and baselines. FO and K1 share the initial state, time, sampler, objective, and request. FO matches K1's *predicted objective-space* movement with $\eta = \|J_{h,t}\Delta x_{\text{K1}}\|_2/\|J_{h,t}J_{h,t}^\top d\|_2$ for nonzero denominator, using the task's scaled coordinates. It reuses K1's Jacobian or corresponding Krylov quantities; this isolates direction, not standalone cost. A separate control matches latent



(a) Audio request **(b)** Modal correction **(c)** Joint success **(d)** Motion predictor

Figure 4: Testing the directional mechanism. (a) Four-region audio request; spectrograms share a color scale (time right, frequency up). (b) FO overweights sensitive modes; inverse responses lie within 5% of each displayed request. (c) At 32 audio cells, FO/K1/K2 attain 4/82/100% success ($n = 50$; 95% intervals). (d) Predicted versus realized cosine gain on 50 character-avoidance seeds ($r = 0.94$, $\rho = 0.92$); red line: linear fit. Appendix D.2 reports the cross-domain analysis.

displacement. SGD, Adam, L-BFGS and domain-specific baselines use fixed, pilot-selected, or evaluation-selected settings as documented in Appendix G; the main table is not a common-budget ranking.

Evaluation. We distinguish request fidelity, application outcomes, and executed behavior. Paired comparisons are used where available. Compute comparisons separately count vector–Jacobian products (VJPs), total VJP/Jacobian–vector-product (JVP) traversals, or network evaluations; these do not establish equal wall-clock cost. Shared-task adaptations are distinguished from native baseline protocols. A matrix-free spectral solve uses up to 16 VJPs and 16 JVPs, whereas each SGD/Adam step uses one VJP; explicit implementations have different derivative representations. Sample sets, uncertainty, solver mappings, and qualitative selection rules are specified with the results and in the appendix.

5 Results and Evaluation

5.1 Correcting modal gains improves structured responses

Figure 4(a,b) connects a four-region audio request to the modal mechanism: the pullback overweights sensitive modes, while inverse responses lie within 5% of the request in the displayed example. The scalar null supplies a falsifiable control: all 935 tested pairs tie to machine precision at $Z = 1$. Benefits emerge with multiple coordinates. At 32 audio cells, joint success is 4% for FO, 82% for K1, and 100% for K2 ($n = 50$), showing improved simultaneous realization and the value of relinearization.

5.2 Local geometry predicts directional gains

On 50 character-motion avoidance seeds, predicted and observed cosine gains correlate at $r = 0.94$, $\rho = 0.92$ (Fig. 4(d)). Across 8,923 instances in 27 vision, audio, and robotics populations within the measured local regime, $r = 0.75$ and $\rho = 0.79$, versus -0.04 and -0.11 for log condition number. Centering within populations retains $\rho = 0.77$, and Pearson association is positive in 26 of 27 populations, supporting an association beyond pooled between-task differences (Appendix D.2). The score is computed before editing, but regime selection is retrospective; this supports a local diagnostic, not a validated acceptance rule.

Table 1: General optimizers across domains. Bold: best displayed value, including rounded ties. Settings and selected budgets vary by task. ns: no statistically significant difference between the two marked methods in the same row. GD/Adam learning rates selected by a hyperparameter search, L-BFGS learning rate selected on the evaluation set. Character optimizers match K4's nominal backward-pass count. Protocols and checkpoints: Appendices G, E, and L.

Data / simulator Diffusion model	Task / endpoint	Unst.	FO	SGD	Adam	L-BFGS	Spectral (K1)	Spectral (best K)
Vision								
CelebA-HQ DDPM EMA-256	Inpainting PSNR ↑	16.55	19.61	25.57	25.98	25.76	23.69	**27.75 (K4)**
	4× SR PSNR ↑	25.05	26.32	27.69	27.79	27.58	27.39	**27.90 (K2)**
	Deblur PSNR ↑	24.03	25.26	27.26	27.92^{ns}	27.90	26.61	**27.94 (K16)**^{ns}
	Multi-constraint PSNR ↑	25.81	26.12	27.69	**28.61**	28.54	26.89	27.75 (K4)
Generated images Stable Diff. 1.5	Structured edit [0-1] ↑	0.000	0.574	0.734	0.774	0.765	0.954	**1.000 (K8)**
Audio								
MAESTRO 150k Generated mel Audio Diffusion	Localized EQ err. ↓	7.000	5.878	6.623	4.692	3.895	5.538	**2.691 (K16)**
	Bandwidth LSD ↓	12.757	13.408	13.151	26.604	13.487	12.733	**12.730 (K4)**
	EQ-16 success ↑	0.000	0.32	1.00	1.00	1.00	0.98	**1.00 (K2)**
	Selective EQ dB ↓	2.5981	1.0883	0.0005	0.0300	0.0005	0.3935	**0.0004 (K4)**
Robotics								
MuJoCo Temporal DDPM	Panda reach err. cm ↓	3.89	2.39	6.54	2.03	**0.63**	1.06	0.74 (K16)
	Panda pose err. cm ↓	5.10	2.28	0.82	0.97	0.79	1.09	**0.70 (K4)**
	Cheetah target cm ↓	16.16	11.57	6.79	**4.18**	5.74	6.67	4.31 (K16)
	Cheetah traj. cm ↓	13.29	9.00	5.57	6.19	5.19	6.44	**4.72 (K16)**
	Panda contact [0-1] ↑	0.000	0.769	0.766	0.851	**0.926**	0.718	0.894 (K16)
Control & planning								
Generated motion MDM	Char. obstacle [0-1] ↑	0.000	0.000	0.960^{ns}	0.880	0.980	0.180	**0.980 (K16)**^{ns}
	Char. reaching [0-1] ↑	0.000	0.000	0.480	0.000	0.960^{ns}	0.140	**0.960 (K4)**^{ns}
2D integrator Temporal DDPM	1-waypoint err. cm ↓	110.86	30.35	37.69	7.48	26.34	7.41	**3.05 (K4)**
	Multi-waypoint [0-1] ↑	0.000	0.223	0.873	0.534	0.945	0.854	**1.000 (K4)**
	Point-mass [0-1] ↑	0.000	0.983	0.999	0.997	1.000^{ns}	0.997	**1.000 (K4)**^{ns}

Matched movement. Across 576 instances with latent norms agreeing within 4×10^{-14}, K1 improves request realization on 100% of inpainting images, 85% of audio excerpts, and 93% of Panda instances. Interpolating toward the inverse at fixed predicted response norm improves reconstruction by 5.34 dB, with improvements on all 50 images at every tested step (Appendix B). These controls isolate a directional benefit. They also separate the proposed mechanism from simply taking a larger step: the latent-norm control fixes intervention size, while the interpolation fixes predicted downstream movement (Appendix B).

5.3 Vision: reconstruction and structured edits

Generic optimizers. In Table 1, K4 reaches 27.75 dB paired inpainting PSNR versus Adam's 25.98. Structured editing gives cosine 0.954 for K1 and 1.000 for K8, versus 0.774 for Adam and 0.765 for L-BFGS. Spectral also leads reported super-resolution PSNR. Adam leads composite restoration; the K16 deblurring comparison is statistically unresolved and costs eight times Adam's network evaluations in that audit.

Residual versus reconstruction. On the separate paired trajectory in Fig. 6, Adam peaks at 25.60 dB near step 54, then loses 1.59 dB by step 256 despite decreasing measurement loss. K4 reaches 28.32 dB: it exceeds Adam@64 by 2.74 dB at equal VJP count and Adam@128 by 3.25 dB at equal total derivative traversals (95% interval $[2.32, 4.35]$; 86% paired wins). K1 trails Adam under both conventions (Appendix F); the benefit here comes from repeated geometry correction.

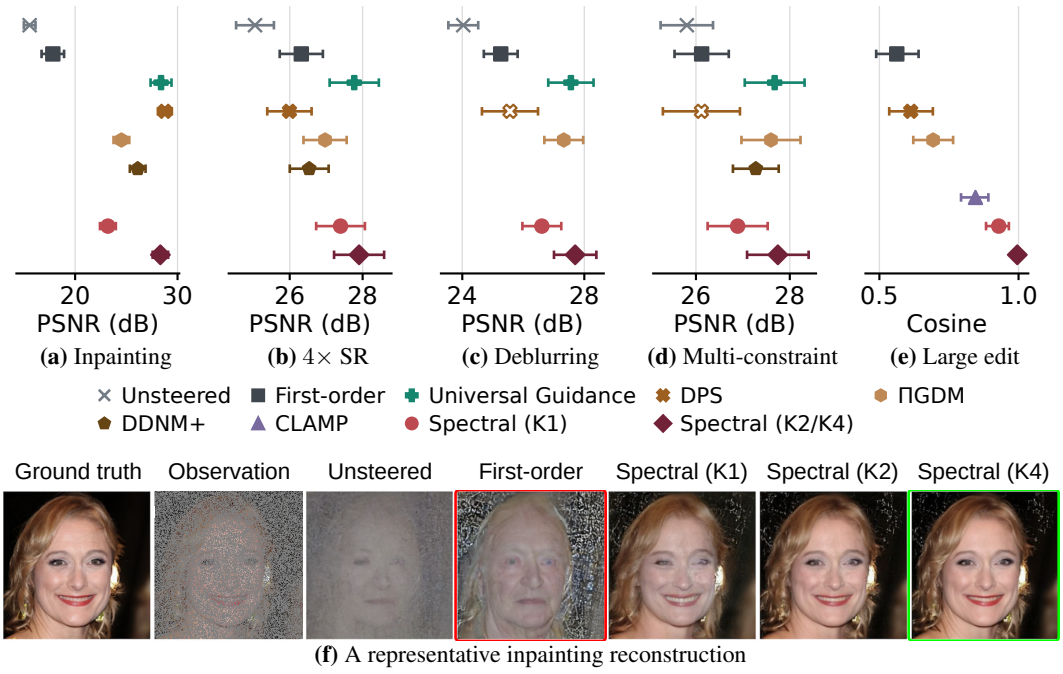


Figure 5: Vision accuracy and reconstruction. (a–d) CelebA-HQ/DDPM PSNR, 50 reference images per task, 95% intervals; hollow DPS uses deblur $n = 25$ and composite $n = 20$. UG uses selected settings; native DPS uses 1000 steps. (e) SD1.5 request cosine at $4\times$ scale (window 10, $n = 20$). Deeper spectral rungs are K2 in (b,e), K4 elsewhere. DPS, ΠGDM, and CLAMP in (e) are objective adaptations. (f) One paired reconstruction. Protocols and differing image sets: Appendices G and J.

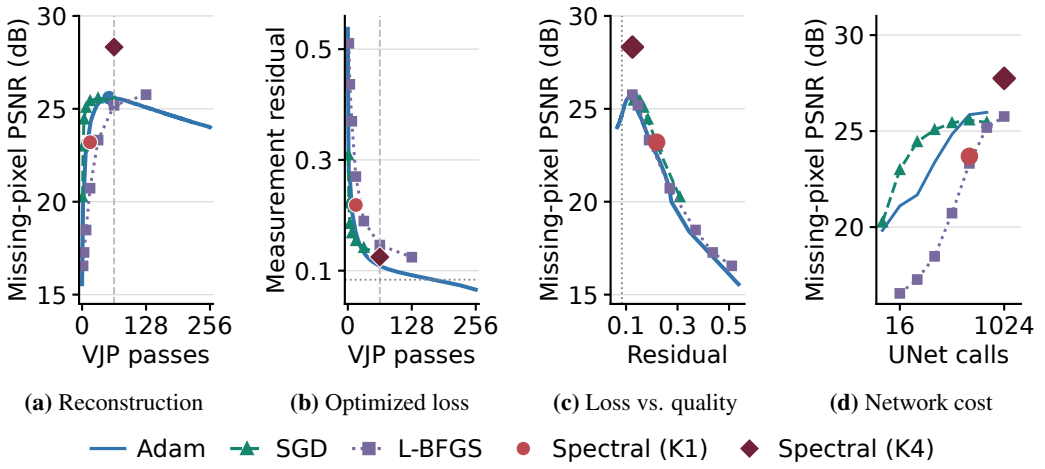


Figure 6: Residual minimization can degrade reconstruction. (a–c) Adam and spectral use 50 paired reference images; SGD/L-BFGS use the separate 50-image optimizer set also in (d). Adam loses 1.59 dB after its peak while residual falls. Dashed guides mark 64 VJPs; dotted guides mark clean-image residual. L-BFGS uses recorded backward counts. (d) Log-scale UNet evaluations charge VJP and JVP traversals. Paired budget comparisons: Appendix F.

Guidance and structured edits. In Fig. 5(a–d), spectral exceeds selected Universal Guidance by 0.133 dB on super-resolution and 0.140 dB on deblurring, with no significant paired difference on inpainting or composite restoration; guidance obtains early gains at lower cost on three operators. At a $4\times$ structured-edit request, K1/K2 reach cosine 0.929/0.996 versus 0.845 for CLAMP. The stress test probes realization of an analytic property beyond reconstruction. These are descriptive objective adaptations; marginal intervals alone do not establish paired significance. Across the

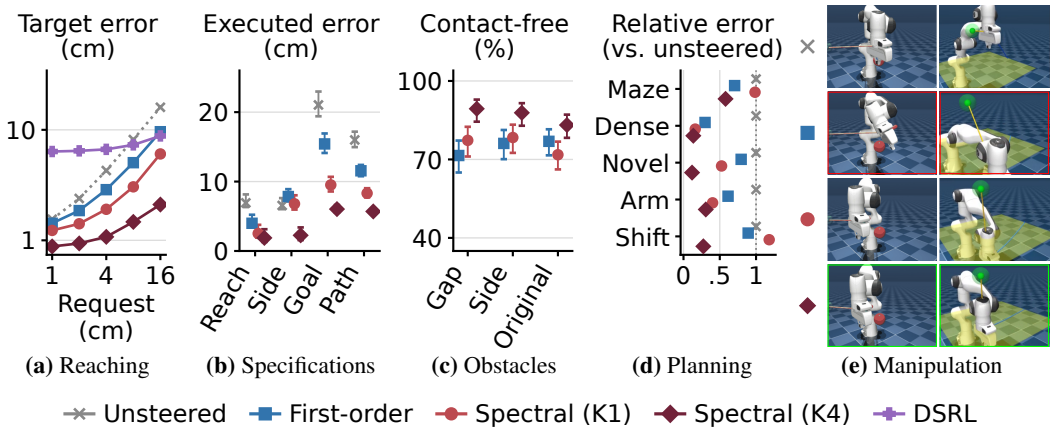


Figure 7: Robotics and planning outcomes. (a,b) Executed error medians and 95% intervals for Panda reaching, orientation/height preservation, and Cheetah goal/path tracking. (c) Contact-free execution rates. Unsteered fails the reach, so its contact-free score is omitted. (d) Endpoints normalized by unsteered (lower is better): Maze2D distance, dense/novel-obstacle collision, arm error, shifted-goal failure. Counts: Appendix G. (e) Paired obstacle-clearance (left) and constant-height (right) examples; symbols identify methods via the shared legend.

evaluated settings, spectral has no below-prior outcomes, while selected unstable guidance settings do. Implementation and sensitivity details are in Appendices G and J.

5.4 Audio: Precise Change and Preservation

Across four practical mel-domain tasks, K2 attains 100% joint success versus 8–82% for FO. At 64 backward passes on 30 selective-EQ clips, K4 achieves 0.00038 dB all-coordinate RMSE versus Adam's 0.02997, winning all pairs. SGD reaches 0.00047 dB with no significant difference from K4 ($p = .63$); all three satisfy the success threshold. The precision comparison remains informative after the binary success threshold saturates: K4 and Adam both pass, but differ in realized coordinate errors. Transition edits further show that meeting the measured request can increase an unoptimized abruptness diagnostic. These are acoustic-control results, without a listening-study claim (Appendix D.1).

Waveform localized EQ reaches 2.691 dB error at K16 versus 3.895 for L-BFGS at the reported settings. Bandwidth extension is budget sensitive: Adam's error is 26.604 dB at 16 steps versus K4's 12.730, but early-stopped Adam reaches 12.525 with no significant paired difference. Table 1 therefore establishes neither uniform superiority nor equal-cost advantages across audio tasks.

5.5 Robotics: Transferring Corrections to Execution

On 276 Panda instances, K1 reduces median executed reaching error from FO's 2.39 cm to 1.06 cm; learned DSRL reaches 6.01 cm after 8,000 environment episodes per seed. DSRL learns from black-box rewards; spectral uses differentiability at inference time, so their access and deployment conditions differ. Figure 7(a–c,e) covers reaching, preservation, clearance, and paired manipulation examples.

Table 1 shows benefits on Panda pose and Gymnasium HalfCheetah trajectory tracking (Mandlekar et al., 2021). On shared reaching instances, K8 ties L-BFGS at 0.63 cm; K16's 0.59 cm is not a significant improvement. Greater clearance need not increase joint reach-and-clear success, and depth can be non-monotone. The unsteered contact-free score is omitted from the plot because it fails the reach. The qualitative rows pair obstacle clearance with a constant-height constraint. Generated fidelity predicts executed improvement more strongly once requests exceed servo resolution, reaching rank correlation 0.77; smaller corrections can be attenuated during tracking. Execution and refinement analyses are in Appendices C and G.

5.6 CONTROL AND PLANNING

Character motion generation. A frozen MDM produces three-second clips with 25 deterministic DDIM transitions remaining at intervention. On 50 paired seeds per task, avoidance ($Z = 13$) and multi-objective reaching ($Z = 77$) combine pelvis or wrist changes with preservation (Fig. 2; Appendix E). K1 improves cosine over movement-matched FO from 0.458 to 0.976 and from 0.701 to 0.931, winning every pair. Relinearization reduces remaining positional error. At K4-matched nominal backward-pass counts, K4 achieves 96% joint success on both tasks versus L-BFGS's 98% and 96%. On reaching, K4 limits bone drift to 1.05 cm versus 5.66 for SGD and 7.20 for Adam. Joint success checks positions and jerk; collision, velocity, and bone drift are separate diagnostics. For example, Adam satisfies the avoidance success gate in 88% of cases but is collision-free in 78%. This distinguishes coordination of measured motion properties from additional physical constraints. The comparison matches nominal backward passes, not standalone runtime. These results support structured motion control while leaving generic optimizers competitive.

Trajectory control and planning. Across 45 executed waypoint pairs, K4 reduces Adam's mean error from 7.48 to 3.05 cm (59%; 41/45 wins; paired difference interval $[2.96, 5.86]$ cm), also improving on fixed GD and L-BFGS-B settings (Appendix G). Figure 7(d) shows shifted-goal success rising from FO's 26.5% to K4's 77.4%, although K1 reaches only 2.6%; unseen-obstacle collision falls from 79.5% to 11.8%. Specialized guidance can achieve lower collision rates. The relative ranking depends on goal accuracy, clearance, and budget (Appendix H).

6 DISCUSSION AND LIMITATIONS

The evidence supports correction of structured requests whose local model remains informative. Inverse corrections and nonlinear optimizers both refine residuals; their relative outcomes depend on task and budget, and relinearization need not improve monotonically. Coordinate scaling determines the geometry; omitted properties receive no protection. Generated fidelity can be attenuated or overwritten during execution and replanning. The predictor's pooled validation uses retrospective regime selection and concerns cosine alignment rather than task success. Derivative counts do not imply runtime gains. Robotics is simulated and audio has no listening study; the appendix analyzes these boundaries alongside optimizer counterexamples.

Toward precise downstream control. A natural direction is to use the local inverse as a precision layer within closed-loop diffusion policies, which already support receding-horizon execution (Chi et al., 2023). New observations would trigger a new geometry estimate and requests for tracking, clearance, or terminal-state changes. Differentiable dynamics and conservation residuals could encode physical-law targets. Such a layer would require execution-aware validation and independent safety enforcement: local request fidelity alone does not guarantee constraint satisfaction. A complementary direction replaces hand-designed properties with learned semantic features, requesting targeted edits while preserving selected attributes. Its value depends on feature calibration, local controllability, and whether the representation measures the intended semantics. Appendix A formulates these directions and their evaluation requirements.

7 CONCLUSION

Spectral steering connects a structured request's modal composition to its distortion through a frozen generator, the predicted benefit of local inverse correction, and the realized outcome. Scalar nulls and matched-movement controls test this connection; cross-domain experiments establish its practical relevance, including complex character motion generation. Classical damped least squares supplies the correction, while fixed-time relinearization extends its use to finite edits. The evidence supports request-dependent geometry as a useful control interface, with benefits bounded by local-model accuracy, objective coverage, and downstream execution. Future work should further characterize how precise generator-space steering transfers to downstream tasks, including closed-loop control under physical and safety constraints and targeted semantic editing with learned representations, while deriving theoretical locality bounds on the maximum reliable steering magnitude.

AI USE STATEMENT

Generative AI tools were used for code assistance, result interpretation, and manuscript drafting/editing. The authors remain responsible for citation accuracy, originality, policy compliance, and the final submission.

ETHICS STATEMENT

This work studies inference-time steering of pretrained generators and inherits their misuse and deployment risks. Robotics experiments are simulator based and do not establish safety guarantees. Improved request fidelity is not guaranteed task success.

REPRODUCIBILITY STATEMENT

The supplement records objectives, checkpoints, matching protocols, optimizer budgets, paired statistics, figure source data, and available qualitative selection rules. The accompanying additional material supports the paper's claims with code and media files. Media files include image, video, and audio samples of the steered models. The appendix specifies implementation and evaluation protocols.

REFERENCES

Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 843–852, 2023. URL https://openaccess.thecvf.com/content/CVPR2023W/GCV/html/Bansal_Universal_Guidance_for_Diffusion_Models_CVPRW_2023_paper.html.

Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-Flow: Differentiating through flows for controlled generation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 3462–3483, 2024. URL https://proceedings.mlr.press/v235/ben-hamu24a.html.

Zander Blasingame and Chen Liu. Greed is good: A unifying perspective on guided generation. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/a565741b6a2b152323452215bd1f4185-Abstract-Conference.html.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin C. M. Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Proceedings of Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.026. URL https://roboticsproceedings.org/rss19/p026.html.

Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/49ad23d1ec9fa4bd8d77d02681df5cfa-Abstract.html.

Maximilian Du and Shuran Song. DynaGuide: Steering diffusion polices with active dynamic guidance. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/3ec6c6fc9065aa57785eb05dffe7c3db-Abstract-Conference.html.

Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. Enabling factorized piano music modeling and generation with the maestro dataset. *arXiv preprint arXiv:1810.12247*, 2018.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, WeiHsiang Liao, Yuki Mitsufuji, Zico Kolter, Ruslan Salakhutdinov, et al. Manifold preserving guided diffusion. In *International Conference on Learning Representations*, volume 2024, pp. 44819–44850, 2024.

Liav Hen, Tom Tirer, Raja Giryes, and Shady Abu-Hussein. Jacobian-aware posterior sampling for inverse problems. *Transactions on Machine Learning Research*, 2026. URL https://arxiv.org/abs/2511.18471.

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

Michael Janner, Yilun Du, Joshua Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9902–9915, 2022. URL https://proceedings.mlr.press/v162/janner22a.html.

Wonsuhk Jung, Utkarsh Aashu Mishra, Nadun Ranawaka Arachchige, Yongxin Chen, Danfei Xu, and Shreyas Kousik. Joint model-based model-free diffusion for planning with constraints. In *Proceedings of The 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pp. 4328–4350, 2025. URL https://proceedings.mlr.press/v305/jung25d.html.

Korrawe Karunratanakul, Konpat Preechakul, Emre Aksan, Thabo Beeler, Supasorn Suwajanakorn, and Siyu Tang. Optimizing diffusion noise can serve as universal motion priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1334–1345, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Karunratanakul_Optimizing_Diffusion_Noise_Can_Serve_As_Universal_Motion_Priors_CVPR_2024_paper.html.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/95504595b6169131b6ed6cd72eb05616-Abstract-Conference.html.

Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944. doi: 10.1090/qam/10666.

Xingchao Liu, Lemeng Wu, Shujian Zhang, Chengyue Gong, Wei Ping, and Qiang Liu. FlowGrad: Controlling the output of generative ODEs with gradients. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24335–24344, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Liu_FlowGrad_Controlling_the_Output_of_Generative_ODEs_With_Gradients_CVPR_2023_paper.html.

Ajay Mandlekar, Danfei Xu, Josiah Wong, Soroush Nasiriany, Chen Wang, Rohun Kulkarni, Li Fei-Fei, Silvio Savarese, Yuke Zhu, and Roberto Martín-Martín. What matters in learning from offline human demonstrations for robot manipulation. In *arXiv preprint arXiv:2108.03298*, 2021.

Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. doi: 10.1137/0111030.

Zachary Novack, Julian Mcauley, Taylor Berg-Kirkpatrick, and Nicholas J. Bryan. DITTO: Diffusion inference-time t-optimization for music generation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 38426–38447, 2024. URL https://proceedings.mlr.press/v235/novack24a.html.

Yong-Hyun Park, Mingi Kwon, Jaewoong Choi, Junghyo Jo, and Youngjung Uh. Understanding the latent space of diffusion models through the lens of riemannian geometry. In *Advances in Neural Information Processing Systems*, volume 36, pp. 24129–24142, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/4bfcebedf7a2967c410b64670f27f904-Abstract.html.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 10674–10685. ieee, 2022.

Seunghyeok Shin, Minwoo Kim, Dabin Kim, and Hongki Lim. Geometry-correct diffusion posterior sampling with denoiser-pullback curvature guidance and manifold-aligned damping. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2605.27990.

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

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023a. URL https://research.nvidia.com/labs/genair/publication/song2023pseudoinverse/.

Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32483–32498, 2023b. URL https://proceedings.mlr.press/v202/song23k.html.

Guy Tevet, Sigal Raab, Brian Gordon, Yoni Shafir, Daniel Cohen-or, and Amit Haim Bermano. Human motion diffusion model. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=SJ1kSyO2jwu.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033, 2012. doi: 10.1109/IROS.2012.6386109.

Andrew Wagenmaker, Yunchu Zhang, Mitsuhiko Nakamoto, Seohong Park, Waleed Yagoub, Anusha Nagabandi, Abhishek Gupta, and Sergey Levine. Steering your diffusion policy with latent space reinforcement learning. In *Proceedings of The 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pp. 258–282, 2025. URL https://proceedings.mlr.press/v305/wagenmaker25a.html.

Bram Wallace, Akash Gokul, Stefano Ermon, and Nikhil Naik. End-to-end diffusion latent optimization improves classifier guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7280–7290, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Wallace_End-to-End_Diffusion_Latent_Optimization_Improves_Classifier_Guidance_ICCV_2023_paper.html.

Luran Wang, Chaoran Cheng, Yizhen Liao, Yanru Qu, and Ge Liu. Training free guided flow-matching with optimal control. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/19a94fdf9e1c5b387830e4c4fef6972a-Abstract-Conference.html.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations*, 2023. URL https://mlanthology.org/iclr/2023/wang2023iclr-zeroshot/.

Weihao Xia, Yujiu Yang, Jing-Hao Xue, and Baoyuan Wu. Tedigan: Text-guided diverse face image generation and manipulation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2256–2265. IEEE, 2021.

Wei Xiao, Johnson (Tsun-Hsuan) Wang, Chuang Gan, Ramin Hasani, Mathias Lechner, and Daniela Rus. SafeDiffuser: Safe planning with diffusion probabilistic models. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/f95606d8e870020085990d9650b4f2a1-Abstract-Conference.html.

Lingxiao Yang, Shutong Ding, Yifan Cai, Jingyi Yu, Jingya Wang, and Ye Shi. Guidance with spherical gaussian constraint for conditional diffusion. *arXiv preprint arXiv:2402.03201*, 2024.

Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. Freedom: Training-free energy-guided conditional diffusion model. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 23117–23127. IEEE, 2023.

APPENDIX CONTENTS

The appendix opens with supporting method and comparison details, followed by additional quantitative analyses, comparison protocols, and solver details. The qualitative gallery then follows the main-table order: vision, audio, robotics, and control and planning, with character motion generation first in the last category. Task objectives and evaluation endpoints are collected at the end. The entries below link to the corresponding sections.

A Supporting Analysis and Comparisons **15**

B The Correction Is a Direction **16**

C Failure Modes and Scope **16**

D Additional Domain Analyses **17**
D.1 Practical audio protocol and remaining boundaries . . . 17
D.2 Predictor scope . . . 18

E Control and Planning: Character Motion Generation **19**

F Paired Adam Budget Audit **21**

G Comparison Protocols **21**

H General Guidance on Executed Obstacles **23**

I Optimizer Trajectories and Resource Accounting **23**
I.1 Solver implementation . . . 23

J Universal Guidance Protocol **25**

K Qualitative Results **25**
K.1 Vision . . . 25
K.2 Audio . . . 29
K.3 Robotics . . . 31
K.3.1 Locomotion . . . 34
K.4 Control and planning . . . 34
K.4.1 Character motion generation . . . 34
K.4.2 Canonical control . . . 35
K.4.3 Trajectory planning . . . 36

L Task Objectives and Evaluation Endpoints **37**
L.1 Request, preservation, and outcome . . . 37
L.2 Vision . . . 38
L.3 Audio . . . 39
L.4 Executed manipulation and locomotion . . . 39
L.5 Control and planning . . . 40
L.5.1 Character motion generation . . . 40
L.5.2 Canonical control and trajectory planning . . . 40

A Supporting Analysis and Comparisons

We give the scalar control and support diagnostic, then summarize the control interpretation and complementary domain findings. Detailed protocols and qualitative examples follow.

Local model and exact scalar null. Section 3 gives the local model and primal/dual solves; the dual system has dimension Z rather than D. For $Z = 1$, write $J_{h,t} = j^\top$ with $j \neq 0$ and $d \neq 0$. The inverse is $jd/(\|j\|_2^2 + \lambda)$. Matching predicted objective-space movement sets FO's step size to $(\|j\|_2^2 + \lambda)^{-1}$, making the latent updates, and hence the outputs under the same sampler randomness, identical. The 935-pair scalar experiment tests this exact null; Appendix B tests the multi-coordinate case.

Damping and numerical support. The coefficient $\gamma = 0.01$ is fixed; $\lambda_k = \gamma\sigma_1(J_k)^2$ is refreshed at each correction. For nonzero d, numerical reachability is

$$A_{\text{reach}}(d) = \frac{\|U_r U_r^\top d\|_2^2}{\|d\|_2^2}. \tag{9}$$

The SD1.5 diagnostic retains $\sigma_i > 10^{-4}\sigma_1$ without truncating the damped correction. Measured real-task requests were fully reachable, so natural rank deficiency is not established here. If $J = 0$, no first-order response is available. Appendix I.1 specifies solver variants, thresholds, and stopping rules.

Local inversion as a control interface. The Jacobian maps perturbations of the full intermediate state to selected properties; its damped inverse maps a request back to an intervention. "Direct" describes this inverse relation: matrix-free solves and finite-edit relinearization still iterate. A future precision layer could re-estimate this geometry from each new observation within a receding-horizon diffusion policy (Chi et al., 2023), using tracking, clearance, energy, or dynamics residuals as properties. For example,

$$h(y) = \left[p_{\text{endpoint}}(y),\ r_{\text{dynamics}}(y),\ E(y),\ c_{\text{clearance}}(y)\right]^\top.$$

A zero dynamics target requires $d_{\text{dynamics}} = -r_{\text{dynamics}}(y)$; zero displacement instead preserves the current residual. Execution-aware validation and independent safety enforcement remain necessary. Learned semantic properties offer a complementary direction, with edit and preservation quality evaluated independently of the optimized features. Neither extension is evaluated here.

Table conventions. Bold marks numerical winners, including rounded ties. The two ns-marked methods in a row have no reported statistically significant difference; this does not establish equivalence. The structured-edit unsteered cosine of 0.000 is a display convention for an undefined zero-edit cosine. Qualitative selection rules appear in Appendix K.

Vision. The selected-guidance comparisons give spectral advantages of 0.133 dB on super-resolution ($p = 4 \times 10^{-8}$) and 0.140 dB on deblurring ($p = .002$). Appendix J reports the setting-specific instability counts, and Appendix F gives the paired optimizer comparisons.

Audio. K1 raises localized mel-EQ success from 8% to 94% and paired-clip consistency from 25% to 100%. On 30 selective-EQ clips at 64 backward passes, K4/Adam/SGD latent displacements are 7.85/19.03/7.02. Appendix D.1 reports coordinate errors and the unoptimized abruptness diagnostic; these are acoustic-control measurements, not listening-study outcomes.

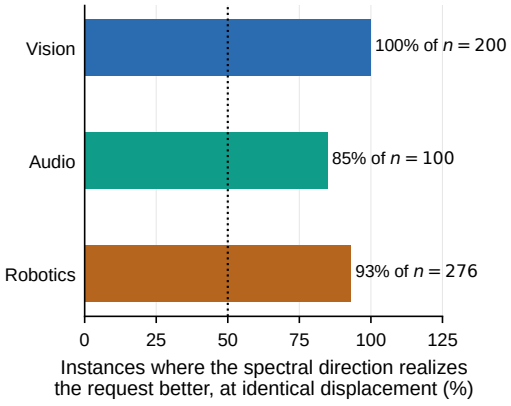
Robotics. DSRL's 6.01 cm reaching error improves its separate 23.6 cm training reference after 8,000 episodes per seed over five training seeds; spectral's unsteered reference is 3.89 cm. DSRL amortizes learned adaptation, while spectral differentiates at inference time. Figure 10 shows how request scale affects transfer to execution. A via-point example improves generated cosine by 0.0674 without detectable executed waypoint improvement ($p = .77$).

Control and planning. After K1, median character errors are 2.92 cm for avoidance and 2.21 cm for reaching, above their 1.66 cm and 1.46 cm tolerances. Directional fidelity therefore need not complete the request (Appendix E). For trajectory control, Appendix G gives paired executed errors; Appendix H compares clearance and goal accuracy. Dropping Panda clearance coordinates reduces target error from 1.30 to 0.70 cm while lowering contact-free execution from 83.2% to 61.9%, illustrating the importance of objective coverage (Appendix C).

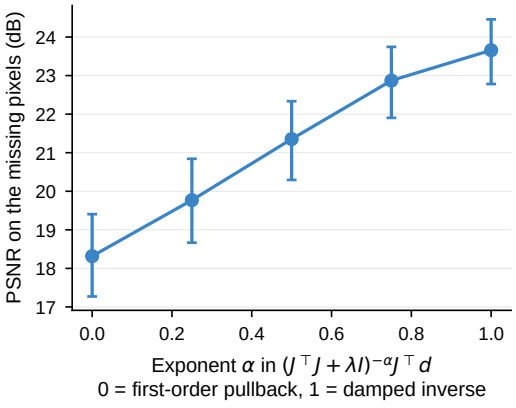
Connection to prior control and guidance. End-to-end optimization, including D-Flow (Ben-Hamu et al., 2024), establishes full-process differentiation and output-gradient geometry; damped least squares and relinearization are classical (Section 2). Here the predictor, scalar null, and movement controls test how request–geometry interactions explain realized directional gains. Finite-edit and execution results test their practical consequences. The predictor requires derivative information and is not a certified step-acceptance rule.

B The Correction Is a Direction

The scalar null in Appendix A predicts identical updates for a single property. The multi-coordinate controls here test whether spectral steering benefits from its update direction after accounting for movement magnitude. Figure 8 first fixes latent displacement, then varies the direction at fixed predicted downstream movement. The two controls separate geometry correction from simply taking a larger step.



(a) Exactly matched latent displacement



(b) Interpolating the direction alone

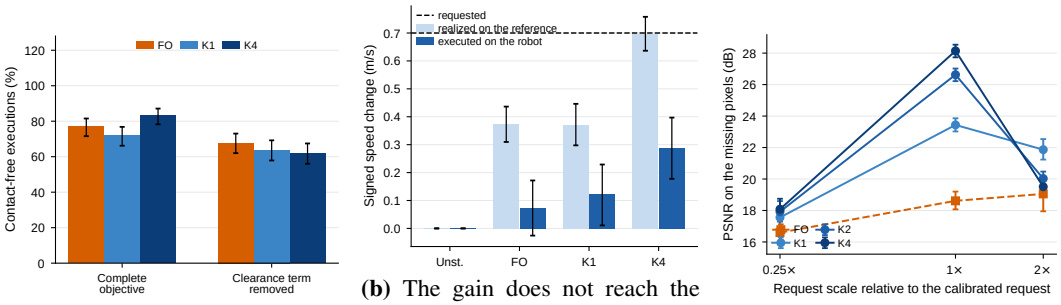
Figure 8: (a) With latent-displacement norms agreeing within 4×10^{-14}, the spectral direction realizes the request better on almost every instance in three modalities. (b) The fractional family $\Delta x(\alpha) = (J_{h,t}^\top J_{h,t} + \lambda I)^{-\alpha} J_{h,t}^\top d$ at exactly matched *predicted objective-space* magnitude $\|J_{h,t}\Delta x\|_2$ interpolates from the first-order pullback ($\alpha = 0$) to the damped inverse ($\alpha = 1$) and moves reconstruction monotonically, improving on 50 of 50 images at every step. Neither panel is a compute comparison; both vary only the direction.

C Failure Modes and Scope

Local request fidelity is only one requirement for useful steering. Figure 9 separates incomplete specifications, loss of a correction during execution, and an excessive requested change. The transfer analysis and optimizer counterexample below identify further limits on the main-paper gains.

A complete optimizer counterexample. On the separate frozen dense-obstacle B2 task ($n = 60$), tuned optimizers outperform the spectral ladder on the preregistered request residual. At about 128 backward passes, SGD and L-BFGS achieve mean residuals 0.0446 and 0.0422 versus K4's 0.6364; L-BFGS also has higher mean minimum clearance (0.09988 versus 0.09294). Thus the executed waypoint advantage does not establish a general planning advantage. An exploratory path-length difference in this campaign is not used to reverse its primary-endpoint conclusion.

Across these cases, useful steering requires objective coverage, a valid local model, and reliable downstream execution.



(a) An incomplete objective **(b)** The gain does not reach the robot **(c)** The request itself is wrong

Figure 9: (a) Dropping the clearance coordinate improves the executed target error and lowers contact-free execution, and the strongest rung loses the most: a more faithful inverse of an incomplete request is a worse outcome. (b) A signed speed request is realized on the generated reference at cosine 0.998 and the robot executes 41% of what was asked; generated and executed quantities are separate bars because they are separate measurements. (c) At a doubled request the deeper rungs pursue the overshoot faithfully and finish below one solve.

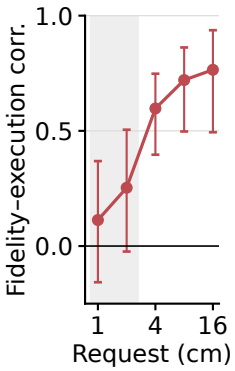


Figure 10: When generated fidelity transfers to robot execution. For Panda reaching, Spearman correlation between the K1-minus-FO gain in generated request cosine and the FO-minus-K1 reduction in executed target error, within each request size. Error bars show 95% bootstrap intervals ($n = 50$–59 per size). The correlation rises from 0.11 at 1 cm to 0.77 at 16 cm; shading marks requests below the servo's resolution. This scale dependence complements the transmission failure in Fig. 9b.

D Additional Domain Analyses

This section supplements the main results with acoustic precision and locality analyses, followed by the pooled directional-prediction audit. It distinguishes an accurately realized request from independent quality criteria and documents how the local-regime population is selected.

D.1 Practical audio protocol and remaining boundaries

The practical audio campaign uses one frozen mel-domain prior with differentiable time–frequency and loudness coordinates. At the headline request, localized EQ, two-clip transitions, reference tonal matching, and paired-clip consistency contain 50, 50, 50, and 40 paired instances. All coordinates must meet the campaign's predefined tolerances. The localized-EQ optimizer comparison uses the first 30 instances of request A1c, with three change and thirteen preservation coordinates. The selective-EQ row in Table 1 reports the mean of the per-instance RMSE over all sixteen coordinates, not a mean of the two block RMSEs. Adam uses learning rate 0.01 and SGD 0.1, selected using 32-step runs on five disjoint pilot seeds (900–904), then frozen for all evaluation budgets. Budgets of 16, 32, and 64 backward passes correspond to K1, K2, and K4 in this explicit-Jacobian audio implementation; these counts are not wall-clock matches.

At 64 passes, change / preservation RMSE is 0.000308 / 0.000392 dB for spectral, 0.034511 / 0.028056 for Adam, and 0.000413 / 0.000477 for SGD. The mean all-coordinate error reduction relative to Adam@64 is 0.02958 dB (95% paired bootstrap interval [0.02760, 0.03144], 30/30 wins; two-sided Wilcoxon $p = 1.86 \times 10^{-9}$). Against SGD@64 the interval includes zero ($p = .63$). SGD wins at smaller budgets and nearly matches spectral at 64 passes; no significant spectral advantage over SGD is claimed. On transitions, every steered variant increases unoptimized energy abruptness from 9.47 to 11.3–11.6 dB. Separate waveform-prior tests show that one solve does not improve bandwidth extension and that preservation requires coordinates that actually isolate the held region:

waveform localized EQ worsens pinned cells by 0.0834 dB at K4. These outcomes do not imply a listening-study or naturalness benefit. Figure 11 examines how request scale changes success and local-model accuracy.

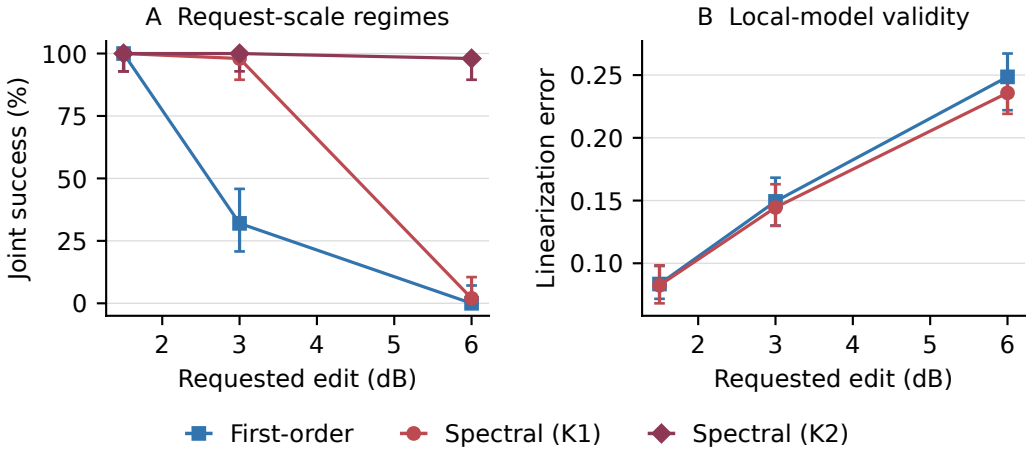


Figure 11: Audio request scale separates direction from locality. At the calibrated edit, one solve improves success; at twice the request, relinearization is needed. FO and spectral incur similar linearization error at matched scales (16-cell diagnostic), so their intermediate-scale difference is directional. These scale sweeps complement the practical tasks in the main text.

D.2 Predictor scope

We compare the predicted and realized gain in request cosine, then examine where the local model ceases to be informative. Figure 12 summarizes the retained populations; Fig. 13 shows the excluded nonlinear regimes. Selection is retrospective, as detailed below.

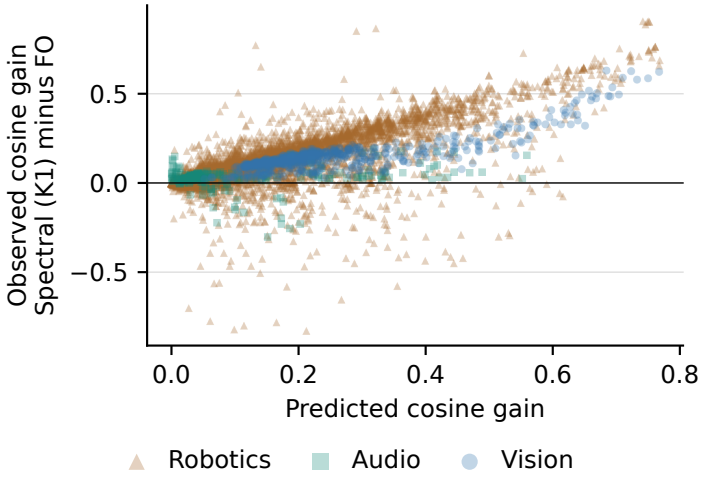


Figure 12: Cross-domain prediction of directional benefit. Each point is one of 8,923 instances from 27 vision, audio, and robotics populations in the measured local regime. The horizontal axis is the pre-edit predicted cosine gain; the vertical axis is the observed cosine gain of K1 over magnitude-matched FO ($r = 0.75$, $\rho = 0.79$). Positive vertical values favor K1. Population-level local-regime selection uses nonlinear responses and is retrospective. The motion illustration in Fig. 4d is a separate diagnostic population and is not included here.

The complete diagnostic collection contains 30 populations and 10,362 instances. Three populations have median one-shot linearization error above one: Panda rotation with position preservation,

Maze2D goal reaching, and the test-time obstacle. Removing those leaves the 27 populations and 8,923 instances in Fig. 12, with positive Pearson correlation in 26 populations. After centering within each population, the pooled correlations remain Pearson 0.74 and Spearman 0.77. Regime assignment uses measured nonlinear responses and is retrospective, so these results do not establish an automatic pre-edit acceptance rule.

On a separate composite-restoration evaluation ($n = 150$), the predictor correlates positively with realized cosine gain ($r = 0.805$, $\rho = 0.766$), but negatively with change-coordinate error reduction ($r = -0.438$, $\rho = -0.424$). Directional fidelity and error reduction are distinct endpoints; neither establishes independent reconstruction quality.

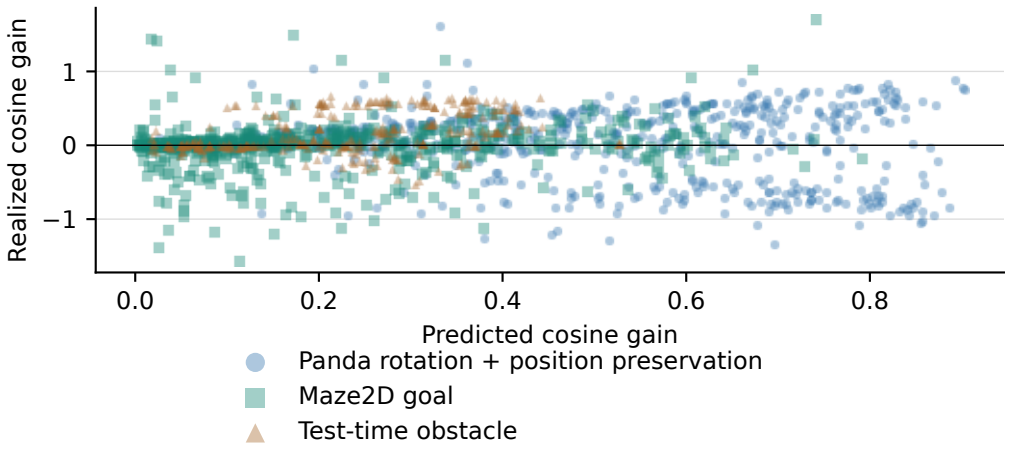


Figure 13: The local-model boundary. The 1,439 instances from the three populations excluded by the retrospective linearization-error criterion. Their broad realized responses show why the headline local-regime correlation should not be extrapolated to large nonlinear edits.

E Control and Planning: Character Motion Generation

Population and generator. We report results on 50 seeds per task, using the same seeds across methods. Comparator hyperparameters were selected on six separate calibration instances, and the jerk threshold was calibrated on 24 additional generator samples. The frozen text-conditioned MDM (Tevet et al., 2023) generator produces 60 frames at 20 fps using a native 50-step sampling schedule and classifier-free guidance 2.5. Steering acts on an intermediate HumanML3D feature state of dimension $263 \times 1 \times 60 = 15{,}780$, with 25 deterministic DDIM transitions remaining (Song et al., 2020). All local corrections act at this same diffusion time.

The main table reports joint success under the positional-target, preserved-position, and jerk criteria defined below, whereas Table 2 reports target-only success. These criteria must not be interchanged. The main-table entries for SGD, Adam, and L-BFGS use the K4-matched nominal backward-pass count; this does not match forward evaluations or standalone runtime.

Solver and first-order control. We differentiate the scaled objectives through the complete remaining sampler and differentiable joint reconstruction. Each correction uses the explicit Jacobian with reduced-SVD filtering and $\lambda_k = 0.01\sigma_1(J_k)^2$, recomputed at each relinearization toward the fixed target $z(x_t^{(0)}) + d$. FO is analytically scaled to match K1's predicted objective-space displacement norm.

Requests and preservation coordinates. Directions are expressed in a subject-centred frame measured from each baseline motion. The requested magnitude is one natural excursion of that motion: 16.6 cm of lateral pelvis travel for avoidance and 14.6 cm of wrist-relative-to-pelvis travel for reaching. Obstacle avoidance uses $Z = 13$: a 20 cm sphere is placed on the body path, the mid-clip pelvis is moved far enough to include a 25% clearance margin, and the final pelvis plus seven

arm-bone lengths are preserved. Multi-objective reaching uses $Z = 77$: the wrist follows a five-frame profile, both feet are pinned at two additional frames, a zero wrist-velocity target is imposed on arrival, and bone lengths are preserved at all waypoint frames. Joint success requires target error below $0.1a$, preserved-position error below $0.2a$, and mean third-order joint-position differences below 0.0417139464 m, computed without time-step normalization. Here, $a = 0.16636$ m for avoidance and $a = 0.14647$ m for reaching. Target error is the pelvis-position error at the avoidance waypoint or the mean wrist-target distance over the five reaching waypoints. Preservation error is the maximum displacement over preserved joint–frame positions. Position and bone-length coordinates are divided by 0.10 m, and velocity coordinates by 2.0 m/s before constructing the objective and its Jacobian.

Table 2: Character-animation relinearization ladder ($n = 50$). Target error is averaged over the 50 seeds and reported in cm; success is the fraction within 10% of the requested amplitude. E_{lin} measures deviation of the realized change from the tangent prediction.

Task	Endpoint	Unst.	FO	K1	K2	K4	K16
Obstacle	error ↓	16.64	10.32	3.54	0.95	0.23	**0.01**
	success % ↑	–	–	18	88	98	**100**
	E_{lin} ↓	–	–	0.164	0.140	0.135	**0.132**
Multi-objective	error ↓	10.62	4.85	2.52	1.40	0.79	**0.19**
	success % ↑	–	–	14	58	96	**98**
	E_{lin} ↓	–	–	0.188	0.204	0.210	**0.185**

Table 2 shows how relinearization reduces positional target error across both tasks.

A separate computational benchmark on the $Z = 77$ task reports 8.61 s for 77 optimizer steps, compared with 3.80 s for sequential Jacobian-row construction and 0.95 s for vectorized construction. The Jacobian rows share one forward graph, whereas optimizer steps evaluate successive states. These component timings characterize Jacobian-construction efficiency; they do not measure the complete standalone runtime of K4.

Table 3: One spectral solve versus magnitude-matched FO ($n = 50$, paired). Both methods start from the same x_t and request and are matched in predicted objective-space displacement norm. Leakage is unintended movement relative to requested movement.

Task	Quantity	FO	K1	Δ	95% CI	W/L	p
Obstacle	fidelity cosine ↑	0.458	**0.976**	$+0.519$	$[+0.492, +0.546]$	50/0	2×10^{-15}
	leakage ↓	1.215	**0.219**	-0.997	$[-1.073, -0.918]$	50/0	2×10^{-15}
	bone drift (m) ↓	**0.011**	**0.011**	$+0.000$	$[-0.000, +0.000]$	27/23	0.67 (ns)
Multi-objective	fidelity cosine ↑	0.701	**0.931**	$+0.229$	$[+0.216, +0.242]$	50/0	2×10^{-15}
	leakage ↓	1.477	**0.507**	-0.970	$[-1.040, -0.894]$	50/0	2×10^{-15}
	bone drift (m) ↓	0.011	**0.010**	-0.000	$[-0.000, -0.000]$	27/23	0.67 (ns)

Figure 14 reports character-motion correlations separately from the 27-population aggregate.

Table 4: Deep refinement and adapted guidance ($n = 50$). A normalized residual above one means the result is farther from $h + d$ than the unsteered baseline.

	Obstacle		Multi-objective	
Method	cosine ↑	residual ↓	cosine ↑	residual ↓
Spectral (K16)	**1.000**	**0.001**	**0.996**	**0.044**
DPS-adapted	0.290	1.605	0.575	1.777
UG-adapted	0.435	1.466	0.325	1.619

Table 4 contrasts deep refinement with adapted guidance. The normalized residuals of matched FO and K1 are respectively 0.972 and 0.260 on avoidance, and 0.738 and 0.379 on multi-objective reaching. Their paired cosine and leakage comparisons appear once, in Table 3.

Intervention time and guidance adaptation. Moving the obstacle intervention from 25 remaining denoising steps to only two improves the normalized residual of DPS-adapted from 1.605 to 0.930

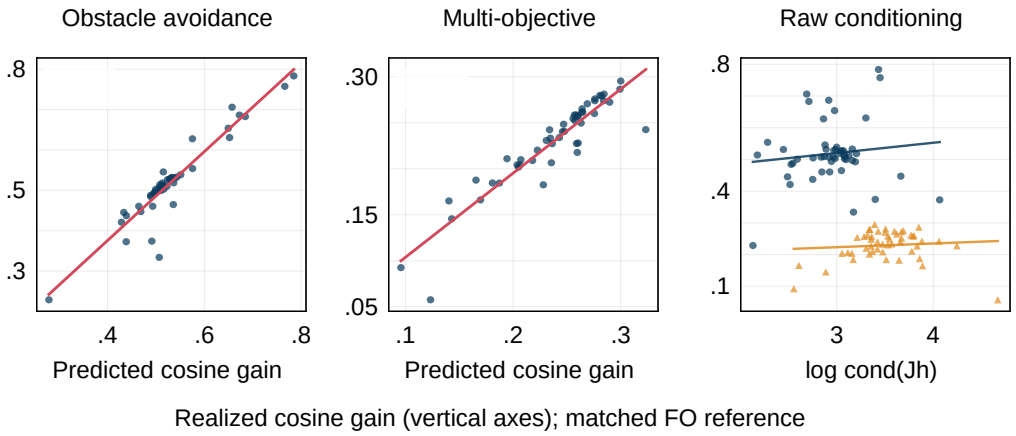


Figure 14: The request-specific predictor transfers to character animation. $G_{\cos}^{\text{pred}}$, computed before editing, predicts the realized cosine gain of K1 over magnitude-matched FO on obstacle avoidance and multi-objective reaching (Pearson $r = 0.94$ and 0.92; Spearman $\rho = 0.92$ for both; $n = 50$ seeds per task). Raw conditioning does not predict that gain within either task (right; $\rho = +0.04$ and $+0.18$). Correlations are computed within each task to avoid between-task offsets. The endpoint is cosine gain.

and Universal Guidance from 1.466 to 0.437, but conjunctive success remains 0% for DPS and rises only to 8% for Universal Guidance. K4 remains at 96–98% over the same sweep, while K1 degrades from 18% to 0%, exposing its sensitivity to a single local model. Universal Guidance operates in the generator's clean motion-feature space. The reported configuration uses five backward-guidance steps and two self-recurrences, for 300 backward passes.

F Paired Adam Budget Audit

The same 50 inpainting cases support two complementary compute comparisons: one matches VJP counts and the other counts both VJPs and JVPs. Table 5 reports reconstruction differences with paired uncertainty. Adam leads one spectral solve at these budgets, whereas K4 leads under both conventions; neither convention establishes equal wall-clock cost.

Table 5: Paired inpainting reconstruction versus Adam at two resource-accounting conventions ($n = 50$). The difference is Spectral minus Adam on missing-pixel PSNR; positive values favor Spectral. Each spectral solve costs 16 VJPs and 16 JVPs, while each Adam iteration costs one VJP. Intervals are paired BCa 95% intervals. The K4 leads survive Holm correction; bold marks the winner in each row.

Accounting	Paired methods	Δ PSNR dB	95% interval	Spectral win %
VJP count	Spectral (K1) vs **Adam@16**	-0.96	$[-1.45, -0.51]$	26
	Spectral (K4) vs Adam@64	$+2.74$	$[+1.78, +3.89]$	78
Total traversals	Spectral (K1) vs **Adam@32**	-2.07	$[-2.73, -1.44]$	18
	Spectral (K4) vs Adam@128	$+3.25$	$[+2.32, +4.35]$	86

G Comparison Protocols

This section specifies sample sets, optimizer selection, and aggregation for the main comparisons, in table order: vision, audio, robotics, and control and planning, beginning with character motion generation.

Vision restoration. The inpainting table uses 50 paired images (760–809); the guidance and optimizer-trajectory figures use images 200–249, with a spectral ladder on each set. On the table's pairs, SGD/Adam/L-BFGS minus K4 PSNR is $-2.18/-1.77/-1.99$ dB, with respective 95% intervals $[-3.31, -1.36]$, $[-2.84, -1.02]$, and $[-3.04, -1.15]$. For SR, deblurring, and composite restoration, SGD/Adam use 50 images and L-BFGS uses 20. L-BFGS selects the best evaluation PSNR over $1, 2, 4, 8, 16, 32, 64$ iterations, choosing 16/64/64 (32/128/128 backward calls).

Structured-edit optimizer row. SD1.5 uses the standard-scale box-average color request (Objective B, window 10), seeds 6000–6019, and realized-request cosine. GD/Adam use pilot-frozen ten-step runs at rates 100/0.01; spectral uses refreshed relative damping and peaks at K8. L-BFGS reaches 0.765 at ten backward passes (five strong-Wolfe iterations), with the same regularized loss as GD/Adam, $\frac{1}{2}\|h(x+\Delta x)-h^\star\|_2^2+\frac{1}{2}10^{-3}\|\Delta x\|_2^2$. Its rate 1.0 is evaluation-selected; rates 0.1/0.3 are ineligible because each stalls on one case. Paired tests use in-process GD/Adam reruns; the table retains their frozen summaries. Figure 5(e) uses $4\times$ requests.

Structured-edit guidance. The $4\times$ stress test uses objective adaptations. CLAMP implements the published equations with stand-in constants, without validation against an official code example. The displayed marginal intervals do not establish paired significance; spectral collateral change was not measured on the same pass.

Waveform audio and budget sensitivity. SGD/Adam use 25 shared excerpts (100–124), with rates selected on the tuning objective using two disjoint excerpts. Table 1 reports 16-step results. Bandwidth extension evaluates high-band reconstruction LSD, not the optimized measurement loss. Selecting the best evaluation LSD over $1, 2, 4, 8, 16$ steps yields SGD@1 (rate 0.3) and Adam@2 (rate 0.01). Adam@2 reaches 12.525 dB versus K4's 12.730, with paired K4-minus-Adam interval $[-0.44, 0.83]$ dB and $p = .37$. This stopping-budget sensitivity test does not compare the table's Adam@16 cell.

Localized EQ reports requested-region error (`on_err_db`). SGD/Adam use rates $3\times10^{-6}/3\times10^{-4}$; SGD's near-zero step reflects its tuning rule. The completed row includes L-BFGS at 3.895 dB and K16 at 2.691 dB. Appendix D.1 describes the separate mel-audio protocol.

Robotics refinement depth. On shared reaching instances, K8 and L-BFGS both have median error 0.63 cm ($p = 1.00$); K16 reaches 0.59 cm without a significant paired difference. HalfCheetah target error worsens at K8 before recovering at K16. On obstacle reaching, K4/K8/K16 joint success is 75.8/75.5/76.2% ($p = 1.00$ for K4 versus K16), while contact-free execution increases from 83.2% at K4 to 89.4% at K16. Panda depth tests are paired within one re-execution.

Character motion generation. The character comparison uses 50 seeds per task, with separate calibration instances and nominal backward-pass accounting. Appendix E defines its requests, matching rule, success criteria, and solver. Target-only ladder success and the main table's joint success are distinct reported endpoints.

Executed one-waypoint control. The additional table row reanalyzes the recorded executed waypoint endpoint from the optimization extension: three frozen missions, fifteen seeds each, with all methods on exactly the same 45 pairs. A six-coordinate request specifies the timed waypoint, terminal position, and terminal velocity at a frozen diffusion intervention. The SGD column is full-batch GD (30 steps, rate 0.5); Adam uses 40 steps at rate 0.3. The L-BFGS column is SciPy L-BFGS-B with no variable bounds, 100 maximum iterations and 300 maximum objective/gradient evaluations. All three minimize the squared nonlinear request error plus $10^{-3}\|\delta x\|_2^2$; their settings are fixed by the source protocol, without a reported learning-rate sweep. Spectral uses the same frozen target with its damped local solves. The recorded count for K4 is 28 Jacobian-building forward/backward calls; the L-BFGS-B objective/gradient count averages 85.3. These counters combine different operations and do not establish a runtime ratio.

Against GD, Adam, and L-BFGS-B, the paired mean executed-error reductions are 34.65, 4.43, and 23.30 cm; the respective percentile bootstrap 95% intervals are [29.46, 40.09], [2.96, 5.86], and [13.38, 34.49] cm (20,000 paired resamples). All three two-sided Wilcoxon tests remain below 10^{-8} after Holm correction over these comparisons. The uncertainty is conditional on the three tested missions and chosen settings. Generated task success is 100% for both K4 and L-BFGS-B, whereas executed success is 84.4% and 53.3%; Adam executes successfully on 71.1%. This distinction motivates reporting execution rather than generated success.

Planning endpoints in Figure 7d. Maze2D uses 553 available evaluations for unsteered, FO, and K1, and 501 for K4. Dense-obstacle and novel-obstacle collision rates use the paired reruns with 315 and 263 executions, respectively. Planar-arm error uses 380 evaluations per method, and

shifted-goal success uses 1,440. The plot divides distance, error, or collision rate by the unsteered value within each task; shifted-goal success is first converted to failure rate. Thus one denotes the unsteered endpoint and smaller values are better, without averaging incompatible physical units. The accompanying source data retain all raw endpoints and counts.

H General Guidance on Executed Obstacles

This planning comparison evaluates general guidance on the same simulator executions as spectral steering. Table 6 reports collision and goal-request error separately, because the multi-coordinate objective can trade off clearance against the requested goal change.

Table 6: Novel test-time obstacle, $n = 263$ paired executions. Collision is physical; request error scores the goal move in generator space. Passes are campaign backward-pass units. Bold marks the best displayed number in each endpoint column; unsteered has the lowest request error here.

Method	Passes	Collision % ↓	Request error ↓
Unsteered	0	100.0	**0.350**
FO	45	79.5	0.522
Spectral (K1)	45	52.5	0.677
Spectral (K4)	180	11.8	0.417
Universal Guidance, fixed $m = 1$	16	9.5	0.756
Universal Guidance, fixed $m = 16$	256	**1.1**	0.442

On the dense-clearance condition ($n = 315$), K4 and the best guidance setting record 1.6% and 0.0% collisions respectively; the gap is small at this low base rate. The novel-obstacle condition supplies the discriminating comparison. Unsteered has lower median goal error than K4 here (0.350 versus 0.417), so the relative K4–guidance request advantage is not an absolute goal-accuracy gain. The 45-coordinate objective contains two goal coordinates, with the remainder for clearance and preservation. Request errors are heavy-tailed; comparisons use medians and paired win rates.

I Optimizer Trajectories and Resource Accounting

Figure 15 puts reconstruction outcomes on a shared network-evaluation axis and distinguishes the evaluated image populations. The accounting below explains how the local inverse and optimizer updates contribute to that cost; the solver summary specifies the corresponding implementations.

One Golub–Kahan iteration requires one VJP and one JVP through the remaining sampler. With the 16-iteration cap, one matrix-free correction uses up to 16 VJPs and 16 JVPs; K corrections use up to $16K$ of each. For the vision configuration with eight remaining denoising steps, this corresponds to 256 derivative-traversal equivalents per correction, versus 128 for 16 SGD/Adam steps. These counts exclude primal forward evaluations, checkpoint recomputation, and endpoint evaluation. Reported timings on the shared GPU are approximately 6.55 s per Krylov iteration and 3.44 s per SGD/Adam step. Magnitude-matched FO reuses K1's computation in the directional control; this shared accounting is not its standalone cost.

I.1 Solver implementation

The same local inverse is evaluated with different derivative representations according to the output dimension and generator. The table maps the experiments to explicit or matrix-free implementations; the paragraphs below specify damping, numerical support, and the iterative stopping rule.

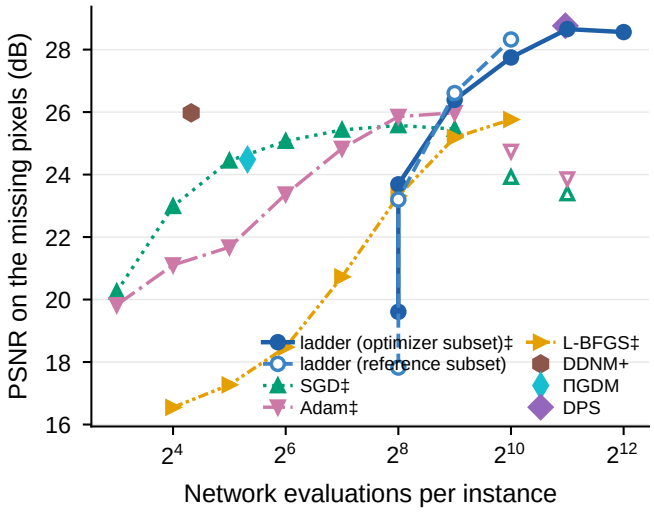


Figure 15: Outcome against compute on one common denominator of network evaluations, on random-mask inpainting. The ladder is drawn on both instance subsets because the optimizer runs used a disjoint set (‡). The specialized inverse methods buy early reconstruction gains at lower network cost. Optimizer lines use the 50-image cohort; hollow optimizer markers are separate 20-image extension endpoints. The paired reconstruction-versus-residual trajectory is in Fig. 6.

Experimental setting	Jacobian and local solve
CelebA-HQ restoration; waveform localized EQ and bandwidth extension	Matrix-free JVPs/VJPs; 16-step Golub–Kahan projection and a direct damped solve in the projected space.
CIFAR and SD1.5 structured edits; mel-audio	Explicit reverse-mode Jacobian; reduced SVD and spectral filtering.
Executed robotics	Explicit Jacobian with batched reverse-mode rows sharing a forward graph; row-wise fallback; reduced SVD.
Early robot geometry diagnostics	Explicit sampler Jacobian composed with the objective Jacobian by the chain rule; reduced SVD.
Character motion generation	Explicit Jacobian through the remaining sampler and joint reconstruction; batched reverse-mode rows with a row-wise fallback; reduced SVD (Appendix E).
Custom trajectory control and standard 2D planning	Explicit reverse-mode Jacobian and reduced SVD; 2D planning uses batched rows with a row-wise fallback.

Damping and support. In the audited implementations, "adaptive" means $\lambda = 0.01\widehat{\sigma}_1^2$, refreshed at every local correction. Explicit solvers use $\widehat{\sigma}_1 = \sigma_1(J)$ and apply $V\,\mathrm{diag}[\sigma_i/(\sigma_i^2+\lambda)]U^\top d$ without hard rank truncation. The SD1.5 reachability diagnostic retains $\sigma_i > 10^{-4}\sigma_1$; that campaign separately excludes $\sigma_1 < 10^{-8}$ (an absolute threshold). The common audio/control/robotics cores log numerical rank using $\sigma_i > 10^{-6}\sigma_1$; this is a separate diagnostic. Named damping ablations vary the relative coefficient: the SD1.5 helper defines $\gamma \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$, and the inpainting runs use $0.1\times$ and $10\times$ the default. The SD1.5 fixed-damping control instead uses $\lambda_0 = 0.01\,\mathrm{median}_{\mathrm{pilot}}(\sigma_1^2)$ per objective and intervention time. None of these rules adds a damping line search.

Matrix-free solve. Starting from $u_1 = d/\|d\|_2$, Golub–Kahan bidiagonalization constructs U, V, B with full reorthogonalization against the stored vectors of both bases. The cap is 16 iterations; early termination occurs only if a newly computed unnormalized basis vector has norm at most $10^{-30}\|d\|_2$. There is no residual-based stopping tolerance or preconditioner. Basis storage and projected algebra use float64; derivative evaluations retain the model's dtype. The implementation sets $\widehat{\sigma}_1 = \sigma_1(B)$ using an SVD of B, then computes

$$z = (B^\top B + \lambda I)^{-1} B^\top (\|d\|_2 e_1), \qquad \Delta x = V z.$$

No separate power iteration is used. The logged residual is that of these *projected* normal equations, not a stopping test or a full-space convergence certificate. A zero request returns a zero correction. The request-started projection is not used to establish full-Jacobian numerical reachability.

J UNIVERSAL GUIDANCE PROTOCOL

We distinguish selected guidance settings from an explicit instability audit. Table 7 records failures at one unstable setting per vision operator; it is not the selected-setting comparison in the main figure. The remaining text specifies the vision, audio, and control-and-planning protocols.

Table 7: Failures below the unguided prior at an unstable guidance setting. Counts for Universal Guidance (UG) use one evaluated unstable setting per operator ($n = 50$); spectral counts pool its three evaluated rungs ($n = 150$). These are robustness counts, not the post-hoc best-setting PSNR comparison shown in Fig. 5.

Vision operator	UG below prior	Spectral below prior
Inpainting	4/50	**0/150**
$4\times$ super-resolution	50/50	**0/150**
Deblurring	40/50	**0/150**
Multi-constraint	50/50	**0/150**

Figure 5 reports Universal Guidance alongside the vision baselines, using the selected settings from the frozen operator audit: fixed $m = 2$ for inpainting, fixed $m = 1$ for super-resolution, and normalized $m = 8$ for deblurring and multi-constraint restoration.

On the 50 reference inpainting images, selected UG reaches 28.39 dB versus K4's 28.32, with no significant paired difference. The dense Adam trajectory peaks at 25.60 dB at 54 steps (Fig. 6).

The separate audio guidance comparison uses the unconstrained $Z = 16$ active-preservation setting. For control and planning, the character motion generation adaptation is described in Appendix E. The constrained-planning results use the same-loop reruns with all methods paired ($n = 315$ dense, $n = 263$ novel obstacle); the separate UG comparisons use fixed $m = 2$ and fixed $m = 16$, respectively. Figure 7d uses these paired reruns.

The audio comparison uses forward guidance plus self-recurrence, with scale selected on separate tuning examples. At $Z = 16$, unconstrained Universal Guidance reaches 0.0501 dB error with latent norm 58.8, versus K1's 0.4481 dB and norm 6.8; displacement matched, it reaches 1.5760 dB. At $Z = 32$, the corresponding values are 0.0327 versus 0.5308 dB unconstrained and 1.5356 versus 0.5308 dB matched. This is evidence that nonlinear guidance can optimize the endpoint and that movement control changes the question.

K QUALITATIVE RESULTS

The gallery complements the examples alongside the domain plots in Figs. 2, 5, and 7. Each panel is an existing experiment output. The organizing question is whether the requested coordinates change while the named preservation coordinates and downstream behavior remain satisfactory.

Selection and interpretation. Captions identify recorded median, quantile, boundary, and high-effect examples. Video-derived control and planning panels use synchronized instants selected for maximum change-from-initial separation between FO and the highest displayed spectral rung. These are illustrations, not estimates of typical performance; printed metrics describe individual cases. Crops preserve aspect ratios. Task definitions and endpoints are in Appendix L; aggregate comparisons appear in the quantitative results.

K.1 VISION

Figures 16–17 show inverse problems whose loss is a measurement residual. Figure 18 shows paired-image and multi-region objectives built from regional color, agreement, and preservation coordinates. Printed metrics describe the displayed instances.

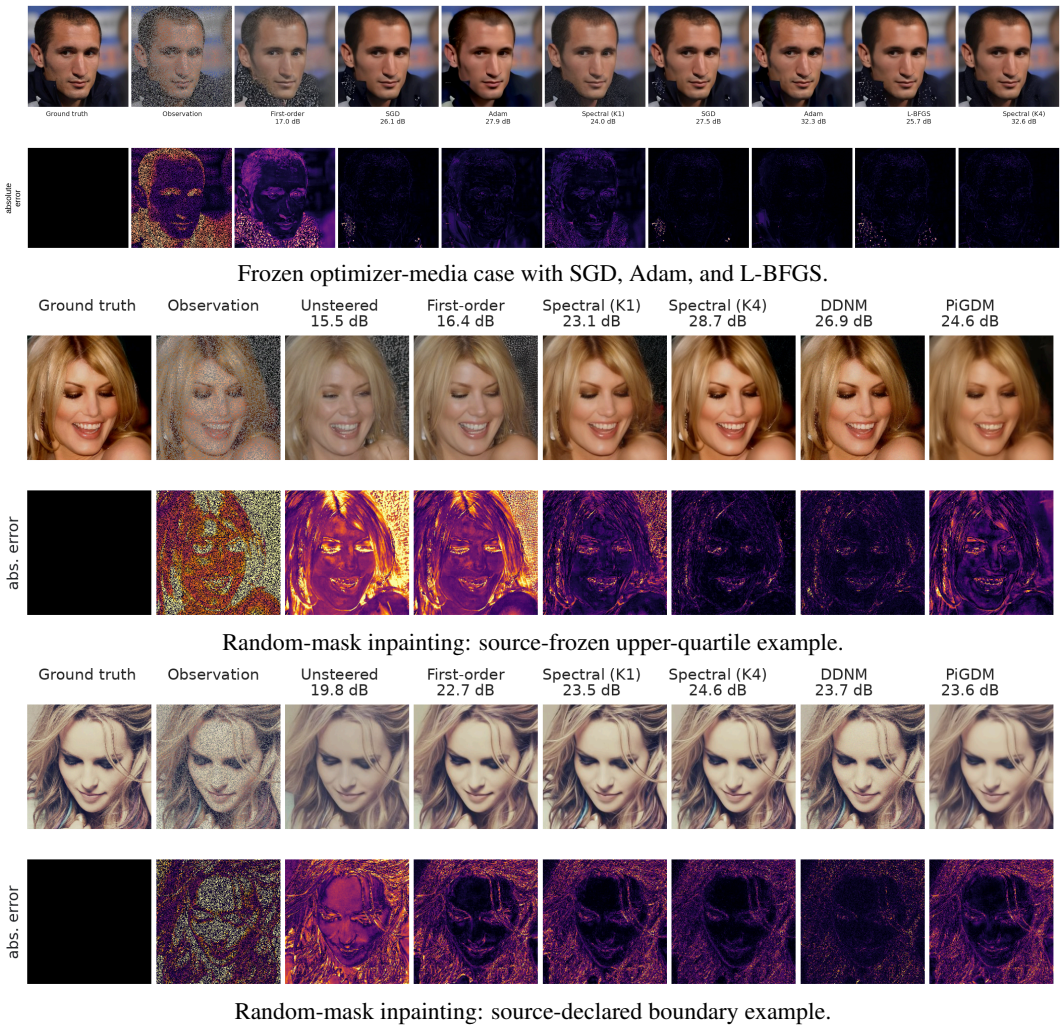
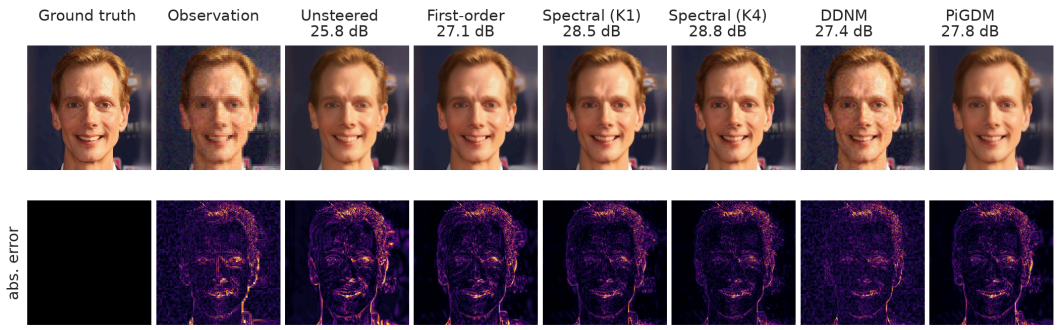
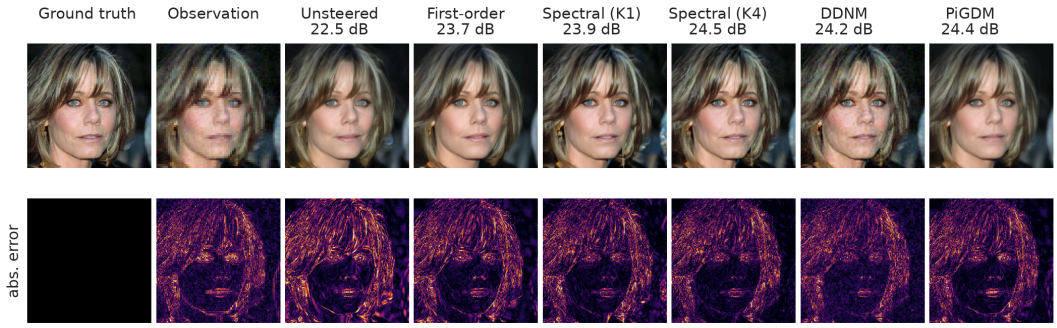


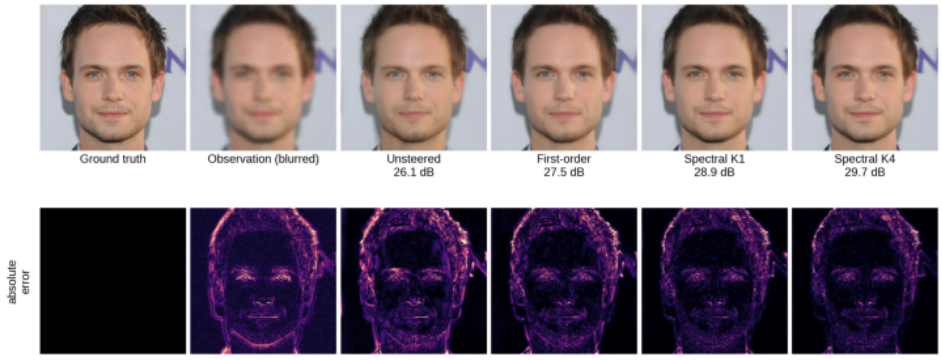
Figure 16: Random-mask inpainting: a frozen optimizer comparison, an upper-quartile case, and a boundary case. Methods share the observation within each example; printed PSNR and error maps describe that instance.



$4\times$ super-resolution: source-frozen upper-quartile example.

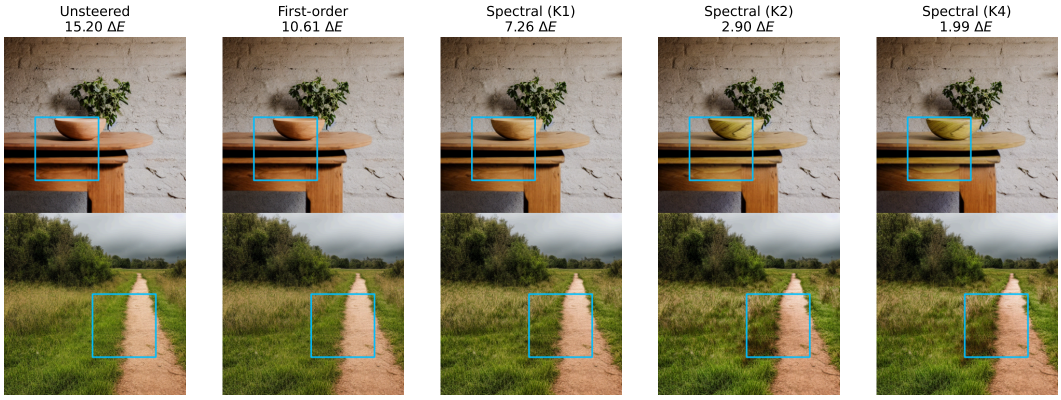


$4\times$ super-resolution: source-declared boundary example.

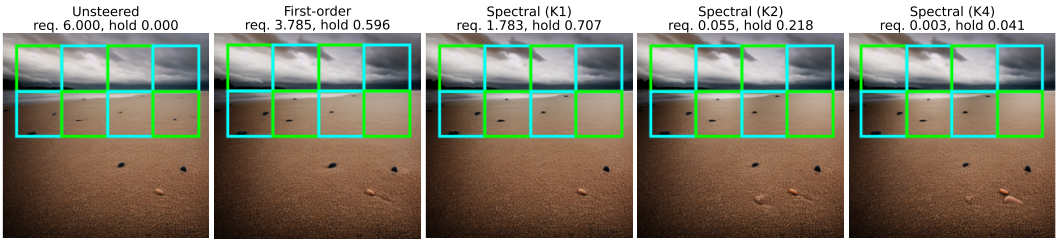


Gaussian deblurring: source-frozen median example.

Figure 17: Vision inverse-problem examples: super-resolution and deblurring. The reconstruction and error maps complement PSNR by locating residual errors. The upper-quartile, boundary, and median cases illustrate variation within the stored evaluations.



Joint edit of two images: the outlined regions are requested to agree while the remaining cells are held fixed.

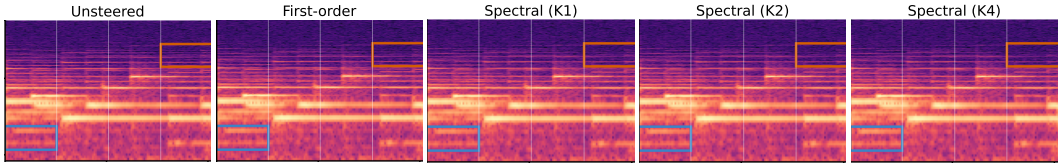


Multi-region appearance control with requested and held CIELab blocks.

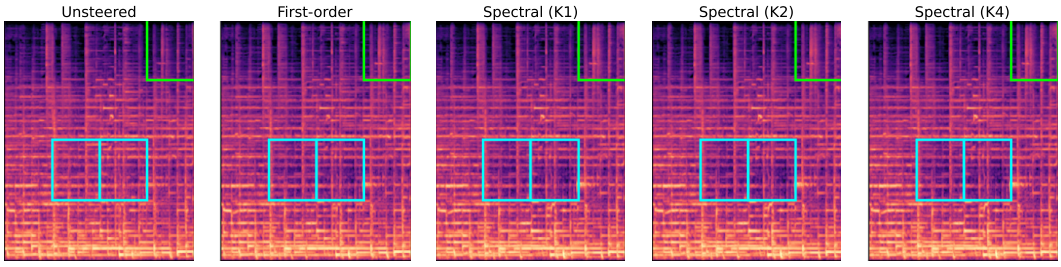
Figure 18: Paired-image and multi-region objectives from the vision extension.

K.2 AUDIO

Figures 19, 20, and 21 organize the audio examples by objective support. Equalization rows expose the time–frequency cells that enter equation 11; restoration rows expose observed and missing frequency or time regions; transition rows expose the temporal quarters, contexts, and seams used by their dB objectives. Spectrograms locate changes in time and frequency. They are visual diagnostics; perceived quality requires listening evidence, and held-cell errors remain part of the quantitative evaluation.

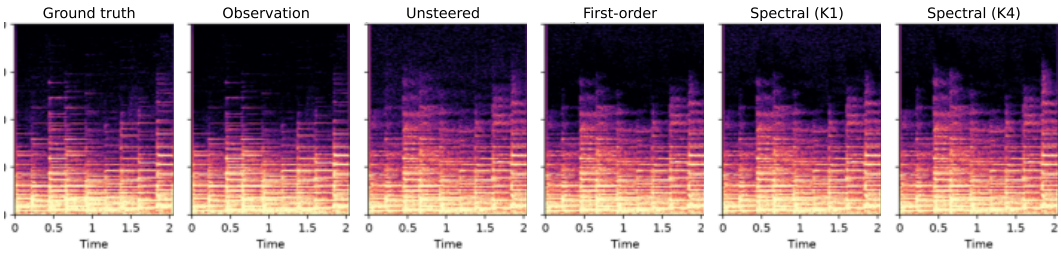


Localized time–frequency equalization shown as spectrograms.



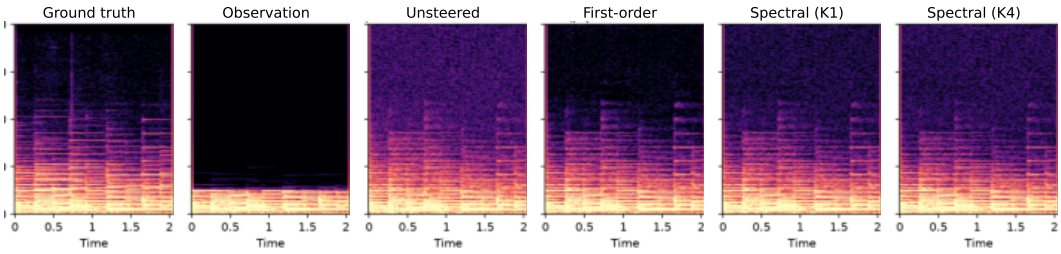
Localized EQ automation: low-mid cut in the middle and high-band boost late in the clip.

Figure 19: Localized and automated audio equalization. Requests identify both edited and held time–frequency regions; Appendix D.1 reports their quantitative errors.

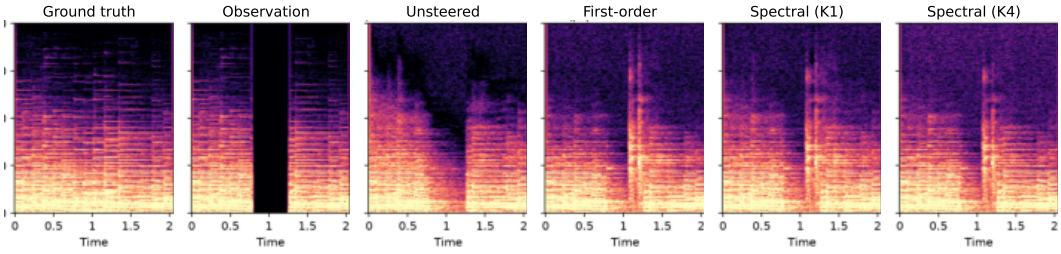


Known-filter equalization restoration.

Figure 19: Audio equalization (continued): restoration toward a supplied target spectrum.

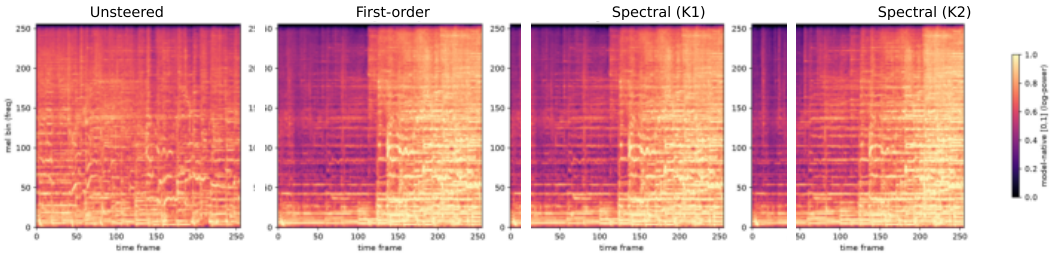


Bandwidth-extension reconstruction.



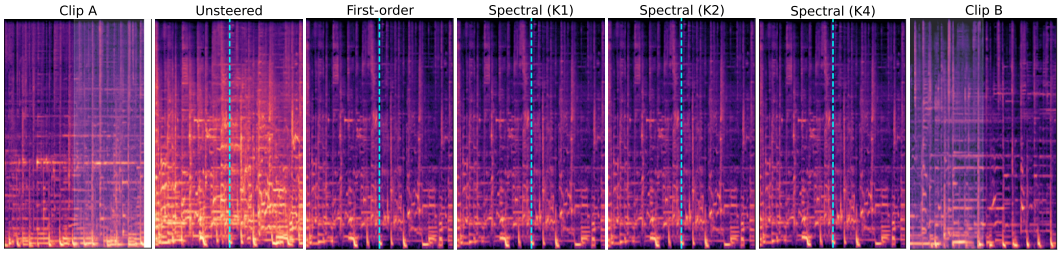
Missing-segment waveform inpainting.

Figure 20: Audio restoration: bandwidth extension and missing-segment inpainting.

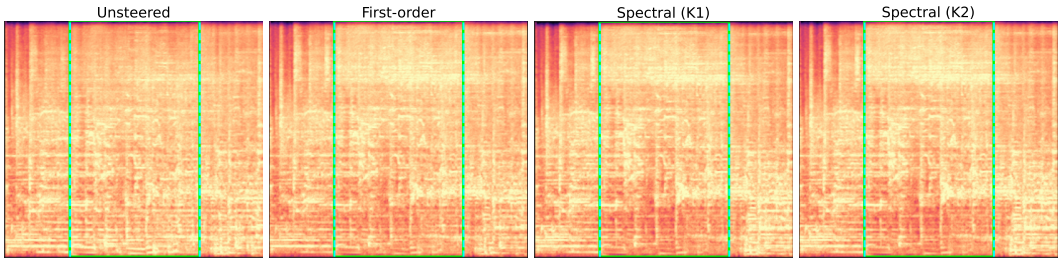


Crescendo temporal-envelope edit shown as spectrograms.

Figure 20: Audio restoration and temporal control (continued): crescendo-envelope editing.



Crossfade from clip A to clip B across four temporal quarters.



Interior high-band boost and low-band cut with both contexts and both seams held.

Figure 21: Audio transition objectives. Dashed or boxed boundaries in the spectrograms mark the temporal regions entering the cost.

K.3 ROBOTICS

Each row is a synchronized frame from the stored execution videos. The first two groups show position, orientation, checkpoint, and plane-preservation objectives (Figs. 22 and 23). The third shows whole-arm clearance constraints, including cases where the first-order rollout contacts the obstacle while a spectral rollout remains clear (Fig. 24). These manipulation examples were selected as high-effect illustrations. A contact-free unsteered rollout that fails to complete the reach does not establish task success.

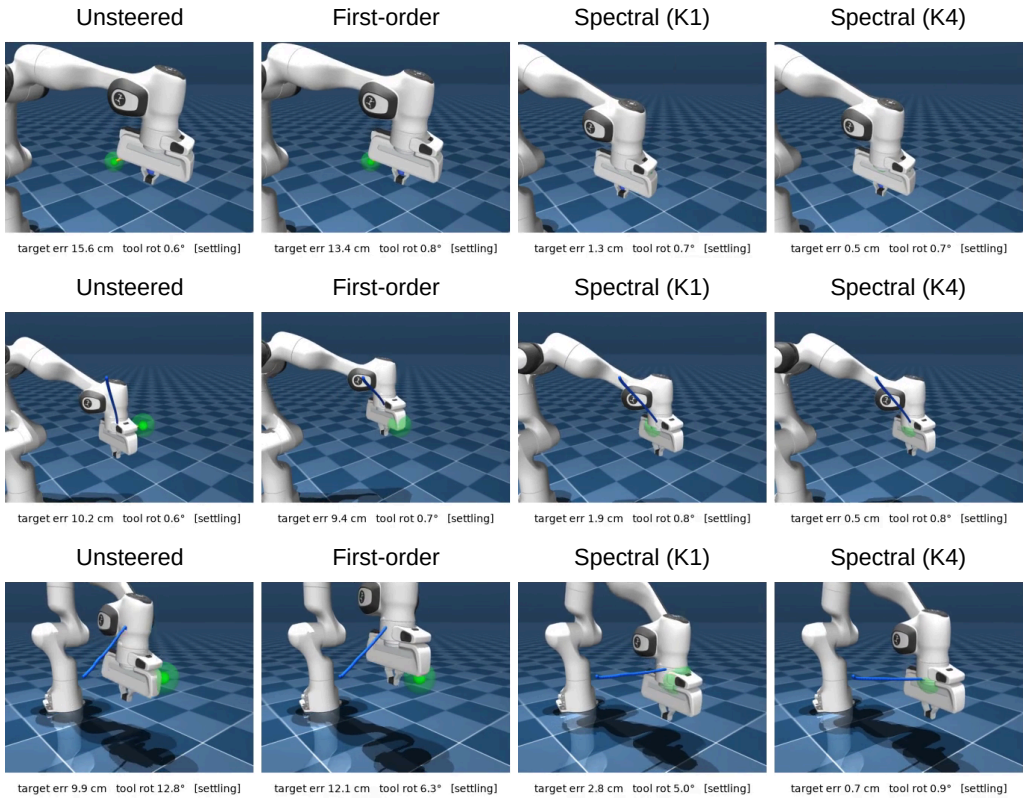


Figure 22: Manipulation, top to bottom: Cartesian reach, XYZ-marker reach, and orientation hold.

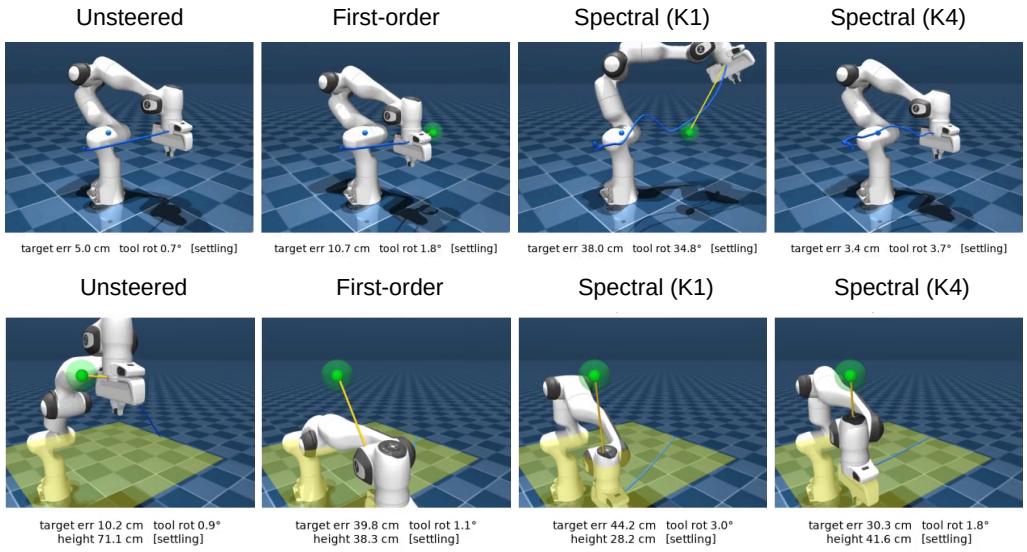


Figure 23: Manipulation, top to bottom: via-point reach and constant-height plane constraint.

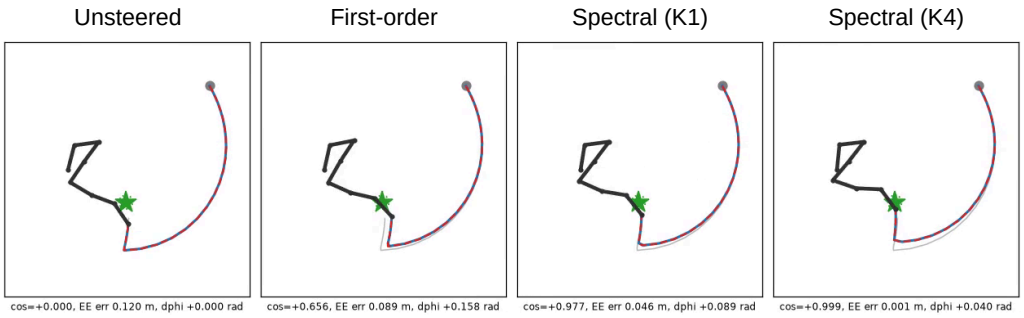


Figure 23: Manipulation (continued): planar-arm pose control.

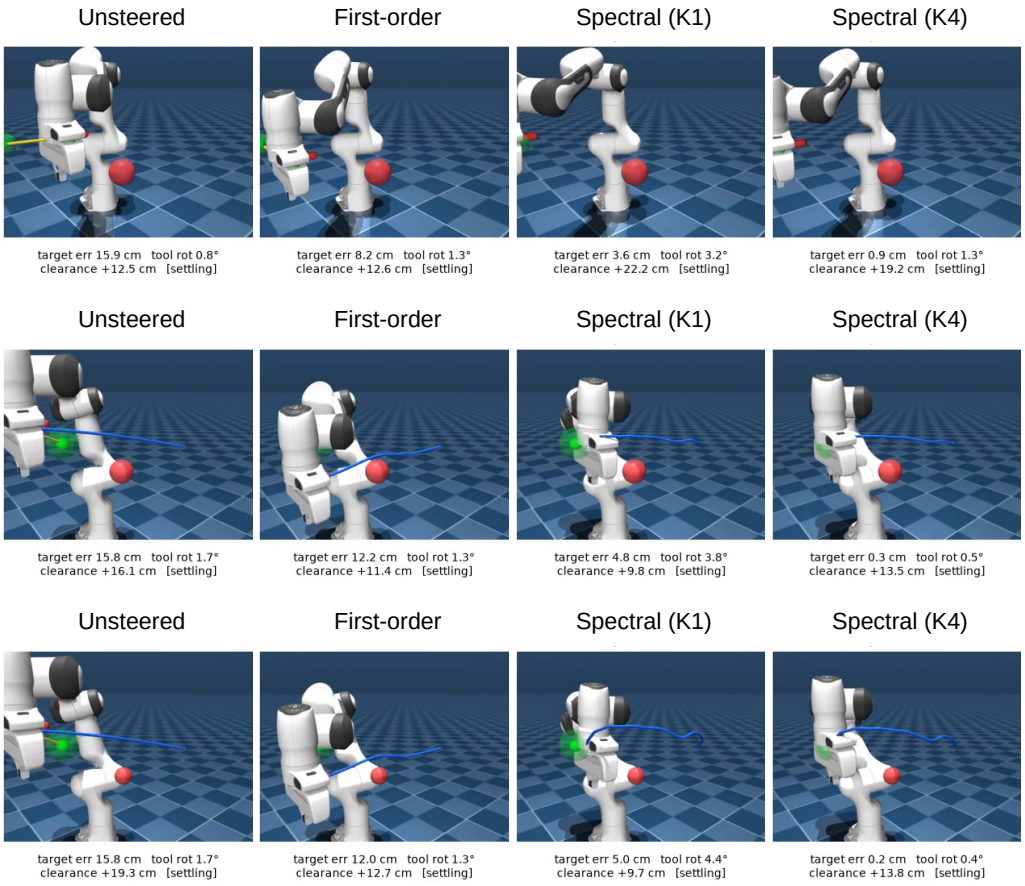


Figure 24: Executed obstacle avoidance, top to bottom: unseen obstacle, lateral clearance, and narrow gap. Selected high-effect cases; contact-free execution must also complete the reach.

K.3.1 LOCOMOTION

Figure 25 retains the rollout renderings and their printed progress and posture outcomes. The cost requests forward progress over eight replanning windows while holding torso height and pitch. This source-declared high-effect example illustrates closed-loop transmission; the quantitative reference-to-execution limitations remain in Appendix C.

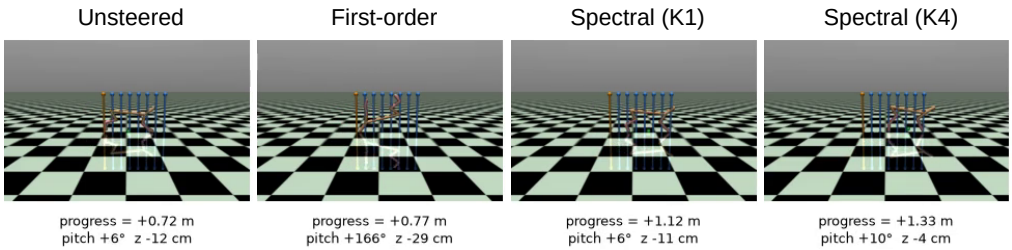


Figure 25: Forward progress over eight planning windows. All references are executed by the same objective-blind tracking controller.

K.4 CONTROL AND PLANNING

The gallery starts with character motion generation, where the specification coordinates several joints over time. Canonical control and trajectory planning follow, showing how the same interface applies to dynamical states and routes.

K.4.1 CHARACTER MOTION GENERATION

Figures 26 and 27 show avoidance and reaching with the same camera and instance across methods.

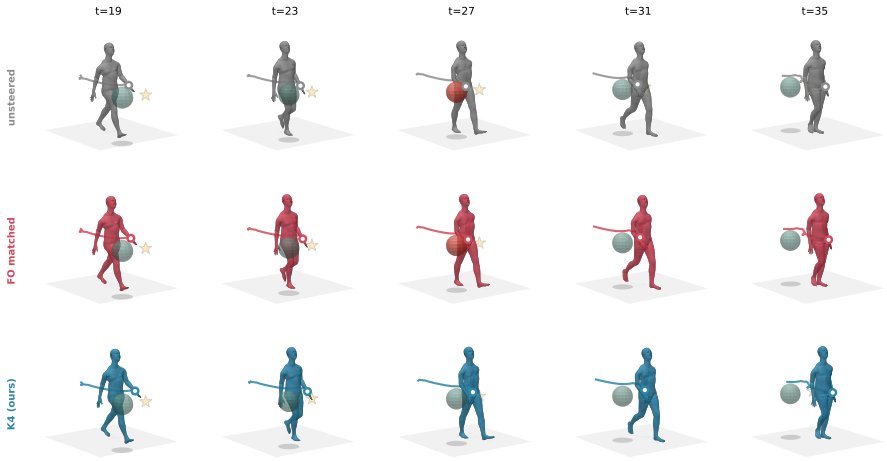


Figure 26: Obstacle-avoidance character sequence. Five instants surround the unsteered motion's deepest penetration. Rows show unsteered, magnitude-matched first-order, and Spectral (K4). A red sphere indicates that any skeleton segment intersects it. Figure 2 uses two instants from this same sequence.

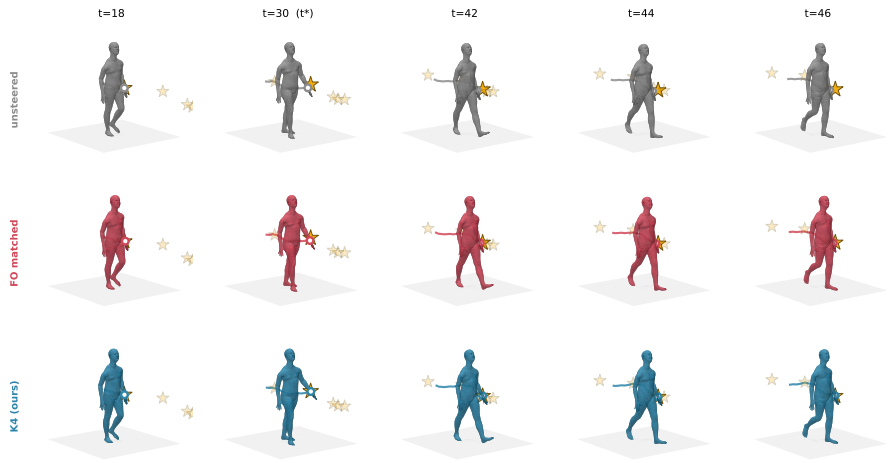


Figure 27: Multi-objective character reaching. Columns show five requested instants; rows show unsteered, magnitude-matched first-order, and Spectral (K4), with a fixed camera. Stars mark requested wrist positions; the active waypoint is solid. Foot positions, zero arrival velocity, and bone lengths enter the objective; velocity and bone drift are separate diagnostics from the reported success criterion. The inherited small sphere is not a clearance objective for this task.

K.4.2 Canonical control

The pendulum request (Fig. 28) specifies checkpoint angles and terminal angular velocity; CartPole requests a positive terminal cart displacement with upright, stationary terminal targets (Fig. 29). The renderings are paired with full state traces, so a selected frame can be checked against the complete rollout. Frame selection follows the visual-separation rule above.

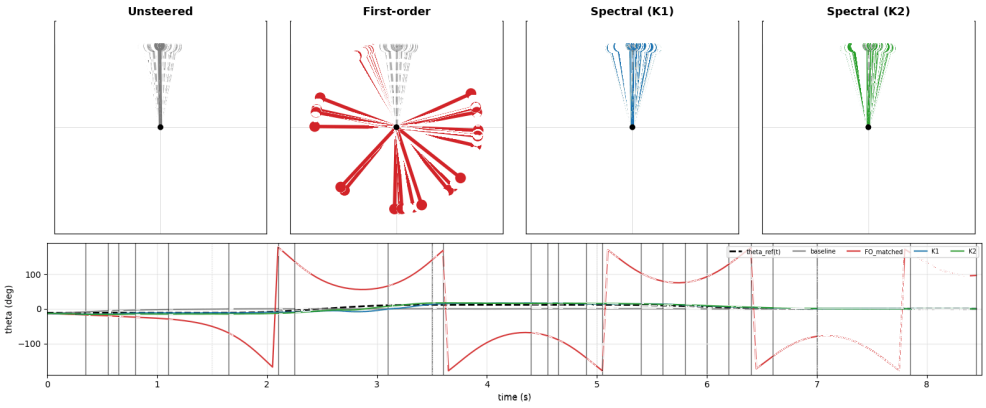
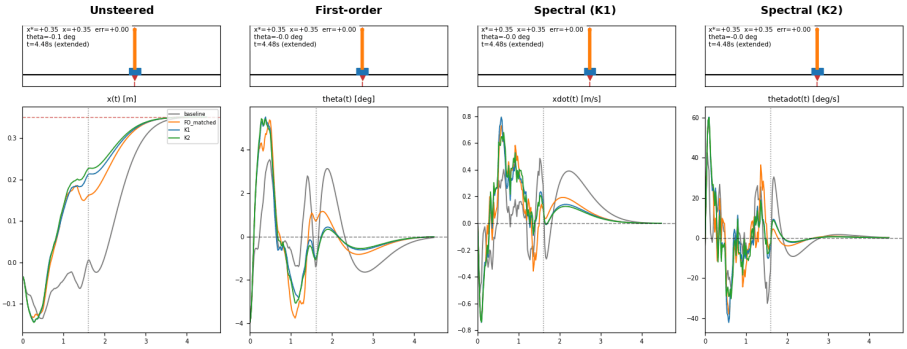


Figure 28: Pendulum reference tracking, shown on a source-declared high-effect example. The angular reference and full traces complement the selected rendering.

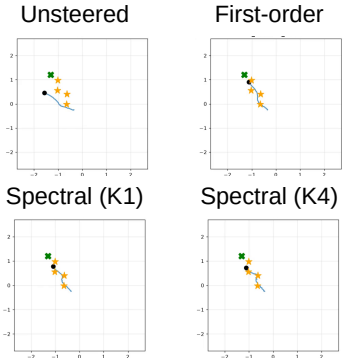


CartPole terminal-position request: selected illustrative rollout.

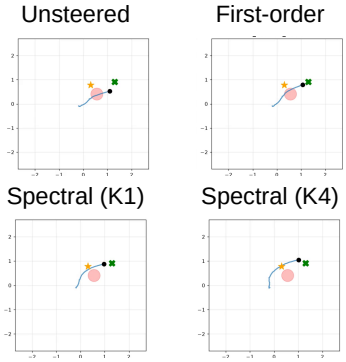
Figure 29: CartPole terminal displacement with upright-angle and zero-velocity terminal coordinates. The upper strip renders the selected state, while the lower plots show the full state trajectories and the requested terminal cart position.

K.4.3 TRAJECTORY PLANNING

Figure 30 shows timed-waypoint tracking and obstacle avoidance with a double integrator. Black dots mark endpoints, orange stars waypoints, green crosses goals, and red disks obstacles. Frames maximize the change-from-initial separation between FO and Spectral (K4), following the gallery selection rule.



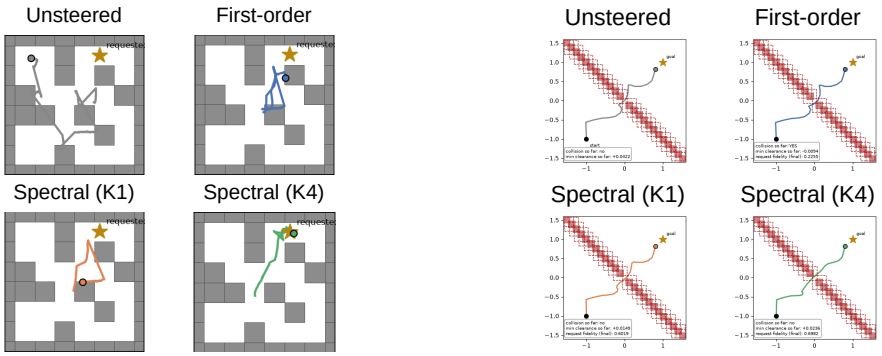
Four timed waypoints.



Waypoint and goal with obstacle avoidance.

Figure 30: Double-integrator trajectory objectives. Method order is fixed as Unsteered, First-order, Spectral (K1), and Spectral (K4).

Figure 31 shows a shifted Maze2D goal and a narrow passage. Path diagnostics report goal error, clearance, and smoothness; negative clearance denotes a violation. Paths are generated references unless execution is explicitly reported.



Maze2D medium maze, shifted goal. Narrow-passage planning.

Figure 31: Goal-conditioned and obstacle-aware diffusion planning. Each trajectory panel uses the same method order: Unsteered, First-order, Spectral (K1), and Spectral (K4).

L Task Objectives and Evaluation Endpoints

The following catalogue connects the common steering interface to the task-specific experiments and preceding gallery. It supplements the quantitative comparisons with the definition of what each method is asked to change, what is explicitly held, and how the outcome is evaluated. Parenthesized source keys identify the corresponding task campaigns.

Datasets and frozen diffusion models. The model names in Table 1 abbreviate the actual frozen checkpoints. Vision restoration in the main table uses `google/ddpm-ema-celebahq-256` (DDPM EMA-256) (Ho et al., 2020) on CelebA-HQ validation images (Xia et al., 2021). The structured vision edits use an additional model: `stable-diffusion-v1-5/stable-diffusion-v1-5` (Stable Diffusion 1.5) (Rombach et al., 2022), with generated 512×512 images from the campaign's frozen prompt/seed pools. This is the model in Fig. 5(e), the spatial and composite editing experiments, and the paired-image/region-transfer qualitative examples. The table uses standard-scale edits; the guidance figure uses $4\times$ requests.

Waveform EQ and bandwidth extension use `harmonai/maestro-150k` (MAESTRO 150k) on MAESTRO v3 test excerpts (Hawthorne et al., 2018). EQ-16 and selective EQ use `teticio/audio-diffusion-ddim-256` (Audio Diffusion), evaluated on generated log-mel spectrograms.

Temporal DDPM denotes a task-specific Diffuser-style temporal U-Net, trained before steering and then frozen. Panda reaching uses `m1_generator_v1.pt`; Panda pose and obstacle-aware reaching use `m2_generator_v1.pt`. Both HalfCheetah rows use `loco_halfcheetah_generator_v2.pt`. Execution uses MuJoCo (Todorov et al., 2012): Panda position servos or a custom tracking wrapper around Gymnasium's HalfCheetah model (Mandlekar et al., 2021). The character extension uses a frozen text-conditioned MDM (Tevet et al., 2023) (60 frames at 20 fps, 50 DDIM steps, classifier-free guidance 2.5), with 25 deterministic transitions remaining at the intervention, acting on the $263 \times 1 \times 60$ HumanML3D feature representation. We use the HumanML3D MDM checkpoint `humanml_enc_512_50steps/model000750000.pt`, with a 512-dimensional transformer encoder and eight layers.

The custom two-dimensional double-integrator experiments use distinct frozen checkpoints: `di_generator.pt` for one-waypoint control, `ds_generator.pt` for multi-waypoint tracking, and `r1_generator_v5.pt` for the point-mass cosine task. Sharing an architecture label does not imply sharing weights across these tasks.

L.1 Request, Preservation, and Outcome

For the frozen remaining sampler in equation 1, write $z_0 = h(F_{t\to 0}(x_t^{(0)}; c))$ and specify an absolute target $z_\star = z_0 + d$. The nonlinear request loss optimized by the generic gradient comparators is

$$\mathcal{L}_{\text{req}}(x_t) = \tfrac{1}{2}\|h(F_{t\to 0}(x_t; c)) - z_\star\|_2^2. \tag{10}$$

Coordinates use the units and any fixed scaling declared by the task; a pixel residual, a dB error, and a normalized physical-state error are not interchangeable quantities. A zero displacement $d_i = 0$ holds a measured coordinate at its unsteered value. An absolute target such as terminal velocity zero instead specifies $d_i = -z_{0,i}$, unless the anchor is already stationary. Coordinates absent from h receive no preservation constraint.

Spectral steering solves the damped local model in equation 4–equation 5; K denotes the number of local solves at the same diffusion time. Generic SGD, Adam, and L-BFGS optimize the nonlinear loss through the remaining sampler, adding a fixed displacement penalty where the source protocol specifies one (as in Appendix G), with the resource conventions in Appendices F and I. The magnitude-matched FO control isolates direction; the external guidance methods retain the operator-specific protocols and disclosed adaptations of Fig. 5.

Request fidelity measures how well Δh realizes d, for example $E_{\mathrm{req}} = \|\Delta h - d\|_2 / \|d\|_2$ for nonzero d. PSNR on unobserved pixels and physical execution provide separate tests of whether this realization is useful. The catalogue labels fidelity diagnostics separately from independent endpoints; neither an accurate request nor a convincing single example establishes a population-level quality or execution advantage.

L.2 VISION

The inverse-problem objectives compare a generated image after a known measurement operator with the stored observation. Structured edits instead reduce the image to analytic regional statistics. Background or boundary preservation is part of the loss only when those coordinates are explicitly included with zero requested change. The vision gallery mirrors these two groups: restoration strips first, then regional, paired-image, and multi-region edits. Table 8 specifies their measurement maps and endpoints.

Table 8: Vision objective and cost functions. Pixel-space inverse problems use the measurement residual itself; structured editing uses analytic color, spatial, boundary, and texture statistics.

Task / source	Measurement map and target	Evaluation
Random-mask inpainting (V0)	Retained pixels $M \odot y$ in DDPM data units. *Target:* $d = y_{\mathrm{obs}} - h(y_0)$; unobserved pixels are omitted, not preservation coordinates	PSNR on unobserved pixels
$4\times$ super-resolution (V0)	DPS bicubic antialiased resizer $R_{1/4}(y)$. *Target:* $d = y_{\mathrm{obs}} - R_{1/4}(y_0)$	Full-image PSNR
Gaussian deblurring (V0)	Reflection-padded 61×61 Gaussian convolution $G_{\sigma=3}y$. *Target:* $d = y_{\mathrm{obs}} - G_{\sigma=3}y_0$	Full-image PSNR
Two-region restoration (V0)	$[\mathrm{avgpool}_4(y\|_{R_1});\ y\|_{R_2}]$. *Target:* Restore low-resolution R_1; preserve exactly observed R_2	PSNR on R_1
Multi-region color control (V0)	Mean CIELab coordinates in each named region, $Z \in \{12, 27, 48\}$. *Target:* Fixed per-region color displacement; included preservation regions request zero change	Dense ΔE on preserved pixels
Spatial-statistic editing (V1)	$[\mu_x, \mu_y, \sigma_x, \sigma_y]$ under the smooth luminance weight $\widetilde{w}_{ij}$. *Target:* Shift the weighted center and spreads; d is calibrated from frozen pilot statistics	Objective fidelity
Combined color–geometry–texture (V1)	$[\bar{R}_B, \bar{G}_B, \bar{B}_B, \mu_x, \mu_y, E_{\mathrm{HP}}]$, where $E_{\mathrm{HP}} = \|K_{\mathrm{Lap}} * Y\|_2^2 / N$. *Target:* Pilot-scaled displacement in the raw objective coordinates; the apple large-edit illustration uses this objective at $4\times$ scale	Objective fidelity and outside-region change
Paired compositing/appearance family (V2)	Analytic CIELab destination cells, luminance spread, inner/outer boundary rings, and background preservation blocks. *Target:* Reference-derived requested blocks; background entries are explicitly zero	Boundary and appearance diagnostics not optimized as semantic scores
Reference patch transfer (V3)	CIELab statistics over 20 destination edit coordinates and 56 background coordinates. *Target:* Match the supplied source patch in the destination rectangle; request zero change outside it	Dense outside-region ΔE
Two-image joint edit (V3)	Concatenated regional CIELab cells for two jointly steered images, followed by preserve cells from both images. *Target:* Drive the two named regions to agreement; request zero change over the remaining cells	Region disagreement and per-image collateral change

The spatial-moment map uses normalized smooth luminance weights $\widetilde{w}_{ij} \propto \mathrm{softplus}(4(Y_{ij} - \bar{Y})) + 10^{-6}$ on an image grid in $[-1, 1]^2$. In the combined objective, pilot standard deviations scale the requested direction before its normalization in the raw measurement space; they do not imply that every loss coordinate was whitened. The separate benchmark in Fig. 5(e) uses the linear RGB box-average map, at diffusion window 10 and request scale $4\times$ (source V4). It is distinct from the combined color–geometry–texture apple illustration.

L.3 AUDIO

For waveform localized equalization, with STFT X, four temporal segments and six log-frequency bands,

$$h(x)_{q,b} = 10 \log_{10} \left(\sum_{(m,f)\in(q,b)} |X_{m,f}|^2 + 10^{-10} \right). \tag{11}$$

The other audio rows in Table 9 either use the same dB construction on a different partition or use the stated linear measurement operator directly. For a dB-cell task, the target is additive: requesting $+6$ dB means that the target coordinate is the unsteered value plus six. Cells named as "held" are included with a zero displacement, so leakage into them increases the same request loss. The spectrogram gallery shows where those requested and held regions lie in time and frequency.

Table 9: Audio objective and cost functions. All dB requests are literal changes in the stated time–frequency cells. No listening-study or perceptual-preference endpoint is claimed.

Task / source	Measurement map and target	Evaluation
Localized waveform EQ (A0)	equation 11, $4 \times 6 = 24$ dB cells. *Target:* Primary request: $+6$ dB in one cell and -8 dB in another; remaining 22 cells request 0 dB	Requested-cell dB error; LSD and SI-SDR to a reference edit
Disjoint-cell EQ ladder (A0)	Per-cell dB on $Z \in \{2, 4, 8, 16, 32\}$ disjoint log-mel cells. *Target:* Alternating ± 3 dB pattern at the declared scales; included hold cells request 0 change	Decoded-signal cell error and feasibility gate
Bandwidth extension (A1)	$A_{\mathrm{LPF}}(x)$, a 601-tap Kaiser low-pass measurement. *Target:* $d = y_{\mathrm{obs}} - A_{\mathrm{LPF}}(x_0)$	High-band LSD above the cutoff
Temporal inpainting (A1)	Observed samples of a centered binary temporal mask. *Target:* Match the observed samples; the missing interval is absent from h	Gap-region SNR and splice-boundary discontinuity
Known-filter EQ restoration (A1)	24 log-spaced band energies $10 \log_{10}(\sum_{f\in b} \lvert X_f\rvert^2 + 10^{-10})$. *Target:* $d = h(x_\star) - h(x_0)$, restoring the clean band balance	LTAS distance and waveform metrics
Missing-segment boundary match (A0)	Eight boundary-band levels in dB. *Target:* Match the generated segment's beginning/end to the adjacent clips; interior coordinates request 0	Abrupt energy jump across the splice
Crossover between two clips (A3)	Four band levels in each of four temporal quarters, plus four broadband quarter levels ($Z = 20$). *Target:* Ramp each band and broadband level from clip A's ending profile to clip B's opening profile	Bandwise ramp error and splice diagnostics
Context-preserving interior edit (A4)	Four interior band levels; four left/right context levels; two seam-continuity coordinates ($Z = 10$). *Target:* Interior low band -6 dB and high band $+6$ dB; both contexts and both seams request no change	Interior error and leakage into contexts/seams
Crescendo envelope (A2)	$[E_1, \ldots, E_5, E_{\mathrm{total}}]$, five equal time-bin mean-dB levels and global mean dB. *Target:* $8[-2, -1, 0, 1, 2, 0]$ dB; middle and total energy request zero change	Realized temporal envelope; no listening preference

L.4 EXECUTED MANIPULATION AND LOCOMOTION

The steering loss is applied to a generated reference trajectory, not directly to the MuJoCo rollout. The same objective-blind controller then executes every method's reference. This separation is why Table 10 distinguishes requested reference coordinates from executed target error, contact, height drift, or world-frame progress.

Table 10: Manipulation and locomotion objectives. The objective is evaluated on the generated reference; the endpoint is measured after an independent controller executes that reference in MuJoCo.

Task / source	Measurement map and target	Evaluation
Panda Cartesian reach (R0)	Position $p_{\mathrm{EE}}(T)/L_P$; orientation error ρ_T/L_R, $\rho_T = \mathrm{rotvec}(R_T, R_0)$; terminal v_{EE}/L_V and ω_{EE}/L_W. *Target:* Move position; preserve orientation and terminal linear/angular velocity	Executed Cartesian target error
Panda checkpoint/via-point reach (R1)	$[p_{\mathrm{EE}}(t_1), \ldots, p_{\mathrm{EE}}(T)]/L_P$, with $Z = 3, 6$, or 9. *Target:* Independent checkpoint displacements or geometry-defined via-points	Executed closest-approach via error and final error
Reach with orientation hold (R0/R1)	$[p_{\mathrm{EE}}(T)/L_P, \mathrm{rotvec}(R_T, R_0)/L_R]$. *Target:* Move position; request zero orientation change	Executed target error and tool rotation
Lateral move at constant height (R0/R1)	Terminal $(x_{\mathrm{EE}}, y_{\mathrm{EE}})/L_P$ and four checkpoint heights $z_{\mathrm{EE}}(t_i)/L_P$. *Target:* Move laterally; hold start height at four times	Executed target error and maximum height drift
Unmodelled obstacle reach (R0)	$[p_{\mathrm{EE}}(T)/L_P, \mathrm{rotvec}\,/L_R, c(\tau)/L_C]$ with soft-minimum whole-arm clearance. *Target:* Request goal and clearance; preserve orientation	Contact-free execution and MuJoCo geom distance
Spatial locomotion target (R0/R1)	Midpoint/final x, terminal v_x, torso height, and pitch, each with its frozen physical scale. *Target:* Request intermediate/final position and speed; preserve height and pitch	Executed world-position error
Locomotion trajectory tracking (R0/R1)	Four $x(t_i)$ checkpoints, terminal v_x, torso height, and pitch, each frozen-scaled. *Target:* Request a smooth world-position profile; preserve posture	Executed trajectory RMSE
Progress and speed control (R0)	Progress $\Delta x/L_X$, terminal speed v_x/L_V, height z/L_Z, and pitch$/L_{\mathrm{pitch}}$. *Target:* Request signed progress/speed change; preserve height and pitch	Executed progress or signed speed change

L.5 Control and planning

The catalogue first defines character motion generation requests, then the canonical control and trajectory-planning objectives. Both combine requested motion with explicitly represented preservation or terminal-state requirements.

L.5.1 Character motion generation

Character steering differentiates joint coordinates recovered by forward kinematics through the remaining motion sampler. Obstacle avoidance specifies a mid-clip pelvis shift around a 20 cm sphere, preserving final pelvis location and seven arm-bone lengths ($Z = 13$). Reaching specifies five wrist waypoints, preserved ankle and hip positions at two additional frames, zero wrist arrival velocity, and bone lengths at waypoint frames ($Z = 77$). Requested displacements use the unsteered subject frame and natural excursion scales. The reported conjunctive success criterion requires positional target accuracy, preservation of joint positions, and jerk below a population-calibrated threshold. Velocity, bone drift, and geometric collision are separate diagnostics, even when the corresponding quantities enter the steering objective. The full campaign definitions and reported thresholds are in Appendix E.

L.5.2 Canonical control and trajectory planning

Table 11 lists the canonical control and planning objectives. Here h is evaluated on a generated state trajectory. Endpoint terms encode the requested final state; checkpoint terms encode the route; and velocity, posture, effort, and clearance entries specify either absolute targets or changes relative to the anchor. Collision and execution diagnostics are evaluated according to each task protocol, so satisfying a sparse endpoint request alone is not reported as obstacle-avoidance success.

Table 11: Control and planning objective and cost functions. State coordinates are normalized by the frozen physical scales named in the source protocols before entering the Euclidean request cost.

Task / source	Measurement map and target	Evaluation
Pendulum stabilization (C0)	$[\sin(\theta_T - \theta^\star),\ 1 - \cos(\theta_T - \theta^\star),\ \dot{\theta}_T,\ E_u]$. *Target:* Target $[0, 0, 0, E_{u,0}]$; control effort is preserved	Real-dynamics replay: angle, speed, return, and effort
Pendulum reference tracking (C0)	Wrapped angle errors at fixed intermediate checkpoints and T, plus $\dot{\theta}_T$. *Target:* Reach the declared reference angles and finish stationary	Real-dynamics replay and tracking error
CartPole target position (C0)	$[x_T, \theta_T, \dot{x}_T, \dot{\theta}_T]$ divided by frozen training-distribution scales. *Target:* Target $[x^\star, 0, 0, 0]$; upright orientation and terminal velocities are explicit zeros	Open-loop true-dynamics survival and terminal errors
Double-integrator waypoints (P0)	$[p_{t_1}, \ldots, p_{t_K}, p_T, v_T]$, $K \in \{1, 2, 4, 8\}$. *Target:* Hit all requested positions; finish at $v_T = 0$	Executed goal/waypoint error and task success
Maze2D goal and via-points (P0)	$[p_{t_1}, \ldots, p_{t_K}, p_T, v_T]$ with A* reference waypoints. *Target:* Reach requested free-space waypoints and goal; finish stationary	Executed goal distance, wall collision, path length
Dense obstacle navigation (P0)	$[p_T, v_T, \mathrm{viol}_{1:H}, s]$, $\mathrm{viol}_t = \sum_j [r_j + r_{\mathrm{safe}} - \|p_t - o_j\|]_+$. *Target:* Request the new goal and zero velocity/clearance violations; preserve smoothness s	Executed collision, clearance, goal error
Planar-arm pose planning (P0)	$[p_{\mathrm{EE,T}}, \cos\phi_T, \sin\phi_T, \dot{p}_{\mathrm{EE,T}}]$; dense body-clearance entries are added for the obstacle task. *Target:* Move the end effector; preserve orientation, terminal velocity, and clearance	Executed pose error and body collision