KazuNet: Invariant Learning for Factor Graphs

Supplementary Material

Yueze Wang
University of California Santa Barbara
yueze@ucsb.edu

This supplement contains proofs and operator definitions (Sec. A), architecture sweeps, validation comparisons, and implementation provenance (Sec. B), complete 19-solver five-step benchmark comparisons and controlled ablations (Sec. C), overhead, memory accounting, and numerical equivalence telemetry (Sec. D), and detailed calibration and feature specifications (Sec. E). Component and controlled ablations with the parameter-free Quotient-LM control appear in Secs. B and C.

A. Mathematical Proofs & Numerical Verification

A.1. Theorem 1: $\mathcal{G}_{\text{scale}}^{\text{struct}}$-Invariance & Upper-Tail Breakdown

Proof. Let $\mathcal{K}_{\text{over}} = \{k : s_{\text{redundancy},k} > 0.05\}$ and $\mathcal{K}_{\text{min}} = \{k : s_{\text{redundancy},k} \le 0.05\}$, and consider $(F_k, g_k, H_k) \mapsto \omega_k(F_k, g_k, H_k)$ ($\omega_k > 0$ arbitrary for $k \in \mathcal{K}_{\text{over}}$; common $\omega_k \equiv \omega_{\text{min}} > 0$ for $k \in \mathcal{K}_{\text{min}}$, which equals full $\mathcal{G}_{\text{scale}} = \mathbb{R}_{>0}^K$ when $\mathcal{K}_{\text{min}} = \emptyset$). Under rank-preserving chart and scale inputs, the equilibrated positive eigenspace U_k is unchanged while $H_k^+ \mapsto \omega_k^{-1} H_k^+$, giving $e_k^{\parallel} \mapsto \omega_k e_k^{\parallel}$ and $e_k^{\perp} \mapsto \omega_k e_k^{\perp}$. Hence F_{stat} and $s_{\text{redundancy},k}$ are degree-0 homogeneous in ω_k, so on $\mathcal{K}_{\text{over}}$ the James–Stein precision scales identically as $\pi_k^{\text{over}} \mapsto \omega_k^{-1} \pi_k^{\text{over}}$. On $\mathcal{K}_{\text{min}}$, pooling all low-redundancy factors (measurements and role$_k$ $= 1$ priors) gives $\pi_0^{\text{min}} = [Q_{0.35}(\{2\omega_{\text{min}} F_j/m_j : j \in \mathcal{K}_{\text{min}}\}) + 10^{-30}]^{-1} = (1/\omega_{\text{min}})\pi_0^{\text{min,clean}}$. Passing $\pi_{0,k}$ into Passes 1–2 yields $\mathcal{G}_{\text{scale}}^{\text{struct}}$-invariant whitened curvature $\check{B}_p$, leverage $\ell_{p,k}$, and unaligned energy ratios $\hat{E}_{p,k}^{\text{raw}}$ (since $1/\omega_{\text{min}}$ cancels ω_{min}). When mixed pools are both $\ge 10\%$, the cross-pool quantile ratio $c_{\text{min}} = Q_{0.35}(\mathcal{K}_{\text{over}})/(Q_{0.35}(\mathcal{K}_{\text{min}}) + 10^{-30})$ is likewise degree-0 homogeneous in ω_k and ω_{min}, leaving $\hat{E}_{p,k}$, $\hat{E}_{\text{med}}$, u_k^{CH}, and $(\check{F}_k, \check{g}_k, \check{H}_k) = \pi_k(F_k, g_k, H_k)$ strictly invariant under $\mathcal{G}_{\text{scale}}^{\text{struct}}$.

For upper-tail contaminated immunity, let $\mathcal{O} \subset \{1, \dots, K\}$ satisfy $|\mathcal{O}| < (1-q)K$, with inliers $k \notin \mathcal{O}$ bounded by $\hat{E}_k \le B_{\text{in}} < \infty$. Since at least $K - |\mathcal{O}| > qK$ inliers satisfy $\hat{E}_k \le B_{\text{in}}$, $Q_q(\{\hat{E}_k\}) \le B_{\text{in}}$ regardless of $F_k \to +\infty$ ($k \in \mathcal{O}$); $q = 0.30$ ($\hat{E}_{\text{med}}$) and $q = 0.35$ (π_0^{min}) confer $\approx 70\%$ and $\approx 65\%$ upper-tail breakdown. *Lower-tail boundary:* setting $\ge 31\%$ of positive active factors to 0 pulls $Q_{0.30} \to 0$, guarded numerically by positive-energy masking ($> 10^{-24}$ max) and $\varepsilon = 10^{-30}$ clamping. □

A.2. Quotient Whitening and Numerical Stability: Observable Quotient Pullback

Proof. Under $\theta(\phi) = \theta_0 + T\phi$ ($T \in \text{GL}(n)$), $B_\phi = T^\top B_\theta T$. Suppose $W_\theta \in \mathbb{R}^{n\times r}$ and $W_\phi \in \mathbb{R}^{n\times r}$ satisfy $W_\theta^\top B_\theta W_\theta = I_r$ and $W_\phi^\top (T^\top B_\theta T) W_\phi = I_r$. Decomposing $TW_\phi = W_\theta Q + N_\perp C$ into range(B_θ) $=$ span(W_θ) and ker(B_θ) $=$ span$(N_\perp)$ ($B_\theta N_\perp = 0$), substitution yields $Q^\top (W_\theta^\top B_\theta W_\theta) Q = Q^\top Q = I_r$, so $Q \in O(r)$. Every Weyl–Procesi scalar invariant and Gram kernel evaluated on whitened contractions $\bar{g}_{k,\phi} = (TW_\phi)^\top \check{g}_{k,\theta} = Q^\top \bar{g}_{k,\theta}$ and $\bar{S}_{k,\phi} = Q^\top \bar{S}_{k,\theta} Q$ is orthogonal-invariant under $Q \in O(r)$, guaranteeing $\xi_\phi = Q^\top \xi_\theta$. Pulling back to state space gives:

$$\begin{aligned} T\Delta\hat{\phi} &= TW_\phi \xi_\phi \\ &= (W_\theta Q + N_\perp C)(Q^\top \xi_\theta) \\ &= \Delta\hat{\theta} + N_\perp (C Q^\top \xi_\theta). \end{aligned}$$

When $r = n$, ker$(B_\theta) = \{0\}$ so $N_\perp = 0$ and exact primal equivariance $T\Delta\hat{\phi} = \Delta\hat{\theta}$ holds. When $r < n$, $B_\theta(T\Delta\hat{\phi} - \Delta\hat{\theta}) = B_\theta N_\perp (CQ^\top \xi_\theta) = 0$: observable quotient equivalence holds identically even though the nullspace coordinate difference $\|T\Delta\hat{\phi} - \Delta\hat{\theta}\|_2 = \|\delta_\perp\|_2$ can be nonzero (verified in our 2D rank-deficient control: observable error $= 6.28\text{e-}16$ vs. nullspace norm $= 0.7071$). The square-root stabilization avoids squaring $\kappa_2(L_{\text{stack}})$ by running chunked TSQR [1] on column-equilibrated L_{stack} followed by exact SVD, achieving $6.54\text{e-}11$ / $4.86\text{e-}11$ equivariance at $\kappa_2(T) = 10^6$. □

A.3. Risk, Descent, and Numerical Verification

In the whitened orthogonal eigenbasis, Tikhonov shrinkage $\hat{x}_i(\lambda_i) = \frac{\sigma_i^2}{\sigma_i^2 + \lambda_i}(x_{\star,i} + \sigma_i^{-1} z_i)$ decouples total MSE risk into the spectral-risk equation in the main paper; differentiating with respect to $\lambda_i \ge 0$ gives zero at $\lambda_i^\star = \sigma^2 / x_{\star,i}^2$. Over unrestricted $\lambda_i \ge 0$, letting $\lambda_i \to \infty$ collapses $\hat{x} \to 0$ with $\mathcal{R}(\lambda) \to \|x_\star\|_2^2$; hence higher isotropic risk floors apply only under non-collapse progress constraints.

For dual-merit descent, $\bar{M}^{(t)} \succ 0$ on $\mathcal{V}_{\text{obs}}^{(t)}$ and oblique dual-cone projection certify $\langle g_{\text{cal}}^{(t)}, \Delta\theta_t \rangle < 0$ and $\langle g_w^{(t)}, \Delta\theta_t \rangle < 0$. The line search tests both frozen merit functions at candidate steps, and $(w_k^{(t)}, \pi_k^{(t)})$ are refreshed at the next outer iteration. For **Proposition 1**, under $T \in \text{Diag}_{>0}(n)$, equilibration transforms as $D_\phi =$

$T^{-1}D_\theta$, so $B_{\text{eq},\phi} = D_\phi T^\top \check{B}_\theta T D_\phi = B_{\text{eq},\theta}$ and truncated mode ratios $(\sigma_i/\sigma_{\max})^2$ are $\text{Diag}_{>0}(n)$-invariant. Under non-diagonal 45° rotation $R \in O(2) \setminus \text{Diag}_{>0}(2)$ on $0.5\omega(0.05 + (a-1)^2 + 10^{-5}(b - 1 + 1000b^2)^2)$, equilibration $D_\phi \neq R^\top D_\theta R$ shifts $(\sigma_i/\sigma_{\max})^2$ across discrete thresholds $\tau \in \{10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}\}$, moving the physical update from $(0.2941, 0.2941)$ to $(0.9999, 0.0)$ (relative difference 1.838), whereas disabling spectral truncation (Strict Mode) restores $O(2)$ equivariance to 7.84e-12 $< 10^{-11}$ (Table S1).

For **Full-Graph Foster Cut-Bridge Calibration**, let $\mathcal{K}_{\min} = \{k : s_{\text{redundancy},k} \leq 0.05, \text{role}_k = 0\}$ and $g_{\min,k} = \mathbb{I}[\text{role}_k = 0]\,\text{clip}(1 - s_{\text{redundancy},k}/0.05, 0, 1)$. Evaluating full-graph $(\mathcal{G}_{\text{scale}}^{\text{struct}})^{-1}$-equivariant probe precisions π_k^{probe} (preserving overdetermined $s_{\text{redundancy},k} > 0.05$ factors and $\text{role}_k = 1$ priors) gives whitened leverage $\ell_{\text{full},k} = \text{tr}(W_{\text{full}}^\top(\pi_k^{\text{probe}} H_k)W_{\text{full}})$ and $\tau_{\text{cut},k} = \text{clip}(\ell_{\text{full},k}/\max(q_k, 1), 0, 1) \cdot g_{\min,k}$. Since $\tau_{\text{cut},k} > 0.85$, dead-reckoning quantile ratio $\rho_{\text{dead-reckoning}} > 10^5$, the Förstner non-interpolating mask $(m_k - \ell_{m,k}) > 0.05 m_k$ (10^{-12} tolerance), smooth C^1 bridge floors, and Stage 4 per-case calibrated energy growth guard $\Phi_{\text{base}}(\theta + \Delta\theta) \leq 6.0\,\Phi_{\text{base}}(\theta)$ rely purely on quotient-invariant scalars and projections, bridge calibration preserves $\text{GL}(n) \times \mathcal{G}_{\text{scale}}^{\text{struct}} \times S_K$ invariance (relative drift on calibrated energy $\check{F}_k$: 2.64e-13, preliminary precision $\pi_{0,k}$: 3.00e-11, KazuNet-Zero / KazuNet sqrt steps: 4.01e-12 / 5.23e-12, dense step: 2.30e-10) with zero false positives on cycle-redundant rings.

A.4. Notation and Operator Definitions

The state dimension is n, the number of active factors is K, and the observable rank is r_{obs}. For factor k, F_k is its energy, g_k its gradient, H_k its positive-semidefinite curvature, m_k its residual dimension, q_k its effective rank, and $\nu_k = \max(m_k - q_k, 0)$ its excess residual dimension. The redundancy score $s_{\text{redundancy},k}$ determines the calibration pool; π_k is the calibration precision and w_k the learned factor weight. A check denotes calibration, $(\check{F}_k, \check{g}_k, \check{H}_k) = \pi_k(F_k, g_k, H_k)$; a bar denotes quotient coordinates, $\bar{g}_k = W^\top \check{g}_k$ and $\bar{S}_k = W^\top \check{H}_k W$. For any scalar u, $[u]_+ = \max(u, 0)$.

Conflict curvature. Let $\bar{d}_0 = -\bar{M}_0^{-1}\bar{g}_w$ and $\xi_k = \bar{g}_k + \bar{S}_k\bar{d}_0$. With all quantities evaluated at the current step,

$$
\begin{aligned}
E_{\text{post},k} &= [2\check{F}_k + 2\bar{g}_k^\top \bar{d}_0 + \bar{d}_0^\top \bar{S}_k \bar{d}_0]_+, \\
\sigma^2_{\text{base},k} &= \begin{cases} \pi_k e_k^\perp / \max(\nu_k, 1), & s_{\text{redundancy},k} > 0, \\ \max(\hat{E}_{\text{med}}, 1), & s_{\text{redundancy},k} = 0, \end{cases} \\
\sigma^2_{\text{noise},k} &= \sigma^2_{\text{base},k} + 0.02(1 + \ell_k), \\
a_k &= \frac{\gamma_k w_k}{\sigma^2_{\text{noise},k} + 10^{-4} E_{\text{post},k} + 10^{-12}}, \\
\widetilde{\Omega} &= \frac{\sum_k a_k \xi_k \xi_k^\top}{1 + 2\mu_{\text{bulk}}}, \qquad \tau_{\text{cap}} = 1.5 + 248.5 \frac{\bar{s}}{\bar{s} + 0.05}, \\
\Omega_\xi &= \tau_{\text{cap}}(\tau_{\text{cap}} I + \widetilde{\Omega})^{-1}\widetilde{\Omega}.
\end{aligned}
$$

Here $\bar{s}$ is the mean redundancy score over active factors and μ_{bulk} is the median eigenvalue of the unnormalized conflict numerator $\sum_k a_k \xi_k \xi_k^\top$ on the observable space (averaging the two middle eigenvalues when r_{obs} is even).

Secant curvature. With the current weights held fixed, define $h = W^\top \check{B} \Delta\theta_{t-1}$ and $z = W^\top \sum_k w_k \pi_k (g_{k,t} - g_{k,t-1})$, with

$$
\begin{aligned}
\chi_{\text{sec}} &= \mathbb{I}\{h^\top z > 10^{-10}\|h\|_2\|z\|_2 + 10^{-24}\}, \\
\Omega_{\text{sec}} &= \frac{\chi_{\text{sec}}\eta_{\text{sec}}}{[h^\top z]_+ + \|z\|_2^2 + 10^{-12}} zz^\top.
\end{aligned}
$$

The network predicts $\eta_{\text{sec}} \in (0, 2)$; the zero-parameter variant fixes $\eta_{\text{sec}} = 0.25$. Without prior step history, $\Omega_{\text{sec}} = 0$.

Oblique projection. For $d = -\bar{M}^{-1}\bar{g}_w$ and $g = \bar{g}_{\text{cal}}$, set $a^2 = g^\top \bar{M}^{-1} g$, $b^2 = -\bar{g}_w^\top d$, $c = -g^\top d/(ab)$, and $\eta = \min\{0.02, (1 - c^2)/2\}$. The oblique projection $\mathcal{P}_{\text{cone}}(d) = d - \frac{[g^\top d + \eta ab]_+}{a^2}\bar{M}^{-1}g$ is skipped (returning d) if $g^\top d \leq -\eta ab$ or if a^2 or b^2 falls below machine numerical rank (10^{-50}), and reverts to the steepest calibrated direction $-(b/a)\bar{M}^{-1}g$ in the nearly collinear limit ($1 - c^2 \leq 10^{-24}$), with safe $O(10^{-30})$ denominator floors throughout.

B. Architecture Sweep, Validation Baselines, & Provenance

Tables S3 and S4 report the internal component/capacity ablation on held-out validation ($N = 256, T = J = 5$) and the full 12-row baseline comparison alongside 32-case $\mathcal{G}_{\text{scale}}^{\text{struct}}$ drift checks. Tables S5 and S6 document the 13-variant architecture/curriculum sweep and upstream external implementations.

Table S1. **Numerical Verification of Equivariance and Derivatives (NVIDIA A100-SXM4-40GB, canonical `gesvd`).**

Numerical verification item	Measured discrepancy (KazuNet-Zero / KazuNet / Oracle)
One-step equivariance ($\mathrm{GL}(n) \times \mathcal{G}^{\mathrm{struct}}_{\mathrm{scale}}$)	Overdetermined: 3.643e-12 / 3.618e-12; Mixed: 1.079e-11 / 1.079e-11
Robustness and duality (Breakdown, Dual, Rank, AD)	Upper-tail bound, convex-duality Lagrangian, Schur identity, surrogate $= 1.00$
Dense whitening (Dense eigendecomp., $\kappa_2(T) = 10^5$)	1.475×10^{-9} relative equivariance error
Trajectory equivariance (5-step trajectory)	Overdetermined: 1.373e-13 / 7.593e-13; Mixed: 1.357e-11 / 8.793e-12
Square-root whitening (TSQR+SVD, $\kappa_2(T) = 10^6$)	KazuNet-Zero: 4.941e-11; KazuNet: 4.857e-11
Gradient consistency (Finite-difference checks)	Raw-BAL/g2o: $\leq$ 6.75e-9; HPatches: $\leq$ 1.90e-8; 3D-Scans: 5.95e-8; Wild-BAL: 5.95e-9; MAT-POWER/PS: $\leq$ 2.62e-8; A3: 1.25e-8
Symmetry boundary (Prop. 1 + 13 variants)	Strict: 7.84e-12 ($O(2)$), 2.64e-11 ($\mathrm{GL}(2)$); Practical: $\mathrm{Diag}_{>0}(2)$ 2.98e-16, 45° break 1.838; 13 variants: $13/13 \leq$ 4.65e-12
Bridge calibration (Foster leverage)	Calibrated $\tilde{F}_k$: 2.64e-13; $\pi_{0,k}$: 3.00e-11; Zero/KazuNet sqrt: 4.01e-12 / 5.23e-12; dense: 2.30e-10; zero ring false positives

Table S2. **Stage-1 Equilibrated Rank (2×2 counterexample) & 3-Regime 32-Case Scale Drift Measurement (A100 `gesvd`).** Top: Stage-1 preserves $q_k = 2$ and $e_k^{\|} = 1$ under $\mathrm{diag}(1, 10^{-8})$ in both modes and under rotated $\kappa_2(T) \leq 10^6$ in $J_{\sqrt{k}}$ mode. Bottom: median / max relative step-2 drift across 32 graphs under $\mathcal{G}^{\mathrm{struct}}_{\mathrm{scale}}$ (mixed), $\mathbb{R}^K_{>0}$ (overdet.), and adversarial out-of-group independent minimal scaling $\omega_k \notin \mathcal{G}^{\mathrm{struct}}_{\mathrm{scale}}$ (Quotient-LM included as theoretical ablation control).

Part A: 2×2 Rank Check	$\mathrm{diag}(1, 10^{-6})$	$\mathrm{diag}(1, 10^{-8})$	Rot. $\kappa_2 = 10^4$	Rot. $\kappa_2 = 10^6$
$J_{\sqrt{k}}$ row-Gram G_k ($\Delta e_k^{\|}$)	0.00e+00	0.00e+00	2.22e-16	2.22e-16
Dense $H_{\mathrm{eq},k}$ ($\Delta e_k^{\|}$)	0.00e+00	0.00e+00	1.39e-09	2.11e-05
Rank q_k ($J_{\sqrt{k}}$ / dense)	$2 \to 2$ / $2 \to 2$	$2 \to 2$ / $2 \to 2$	$2 \to 2$ / $2 \to 2$	$2 \to 2$ / $2 \to 2$
Part B: 32-Case Scale Drift (med / max)	LM	Q-LM (Ctrl)	KazuNet (Strict)	KazuNet (Pract.)
$\mathcal{G}^{\mathrm{struct}}_{\mathrm{scale}}$ (Mixed)	5.96e-1 / 47.4	**6.76e-13** / 2.38e-9	**1.19e-11** / 5.15e-6	1.20e-10 / 5.68e-2
$\mathbb{R}^K_{>0}$ (Overdet.)	4.53e-1 / 80.9	**6.26e-13** / 1.18e-10	**4.88e-12** / 3.29e-7	1.15e-11 / 6.24e-2
Adv. Indep. ($\omega_k \notin \mathcal{G}^{\mathrm{struct}}_{\mathrm{scale}}$)	6.26e-1 / 45.6	2.84e-1 / 95.4	5.83e-1 / 86.9	5.54e-1 / 86.9

C. Complete Benchmark Comparisons and Controlled Ablations ($T = J = 5$ and $T = J = 20$)

Table S7 reports DeepLM, completing the 19-method five-step comparison. Tables S8–S13 provide the remaining 18-method benchmark evaluations at $T = J = 5$ and $T = J = 20$ (Tables S8–S12), the expanded mean/median summary (Table S10), and the Strict vs. Practical mode ablation (Table S13). Tables S14–S15 isolate all 19 solvers (including Quotient-LM strictly as an ablation control) across the three controlled ablations A1–A3. Results use double precision (`float64`), canonical `gesvd`, and an NVIDIA A100 GPU.

DeepLM implementation and results. We use the author DeepLM solver [2] (`hjwdzh/DeepLM`, commit `a5536cb5`) with the original benchmark residuals, analytic Jacobians, initial states, and evaluation metrics, in double precision on an A100 GPU. The adapter preserves native PCG, block preconditioning, column normalization, damping, and acceptance; SE(3) uses the benchmark's right retraction in a recentered local chart. The solver source is unchanged, with only CUDA/PyTorch build-compatibility edits. The five-outer-attempt budget uses `maxNumIterations=6` because the author loop exits before its equality iteration, `maxSuccessIterations=5`, and zero parameter/function stopping tolerances. Native gradient/radius stopping ends 11 HPatches cases after four attempts; all other cases complete five. No learned weights or training are used. Table S7 gives all 1,371 outcomes; the existing LM and KazuNet references use their frozen five-step results.

D. Zero/Low Network Overhead, Oracle Budgets (T, J) & Equivalence Measurements

D.1. Numerical Equivalence and Hardware Measurements

Table S16 summarizes trainable neural overhead (parameter counts, `float64` weight footprints, and per-step neural MACs at $K = 64$), exact Jacobian oracle accounting ($J = T$, $\Delta J = 0$ in backtracking), and complete Ablation A3 sparse-clique curvature memory scaling. Table S17 and Figure S1 report NVIDIA L4 three-backend SVD numerical comparisons on UW-BAL and Wild-BAL alongside NVIDIA A100-SXM4-40GB trajectory equivalence and GPU power telemetry across 7,532 five-step runs per configuration: C1 serves as the canonical trajectory reference, C2 increases batch size $4\times$, C3 decouples worker processes, C4 enables NVIDIA CUDA MPS on isolated workers (preserving $0.00\text{e+}00 < 10^{-10}$ exact equivalence), and C5 also batches external backbones $4\times$. Two-band approximate SVD causes a $+16.94\%$ regression on rank-borderline Wild-BAL trajectories.

Table S3. **Component and capacity ablations on held-out validation ($N = 256$, $T = J = 5$).** Paired W/T/L counts and one-sided Wilcoxon tests compare each learned model with Quotient-LM (ablation-only analytical control). Scale drift reports median / max $\mathcal{G}^{\text{struct}}_{\text{scale}}$ step-2 equivariance.

Model	Parameters	Mean	Median	W/T/L (p vs. Q-LM)	Scale drift ($\mathcal{G}^{\text{struct}}_{\text{scale}}$)
Quotient-LM (ablation control)	0	0.4192	0.0547	—	6.76e-13
KazuNet-Zero	0	0.3811	0.1867	—	1.71e-11
KazuNet	45,294	**0.2065**	0.0432	164/0/92 (1.13e-12)	1.20e-10
KazuNet-M	146,742	0.2111	0.0433	167/0/89 (7.10e-14)	3.57e-12 / 1.24e-11
KazuNet-L	239,078	0.2141	**0.0423**	162/0/94 (8.00e-13)	2.37e-11 / 2.93e-11

Table S4. **Held-out validation baseline comparison ($N = 256$, $T = J = 5$) and 32-case scale drift.** Paired W/T/L counts and one-sided Wilcoxon tests compare KazuNet against each baseline [9].

Solver	Params	Mean	Median	KazuNet W/T/L (p)	$\mathcal{G}^{\text{struct}}_{\text{scale}}$ Drift (32 cases)
KazuNet (Strict)	**45,294**	**0.2050**	0.0432	— (Overdet: 4.88e-12)	**1.19e-11**
KazuNet (Practical)	**45,294**	0.2065	0.0432	— (Overdet: 1.15e-11)	1.20e-10
KazuNet-M / KazuNet-L	146.7K / 239.1K	0.2111 / 0.2141	0.0433 / **0.0423**	—	3.57e-12 / 2.37e-11
KazuNet-Zero	0	0.3811	0.1867	199/0/57 (2.10e-24)	1.71e-11
CN-IRLS [6]	44,611	0.5039	0.3372	232/0/24 (9.27e-32)	uncal. (3.22× ind.)
PointNet-LM (in-repo)	45,389	0.5263	0.3770	231/0/25 (6.88e-33)	uncal. (2.14× ind.)
FGNN [12]	98,613	0.7880	0.8527	241/0/15 (7.92e-39)	uncal. (1.40× ind.)
Euclidean-Net (in-repo)	44,749	0.7997	0.8977	246/0/10 (1.21e-39)	uncal. (1.32× ind.)
LFGC-DFE [10]	405,717	0.8459	0.8897	249/0/7 (2.13e-40)	uncal. (1.40× ind.)
OANet [11]	55,477	0.8629	0.8762	246/0/10 (3.30e-40)	uncal. (1.33× ind.)
CLNet [13]	154,262	0.8848	0.8943	247/0/9 (1.95e-40)	uncal. (1.19× ind.)
NCMNet [3]	350,758	1.5419	1.1992	252/0/4 (1.07e-43)	uncal. (1.49× ind.)
LM (Ceres-diag) [4]	0	1.9922	1.5388	253/0/3 (1.57e-43)	5.96e-01

E. Calibration, Bridge, and Architecture Details

E.1. Rank estimation details

To prevent coordinate-scaling rank flips under raw eigenthresholding (*e.g.*, $H = I_2, T = \text{diag}(1, 10^{-6}) \implies T^\top H T = \text{diag}(1, 10^{-12})$, dropping q_k from 2 to 1 if raw $\lambda_{k,i} > 10^{-11}\lambda_{k,\max}$ were used directly), Stage 1 evaluates effective rank q_k and tangent energy in an **equilibrated space**: when $J_{\sqrt{k}}$ is supplied, it decomposes the $\text{GL}(n)$-invariant row-Gram $G_k = (J_{\sqrt{k}}W_0)(J_{\sqrt{k}}W_0)^\top$ ($\sigma_{k,i} > 10^{-12}\sigma_{k,\max}$, where W_0 spans normalized total curvature); otherwise it eigendecomposes Jacobi-equilibrated $H_{\text{eq},k} = D_k H_k D_k$ ($\lambda_{\text{eq},k,i} > 10^{-11}\lambda_{\text{eq},k,\max}$, $D_k = \text{diag}((H_{k,ii})^{-1/2})$). On CUDA 2×2 controls, this preserves $q_k = 2 \to 2$ and $e_k^{\parallel} = 1.000000$ to 0.00e+00–2.22e-16 for $\text{diag}(1, 10^{-8})$ in both modes and across rotated non-diagonal T up to $\kappa_2(T) = 10^6$ in $J_{\sqrt{k}}$ mode (2.11e-05 in dense $H_{\text{eq},k}$).

E.2. Calibration coefficients and bridge treatment

A smooth $\text{GL}(n) \times \mathcal{G}^{\text{struct}}_{\text{scale}}$-invariant redundancy score $s_{\text{redundancy},k} = \psi_{\text{redundancy},k}\frac{\max(\nu_k-1,0)}{\max(\nu_k-1,0)+q_k+4} \in [0, 1)$ routes overdetermined measurements ($s_{\text{redundancy},k} > 0.05$) through effective DOF $\nu_{\text{effective}} = \min(\max(\nu_k, 1), 4q_k + 2\sqrt{\max(\nu_k, 1)})$ and James–Stein shrinkage $w_{\text{shrink}} = (F_{\text{stat}} + 0.65)/(F_{\text{stat}} + 1)$ to yield $\pi_k^{\text{over}} = w_{\text{shrink}}\nu_{\text{effective}}/(e_k^{\perp} + 0.02(2F_k) + 10^{-30})$ (or $m_k/(2F_k + 10^{-30})$ for priors), while all low-redundancy factors ($s_{\text{redundancy},k} \le 0.05$, $\omega_k \equiv \omega_{\min}$) are pooled at $\pi_0^{\min} = [Q_{0.35}(\{2F_j/m_j : s_{\text{redundancy},j} \le 0.05\}) + 10^{-30}]^{-1}$.

To protect cut-edges without altering cycle-redundant or overdetermined graphs, Stage 1 runs full-graph probe whitening yielding Foster leverages $\tau_{\text{cut},k} = \text{clip}(\ell_{\text{full},k}/\max(q_k, 1), 0, 1)g_{\min,k}$ ($g_{\min,k} = \mathbb{I}[\text{role}_k = 0]\text{clip}(1 - s_{\text{redundancy},k}/0.05, 0, 1)$) and dead-reckoning ratio $\rho_{\text{dead-reckoning}} = \max_{\mathcal{K}_{\min}}(2F_k/m_k)/(Q_{0.35}(\mathcal{K}_{\min}) + 10^{-30})$; whenever $\max_k \tau_{\text{cut},k} > 0.85$ or $\rho_{\text{dead-reckoning}} > 10^5$ (indicating a bridge factor), one-step minimal-pool W_m projection with Förstner mask $(m_k - \ell_{m,k}) > 0.05m_k$ and C^1 Hermite floors anchor bridge factors cleanly.

E.3. Network features and coefficient ranges

In the r_{obs}-dimensional quotient space (main Fig. 2), KazuNet (45,294 `float64` parameters, 353.9 KB: three-layer Gram-coupled Transformer, $d = 32$, 4 heads, $d_{\text{ff}} = 72$, alongside zero-shot capacity variants KazuNet-M at 147K and KazuNet-L at 239K) consumes 10 Weyl–Procesi $\text{GL}(n) \times \mathcal{G}^{\text{struct}}_{\text{scale}}$-invariant node scalars [5, 7, 8]: normalized leverage ℓ_k/q_k, leverage concentration $\|\bar{S}_k\|_F^2/\ell_k^2$, relative redundancy ν_k/m_k, redundancy score $s_{\text{redundancy},k}$, calibrated log-energy $\log(1 + \check{F}_k/\tilde{F})$, explained ratio $\text{clip}(\pi_k e_k^{\parallel}/(2\check{F}_k), 0, 1)$, whitened gradient norm $\log(1 + \|\bar{g}_k\|_2/\tilde{g})$, global force alignment $\cos(\bar{g}_k, \bar{g}_{\text{tot}})$, factor role

Table S5. **Architecture and curriculum variants (13 KazuNet candidates).** Held Synth is the sweep synthetic validation error. Geom vs. KazuNet reports the normalized error ratio across the seven sweep suites (UW-BAL, 3D-Scans, Wild-BAL, A1, A2a, A2b, A3); values below one favor the variant. W/L counts suites with lower/higher error, and the remaining columns verify coordinate-scale and factor-permutation invariance.

Rank	Model ID	Family	Params	Held Synth	GL+Scale Step-2	Perm. S_K	Geom vs. KazuNet (W/L)
1	KazuNet-L	Cap.+Recur.	239,078	0.2171	1.043e-11	3.86e-14	**0.7475** (5/2)
2	KazuNet-M	Architecture	146,742	0.2050	1.069e-11	4.01e-14	**0.7932** (5/2)
3	KazuNet-Wide	Capacity	214,118	0.2256	1.070e-11	4.02e-14	0.8285 (4/3)
4	KazuNet-Deep	Capacity	151,270	0.2072	1.065e-11	4.00e-14	0.8387 (6/1)
5	KazuNet-48	Capacity	94,694	0.2243	1.104e-11	4.28e-14	0.8484 (5/2)
6	KazuNet-Tiny	Capacity	18,934	0.2295	1.088e-11	4.05e-14	0.8740 (4/3)
7	KazuNet-DeepSets	Architecture	85,286	0.2139	1.104e-11	4.08e-14	0.8841 (3/4)
8	KazuNet-Features	Input/Curr.	45,838	0.2084	1.088e-11	4.13e-14	0.9528 (4/3)
9	KazuNet-Curriculum	Input/Curr.	45,838	0.2150	1.091e-11	4.13e-14	0.9635 (2/5)
10	**KazuNet**	Core Shared	45,294	0.2221	1.922e-12	8.33e-15	1.0000 (—)
11	KazuNet-PolyGNN	Architecture	131,942	**0.2028**	1.080e-11	3.88e-14	1.1613 (2/5)
12	KazuNet-Recurrent	Recurrence	52,174	0.2177	1.075e-11	3.98e-14	1.1779 (2/5)
13	KazuNet-Retrain	Input/Curr.	45,294	0.2130	1.073e-11	4.08e-14	1.2168 (4/3)

Table S6. **Model implementations and upstream baseline provenance.** All learned solvers share the exact 15,360-graph training protocol.

Solver	Architecture / Upstream Repository	Upstream Commit	Params	Train Graphs	License / Origin
KazuNet	In-repo (3-layer quotient Transformer)	—	45,294	15,360	Anonymous submission
Euclidean-Net	In-repo (unwhitened quotient baseline)	—	44,749	15,360	Anonymous submission
PointNet-LM	In-repo (distinct from DeepLM [2])	—	45,389	15,360	Anonymous submission
CN-IRLS	In-repo adapted [6]	—	44,611	15,360	Anonymous submission
OANet [11]	`zjhthu/OANet`	`5d509bd8`	55,477	15,360	MIT / EPFL
FGNN [12]	`zzhang1987/Factor-Graph-NN`	`1860e3d6`	98,613	15,360	MIT
CLNet [13]	`sailor-z/CLNet`	`a15329a8`	154,262	15,360	GPL-3.0
NCMNet [3]	`xinliu29/NCMNet`	`c22db5fc`	350,758	15,360	Upstream research
LFGC-DFE [10]	`isl-org/dfe`	`7cb8514d`	405,717	15,360	MIT

Table S7. **Author DeepLM under a five-step budget.** Native physical error; lower is better.

Benchmark (unit)	N	Mean	Median
Raw-BAL (px)	140	0.7383	0.6339
Raw-g2o (m)	160	0.2218	0.0443
HPatches (px)	231	15.9018	1.5269
UW-BAL (px)	150	6.7660	0.9925
Wild-BAL (px)	150	33.3635	29.4269
3D-Scans (mm)	180	29.2858	25.1049
Photo-Stereo (deg)	180	34.3275	31.8036
MATPOWER (pu)	180	0.1820	0.1617

$2\,\mathrm{role}_k - 1$, and tangent conflict $\log(1 + \hat{E}_k/\hat{E}_{\mathrm{med}})$, alongside force/curvature Grams $A_{\mathrm{force}}(j,k)$, $A_{\mathrm{curv}}(j,k) = \mathrm{tr}(\bar{S}_j\bar{S}_k)/(\|\bar{S}_j\|_F\|\bar{S}_k\|_F)$ and an 8-D spectrum summary. Its cost is 4.35M neural MACs per step at $K = 64$, scaling as $O(K^2 d)$ *independent of DOF* M *and state dimension* n. A consensus shrinkage head contracts log-weights toward $w_k \equiv 1$ on clean graphs while predicting $w_k \in [10^{-4}, 10^2]$, conflict gains $\gamma_k \in (0, 5)$, isotropic mix $\rho \in (0, 0.40)$, $\lambda_{\mathrm{rel}} \in [10^{-6}, 10]$, and secant gain $\eta_{\mathrm{sec}} \in (0, 2)$ (parameter-free KazuNet-Zero fixes $w_k \equiv 1$, $\gamma_k \equiv 0.5$, $\rho \equiv 0.02$, $\lambda_{\mathrm{rel}} \equiv 10^{-4}$, and $\eta_{\mathrm{sec}} \equiv 0.25$).

Table S8. **Mean errors for the other 18 methods on the eight benchmarks and subgroups ($T = J = 5$, canonical `gesvd`).** Native units. Part 1 of 2: classical solvers and in-repo learned baselines.

Benchmark / Subgroup	LM	Dog.	IRLS	BAR	GNC	HVCE	CN	PNet	Euc.
Raw-BAL (px)	**0.773**	1.059	0.903	0.983	0.883	1.064	0.934	0.863	0.844
Raw-g2o (m)	**0.205**	1.488	1.447	1.940	1.293	4.634	2.120	2.120	2.212
SE(3) (m)	0.267	0.258	0.225	0.385	**0.197**	3.907	0.464	0.390	0.330
SE(2) (m)	**0.144**	2.719	2.670	3.495	2.389	5.361	3.776	3.849	4.093
HPat-SIFT (px)	16.46	14.19	5.614	7.691	**5.398**	7.493	7.366	6.910	7.741
HP 6F (px)	27.61	24.11	8.302	10.19	8.015	9.840	9.982	8.783	10.38
HP 3F (px)	5.411	4.350	2.949	5.210	**2.803**	5.167	4.772	5.053	5.127
HP Illum	12.30	8.399	4.566	7.841	**4.320**	7.822	7.079	6.981	7.821
HP View	20.52	19.82	6.635	7.544	**6.449**	7.173	7.646	6.842	7.663
UW-BAL (px)	2.893	**2.198**	2.386	2.911	2.469	5.546	3.471	3.171	3.848
Wild-BAL (px)	29.46	76.91	18.70	27.50	20.17	56.32	8.184	3.562	14.16
3D-Scans (mm)	31.20	26.06	1.555	1.847	1.758	1.873	2.844	2.490	9.187
Photo-St. (deg)	31.91	24.13	10.77	10.03	17.47	12.59	**3.434**	3.615	20.58
MATPOWER (pu)	0.164	0.162	0.054	0.036	0.126	**0.021**	0.032	0.035	0.130
IEEE-14	0.144	0.143	0.046	0.026	0.112	0.012	0.021	0.018	0.111
IEEE-30	0.181	0.178	0.071	0.049	0.144	**0.031**	0.048	0.056	0.145
IEEE-57	0.168	0.164	0.046	0.033	0.121	**0.019**	0.028	0.030	0.132

Table S9. **Mean errors for the other 18 methods on the eight benchmarks and subgroups ($T = J = 5$, canonical `gesvd`).** Native units. Part 2 of 2: external backbones and KazuNet variants (bold indicates lowest mean across these 18 methods).

Benchmark / Subgroup	OANet	FGNN	CLNet	NCM	DFE	Zero	**KazuNet**	**KazuNet-M**	**KazuNet-L**
Raw-BAL (px)	0.970	0.932	0.955	1.805	0.966	0.860	0.827	0.831	0.820
Raw-g2o (m)	1.274	1.799	1.528	2.272	1.699	0.305	0.495	0.488	0.490
SE(3) (m)	0.372	0.377	0.489	1.323	0.332	0.285	0.352	0.669	0.452
SE(2) (m)	2.176	3.222	2.567	3.222	3.065	0.325	0.638	0.307	0.528
HPat-SIFT (px)	5.694	6.486	6.131	13.06	5.831	6.534	6.492	6.556	6.506
HP 6F (px)	8.569	8.985	9.309	22.52	8.821	8.102	**7.944**	8.053	8.023
HP 3F (px)	2.844	4.008	2.980	3.675	2.867	4.980	5.053	5.071	5.003
HP Illum	4.539	6.042	4.709	7.649	4.656	6.355	6.387	6.484	6.381
HP View	6.820	6.918	7.516	18.33	6.976	6.708	6.594	6.626	6.628
UW-BAL (px)	3.035	16.82	4.256	9.051	5.333	16.51	3.414	3.297	2.564
Wild-BAL (px)	39.57	17.60	29.89	46.51	34.57	220.4	7.749	1.962	**1.632**
3D-Scans (mm)	9.489	5.834	5.608	28.71	11.75	1.489	0.856	**0.812**	0.827
Photo-St. (deg)	14.05	13.91	24.33	31.71	20.20	3.935	3.832	5.297	3.479
MATPOWER (pu)	0.085	0.076	0.095	0.172	0.088	0.033	0.026	0.026	0.026
IEEE-14	0.075	0.069	0.093	0.157	0.085	0.017	0.010	**0.009**	0.010
IEEE-30	0.105	0.096	0.110	0.178	0.107	0.050	0.046	0.046	0.044
IEEE-57	0.074	0.063	0.081	0.180	0.074	0.033	0.022	0.023	0.024

Table S10. **Expanded eight-task benchmark summary at $T = J = 5$ (mean / median).** Native units across all formal benchmarks and key subgroups. Bold marks the lowest mean error in each row.

Benchmark / Subgroup (N, Unit)	Init	Best Class.	PointNet-LM	CN-IRLS	OANet	FGNN	LFGC-DFE	KazuNet-Zero	KazuNet	KazuNet-L / -M
Raw-BAL (140, px)	3.19	**LM** (0.773)	0.863 / 0.808	0.934 / 0.895	0.970 / 0.928	0.932 / 0.916	0.966 / 0.938	0.860 / 0.726	0.827 / 0.718	0.820 / 0.831
Raw-g2o (160, m)	4.89	**LM** (0.205)	2.120 / 0.543	2.120 / 0.501	1.274 / 0.162	1.799 / 0.414	1.699 / 0.175	0.305 / 0.056	0.495 / 0.135	0.490 / 0.488
Raw-g2o SE(3) (80, m)	4.21	**GNC** (0.197)	0.390 / 0.184	0.464 / 0.252	0.372 / 0.165	0.377 / 0.199	0.332 / 0.167	0.285 / 0.125	0.352 / 0.155	0.452 / 0.669
Raw-g2o SE(2) (80, m)	5.58	**LM** (0.144)	3.849 / 2.016	3.776 / 1.894	2.176 / 0.124	3.222 / 1.481	3.065 / 0.258	0.325 / 0.017	0.638 / 0.045	0.528 / 0.307
HPatches-SIFT (231, px)	8.09	**GNC** (5.40)	6.910 / 1.160	7.366 / 1.228	5.694 / 1.197	6.486 / 1.302	5.831 / 1.198	6.534 / 1.176	6.492 / 1.122	6.506 / 6.556
HPatches 6F (115, px)	11.15	GNC (8.02)	8.783 / 1.357	9.982 / 1.528	8.569 / 1.535	8.985 / 1.554	8.821 / 1.435	8.102 / 1.338	**7.944** / 1.322	8.023 / 8.053
HPatches 3F (116, px)	5.06	**GNC** (2.80)	5.053 / 0.987	4.772 / 1.127	2.844 / 1.076	4.008 / 1.106	2.867 / 1.045	4.980 / 1.017	5.053 / 1.003	5.003 / 5.071
UW-BAL (150, px)	130.8	**Dog.** (2.20)	3.171 / 1.629	3.471 / 2.091	3.035 / 2.160	16.82 / 13.73	5.333 / 3.445	16.51 / 6.540	3.414 / 0.917	2.564 / 3.297
Wild-BAL (150, px)	130.8	IRLS (18.7)	3.562 / 2.881	8.184 / 7.087	39.57 / 32.08	17.60 / 16.45	34.57 / 32.34	220.4 / 14.83	7.749 / 5.229	**1.632** / 1.962
3D-Scans (180, mm)	17.46	IRLS (1.55)	2.490 / 1.983	2.844 / 2.185	9.489 / 8.411	5.834 / 5.119	11.75 / 12.02	1.489 / 1.401	0.856 / 0.782	0.827 / **0.812**
Photo-Stereo (180, deg)	28.30	BAR (10.03)	3.615 / 1.364	**3.434** / 1.243	14.05 / 13.78	13.91 / 13.83	20.20 / 20.19	3.935 / 2.198	3.832 / 1.794	3.479 / 5.297
MATPOWER (180, pu)	0.259	**HVCE** (0.0208)	0.0347 / 0.020	0.0324 / 0.018	0.0846 / 0.063	0.0760 / 0.059	0.0882 / 0.064	0.0333 / 0.022	0.0259 / 0.013	0.0260 / 0.0260
IEEE-14 (60, pu)	0.262	HVCE (0.0123)	0.0180 / 0.011	0.0210 / 0.012	0.0750 / 0.050	0.0687 / 0.052	0.0848 / 0.054	0.0167 / 0.007	0.0103 / 0.007	0.0099 / **0.0093**
IEEE-57 (60, pu)	0.264	**HVCE** (0.0188)	0.0302 / 0.024	0.0281 / 0.022	0.0736 / 0.063	0.0629 / 0.055	0.0735 / 0.063	0.0333 / 0.025	0.0219 / 0.015	0.0238 / 0.0226

Table S11. **Anytime convergence at $T = J = 20$ across 18 methods on the 8 Formal Benchmarks (canonical `gesvd`).** Part 1 of 2: classical solvers and in-repo learned baselines.

Benchmark / Subgroup	LM	Dog.	IRLS	BAR	GNC	HVCE	CN	PNet	Euc.
Raw-BAL (px)	**0.740**	0.854	0.846	1.005	0.827	0.916	0.977	0.887	0.848
Raw-g2o (m)	0.196	**0.129**	1.425	1.852	1.179	4.170	1.529	2.036	2.137
Raw-g2o SE(2)	0.135	**0.030**	2.635	3.491	2.124	4.940	2.688	3.721	3.934
HPat-SIFT (px)	19.60	18.94	**5.586**	7.503	5.655	7.185	6.522	6.663	7.311
HP 6F (px)	31.94	30.63	8.176	9.787	7.990	9.293	9.903	8.265	9.632
UW-BAL (px)	2.874	1.963	1.761	2.071	2.020	2.308	2.648	2.541	3.573
Wild-BAL (px)	29.52	27.10	10.29	16.08	20.89	5.079	4.042	2.732	10.58
3D-Scans (mm)	32.55	32.40	1.516	1.672	1.448	1.889	2.889	2.698	7.595
Photo-St. (deg)	30.36	29.04	4.414	4.181	15.01	7.920	3.306	4.459	19.25
MATPOWER (pu)	0.166	0.167	0.048	0.036	0.125	**0.024**	0.034	0.040	0.137
IEEE-14	0.144	0.144	0.040	0.024	0.109	0.010	0.021	0.026	0.116
IEEE-57	0.163	0.166	0.037	0.034	0.119	**0.017**	0.029	0.033	0.141

Table S12. **Anytime convergence at $T = J = 20$ across 18 methods on the 8 Formal Benchmarks (canonical `gesvd`).** Part 2 of 2: external backbones and KazuNet variants.

Benchmark / Subgroup	OANet	FGNN	CLNet	NCM	DFE	Zero	**KazuNet**	**KazuNet-M**	**KazuNet-L**
Raw-BAL (px)	0.930	0.870	0.906	1.452	0.928	0.787	0.792	0.774	0.768
Raw-g2o (m)	0.699	1.333	0.966	1.172	1.120	0.241	0.390	0.287	0.294
Raw-g2o SE(2)	1.059	2.354	1.558	1.980	1.938	0.202	0.440	0.225	0.249
HPat-SIFT (px)	5.866	5.800	6.173	16.23	5.763	6.447	6.445	6.471	6.496
HP 6F (px)	8.746	8.533	9.449	28.05	8.670	7.985	**7.952**	8.045	8.024
UW-BAL (px)	2.390	3.863	2.710	6.067	3.858	9.765	1.304	**0.991**	1.315
Wild-BAL (px)	16.11	14.00	13.68	34.04	17.18	256.8	18.59	**0.331**	3.888
3D-Scans (mm)	5.507	4.365	4.060	30.95	8.995	1.367	0.853	**0.810**	0.822
Photo-St. (deg)	12.60	12.90	23.49	33.98	20.15	**2.656**	3.840	5.208	3.276
MATPOWER (pu)	0.091	0.088	0.110	0.173	0.099	0.034	0.026	0.025	0.027
IEEE-14	0.082	0.082	0.103	0.154	0.095	0.015	0.010	**0.009**	0.010
IEEE-57	0.074	0.070	0.092	0.173	0.081	0.033	0.022	0.020	0.023

Table S13. **Strict and Practical mode ablation ($T = J = 5$).** Each cell reports Strict / Practical mean error for the indicated KazuNet variant across all benchmarks and controlled ablations.

Benchmark / Ablation (N, unit)	Zero (Strict / Pract.)	KazuNet (Strict / Pract.)	KazuNet-M (Strict / Pract.)	KazuNet-L (Strict / Pract.)
Raw-BAL (140, px)	0.8798 / 0.8599	0.8332 / 0.8275	0.8252 / 0.8309	**0.8051** / 0.8197
Raw-g2o (160, m)	0.3492 / **0.3053**	0.5295 / 0.4947	0.4882 / 0.4879	0.5156 / 0.4902
HPat-SIFT (231, px)	6.4974 / 6.5340	**6.3527** / 6.4922	6.5525 / 6.5557	6.4906 / 6.5063
HP 3F (116, px)	4.8668 / 4.9800	**4.7696** / 5.0532	5.0710 / 5.0710	5.0026 / 5.0026
UW-BAL (150, px)	35.28 / 16.51	5.3735 / 3.4142	6.5334 / 3.2966	6.3247 / **2.5643**
Wild-BAL (150, px)	1591 / 220.4	10.422 / 7.7495	3.2759 / 1.9623	4.1633 / **1.6319**
3D-Scans (180, mm)	1.4887 / 1.4887	0.8562 / 0.8562	**0.8116 / 0.8116**	0.8274 / 0.8274
Photo-St. (180, deg)	3.9349 / 3.9349	3.8318 / 3.8318	5.2966 / 5.2966	**3.4793 / 3.4793**
MATPOWER (180, pu)	0.0333 / 0.0333	**0.0259 / 0.0259**	0.0260 / 0.0260	0.0260 / 0.0260
A1 Scale-1e6 (200, m)	**0.0627 / 0.0627**	0.0630 / 0.0630	0.0640 / 0.0640	0.0649 / 0.0649
A2a Gauge-ARAP (120)	1.7802 / 1.7802	1.7203 / 1.7203	1.5653 / 1.5653	**1.4861** / 1.4870
A2b Gauge-BAL (120)	70.183 / 4.8651	64.933 / **4.3657**	64.921 / 4.5889	63.794 / 4.3703
A2c Gauge-Dim (240)	0.4900 / 0.0951	0.4014 / 0.0837	0.3990 / 0.0833	0.3766 / **0.0794**
A3 Sparse (160, mm)	7.7530 / 7.7530	3.5242 / 3.5242	**3.0916 / 3.0916**	3.3743 / 3.3743

Table S14. **Controlled Ablations (A1–A3, $T = J = 5$, canonical `gesvd`) across all 19 solvers.** Part 1 of 2: classical solvers and in-repo learned baselines.

Ablation Study (N, unit)	LM	Dog.	IRLS	BAR	GNC	HVCE	CN	PNet	Euc.
A1 Scale-1e6 (200, m)	1.230	1.173	0.506	0.172	0.845	0.124	0.524	0.497	0.329
A2a Gauge-ARAP (120, mm)	38.78	23.47	10.38	10.96	10.87	9.891	12.24	9.874	14.63
A2b Gauge-BAL (120, px)	6.007	25.80	7.224	8.364	6.803	13.59	4.311	**3.105**	4.490
A2c Gauge-Dim (240)	1.256	0.894	0.357	0.380	0.372	0.381	0.387	0.307	0.462
A3 Sparse-Scale (160, mm)	31.67	22.72	12.90	5.662	21.40	**1.459**	20.25	7.858	22.89

Table S15. **Controlled Ablations (A1–A3, $T = J = 5$, canonical `gesvd`) across all 19 solvers.** Part 2 of 2: external backbones, ablation-only Quotient-LM control, and KazuNet variants (bold indicates lowest mean across all 19 solvers). Note that on $C = 30$ conditioning grids (`ARAP_30`, main Fig. 3), means are Q-LM: 1.693, KazuNet: 1.914, KazuNet-M: 1.672, KazuNet-L: 1.538 mm.

Ablation Study (N, unit)	OANet	FGNN	CLNet	NCM	DFE	Q-LM (Ctrl)	Zero	**KazuNet**	**KazuNet-M**	**KazuNet-L**
A1 Scale-1e6 (200, m)	0.815	0.489	0.634	0.771	0.620	**0.061**	0.063	0.063	0.064	0.065
A2a Gauge-ARAP (120, mm)	13.75	13.50	14.07	30.63	14.73	1.679	1.780	1.720	1.565	**1.487**
A2b Gauge-BAL (120, px)	9.912	9.417	11.46	16.12	9.691	65.75	4.865	4.366	4.589	4.370
A2c Gauge-Dim (240)	0.469	0.462	0.485	1.082	0.503	0.415	0.095	0.084	0.083	**0.079**
A3 Sparse-Scale (160, mm)	23.64	20.11	19.68	30.22	22.39	7.498	7.753	3.524	3.092	3.374

Table S16. **Network parameter/MAC overhead, exact oracle accounting ($J = T, \Delta J = 0$), and Ablation A3 curvature memory scaling.**

Solver Family	Params	Weights (`float64`)	MACs ($K = 64$)	Neural Scaling Law	Extra LS (ΔJ)	Backend
Classical (6 solvers: LM..HVCE)	0	0 KB	0	—	0	solve/eigh
KazuNet-Zero	**0**	**0 KB**	**0**	—	**0**	`gesvd`/PCG
KazuNet (Default Shared Policy)	**45,294**	**353.9 KB**	**4.35M**	$O(K^2d)$ indep. of (n, M)	**0**	`gesvd`/PCG
KazuNet-M / KazuNet-L	146,742 / 239,078	1.12 / 1.82 MB	$O(K^2d)$	$O(K^2d)$ indep. of (n, M)	**0**	`gesvd`/PCG
In-Repo (CN-IRLS, EucNet, PNet)	44.6K–45.4K	~ 350 KB	Euclidean MLP	Coupled to raw scale	0	Cholesky
External (5 GNNs: OANet..DFE)	55.5K–405.7K	0.42–3.10 MB	18-D GNN	Up to $8.9\times$ params	0	IRLS-LM

Ablation A3 Sparse-Clique Curvature Tensor Memory Scaling ($n \leq 2,436$, Batch $B = 20$, `float64`):

Mesh Grid (V, n, K)	**KazuNet ($T = 5$)**	**Zero ($T = 5$)**	**Operator Complexity**	**Peak Sparse VRAM (Δ)**	**Sparse H_{clique}**	**Dense H_k**
10×10 ($V = 100, n = 300, K = 32$)	**3.5284 mm**	7.1097 mm	$O(Kc_{\max}^3)$ vs. $O(n^3)$	505.8 MB	**11.25 MB**	439.45 MB ($39.1\times$)
14×15 ($V = 210, n = 630, K = 84$)	**1.3705 mm**	4.8358 mm	$O(Kc_{\max}^3)$ + PCG	652.4 MB	**29.53 MB**	5,087.22 MB ($172.3\times$)
20×20 ($V = 400, n = 1200, K = 72$)	**4.4453 mm**	9.4659 mm	Sparse Woodbury + PCG	976.9 MB	**61.80 MB**	15,820.31 MB ($256.0\times$)
28×29 ($V = 812, n = 2436, K = 98$)	**4.7525 mm**	9.6005 mm	Zero $n \times n$ alloc.	2,474.1 MB	**174.42 MB**	**88,736.18 MB** ($508.8\times$)

Table S17. **NVIDIA L4 (72 W TDP) SVD backend comparison ($T = J = 5$) and NVIDIA A100-SXM4-40GB (400 W TDP, 61.8 W idle) C1–C5 trajectory equivalence and power telemetry (7,532 trajectories/config).**

Platform / Suite	Worker Config / SVD Backend	Mean / Δ%	Mean Power (+Dyn, %TDP)	P90 / Max Power	Equiv. Error ($< 10^{-10}$)
L4 (UW-BAL, KazuNet)	Canonical `gesvd`	3.4164 px (Ref)	41.4 W (+14.1, 57.4%)	51.1 / 61.4 W (7,632 MiB)	Reference
L4 (UW-BAL, KazuNet)	Exact TSQR + CPU `gesdd`	3.4314 (+0.44%)	47.1 W (+19.8, 65.4%)	57.3 / 65.4 W (7,750 MiB)	Above tol. (4.38e-3)
L4 (UW-BAL, KazuNet)	Two-band approx. SVD	3.3828 (-0.98%)	47.5 W (+20.2, 65.9%)	56.8 / 65.6 W (7,812 MiB)	Above tol. (9.84e-3)
L4 (Wild-BAL, KazuNet)	Canonical `gesvd`	7.7871 px (Ref)	43.1 W (+15.8, 59.8%)	52.8 / 62.9 W	Reference
L4 (Wild-BAL, KazuNet)	Exact TSQR + CPU `gesdd`	7.7939 (+0.09%)	43.8 W (+16.5, 60.8%)	—	Above tol. (8.77e-4)
L4 (Wild-BAL, KazuNet)	Two-band approx. SVD	**9.1063 (+16.94%)**	44.2 W (+16.9, 61.4%)	—	Above tol. (1.69e-1)
A100 C1	3 coupled workers ($1\times$ batch, 30% VRAM cap)	Ref (0.00e+00)	89.6 W (+27.8, 22.4%)	102.6 / 160.8 W (11,685 MiB)	Reference
A100 C2	3 coupled workers ($4\times$ batch)	KazuNet-M UW **0.760**	92.5 W (+30.7, 23.1%)	110.7 / 192.2 W (22,711 MiB)	Above tol. (7.60e-1)
A100 C3	8 decoupled workers ($1\times$ batch)	0.00e+00	90.6 W (+28.8, 22.6%)	104.1 / 139.2 W (21,985 MiB)	Within tol. (0.00e+00)
A100 C4	8 decoupled workers ($1\times$ batch, CUDA MPS)	**0.00e+00**	**100.0 W (+38.2, 25.0%)**	127.1 / 216.1 W (21,445 MiB)	**Within tol.** (0.00e+00)
A100 C5	8 decoupled workers ($4\times$ ext. batch, MPS)	FGNN 3D **1.34e-5**	100.6 W (+38.8, 25.1%)	134.3 / 248.7 W (25,461 MiB)	Above tol. (1.34e-5)

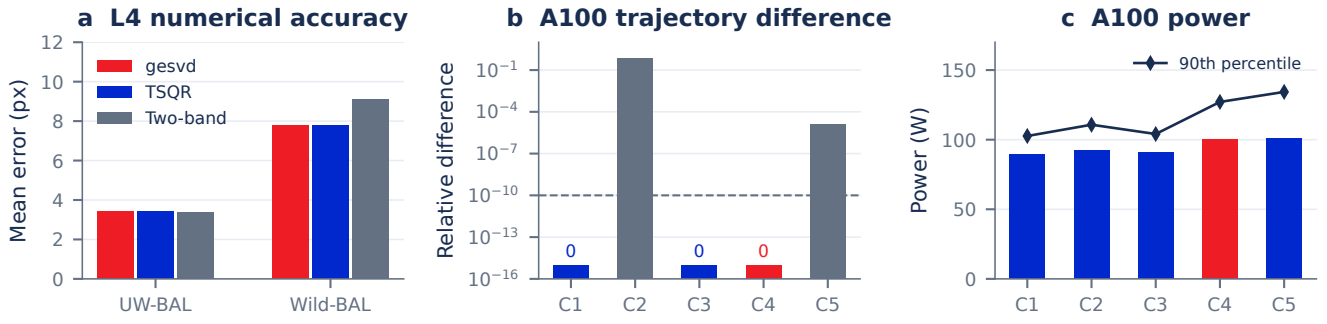


Figure S1. **Numerical accuracy and hardware measurements.** (a) L4 error under the three recorded SVD backends. (b) A100 trajectory discrepancies for configurations C1–C5; the dashed line is 10^{-10}. (c) Mean and 90th-percentile A100 power. Table S17 lists complete measurements.

References

[1] James Demmel, Laura Grigori, Mark Hoemmen, and Julien Langou. Communication-optimal parallel and sequential QR and LU factorizations. *SIAM Journal on Scientific Computing*, 34(1), 2012. 1

[2] Jingwei Huang, Shan Huang, and Mingwei Sun. DeepLM: Large-scale nonlinear least squares on deep learning frameworks using stochastic domain decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10308–10317, 2021. 3, 5

[3] Xin Liu and Jufeng Yang. Progressive neighbor consistency mining for correspondence pruning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9527–9537, 2023. 4, 5

[4] Jorge J. Moré. The Levenberg–Marquardt algorithm: Implementation and theory. In *Numerical Analysis*, pages 105–116. Springer, 1978. 4

[5] Claudio Procesi. The invariant theory of $n \times n$ matrices. *Advances in Mathematics*, 19(3):306–381, 1976. 4

[6] René Ranftl and Vladlen Koltun. Deep fundamental matrix estimation. In *European Conference on Computer Vision (ECCV)*, pages 284–299, 2018. 4, 5

[7] Soledad Villar, David W. Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 4

[8] Hermann Weyl. *The Classical Groups: Their Invariants and Representations*. Princeton University Press, 1946. 4

[9] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. 4

[10] Kwang Moo Yi, Eduard Trulls, Yuki Ono, Vincent Lepetit, Mathieu Salzmann, and Pascal Fua. Learning to find good correspondences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2666–2674, 2018. 4, 5

[11] Jiahui Zhang, Dawei Sun, Zixin Luo, Anbang Yao, Lei Zhou, Tianwei Shen, Yurong Chen, Long Quan, and Hongen Liao. Learning two-view correspondences and geometry using order-aware network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5845–5854, 2019. 4, 5

[12] Zhen Zhang, Fan Wu, and Wee Sun Lee. Factor graph neural network. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 10720–10730, 2020. 4, 5

[13] Chen Zhao, Yixiao Ge, Feng Zhu, Rui Zhao, Hongsheng Li, and Mathieu Salzmann. Progressive correspondence pruning by consensus learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6464–6473, 2021. 4, 5