Final Flexformer / Encoder–Decoder / GP–KRR / PDE Framework Report

MMNN certified benchmark synthesis

July 2, 2026

1 Status

Item	Status	Evidence
Frozen decoder certificates	verified	GP violations $0.0e+00, 0.0e+00$; PDE $0.0e+00$
Finite-rank GP/KRR recast	verified	primal–dual gap $1.443e-14$
PDE encoder–decoder recast	verified	z MSE $6.136e-08$, A rel $6.076e-05$
Scaling sweep	verified	96 rows, max violation $0.0e+00$
Spectra-only closure	disproved	depth-1 risks $1.5e-33$ vs $1.1e-01$
Joint overlap logging	verified	108 rows, reconstruction error $1.1e-12$
Exact nonlinear finite-set dynamics	verified	∇_θ error $1.0e-15$; ∇_P error $1.7e-16$
Flexformer local NTK curves	verified locally	decoder $5.6e-03$; GP/KRR $6.8e-04$
Flexformer global outer-training theory	open	needs overlap/joint-law closure

2 Exact Frozen Optimization

$$x_{\ell+1} = x_\ell + P_\theta(d - Hx_\ell), \tag{1}$$

$$x^\star = H^{-1}d, \qquad e_\ell = x_\ell - x^\star, \tag{2}$$

$$\boxed{e_L} = \boxed{(I - P_\theta H)^L e_0}, \tag{3}$$

$$\|e_L\|_2 \le \|I - P_\theta H\|_2^L \|e_0\|_2. \tag{4}$$

This is exact for every frozen checkpoint and every preconditioner class tested.

3 Generalization Certificate

$$R_c \le \widehat{R}_c + \sqrt{\frac{2\widehat{V}_c \log(2/\delta)}{n}} + \frac{7c\log(2/\delta)}{3(n-1)}. \tag{5}$$

4 Finite-Rank GP/KRR

$$z \sim \mathcal{N}(0, \Lambda), \qquad f(x) = \phi(x)^\top z, \tag{6}$$

$$A = \Lambda^{-1} + \sigma^{-2}\Phi^\top\Phi, \qquad d = \sigma^{-2}\Phi^\top y, \tag{7}$$

$$z_{\text{post}} = A^{-1}d, \tag{8}$$

$$k(x,x') = \phi(x)^\top \Lambda \phi(x'), \qquad \phi_q^\top z_{\text{post}} = k_q^\top (K + \sigma^2 I)^{-1} y. \tag{9}$$

Best final GP method: `free_linear`, risk $3.317e-07$.

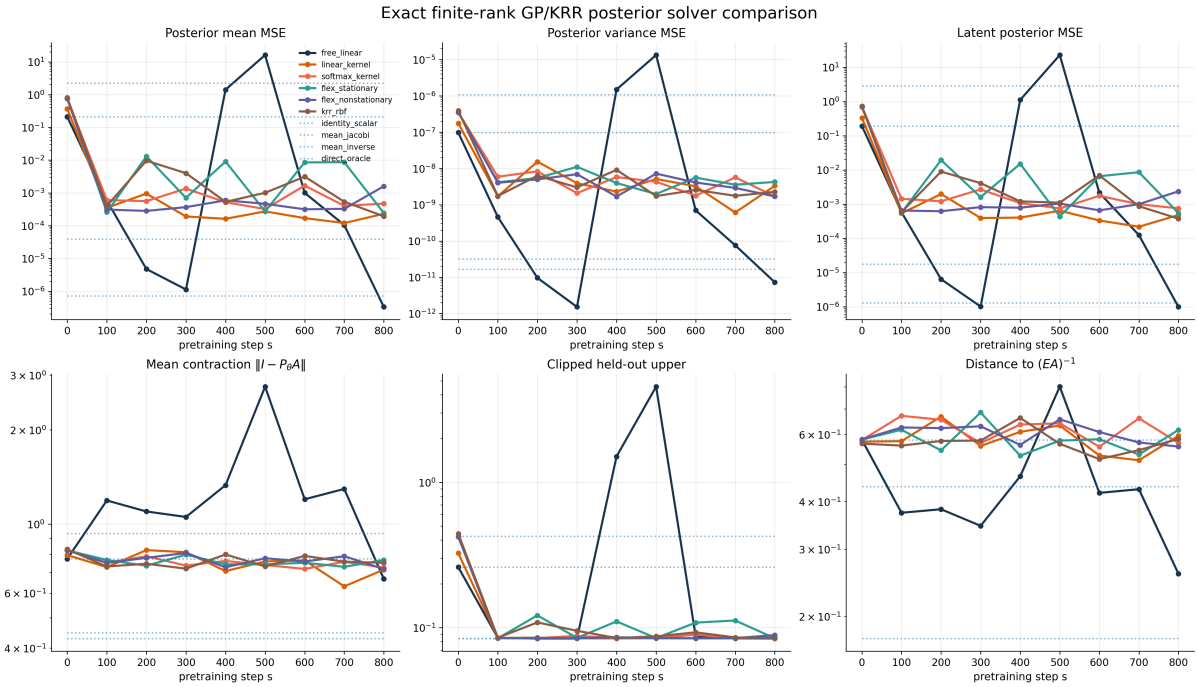

Figure 1: GP/KRR posterior framework comparison.

5 PDE Encoder–Decoder

$$A(z) = A_0 + \sum_{k=1}^{K} z_k A_k, \qquad A(z)u_i = f_i, \tag{10}$$

$$G_{i,r,k} = \langle A_k u_i, v_r \rangle, \qquad b_{i,r} = \langle f_i - A_0 u_i, v_r \rangle, \tag{11}$$

$$z_{\text{enc}} = (G^\top G + \lambda_z I)^{-1} G^\top b, \tag{12}$$

$$u_{\ell+1} = u_\ell + P_\theta(f_\star - A(z_{\text{enc}})u_\ell). \tag{13}$$

Best final PDE method: `free_linear`, MSE $5.976e-06$.

6 Scaling

GP prompt sizes: `[16, 64]`. PDE prompt sizes: `[4, 16]`.

7 Spectral Closure Obstruction

$$R_L(P) = K^{-1} \operatorname{Tr}\{(I - PH)^L S (I - PH)^{L\top}\}. \tag{14}$$

The marginal spectrum of P is insufficient. Same eigenvalues and singular values can give different risks and training derivatives because the eigenbasis overlap with H, S differs. Eigenvalue mismatch $0.0e+00$, singular-value mismatch $4.4e-16$. Depth-1 risks $1.541e-33$ versus $1.090e-01$.

$$\operatorname{Tr}(PH), \quad \operatorname{Tr}(PHPH), \quad \operatorname{Tr}(PHSH^\top P^\top) \tag{15}$$

are examples of required overlap order parameters.

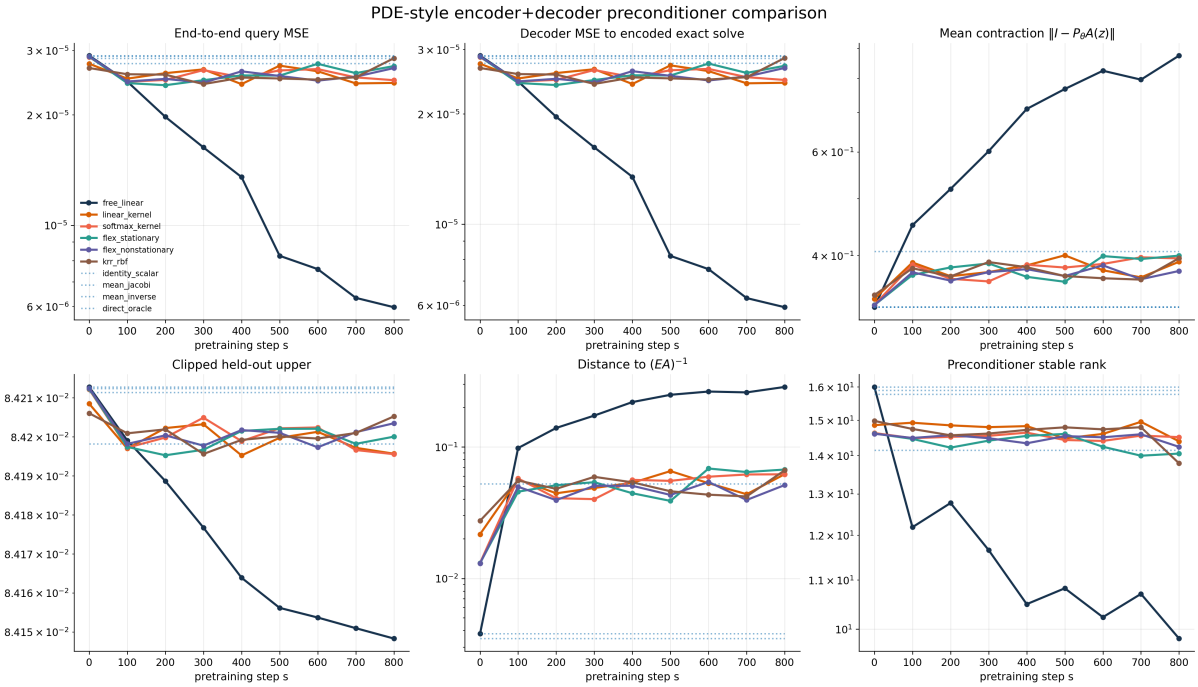

Figure 2: PDE encoder–decoder framework comparison.

8 Joint Order-Parameter Dynamics

For actual GP and PDE checkpoint trajectories, the logged overlap state verifies

$$R_L(P) = K^{-1}\mathbb{E}\|(I - PH)^L x^\star\|^2. \tag{16}$$

Rows: 108. Maximum reconstruction absolute error: $1.137e-12$. Maximum bound violation rate: $0.0e+00$.

$$\mathrm{Tr}(P\bar{H})/K, \quad \mathrm{Tr}(P\bar{H}P\bar{H})/K, \quad \mathrm{Tr}(P\bar{H}\bar{S}\bar{H}^\top P^\top)/K. \tag{17}$$

9 Flexformer Training Dynamics

$$A_\theta(i,j) = \frac{k_\theta(q_i, k_j)}{\sum_a k_\theta(q_i, k_a)}, \qquad P_\theta = D_a A_\theta D_v + D_s, \tag{18}$$

$$\mathcal{L}(\theta) = \mathbb{E}\|x_L(P_\theta) - x^\star\|^2. \tag{19}$$

The frozen certificate is exact. The global nonlinear outer-training dynamics are not closed by the quadratic free-P theory. For a fixed finite training set, the exact nonlinear parameter

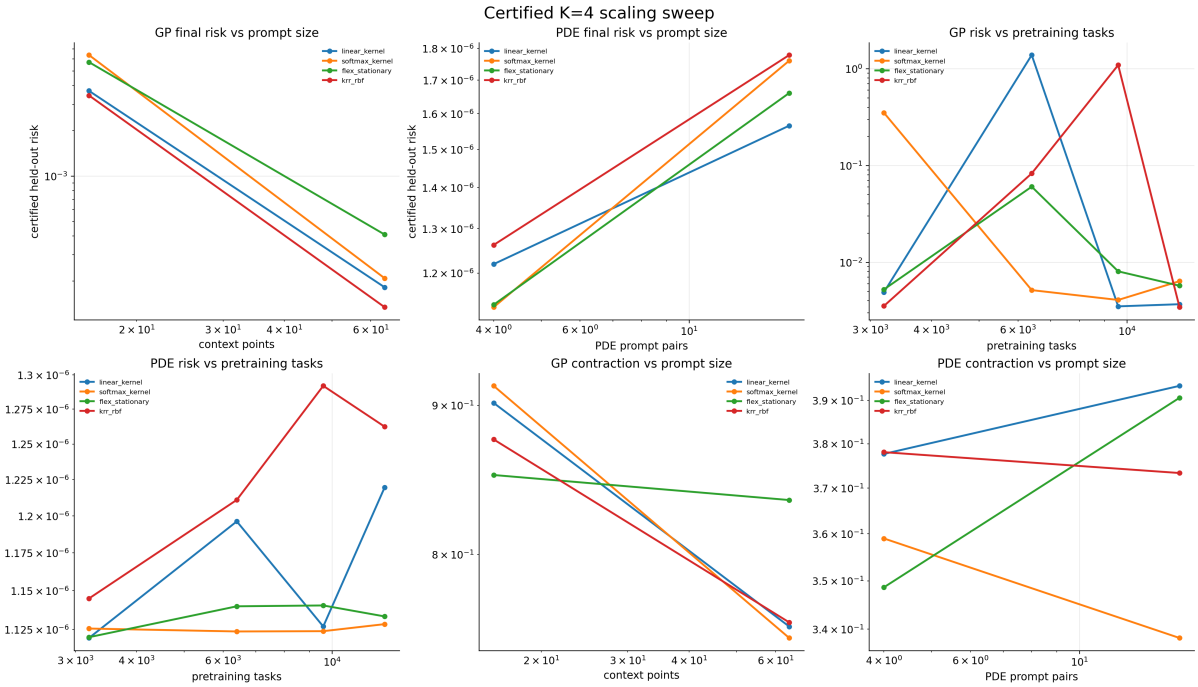

Figure 3: Certified K=4 scaling sweep.

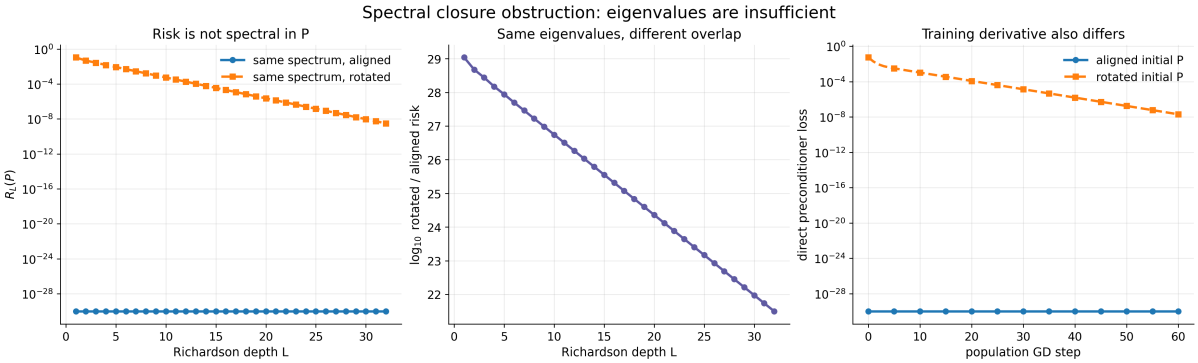

Figure 4: Spectra-only closure obstruction.

recursion is

$$z_{i,0} = 0, \qquad r_{i,\ell} = d_i - H_i z_{i,\ell}, \qquad z_{i,\ell+1} = z_{i,\ell} + P_\theta r_{i,\ell}, \tag{20}$$

$$\mathcal{L}_N(\theta) = \frac{1}{2NK} \sum_{i=1}^{N} \|z_{i,L}(P_\theta) - z_i^\star\|_2^2, \tag{21}$$

$$\lambda_{i,L} = \frac{z_{i,L} - z_i^\star}{NK}, \qquad \lambda_{i,\ell} = (I - P_\theta H_i)^\top \lambda_{i,\ell+1}, \tag{22}$$

$$\nabla_P \mathcal{L}_N(P_\theta) = \sum_{i=1}^{N} \sum_{\ell=0}^{L-1} \lambda_{i,\ell+1} r_{i,\ell}^\top, \tag{23}$$

$$\theta_{s+1} = \theta_s - \eta\, DP^*_{\theta_s}[\nabla_P \mathcal{L}_N(P_{\theta_s})]\,. \tag{24}$$

Exact-gradient validation: dense-P relative error $1.727e-16$, parameter-gradient relative error $1.048e-15$.

$$r_s^{\mathrm{NTK}} = \left(I - \eta J_0 J_0^\top/m\right)^s r_0, \qquad \mathcal{L}_s^{\mathrm{NTK}} = \frac{1}{2m}\|r_s^{\mathrm{NTK}}\|_2^2. \tag{25}$$

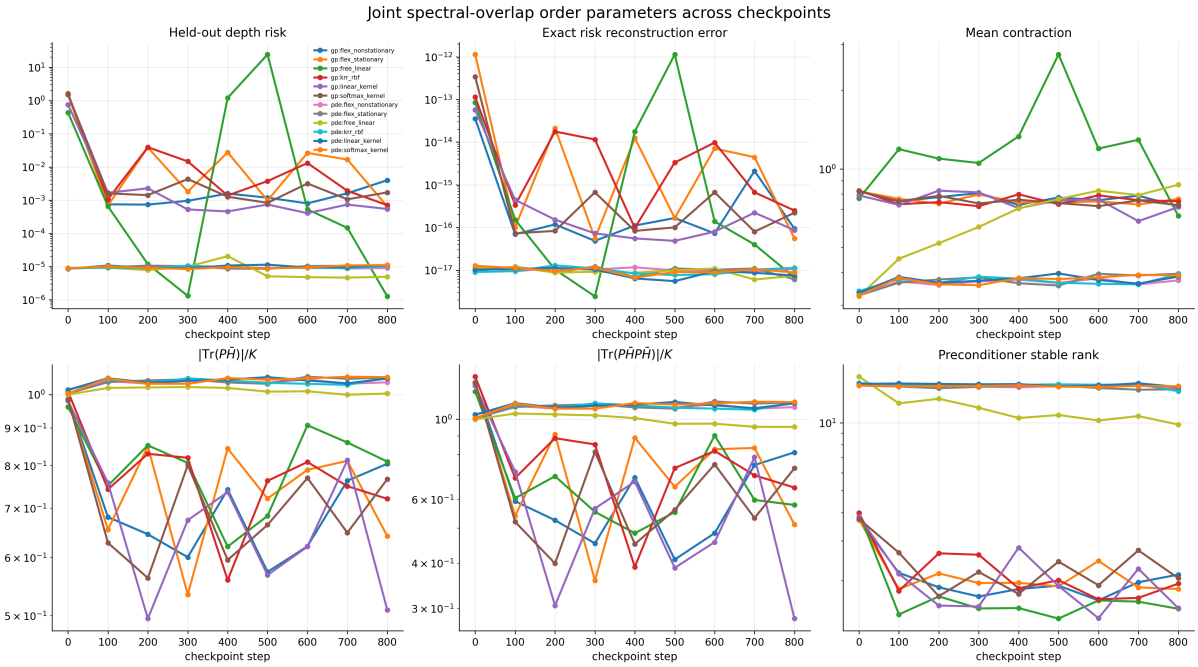

Figure 5: Joint spectral-overlap order parameters along checkpoint trajectories.

Positive-control NTK relative error: $5.572e-03$. Longer local run: $1.906e-02$. For finite-rank GP/KRR posterior training, the checked local model is

$$f_\theta = (z_L(P_\theta; A_i, d_i))_{i=1}^N, \qquad y = (z_{\mathrm{post},i})_{i=1}^N, \tag{26}$$

$$r_s^{\mathrm{NTK}} = \left(I - \eta J_0 J_0^\top / m\right)^s r_0. \tag{27}$$

GP/KRR Flexformer NTK relative error: $6.829e-04$. Local-diagnostic held-out posterior mean MSE: $7.309e+00$.

10 Audit Conclusion

Achieved: exact frozen certificates, finite-rank GP/KRR recast, PDE encoder–decoder recast, K=4 CPU comparisons, scaling plots, checkpointed runs, GPU launchers. Not achieved: exact global replica/mean-field closed form for nonlinear Flexformer outer training. Spectra-only closure is insufficient; an exact closure must include overlap/joint-law order parameters.

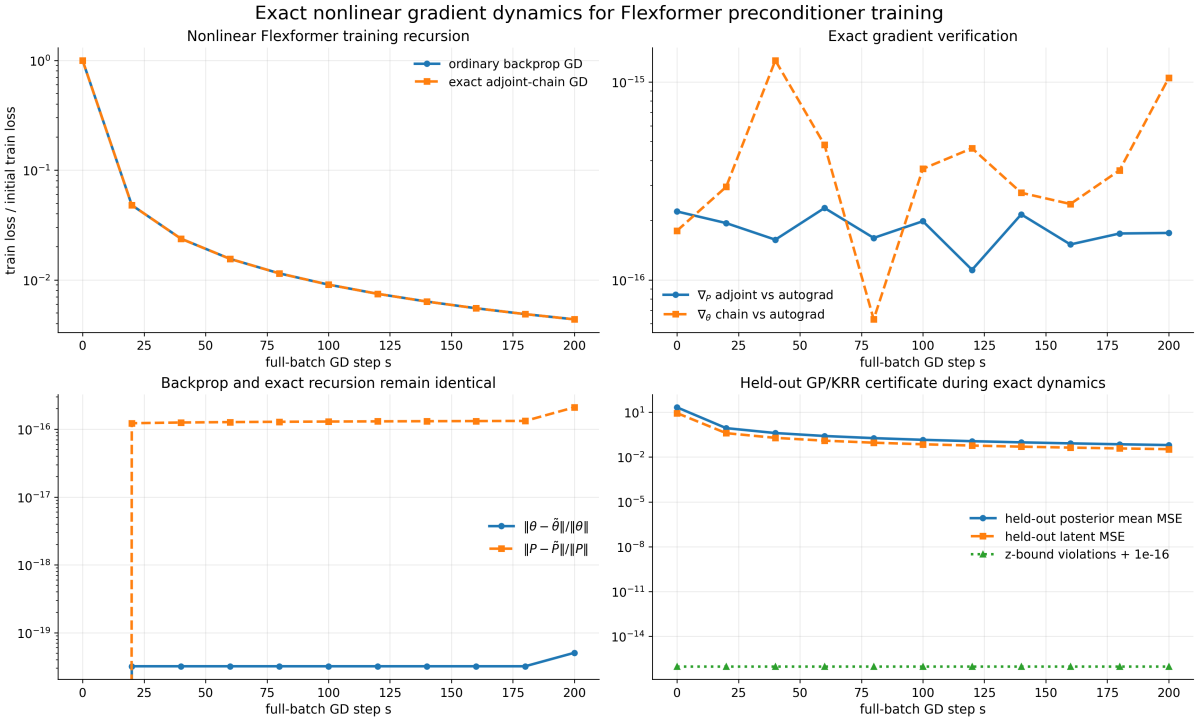

Figure 6: Exact nonlinear finite-set Flexformer gradient dynamics.

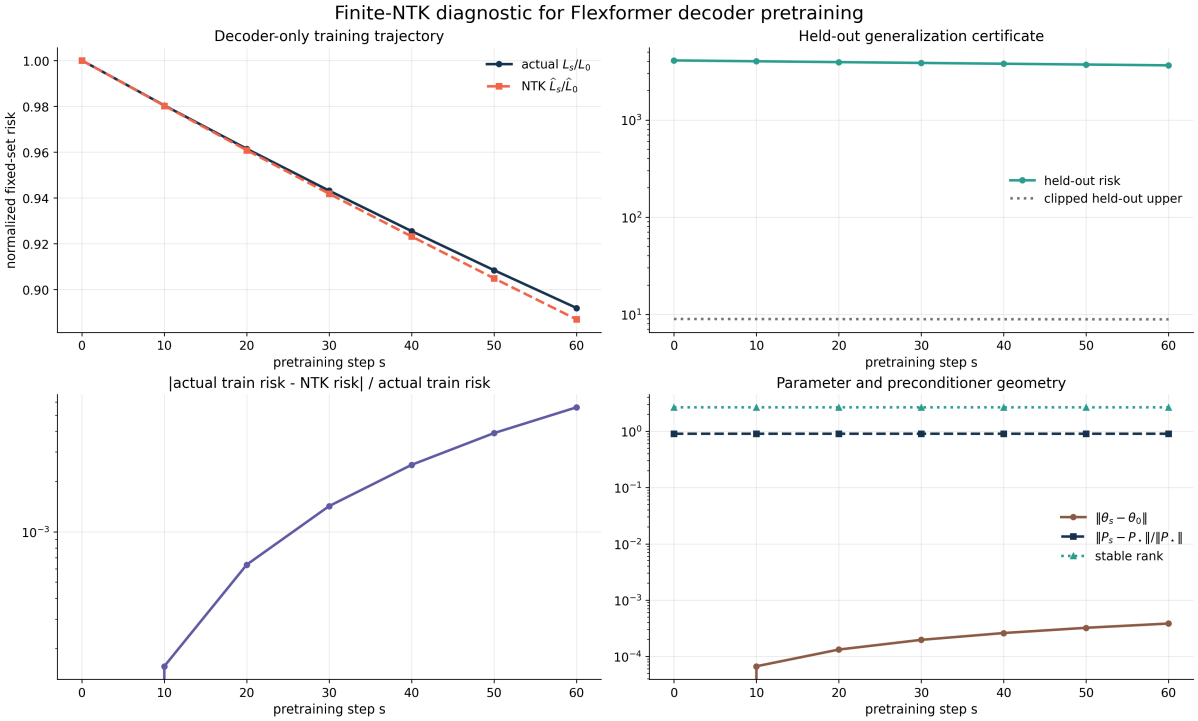

Figure 7: Finite-NTK positive-control curve.

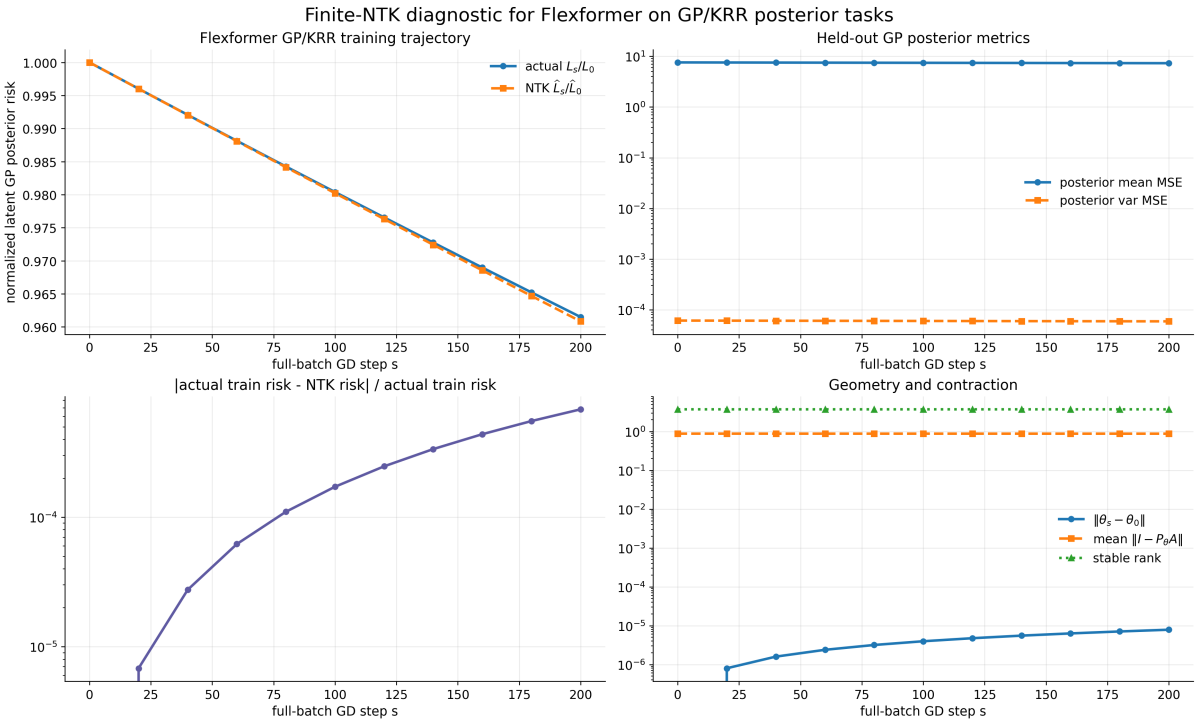

Figure 8: Finite-NTK curve for Flexformer preconditioner training on the GP/KRR posterior latent solve.