LEARNING COMPACT LANCZOS CONTINUATIONS FOR EFFICIENT PDE RESPONSE EVALUATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate evaluation of scalar responses of partial differential equations (PDEs) can be computationally demanding, particularly in high-contrast heterogeneous media. Limiting the number of Lanczos iterations reduces this computational cost, but leaves unresolved contributions that compromise response accuracy. We address this trade-off by learning a compact continuation of the Lanczos model, improving its response approximation without additional Lanczos iterations. For a class of linear second-order elliptic PDEs, we establish an existence guarantee for finite-dimensional continuation: accurate energy response approximation over an entire prescribed high contrast interval is possible with a continuation dimension independent of the full discretization size. At the same reduced dimension, we derive an explicit improvement factor over the classical Lanczos response bound under stated high-contrast conditions. Experiments on heterogeneous conductivity problems demonstrate lower computational cost at comparable attained response accuracy on the evaluated benchmarks, alongside improved accuracy in the tested contrast-extrapolation regimes. The sharper guarantee comes from adapting the approximation space to the inverse-response kernel while retaining the information encoded by the computed Lanczos prefix.

1 INTRODUCTION

Many scientific computing workflows repeatedly evaluate a low-dimensional response of a partial differential equation (PDE) over varying coefficient fields and parameter values. Examples include transport, elasticity, and wave propagation, with scalar outputs such as effective transport coefficients, structural compliance, and scattering cross-sections, respectively. Here we focus on boundary-driven energy responses of two-phase affine elliptic PDEs. Although the requested output is often only a scalar, evaluating it conventionally requires solving a large spatial system for every new instance and every parameter query. High material contrast can make the iterative work particularly costly. The practical objective is therefore to reduce the computation needed to reach a specified response accuracy. Figure 1 illustrates how a single-pixel change in a high-contrast microstructure can alter the response by several orders of magnitude.

A short numerical calculation provides information that a learned surrogate can exploit. For the symmetric affine systems considered here, m Lanczos iterations produce a tridiagonal matrix J_m describing the operator as seen by the prescribed loading (Golub & Meurant, 1994; Li et al., 2016). This reduced model is inexpensive to evaluate at different contrasts. Its accuracy, however, is limited by the directions not yet explored. Increasing m captures more of the unresolved response but increases computational cost. We ask whether a learned, low-dimensional continuation can recover enough of that influence to improve the response effectively.

The unresolved quantity has a useful structure. Block elimination expresses the effect of the remaining operator through a scalar inverse response function. A short structured tridiagonal continuation can model this function and be attached to the computed prefix. The central question is whether it can preserve the computed information while meeting a prescribed response tolerance over an entire finite contrast interval.

Our main theorem answers this representation question for boundary-driven two-phase energy problems considered. For every admissible instance and every finite contrast interval $[0, S]$, a contrast

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

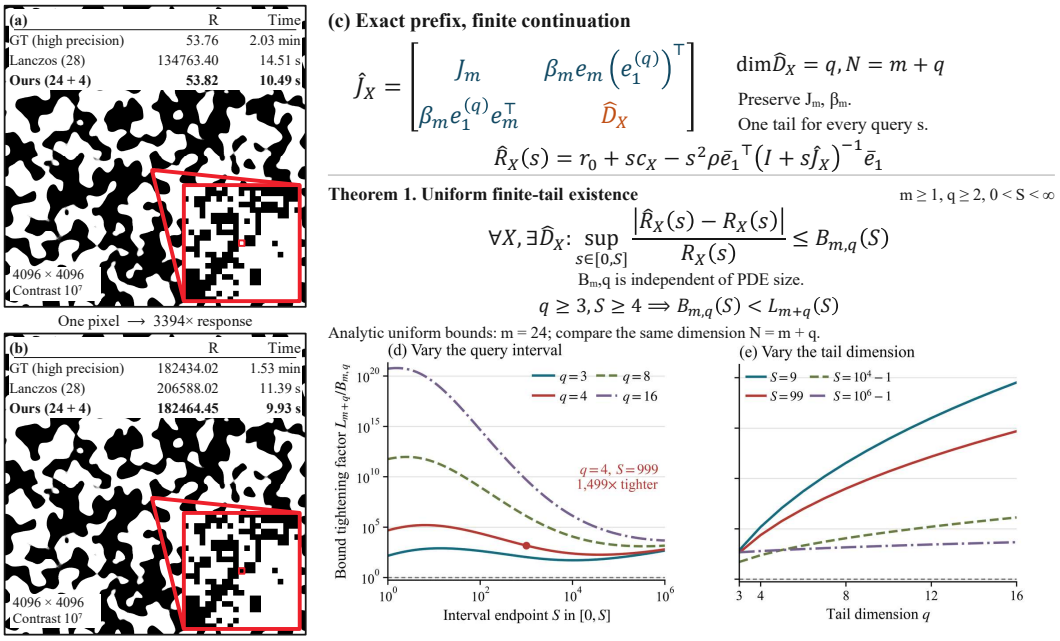


Figure 1: **Microstructure-sensitive responses and finite Lanczos continuation.** (a,b) Two 4096^2 conductivity fields at $s = 10^7 - 1$ differ at one highlighted pixel, changing the reference response by approximately $3394\times$. Black and white denote conductivities 10^7 and 1; insets show the same 21^2 neighborhood. Tables report responses and times for the reference (GT), 28-step Lanczos, and a learned $(24+4)$ continuation. (c) A finite tail retains the exact prefix J_m, β_m and serves all queries. (d,e) The analytic bound-tightening factor $L_{m+q}(S)/\mathcal{B}_{m,q}(S) = \Gamma_{m,q}(S)^{-1}$ at $m = 24$, varying S or q, with logarithmic vertical axes.

independent continuation exists with an explicit uniform relative error bound $\mathcal{B}_{m,q}(S)$. The sufficient tail dimension q depends on the prefix budget, interval and tolerance, independently of the full discretization dimension. We then derive the corresponding ordinary $(m+q)$-step Lanczos energy response bound from the classical CG Chebyshev estimate (Saad, 2003). Their ratio gives an explicit sufficient condition for a tighter uniform bound at the same reduced dimension. The improvement comes from replacing polynomial approximation by an interval-adapted rational space while retaining the computed prefix.

We build the learned model by predicting a compact continuation of the computed Lanczos prefix. Specifically, an encoder maps the computed numerical data and inexpensive instance descriptors to a short tail. A differentiable decoder retains the prefix and enforces the reduced energy inequalities, together with applicable known endpoint conditions. The resulting model requires no further PDE-space iterations after the prefix. Controlled experiments and end-to-end wall-clock timings assess its accuracy–cost trade-off.

Our contributions are:

- **A finite-continuation existence theorem and a quantitative Lanczos comparison.** We prove a discretization-independent uniform response bound while preserving a computed prefix, and derive an explicit ratio to the ordinary Lanczos bound at the same total dimension.
- **A response-level learning method.** We learn a query-independent structured tridiagonal continuation through an energy-feasible decoder, using scalar response supervision without reconstructing the full solution or the true uncomputed operator.
- **An evaluation of the resulting accuracy–cost trade-off.** Controlled conductivity comparisons measure fixed-budget accuracy and end-to-end wall-clock time at comparable attained errors and contrast extrapolation behavior.

2 PRELIMINARIES: PDEs AND SCALAR RESPONSES

Affine elliptic problems. Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with $\partial\Omega = \overline{\Gamma_D} \cup \overline{\Gamma_N}$, partitioned into Dirichlet and Neumann parts. An instance X specifies the spatial coefficient field and loading. We consider a family of self-adjoint, linear second-order elliptic PDEs

$$\begin{aligned} -\nabla\cdot(K_s\nabla u_s) + \chi_s u_s &= F \quad \text{in } \Omega, & K_s &= K_0 + sK_X, \\ u_s &= g_D \quad \text{on } \Gamma_D, & \chi_s &= \chi_0 + s\chi_X, \\ \mathbf{n}^\top K_s \nabla u_s &= g_N \quad \text{on } \Gamma_N, & s &\geq 0. \end{aligned} \tag{1}$$

Here $\mathbf{n}$ is the outward unit normal, $K_0(x)$ is symmetric and uniformly positive definite, and $\chi_0 \geq 0$; the reference coefficients are bounded. Each instance X specifies a phase indicator $\eta_X : \Omega \to \{0,1\}$, with $K_X = \eta_X K_0$ and $\chi_X = \eta_X \chi_0$. The dimensionless parameter $s \geq 0$ controls the perturbation amplitude relative to the reference coefficients. At $s = 0$, the reference problem is recovered. In the phase $\eta_X = 1$, both coefficients are multiplied by $1+s$, while in the phase $\eta_X = 0$ they retain their reference values.

Let $V = \{v \in H^1(\Omega) : v|_{\Gamma_D} = 0\}$ and let $u_{D,X}$ be a fixed lifting of g_D. Define

$$a_i(u,v) = \int_\Omega (\nabla u^\top K_i \nabla v + \chi_i uv)\,dx, \quad i \in \{0, X\}, \qquad \ell(v) = \int_\Omega Fv\,dx + \int_{\Gamma_N} g_N v\,dS. \tag{2}$$

Since $\eta_X \in \{0,1\}$,

$$0 \leq a_X(w,w) \leq a_0(w,w) \qquad (w \in H^1(\Omega)). \tag{3}$$

We assume a_0 is coercive on V and ℓ is continuous. The weak problem is to find $u_s \in u_{D,X} + V$ such that $(a_0 + sa_X)(u_s, v) = \ell(v)$ for every $v \in V$. It has a unique solution for every $s \geq 0$. A conforming space $V_h \subset V$, with basis $\{\varphi_i\}_{i=1}^{n_h}$ and lifting $u_{D,h,X}$, gives

$$(A_i)_{jk} = a_i(\varphi_j, \varphi_k), \qquad A(s) = A_0 + sA_X, \qquad A_0 \succ 0, \quad A_X \succeq 0. \tag{4}$$

All matrix identities and error guarantees below concern this discrete problem; h is the spatial discretization parameter.

Scalar energy response. We consider boundary-driven problems with $F = 0$, $g_N = 0$ and the energy

$$R_X(s) = \min_{u \in u_{D,h,X} + V_h} a_s(u,u) = a_s(u_{s,h,X}, u_{s,h,X}), \qquad a_s = a_0 + sa_X. \tag{5}$$

Let $u_{0,h,X}$ be the reference minimizer and define the induced load and two scalar energies by

$$(f_X)_i = a_X(u_{0,h,X}, \varphi_i), \qquad r_{0,X} = a_0(u_{0,h,X}, u_{0,h,X}), \qquad c_X = a_X(u_{0,h,X}, u_{0,h,X}). \tag{6}$$

Writing $u_{s,h,X} = u_{0,h,X} + \sum_i (v_s)_i \varphi_i$ and using $a_0(u_{0,h,X}, v) = 0$ for $v \in V_h$ yields

$$A(s)v_s = -sf_X, \qquad R_X(s) = r_{0,X} + sc_X - s^2 g_X(s), \qquad g_X(s) = f_X^\top A(s)^{-1} f_X. \tag{7}$$

Thus evaluating $R_X(s)$ reduces to the inverse response $g_X(s)$, whose induced load f_X is fixed as s varies. The subsequent relative energy bounds and decoder constraints concern R_X. We consider instances with $r_{0,X} > 0$, so that $R_X(s) \geq r_{0,X} > 0$ for $s \geq 0$ and relative errors are well defined. If $r_{0,X} = 0$, the form inequalities imply $c_X = 0$ and $f_X = 0$, hence $R_X \equiv 0$, which is evaluated exactly without relative normalization. Since $A_0 \succ 0$, define

$$T = A_0^{-1/2} A_X A_0^{-1/2}, \qquad b = A_0^{-1/2} f_X, \qquad \rho = b^\top b. \tag{8}$$

The definition of $\eta_X \in \{0,1\}$ implies $0 \preceq T \preceq I$. Factoring $A(s) = A_0^{1/2}(I + sT)A_0^{1/2}$ gives

$$g_X(s) = f_X^\top A(s)^{-1} f_X = b^\top (I + sT)^{-1} b. \tag{9}$$

The Lanczos construction below uses (T, b). These quantities depend on X; for simplicity, their subscripts are suppressed. If $\rho = 0$, then $f_X = 0$ and $R_X(s) = r_{0,X} + sc_X$ is evaluated exactly without Lanczos.

Lanczos prefix and unresolved tail. For $\rho = b^\top b > 0$, Lanczos starts from $v_1 = b/\sqrt{\rho}$ and orthogonalizes successive applications of T into $Q_m = [v_1, \dots, v_m]$. The projection $J_m = Q_m^\top T Q_m$ is tridiagonal, with diagonal entries α_j and off-diagonal entries β_j. The mth step also computes the terminal residual norm $\beta_m = \|Tv_m - Q_m J_m e_m^{(m)}\|$, where $e_m^{(m)}$ is the last coordinate vector in $\mathbb{R}^m$. If $\beta_m = 0$ (or termination occurs earlier), the computed invariant subspace already gives the exact response and no tail is needed.

Let L be the dimension of the full cyclic Lanczos space, and let $e_j^{(d)}$ denote the jth coordinate vector in $\mathbb{R}^d$. If $\beta_m > 0$, continuing Lanczos to its exact termination would give

$$J_X^{\text{full}} = \begin{bmatrix} J_m & \beta_m e_m^{(m)} (e_1^{(L-m)})^\top \\ \beta_m e_1^{(L-m)} (e_m^{(m)})^\top & D_X \end{bmatrix}, \qquad g_X(s) = \rho\big[(I + sJ_X^{\text{full}})^{-1}\big]_{11}, \tag{10}$$

where $D_X \in \mathbb{R}^{(L-m)\times(L-m)}$ is the true unresolved tail. We replace the uncomputed tail D_X by an effective symmetric tridiagonal Jacobi matrix $\widehat{D}_{X,S}$ of dimension at most q, for a prescribed interval endpoint S, retaining J_m and β_m. The resulting matrix $\widehat{J}_{X,S}$ gives

$$\widehat{R}_{X,S}(s) = r_{0,X} + sc_X - s^2\rho\big[(I + s\widehat{J}_{X,S})^{-1}\big]_{11}. \tag{11}$$

3 Finite continuation and comparison with Lanczos

A dimension-independent existence guarantee. Let $\mathcal{X}$ be the deterministic class of finite-dimensional boundary-driven instances in Sec. 2 with $r_{0,X} > 0$. For a completed matrix $\widehat{J}$, let $\widehat{R}_X(s;\widehat{J})$ denote its response from equation 11, and define

$$e_X(s;\widehat{J}) := \frac{|\widehat{R}_X(s;\widehat{J}) - R_X(s)|}{R_X(s)}, \qquad \varepsilon_X(\widehat{J};S) := \sup_{0\le s\le S} e_X(s;\widehat{J}). \tag{12}$$

e_X is the pointwise absolute relative error, and ε_X is its supremum over s for one fixed instance. For integers $m \ge 1$, $q \ge 2$, let $N = m + q$ and define

$$\begin{aligned} &a = \frac{\sqrt{1+S}-1}{\sqrt{1+S}+1}, && \nu = m+2, && k = q-2, \\ &\Delta = \nu(1+a^2) + k(1-a^2), && z_\star = \frac{2\nu a}{\Delta + \sqrt{\Delta^2 - 4\nu^2 a^2}}, \\ &H = \begin{cases} a^{m+2}, & q = 2, \\ z_\star^\nu \left(\dfrac{a - z_\star}{1 - a z_\star}\right)^k, & q \ge 3, \end{cases} && \mathcal{B}_{m,q}(S) = \left(\frac{2H}{1+H^2}\right)^2. \end{aligned} \tag{13}$$

Theorem 1 (Uniform finite-tail existence)**.** *For every admissible instance $X \in \mathcal{X}$, its exact m-step Lanczos prefix with $\rho > 0$ and $\beta_m > 0$, every integer $q \ge 2$, and every fixed $0 < S < \infty$, there exists a Jacobi tail $\widehat{D}_{X,S}$ of dimension at most q, retaining J_m, β_m and independent of the query $s \in [0, S]$, such that*

$$\boxed{\varepsilon_X(\widehat{J}_{X,S};S) \le \mathcal{B}_{m,q}(S).} \tag{14}$$

For fixed S, the bound $\mathcal{B}_{m,q}(S)$ decreases as either the number m of computed Lanczos steps or the tail dimension q increases.

Corollary 1.1 (A sufficient tail dimension)**.** *For fixed $m \ge 1$, $0 < S < \infty$, and any tolerance $\delta > 0$, it suffices to take*

$$q \ge \max\left\{2, \left\lceil \frac{\log(4/\delta)}{2\log(1/a)} \right\rceil - m\right\} \tag{15}$$

to obtain $\varepsilon_X(\widehat{J}_{X,S};S) < \delta$ for every $X \in \mathcal{X}$, independently of the full discretization dimension.

Proof idea. Let $\lambda \in [0, 1]$ denote the spectral variable of T. Ordinary Lanczos uses the polynomial space $\mathbb{P}_{N-1}$. Our construction rewrites the energy as a nonnegative affine term plus a positive inverse-response integral, then uses $\{p(\lambda)/(1 + S\lambda)^{q-2} : p \in \mathbb{P}_{N-1}\}$. This changes the functions representable at fixed dimension: for $q \geq 3$, the new space contains $(1 + S\lambda)^{-1}$ exactly. Positive quadrature for the reweighted measure produces a Jacobi continuation preserving the prefix and both endpoint inverse moments. A rational residual bounds its error throughout $[0, S]$, with maximum determined by $z_\star$. For $q \geq 3$, the continuation constructed in the proof satisfies $\widehat{R}_{X,S}(S) = R_X(S)$. Appendix A gives the proof.

Ordinary Lanczos at the same dimension. Let J_N result from $N = m + q$ ordinary Lanczos steps on the original (T, b), with response

$$R^{\mathrm{L}}_{N,X}(s) = r_{0,X} + sc_X - s^2\rho(e_1^{(N)})^\top(I + sJ_N)^{-1}e_1^{(N)}. \tag{16}$$

The classical CG Chebyshev bound (Saad, 2003, Section 6.11.3) and a Schur-complement conversion give a bound on the same energy response R_X, accounting for its reconstruction from g_X.

Proposition 1 (Ordinary Lanczos response bound)**.** *Under the assumptions of Theorem 1, exact N-step Lanczos satisfies*

$$0 \leq \frac{|R^{\mathrm{L}}_{N,X}(s) - R_X(s)|}{R_X(s)} \leq L_N(S) := \frac{4a^{2N+2}}{(1 - a^{2N+2})^2}, \qquad 0 \leq s \leq S. \tag{17}$$

Corollary 1.2 (A tighter guarantee at the same dimension)**.** *For $N = m+q$, the ratio of the uniform bounds is*

$$\Gamma_{m,q}(S) := \frac{\mathcal{B}_{m,q}(S)}{L_N(S)} = \left[\frac{H(1 - a^{2N+2})}{a^{N+1}(1 + H^2)}\right]^2. \tag{18}$$

If $q \geq 3$ and $S \geq 4$, then $0 < \Gamma_{m,q}(S) < 1$: the finite-continuation existence bound is strictly tighter at total dimension N.

Appendix B proves both statements. Figure 1(d,e) shows the reciprocal $\Gamma^{-1}_{m,q} = L_{m+q}/\mathcal{B}_{m,q}$ as the interval or tail dimension varies at fixed prefix length.

4 Learning the unresolved response

Theorem 1 establishes that a sufficiently large continuation can meet any positive relative tolerance on a fixed finite interval. Its construction uses spectral information from the complete instance. Inspired by Theorem 1, we train a network to select an effective continuation from the computed Lanczos data and inexpensive material descriptors. The prefix J_m remains fixed during training and inference.

From numerical inputs to a response curve. For each instance X, the numerical stage computes $(J_m, \beta_m, \rho, c_X, r_{0,X})$. A shared network $\mathcal{N}_\theta$ uses the prefix coefficients, scalars derived from this computation, the phase fraction, the loading direction, and phase-connectivity indicators to predict tail coordinates $\xi_X \in \mathbb{R}^{2q-1}$. A differentiable decoder uses these coordinates and the endpoint indicators described below to construct a structured Jacobi tail $\widehat{D}_{\theta,X}$. The completed matrix retains the Lanczos prefix and its terminal coupling, replacing the uncomputed tail:

$$\widehat{J}_{\theta,X} = \begin{bmatrix} J_m & \beta_m e_m^{(m)}(e_1^{(q)})^\top \\ \beta_m e_1^{(q)}(e_m^{(m)})^\top & \widehat{D}_{\theta,X} \end{bmatrix}, \tag{19}$$

The learned tail $\widehat{D}_{\theta,X} \in \mathbb{R}^{q\times q}$ uses the same block form as Figure 1(c); its training interval is fixed and suppressed in this notation for simplicity. Substituting $\widehat{J}_{\theta,X}$ into equation 11 gives the response at any s from a fixed continuation per instance X. Training uses response labels without requiring the true tail D_X.

Enforcing the energy inequalities. For $\rho > 0$, the two nonnegative energies give $0 < c_X < r_{0,X}$. For the full Lanczos basis $Q = [v_1, \dots, v_L]$, $v_1 = b/\sqrt{\rho}$, let $\psi_j \in V_h$ have coefficients $A_0^{-1/2} v_j$ and set $J_X^{\text{full}} = Q^\top T Q$. Recall $c_X = a_X(u_{0,h,X}, u_{0,h,X})$, while $t \in \mathbb{R}$ and $z \in \mathbb{R}^L$ are arbitrary weights in $w = t u_{0,h,X} + \sum_j z_j \psi_j$. Expanding the forms using $a_0(u_{0,h,X}, v) = 0$ for $v \in V_h$, $a_0(\psi_i, \psi_j) = \delta_{ij}$, and $a_X(u_{0,h,X}, \psi_j) = \sqrt{\rho}\,\delta_{1j}$ gives

$$a_X(w,w) = c_X t^2 + 2\sqrt{\rho}\, t z_1 + z^\top J_X^{\text{full}} z \geq 0,$$
$$(a_0 - a_X)(w,w) = (r_{0,X} - c_X) t^2 - 2\sqrt{\rho}\, t z_1 + z^\top (I - J_X^{\text{full}}) z \geq 0.$$

Replace J_X^{full} by $\widehat{J} \in \mathbb{R}^{N \times N}$, $N = m + q$. For each fixed $z \in \mathbb{R}^N$, both energies are convex quadratics in t. Requiring these minima to be nonnegative for every z gives

$$\widehat{J} - \frac{\rho}{c_X} e_1^{(N)} (e_1^{(N)})^\top \succeq 0, \qquad I - \widehat{J} - \frac{\rho}{r_{0,X} - c_X} e_1^{(N)} (e_1^{(N)})^\top \succeq 0. \tag{20}$$

These full-matrix inequalities jointly constrain all q diagonals and $q - 1$ couplings of $\widehat{D}_{\theta,X}$.

For the first tail diagonal $\widehat{\alpha}_{m+1} = (\widehat{D}_{\theta,X})_{11}$, the leading $(m+1) \times (m+1)$ blocks of equation 20 are

$$\begin{bmatrix} A_- & \beta_m e_m^{(m)} \\ \beta_m (e_m^{(m)})^\top & \widehat{\alpha}_{m+1} \end{bmatrix} \succeq 0, \qquad \begin{bmatrix} A_+ & -\beta_m e_m^{(m)} \\ -\beta_m (e_m^{(m)})^\top & 1 - \widehat{\alpha}_{m+1} \end{bmatrix} \succeq 0,$$

where $A_- = J_m - (\rho/c_X) e_1^{(m)} (e_1^{(m)})^\top$ and $A_+ = I - J_m - (\rho/(r_{0,X} - c_X)) e_1^{(m)} (e_1^{(m)})^\top$. Minimizing over the prefix coordinates gives the bounds $\ell \leq \widehat{\alpha}_{m+1} \leq u$:

$$\ell = \beta_m^2 (e_m^{(m)})^\top A_-^{-1} e_m^{(m)}, \qquad u = 1 - \beta_m^2 (e_m^{(m)})^\top A_+^{-1} e_m^{(m)}, \qquad \widehat{\alpha}_{m+1} = \ell + \sigma(\xi_1)(u - \ell), \tag{21}$$

where $\sigma(t) = (1 + e^{-t})^{-1}$. This display shows only the update for the first tail diagonal; each later logit pair selects a coupling and diagonal with updated bounds, preserving both inequalities for the completed $\widehat{J}$ (Appendix D).

Preserving asymptotic growth and zero-energy conditions. The decoder also preserves known zeros of the two energies. Define

$$d_\infty := \min_{w \in u_{0,h,X} + V_h} a_X(w,w),$$
$$e_- := \min_{w \in u_{0,h,X} + V_h} (a_0 - a_X)(w,w). \tag{22}$$

The discrete energy principle gives $R_X(s) = d_\infty s + O(1)$ as $s \to \infty$: $d_\infty > 0$ yields asymptotically linear growth, whereas $d_\infty = 0$ yields a bounded response. The complementary minimum e_- is the minimum energy at the formal degenerate endpoint $s = -1$, where only phase $\eta_X = 0$ contributes to the energy. This endpoint lies outside the contrast query interval and is only used to impose structural constraints on the decoder. The two indicators above record whether these phase energies vanish. When endpoint information is unavailable, the energy-only decoder can retain the energy inequalities without imposing endpoint equalities.

Setting $t = 1$ in the preceding reduced energies fixes the boundary data; minimizing over z gives

$$\widehat{d}_\infty = c_X - \rho (e_1^{(N)})^\top \widehat{J}^{-1} e_1^{(N)},$$
$$\widehat{e}_- = r_{0,X} - c_X - \rho (e_1^{(N)})^\top (I - \widehat{J})^{-1} e_1^{(N)}. \tag{23}$$

By equation 20 these minima are nonnegative, and $\widehat{R}(s) = \widehat{d}_\infty s + O(1)$. A known zero is imposed on the corresponding predicted minimum; a known positive minimum is kept positive, with its magnitude learned. With unknown indicators, interior sigmoid coordinates ensure nonnegative energies but do not certify the true endpoint behavior.

Let $[\ell_N, u_N]$ be the admissible interval for the final diagonal $\widehat{\alpha}_N = (\widehat{D}_{\theta,X})_{qq}$, obtained from the preceding recurrence that enforces both energy inequalities. The known zero-energy endpoint conditions are imposed at this final step. For a coupled completion, the final energy coefficients are $\widehat{\alpha}_N - \ell_N$ and $u_N - \widehat{\alpha}_N$, giving

$$\widehat{d}_\infty = 0 \iff \widehat{\alpha}_N = \ell_N, \qquad \widehat{e}_- = 0 \iff \widehat{\alpha}_N = u_N. \tag{24}$$

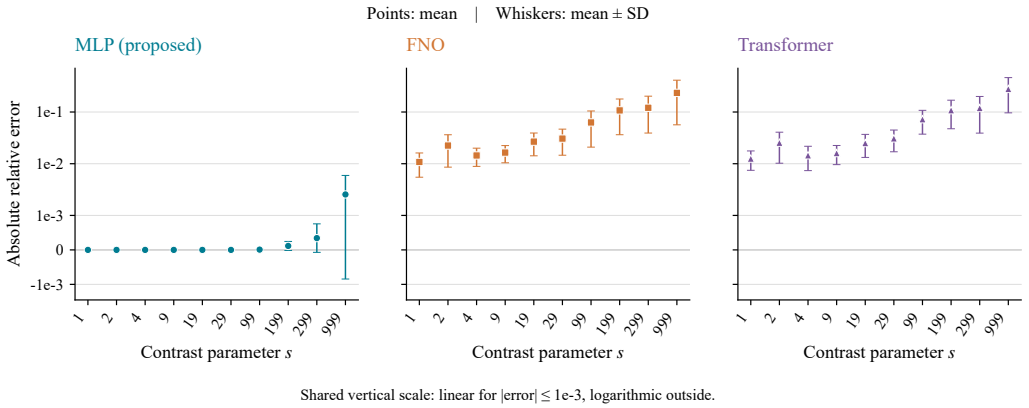


Figure 2: **In-domain generalization across held-out structures.** The validation-selected proposed MLP ($m = 24$, $q = 4$), direct field FNO and Transformer are evaluated on 40 held-out structures. Points show mean absolute relative error; whiskers show $\pm$ one sample SD across structures, after averaging errors over two directions and three fitted seeds within each structure. Queries $s = 2, 19, 199$ are absent from training; the other seven are training values.

Appendix D.2 gives the derivation and coordinate rules. Under the above constraints, for M labeled instances and K training queries per instance, we minimize relative response error:

$$\mathcal{L}(\theta) = \frac{1}{MK} \sum_{i=1}^{M} \sum_{j=1}^{K} \left(\frac{\widehat{R}_{\theta, X_i}(s_j) - R_{X_i}(s_j)}{R_{X_i}(s_j)} \right)^2 . \tag{25}$$

5 Experiments

We test whether a learned continuation improves the response after limited numerical work and reduces query cost. The proposed constrained continuations use $m = 24$, $q = 4$, and query-independent MLP, Transformer (Vaswani et al., 2017), or sequence-FNO (Li et al., 2021) encoders. We also compare with two preconditioned conjugate gradient (PCG) solvers (Saad, 2003). FFT-PCG uses the inverse of the homogeneous reference operator A_0 for preconditioning, applied through fast sine and cosine transforms. AMG-PCG uses a classical Ruge–Stüben algebraic multigrid V-cycle (Xu & Zikatanov, 2017), implemented in PyAMG (Bell et al., 2023). Both solve the same discretized PDE and evaluate the corresponding scalar energy response.

Protocol. Training and validation contain 512 and 128 structures at 128^2, respectively, with phase fraction $\phi \in [0.30, 0.70]$ and $s \in \{1, 4, 9, 29, 99, 299, 999\}$. We report the absolute relative error e_X defined in equation 12, against independent same-grid high precision PDE solver. At each contrast, we first average errors over two loading directions and then report the mean and sample standard deviation across test structures. For the large-scale experiments in Figure 1 and Table 3, we separately train an MLP on 4096×4096 conductivity fields at a fixed conductivity contrast of 10^7 ($s = 10^7 - 1$), using $(m, q) = (24, 4)$. Its parameters are frozen during evaluation. Appendix E gives the complete protocols, hardware and populations.

Accuracy and efficiency of learned continuation. Table 1 compares contrast extrapolation for the constrained MLP, Transformer and sequence FNO with ordinary Lanczos and matched-encoder free continuations and scalar-residual heads. All three constrained encoders implement the proposed method. Methods are grouped by encoder, and each query is reported separately. Ordinary 28-step Lanczos uses the same reduced dimension but performs four additional PDE-space iterations for a matched dimension fair comparison (Appendix E).

Figure 2 compares the validation-selected proposed MLP with direct field FNO and Transformer models on all 40 held-out structures over $1 \leq s \leq 999$. Points and error bars show the geometry-level mean and sample SD of the absolute relative error. The ten queries include the seven training

Table 1: **Contrast extrapolation at fixed numerical budgets.** Entries report the mean ± sample standard deviation of the absolute relative response error e_X, expressed in percent, across 40 held-out test structures. The training queries satisfy $1 \leq s \leq 999$. Proposed models learn a continuation through the constrained decoder; "free ext." learns continuation coefficients without enforcing the energy or endpoint constraints; "residual" learns an unrestricted, query-dependent additive correction to the Lanczos response. Field models directly predict the scalar response from the coefficient field, loading and query, without Lanczos inputs. Seq. FNO operates on the Lanczos coefficient sequence. All learned prefix models use $m = 24$, with $q = 4$ appended nodes for continuation heads; ordinary $m = 28$ performs 28 actual Lanczos steps. Bold entries identify the lowest mean in each column.

Method	$s = 2999$ (%)	$s = 9999$ (%)	$s = 99999$ (%)
Lanczos, $m = 24$	14.4 ± 18.8	54.1 ± 78.1	555 ± 863
Lanczos, $m = 28$	8.34 ± 11.9	33.5 ± 52.8	353 ± 600
MLP (proposed)	$\mathbf{0.61 \pm 0.88}$	$\mathbf{1.19 \pm 2.07}$	$\mathbf{1.81 \pm 3.18}$
MLP free ext.	1.81 ± 2.99	10.07 ± 18.08	117 ± 234
MLP residual	2.88 ± 3.14	11.7 ± 11.6	88.4 ± 137
Transformer (proposed)	0.75 ± 1.10	1.43 ± 2.37	1.98 ± 3.56
Transformer free ext.	2.14 ± 3.59	10.8 ± 20.5	125 ± 256
Transformer residual	4.11 ± 4.39	18.8 ± 22.3	177 ± 285
Seq. FNO (proposed)	0.71 ± 1.03	1.36 ± 2.38	1.91 ± 3.67
Seq. FNO free ext.	1.97 ± 3.54	10.1 ± 20.1	117 ± 252
Seq. FNO residual	4.25 ± 4.78	20.1 ± 26.6	199 ± 339
Field FNO	44.9 ± 26.2	92.9 ± 40.1	265 ± 205
Field Transformer	44.0 ± 25.3	73.3 ± 26.8	158 ± 83.7

Table 2: **Constraint ablation under contrast extrapolation.** Energy denotes the two phase-energy inequalities; Endpoints denotes the known zero endpoint equalities. All variants use the same training data and MLP architecture with $(m, q) = (24, 4)$. Columns give contrast parameter s queried; training ends at $s = 999$. Entries are mean ± sample SD of absolute relative error in percent across 40 held-out structures, after averaging two loading directions and three fitted seeds within each structure.

Energy	Endpoints	$s = 2999$ (%)	$s = 9999$ (%)	$s = 99999$ (%)
✓	✓	0.61 ± 0.88	1.19 ± 2.07	1.81 ± 3.18
✓	×	1.77 ± 2.98	9.71 ± 18.08	115.49 ± 231.10
×	✓	0.63 ± 0.97	1.22 ± 2.15	1.92 ± 3.38
×	×	1.81 ± 2.99	10.07 ± 19.10	116.94 ± 238.77

parameters and three unseen in-range values, $2, 19, 199$. The error definition and averaging agree with Table 1; the figure assesses in-domain generalization, while the table reports contrast extrapolation.

Across twenty fixed geometries, the constrained continuation achieves lower complete query cost than FFT-PCG at comparable response errors (Table 3). These savings remain after accounting for instance-dependent assembly, numerical work, feature computation and response evaluation.

Table 2 shows that endpoint equalities become increasingly important as material contrast grows. With the energy inequalities retained, removing the endpoint equalities significantly raises the mean relative error. Removing only the energy inequalities also slightly increases mean errors at all three contrasts.

6 Related work

Krylov approximation and scalar responses. Lanczos methods connect tridiagonal matrices, spectral moments and Gaussian quadrature, providing approximations of scalar matrix functions

Table 3: Runtime and accuracy at a comparable response error. Twenty independent native 4096^2 structures use conductivity contrast 10^7, x loading and $\phi = 0.5$. Entries are mean $\pm$ sample SD across structures. Errors are dimensionless. The frozen MLP uses $(m, q) = (24, 4)$.

Method	Complete time (s)	Relative error
MLP (proposed)	11.72 ± 1.95	$(3.83 \pm 0.425) \times 10^{-4}$
Ordinary Lanczos	28.77 ± 17.06	$(9.25 \pm 0.731) \times 10^{-4}$
FFT-PCG	31.24 ± 14.82	$(8.21 \pm 1.91) \times 10^{-4}$
AMG-PCG	113.54 ± 90.43	$(4.03 \pm 4.65) \times 10^{-4}$

without computing a full solution (Golub & Meurant, 1994; Li et al., 2016). Rational quadrature extends exactness to functions with prescribed poles by modifying the integration measure (Gautschi, 1993; López Lagomasino et al., 2008). Our existence construction builds on these ideas while retaining a previously computed Lanczos prefix and its terminal coupling. The resulting bound concerns one continuation over an entire contrast interval and admits an explicit comparison with ordinary Lanczos at equal reduced dimension.

Learning components of numerical solvers. Hsieh et al. (2019) learn corrections to linear iterative updates while preserving the desired solution as a fixed point. Greenfeld et al. (2019) learn multigrid prolongation rules using an error-propagation objective, and Li et al. (2023) learn sparse factors for preconditioned conjugate gradients. HINTS combines neural-operator corrections with relaxation to address complementary error modes (Zhang et al., 2024). These methods accelerate iterative solution of the spatial problem. We target a scalar response after a fixed Lanczos calculation, learning a reduced continuation that can be evaluated without further spatial iterations.

Operator learning. Fourier neural operators learn transformations between functions in Fourier space (Li et al., 2021); DeepONet separately encodes an input function and output locations (Lu et al., 2021). These models can predict spatial fields across PDE instances. Our predictor uses instance-specific numerical information and learns only the unresolved scalar response. The contrast query enters the reduced solve after the continuation has been predicted. Our field-based scalar baselines therefore compare complete pipelines with different numerical inputs and representations.

Reduced-output error models. ROMES maps inexpensive indicators to a distribution over reduced-model error, including output errors informed by dual-weighted residuals (Drohmann & Carlberg, 2015). Our continuation also improves a reduced output, with the additional constraint that all queries arise from one energy-feasible matrix. The matched-input scalar-residual baselines assess this choice of representation using the same encoder families. Since those residual heads do not impose the same hard endpoint conditions, the experiments assess the combined representation and constraints.

7 Conclusion

For boundary-driven two-phase elliptic problems, we establish the existence of compact Lanczos continuations with discretization-independent uniform response bounds over prescribed finite contrast intervals. Adapting the approximation space to the inverse-response kernel yields a tighter uniform guarantee than ordinary Lanczos at the same reduced dimension under the stated conditions. A constrained decoder makes this continuation representation trainable without additional Lanczos iterations. Conductivity experiments show improved fixed-budget accuracy and lower complete query cost at comparable attained errors on the evaluated benchmarks. The ablation highlights the importance of endpoint equalities for contrast extrapolation, while energy inequalities enforce structural admissibility. Extending these gains to broader PDE families and spatial distributions, and maintaining accuracy at tighter tolerances, remain directions for future work.

AI USE STATEMENT

OpenAI Codex was used to assist with developing the mathematical formulation and claims, implementing the methods, and drafting the manuscript. We reviewed all code, mathematical proofs, and experimental data, throughout the study.

ETHICS STATEMENT

The study concerns synthetic coefficient fields and numerical PDE responses, with no human participants or personal data. The work is presented in accordance with the principles of transparent and reproducible reporting in the ICLR Code of Ethics.

REPRODUCIBILITY STATEMENT

Section 2 states the mathematical assumptions; Appendices A and B provide the existence proof and the conversion of the classical CG estimate to the response bound. Section 4 and Appendix D specify the predictor, decoder and numerical safeguards. Appendix E documents the experimental data splits, model configurations, training objectives, error aggregation, hardware, and timing protocol. Accuracy is assessed against independently computed reference responses for the same discretized PDE. The theoretical guarantees concern exact arithmetic.

REFERENCES

Nathan Bell, Luke N. Olson, Jacob Schroder, and Ben Southworth. PyAMG: Algebraic multigrid solvers in Python. *Journal of Open Source Software*, 8(87):5495, 2023. doi: 10.21105/joss.05495.

Martin Drohmann and Kevin Carlberg. The ROMES method for statistical modeling of reduced-order-model error. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):116–145, 2015. URL https://kevintcarlberg.net/publication/drohmann/.

Walter Gautschi. Gauss-type quadrature rules for rational functions. In Helmut Brass and Günther Hämmerlin (eds.), *Numerical Integration IV*, volume 112 of *International Series of Numerical Mathematics*, pp. 111–130. Birkhäuser, 1993. URL https://arxiv.org/abs/math/9307223.

Gene H. Golub and Gérard Meurant. Matrices, moments and quadrature. In D. F. Griffiths and G. A. Watson (eds.), *Numerical Analysis 1993*, volume 303 of *Pitman Research Notes in Mathematics*. Longman, 1994. URL https://gerard-meurant.fr/mmq_1993.pdf.

Daniel Greenfeld, Meirav Galun, Ronen Basri, Irad Yavneh, and Ron Kimmel. Learning to optimize multigrid PDE solvers. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2415–2423, 2019. URL https://proceedings.mlr.press/v97/greenfeld19a.html.

Jun-Ting Hsieh, Shengjia Zhao, Stephan Eismann, Lucia Mirabella, and Stefano Ermon. Learning neural pde solvers with convergence guarantees, 2019. URL https://arxiv.org/abs/1906.01200.

Chengtao Li, Suvrit Sra, and Stefanie Jegelka. Gaussian quadrature for matrix inverse forms with applications. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pp. 1766–1775, 2016. URL https://proceedings.mlr.press/v48/lig16.html.

Yichen Li, Peter Yichen Chen, Tao Du, and Wojciech Matusik. Learning preconditioners for conjugate gradient PDE solvers. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 19425–19439, 2023. URL https://proceedings.mlr.press/v202/li23e.html.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.08895.

G. López Lagomasino, L. Reichel, and L. Wunderlich. Matrices, moments, and rational quadrature. *Linear Algebra and its Applications*, 429(10):2540–2554, 2008. doi: 10.1016/j.laa.2008.04.047. URL https://www.math.kent.edu/~reichel/publications/ratgauss.pdf.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3:218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Yousef Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2 edition, 2003. doi: 10.1137/1.9780898718003. URL https://www-users.cse.umn.edu/~saad/IterMethBook_2ndEd.pdf.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf.

Jinchao Xu and Ludmil T. Zikatanov. Algebraic multigrid methods. *Acta Numerica*, 26:591–721, 2017. URL https://arxiv.org/abs/1611.01917.

Enrui Zhang, Adar Kahana, Alena Kopaničáková, Eli Turkel, Rishikesh Ranade, Jay Pathak, and George Em Karniadakis. Blending neural operators and relaxation methods in PDE numerical solvers. *Nature Machine Intelligence*, 6:1303–1313, 2024. doi: 10.1038/s42256-024-00910-x.

A PROOF OF CONTINUOUS-INTERVAL TAIL EXISTENCE

All quantities in this appendix are exact. For brevity, write $V = V_h$ and $u_0 = u_{0,h,X}$. We use the bound parameters $a, \nu, k, \Delta, z_\star, H$ from equation 13, and write $N = m + q = \nu + k$.

Proof of Theorem 1. Fix an instance X and suppress its subscript. If $\rho = 0$, the inverse contribution in equation 7 vanishes. If Lanczos terminates at or before step m, its computed invariant space already gives the exact response. These cases require no further tail. Otherwise, let μ be the spectral probability measure of T at $b/\sqrt{\rho}$, so that $g(s) = \rho \int (1 + s\lambda)^{-1}\, d\mu(\lambda)$.

Step 1: a positive representation of the response. Evaluate the two nonnegative forms in equation 3 on the reference minimizer u_0 and normalized free coordinates. Since $a_0(u_0, v) = 0$ for $v \in V$, this gives

$$\begin{bmatrix} c & b^T \\ b & T \end{bmatrix} \succeq 0, \qquad \begin{bmatrix} r_{0,X} - c & -b^T \\ -b & I - T \end{bmatrix} \succeq 0. \tag{26}$$

Positivity forces $b \perp \ker T$ and $b \perp \ker(I - T)$. The generalized Schur complements therefore imply

$$M_- := \int \lambda^{-1}\, d\mu \le \frac{c}{\rho}, \qquad M_+ := \int (1 - \lambda)^{-1}\, d\mu \le \frac{r_{0,X} - c}{\rho}. \tag{27}$$

Thus the finite loaded spectrum lies in $(0, 1)$, and both inverse moments are finite. Define

$$d\eta = \frac{d\mu}{\lambda(1 - \lambda)}, \qquad F(s) = \int \frac{d\eta}{1 + s\lambda}, \qquad d_\infty = c - \rho M_-, \qquad e_- = r_{0,X} - c - \rho M_+. \tag{28}$$

Here η is finite because its mass is $M_- + M_+$, and $d_\infty, e_- \ge 0$. The scalar identity

$$\frac{1 + s}{\lambda(1 - \lambda)(1 + s\lambda)} - \frac{1 + s}{\lambda} - \frac{1}{1 - \lambda} = -\frac{s^2}{1 + s\lambda}$$

converts equation 7 into

$$R(s) = (1 + s)d_\infty + e_- + \rho(1 + s)F(s) \ge \rho(1 + s)F(s). \tag{29}$$

Step 2: a finite tail retaining the computed prefix. Fix the polynomial and positive measure

$$d(\lambda) = (1 + S\lambda)^k, \qquad d\zeta = \frac{d\eta}{d(\lambda)^2}. \tag{30}$$

If μ has $L \le N$ support points, retain its exact cyclic Jacobi matrix; its remaining tail has at most q coordinates. Otherwise, take the N-node Gaussian rule for ζ, with distinct nodes $\lambda_i \in (0, 1)$, positive weights w_i, and polynomial exactness through degree $2N - 1$ (Gautschi, 1993). Set

$$\widehat{\mu} = \sum_{i=1}^{N} w_i \lambda_i (1 - \lambda_i) d(\lambda_i)^2 \delta_{\lambda_i}. \tag{31}$$

Exactness for $\lambda^{j+1}(1 - \lambda)d^2$ gives

$$\int \lambda^j\, d\widehat{\mu} = \int \lambda^j\, d\mu, \qquad 0 \le j \le 2N - 2k - 3 = 2m + 1. \tag{32}$$

In particular, $\widehat{\mu}$ is a positive probability measure. For monic orthogonal polynomials p_j, put $h_j = \int p_j^2\, d\mu$. The Jacobi coefficients satisfy

$$\alpha_j = \frac{\int \lambda p_{j-1}^2\, d\mu}{h_{j-1}}, \qquad \beta_j^2 = \frac{h_j}{h_{j-1}}.$$

Consequently moments through degree $2m$ determine J_m, β_m. The N-dimensional Jacobi matrix of $\widehat{\mu}$ retains these data and supplies the desired q-dimensional trailing block $\widehat{D}$.

The polynomials $(1-\lambda)d^2$ and λd^2 have degree $2k+1 \le 2N-1$, so the same rule preserves M_-, M_+ exactly. Thus equation 29 also holds for $\widehat{R}$, with the same d_∞, e_- and

$$\widehat{F}(s) = \int \frac{d\widehat{\mu}}{\lambda(1-\lambda)(1+s\lambda)} = \sum_i \frac{w_i d(\lambda_i)^2}{1+s\lambda_i}. \tag{33}$$

The preserved inverse moments imply both energy inequalities for the completed matrix by the Schur complements in equation 26. Any zero endpoint equality associated with either inverse-moment constraint is also preserved. The construction depends on the prescribed interval endpoint S and is fixed for all queries in that interval.

Step 3: an exact Galerkin error identity. Let $\mathbb{P}_j$ denote the real polynomials of degree at most j. In $L^2(\eta)$ define

$$u_s(\lambda) = \frac{1}{1+s\lambda}, \qquad \mathcal{V} = \{p/d : p \in \mathbb{P}_{N-1}\}, \qquad \langle u, v\rangle_s = \int (1+s\lambda)uv\, d\eta. \tag{34}$$

The orthogonal projection $u_\mathcal{V} = p/d$ of u_s onto this fixed space satisfies

$$\int (1+s\lambda)ph\, d\zeta = \int hd\, d\zeta \qquad (h \in \mathbb{P}_{N-1}).$$

Both integrands have degree at most $2N-1$. Gaussian exactness and cardinal test polynomials at the nodes therefore give $p(\lambda_i) = d(\lambda_i)/(1+s\lambda_i)$. Applying exactness to pd yields $\widehat{F}(s) = \int u_\mathcal{V}\, d\eta$. For $e = u_s - u_\mathcal{V}$, orthogonality gives

$$F - \widehat{F} = \int e\, d\eta = \langle u_s, e\rangle_s = \langle e, e\rangle_s.$$

The projection minimizes this squared norm, hence

$$0 \le F - \widehat{F} = \min_{p\in\mathbb{P}_{N-1}} \int \frac{[1-(1+s\lambda)p(\lambda)/d(\lambda)]^2}{1+s\lambda}\, d\eta. \tag{35}$$

For $0 < s < S$, any rational residual $v_s = P/d$ with $P \in \mathbb{P}_N$ and $v_s(-1/s) = 1$ is admissible: the polynomial $d - P$ vanishes at $-1/s$, so $(d-P)/(1+s\lambda)$ belongs to $\mathbb{P}_{N-1}$. Equations equation 29 and equation 35 then imply the following bound. For this fixed witness, abbreviate $e_X(s; \widehat{J}_{X,S})$ by $e_X(s)$:

$$e_X(s) = \frac{R(s) - \widehat{R}(s)}{R(s)} = \frac{\rho(1+s)(F-\widehat{F})}{R(s)} \le \frac{F-\widehat{F}}{F} \le \|v_s\|^2_{L^\infty([0,1])}. \tag{36}$$

Step 4: a rational residual of the required degree. Use $\lambda = (w + w^{-1} + 2)/4$ and set

$$\Phi(w) = w^\nu \left(\frac{w+a}{1+aw}\right)^k, \qquad C(\lambda) = \frac{\Phi(w) + \Phi(w)^{-1}}{2}. \tag{37}$$

Since $\Phi(1/w) = \Phi(w)^{-1}$, C is independent of the two branches for w. On $\lambda \in [0,1]$ choose $|w| = 1$; then $|w+a| = |1+aw|$, so $|\Phi(w)| = 1$ and $|C(\lambda)| \le 1$. The identity $S = 4a/(1-a)^2$ also gives

$$1 + S\lambda = \frac{(w+a)(1+aw)}{(1-a)^2 w}.$$

Multiplying equation 37 by d, expanding, and pairing opposite powers of w shows that

$$C(\lambda) = \frac{\displaystyle\sum_{j=0}^{2k} \binom{2k}{j} a^{2k-j} \mathrm{T}_{|\nu-k+j|}(2\lambda-1)}{(1-a)^{2k} d(\lambda)}, \tag{38}$$

where $\mathrm{T}_j(\cos t) = \cos(jt)$. Every index is at most $\nu + k = N$, including when $k > \nu$, so the numerator has degree at most N.

For $0 < s < S$, define

$$y = \frac{\sqrt{1+s}-1}{\sqrt{1+s}+1} \in (0, a), \qquad \beta(y) = y^{\nu}\left(\frac{a-y}{1-ay}\right)^{k}. \tag{39}$$

The point $\lambda = -1/s$ corresponds to $w = -y$. Since $\Phi(-y)$ is real with magnitude $0 < \beta(y) < 1$, $|C(-1/s)| = [\beta(y) + \beta(y)^{-1}]/2$. The residual $v_s(\lambda) = C(\lambda)/C(-1/s)$ is therefore admissible in equation 36, and

$$e_X(s) \le \left(\frac{2\beta(y)}{1+\beta(y)^2}\right)^2. \tag{40}$$

At $s = 0$ both responses equal $r_{0,X}$. For $k \ge 1$, $u_S = (1 + S\lambda)^{k-1}/d \in \mathcal{V}$, so the response is exact at $s = S$ as well. Equation equation 40 thus extends to both endpoints. When $k = 0$ ($q = 2$), the ordinary Chebyshev expression applies at $s = S$ with $\beta(a) = a^{\nu}$.

Step 5: compute the interval maximum. The function $[2h/(1 + h^2)]^2$ is increasing for $0 \le h < 1$. For $k = 0$, $\max_{[0,a]} \beta = a^{\nu} = H$ and $z_\star = a$. For $k > 0$, β vanishes at both endpoints, and

$$\frac{d}{dy}\log\beta(y) = \frac{\nu}{y} - \frac{k(1-a^2)}{(a-y)(1-ay)}.$$

Its critical-point equation is

$$\nu a\, y^2 - \big[\nu(1+a^2) + k(1-a^2)\big]y + \nu a = 0. \tag{41}$$

The polynomial is positive at 0 and negative at a; its roots have product 1. Thus exactly one root lies in $(0, a)$, namely $z_\star$ in equation 13, and $\max_{[0,a]} \beta = \beta(z_\star) = H$. Taking the supremum of equation 40 over $s \in [0, S]$ gives $\varepsilon_X(\widehat{J}_{X,S}; S) \le \mathcal{B}_{m,q}(S)$. The right side is independent of X, so the same bound holds for every instance in $\mathcal{X}$. □

Monotonicity and finite budgets. For convergence of the existence bound, consider the auxiliary polynomial Chebyshev expression

$$\mathcal{B}^{\mathrm{poly}}_{m,q}(S) = \left(\frac{2a^{m+q}}{1+a^{2(m+q)}}\right)^2. \tag{42}$$

For $0 \le y \le a < 1$, one has $0 \le (a - y)/(1 - ay) \le a$. If $q \ge 3$, the maximizer $z_\star$ is interior, and therefore

$$H = z_\star^{\nu}\left(\frac{a-z_\star}{1-az_\star}\right)^{k} < a^{\nu+k} = a^{m+q}.$$

Monotonicity of $[2h/(1 + h^2)]^2$ proves $\mathcal{B}_{m,q}(S) < \mathcal{B}^{\mathrm{poly}}_{m,q}(S)$; for $q = 2$ the bounds coincide. This auxiliary expression is not identified with the ordinary Lanczos energy-response bound. The latter requires the conversion proved in Appendix B and is $L_{m+q}(S)$. Increasing m multiplies $\beta(y)$ by y, and increasing q multiplies it by $(a - y)/(1 - ay)$. These factors are strictly below one where the new maximum is positive, so $\mathcal{B}_{m,q}(S)$ decreases with either budget. Finally,

$$\mathcal{B}_{m,q}(S) \le 4a^{2(m+q)} \longrightarrow 0 \qquad (q \to \infty). \tag{43}$$

Every positive tolerance therefore admits a finite tail budget depending on m, S and the tolerance, independently of the full discretization dimension.

The range dependence is also explicit: increasing S increases a, enlarges $[0, a]$, and increases $(a - y)/(1 - ay)$ at each fixed y. Thus $\mathcal{B}_{m,q}(S)$ is nondecreasing in S.

B Ordinary Lanczos and the bound comparison

This appendix proves Proposition 1 and Corollary 1.2 in exact arithmetic. The classical ingredient is the CG energy-norm Chebyshev bound (Saad, 2003, Theorem 6.29). We give the conversion to the PDE energy response explicitly, rather than identifying a bound on the inverse form g_X with a bound on R_X. The connection between Lanczos, polynomial Galerkin approximation and Gaussian quadrature is standard (Golub & Meurant, 1994).

B.1 From the classical CG bound to the energy response

Proof of the upper bound in Proposition 1. Fix an instance and suppress X. The cases $\rho = 0$, $s = 0$, and exact termination are immediate. Otherwise the two phase inequalities equation 26, after scaling the lifting coordinate by $r_{0,X}^{-1/2}$, imply

$$K = \begin{bmatrix} c/r_{0,X} & b^\top/\sqrt{r_{0,X}} \\ b/\sqrt{r_{0,X}} & T \end{bmatrix}, \qquad 0 \preceq K \preceq I. \tag{44}$$

Thus $C_s = I + sK$ is positive definite with spectrum in $[1, 1+s]$. Write $e_0 = (1, 0, \dots, 0)^\top$ for its lifting coordinate. The Schur complement gives

$$F_s := e_0^\top C_s^{-1} e_0 = \frac{r_{0,X}}{r_{0,X} + sc - s^2 b^\top (I + sT)^{-1} b} = \frac{r_{0,X}}{R(s)}. \tag{45}$$

If Q_N is the original N-step Lanczos basis, then for $s > 0$

$$\mathcal{K}_{N+1}(C_s, e_0) = \operatorname{span}\{e_0, (0,b), (0,Tb), \dots, (0,T^{N-1}b)\} = \operatorname{range}\begin{bmatrix} 1 & 0 \\ 0 & Q_N \end{bmatrix}. \tag{46}$$

Indeed, multiplication by K adds one power of T in the free coordinates and changes only lower-degree terms and the lifting coordinate. Projection onto this space has Schur complement $R_N^{\mathrm{L}}(s)/r_{0,X}$. Hence the CG approximation after $j = N+1$ steps, started at zero, has inverse form

$$F_{s,j} = \frac{r_{0,X}}{R_N^{\mathrm{L}}(s)}. \tag{47}$$

The extra dimension here is the fixed lifting coordinate; the original reduced operator J_N still has dimension N.

Let $x = C_s^{-1} e_0$ and let x_j be this Galerkin approximation. Orthogonality yields

$$F_s - F_{s,j} = \|x - x_j\|_{C_s}^2, \qquad F_s = \|x\|_{C_s}^2.$$

For a positive definite matrix with spectrum in $[1, 1+s]$, the classical Chebyshev estimate is

$$\frac{\|x - x_j\|_{C_s}}{\|x\|_{C_s}} \le \frac{1}{\mathrm{T}_j(1 + 2/s)} = \frac{2y^j}{1 + y^{2j}}, \qquad y = \frac{\sqrt{1+s} - 1}{\sqrt{1+s} + 1}, \tag{48}$$

where T_j is the first-kind Chebyshev polynomial. Consequently, with $t = (F_s - F_{s,j})/F_s$ and $\delta_j = 2y^j/(1 + y^{2j}) < 1$,

$$0 \le \frac{|R_N^{\mathrm{L}}(s) - R(s)|}{R(s)} = \frac{F_s}{F_{s,j}} - 1 = \frac{t}{1-t} \le \frac{\delta_j^2}{1 - \delta_j^2} = \frac{4y^{2N+2}}{(1 - y^{2N+2})^2}. \tag{49}$$

The last expression increases with y, and y increases with s. Its maximum on $[0, S]$ is $L_N(S)$. □

B.2 Attainability on the complete form class

The following finite-dimensional construction establishes sharpness of the uniform bound. It does not assert that the equality case is realizable by a particular two-phase Q1 geometry.

Attainability on the abstract form class. Set $M = N + 1$. For $j = 0, \dots, M-1$, define

$$\theta_j = \frac{(2j+1)\pi}{2M}, \quad x_j = \cos\theta_j, \quad \lambda_j = \frac{1 + x_j}{2}, \quad \mu_j = \frac{2\sin^2\theta_j}{M}. \tag{50}$$

For $N \ge 1$, these positive weights satisfy

$$\sum_j \mu_j = 1, \qquad \sum_j \frac{\mu_j}{\lambda_j} = \sum_j \frac{\mu_j}{1 - \lambda_j} = 4, \qquad \eta_j := \frac{\mu_j}{\lambda_j(1 - \lambda_j)} = \frac{8}{M}. \tag{51}$$

On $W = \mathbb{R} \oplus \mathbb{R}^M$, use lifting $(1, 0)$ and variation space $\{0\} \oplus \mathbb{R}^M$. Represent the forms by

$$A_0^{\mathrm{full}} = I_{M+1}, \qquad A_X^{\mathrm{full}} = \begin{bmatrix} 1/2 & b^\top \\ b & T \end{bmatrix}, \qquad T = \operatorname{diag}(\lambda_j), \quad b_j = \sqrt{\mu_j/8}. \tag{52}$$

Both phase Schur complements vanish by equation 51. Therefore $0 \preceq A_X^{\mathrm{full}} \preceq I$, so this is an abstract form instance with $r_{0,X} = 1$, $c = 1/2$, $\rho = 1/8$, and $d_\infty = e_- = 0$.

To compute its Lanczos error, recall the polynomial Galerkin identity

$$R_N^{\mathrm{L}}(s) - R(s) = s^2 \rho \min_{\substack{p \in \mathbb{P}_N \\ p(-1/s)=1}} \int \frac{p(\lambda)^2}{1 + s\lambda}\, d\mu(\lambda). \tag{53}$$

The minimizing residual is orthogonal under μ to every polynomial of degree at most $N-1$: admissible variations have the form $(1+s\lambda)h(\lambda)$, and differentiating the quadratic objective gives $\int ph\, d\mu = 0$.

Let U_N be the second-kind Chebyshev polynomial, with $U_N(\cos\theta) = \sin((N+1)\theta)/\sin\theta$. The nodes x_j are the roots of T_M, and $\mathrm{T}'_M = MU_N$. Polynomial interpolation at simple roots gives

$$\sum_j \frac{h(x_j)}{\mathrm{T}'_M(x_j)} = 0 \qquad (\deg h \le M-2).$$

Since $U_N(x_j) = (-1)^j/\sin\theta_j$ and $\mu_j U_N(x_j) = 2/\mathrm{T}'_M(x_j)$, the normalized polynomial

$$p_s(\lambda) = \frac{U_N(2\lambda - 1)}{U_N(-1 - 2/s)} \tag{54}$$

is precisely the Galerkin residual in equation 53. Moreover, at every support point, $\lambda_j(1-\lambda_j)U_N(x_j)^2 = 1/4$. Using the positive response representation equation 29 with $d_\infty = e_- = 0$ therefore gives

$$\frac{R_N^{\mathrm{L}}(s) - R(s)}{R(s)} = \frac{s^2}{4(1+s)U_N(1+2/s)^2} = \frac{4y^{2N+2}}{(1-y^{2N+2})^2}. \tag{55}$$

The last equality follows from $U_N(1+2/s) = y^{-N}(1-y^{2N+2})/(1-y^2)$ and $s = 4y/(1-y)^2$. Thus the same abstract form instance attains the pointwise bound for every $s > 0$, and in particular attains $L_N(S)$ at $s = S$. □

B.3 The explicit improvement factor

Proof of Corollary 1.2. Dividing equation 13 by equation 17 gives equation 18. It remains to establish a sufficient condition for strict improvement. With $\nu = m+2$, $k = q-2$, and $N = \nu + k$, the proof of Theorem 1 gives

$$\begin{aligned} \frac{H}{a^N} &= \max_{0 \le t \le 1} t^\nu \left(\frac{1-t}{1-a^2 t}\right)^k \\ &\le \max_{0 \le t \le 1} \frac{t(1-t)}{1-a^2 t} = \frac{1}{(1+\sqrt{1-a^2})^2} < \frac{1}{(2-a)^2}. \end{aligned} \tag{56}$$

Here both factors inside the maximum lie in $[0,1]$, $\nu, k \ge 1$, and the strict inequality uses $0 < a < 1$. For $S \ge 4$ one has $a \ge (3-\sqrt{5})/2$, and

$$a(2-a)^2 - 1 = (a-1)(a^2 - 3a + 1) \ge 0.$$

Hence $H < a^{N+1}$ and

$$\mathcal{B}_{m,q}(S) = \frac{4H^2}{(1+H^2)^2} < 4a^{2N+2} < \frac{4a^{2N+2}}{(1-a^{2N+2})^2} = L_N(S). \tag{57}$$

This proves $0 < \Gamma_{m,q}(S) < 1$. □

The conditions are sufficient rather than necessary; the explicit ratio determines the comparison outside this regime. No unconditional ordering for all $S > 0$ is asserted.

The two upper bounds remain valid on every admissible subclass or target distribution, whereas sharpness of L_N is established only for the complete form class. The comparison neither guarantees that a trained network finds the existence witness nor compares the costs of constructing the two reduced matrices.

C INTERPRETING THE BOUND CURVES AND THEIR TURNING POINTS

Figure 1(d,e) fixes $m = 24$. Panel (d) varies the contrast-parameter endpoint S of the whole interval $s \in [0, S]$; panel (e) varies q for fixed intervals. Both vertical axes show the bound-tightening factor $L_{m+q}(S)/\mathcal{B}_{m,q}(S) = \Gamma_{m,q}(S)^{-1}$ on a logarithmic scale. The formulas below analyze Γ; its reciprocal has the same stationary points with maxima and minima interchanged. These curves compare analytic guarantees, not learned errors.

Exact derivatives of the ratio. To analyze a change of trend, extend the positive budgets continuously and take $q > 2$. Use $a, z_\star, H$ from equation 13, and write

$$t = -\log a, \quad C = m+q+1, \quad h = -\log H, \quad w_\star = \frac{a - z_\star}{1 - a z_\star}, \quad v = -\log z_\star, \quad u = -\log w_\star. \tag{58}$$

All these logarithmic quantities are positive. The two bounds and their ratio have the equivalent expressions

$$\mathcal{B}_{m,q} = \frac{1}{\cosh^2 h}, \qquad L_{m+q} = \frac{1}{\sinh^2(Ct)}, \qquad \Gamma_{m,q} = \frac{\sinh^2(Ct)}{\cosh^2 h}. \tag{59}$$

The maximizer in the proof satisfies

$$\log H = \max_{0<y<a} \left[(m+2) \log y + (q-2) \log \frac{a-y}{1-ay} \right].$$

Its derivative with respect to y vanishes at $z_\star$. Differentiating the optimized value therefore gives $\partial_m \log H = \log z_\star$ and $\partial_q \log H = \log w_\star$. For fixed S, it follows that

$$\left. \frac{\partial \log \Gamma}{\partial m} \right|_q = 2\big[t \coth(Ct) - v \tanh h \big], \tag{60}$$

$$\left. \frac{\partial \log \Gamma}{\partial q} \right|_m = 2\big[t \coth(Ct) - u \tanh h \big]. \tag{61}$$

Thus increasing the computed prefix and increasing the learned continuation are different comparisons, even when both increase $N = m + q$.

Where a curve changes direction. At fixed q, varying N is equivalent to varying m, and a stationary point solves

$$v \tanh h = t \coth(Ct). \tag{62}$$

The ratio increases when the left side is smaller and decreases when it is larger. At fixed m, the corresponding condition replaces v by u in equation 62. A maximum occurs when the derivative changes from positive to negative; this is a turning point, not necessarily an inflection point. Its location depends jointly on m, q, S, rather than on one universal contrast threshold. The actual budgets are integers, so adjacent integer values determine the discrete maximum.

Both $\mathcal{B}$ and L decrease as either budget increases, but their quotient need not decrease. Its logarithmic derivative is the difference of their relative rates of change. For example, at fixed budgets and increasingly large S,

$$\mathcal{B}_{m,q}(S) \to 1, \qquad L_{m+q}(S) \sim \frac{S}{4(m+q+1)^2}, \qquad \Gamma_{m,q}(S) \sim \frac{4(m+q+1)^2}{S}. \tag{63}$$

In this regime, reducing the ordinary Lanczos bound can initially increase the ratio while both absolute bounds improve. The existence bound can then decrease faster at larger budgets, reversing the direction. These fixed-budget asymptotics explain the mechanism; finite-parameter turning points must be computed from equation 60 or equation 61. A small ratio alone is not a small absolute-error guarantee.

Numerical evaluation and checks. We evaluate $\log \mathcal{B}$, $\log L$ and their difference before exponentiation. The stable discriminant identity

$$\Delta^2 - 4\nu^2 a^2 = (1-a^2)\big[4\nu k + (\nu-k)^2(1-a^2)\big]$$

avoids cancellation; `log1p` and `expm1` handle factors close to zero or one. Independent 90-digit evaluations check the bounds and both analytic derivatives, with the derivatives also compared against numerical differentiation at that precision. The plotting record includes every sampled value, the stationary-point equations and their solved locations for the earlier fixed-q comparison. For each interval shown in Figure 1(e), direct checks give $\Gamma^{-1}_{m,q+1} > \Gamma^{-1}_{m,q}$ over the plotted integer budgets. This observation is restricted to those curves and is not asserted as a global monotonicity theorem.

D THE COMPLETION DECODER

D.1 ENERGY-COEFFICIENT RECURRENCE AND THE REPRESENTED CLASS

Write the prefix diagonals as $\alpha_1, \dots, \alpha_m$ and its couplings as $\beta_1, \dots, \beta_m$, including the measured terminal coupling. For $\rho > 0$, eliminating the fixed prefix coordinates of the two energy matrices in equation 20 gives

$$\begin{aligned} p_{-,1} &= \alpha_1 - \rho/c_X, & p_{+,1} &= 1 - \alpha_1 - \rho/(r_{0,X} - c_X), \\ p_{-,k} &= \alpha_k - \beta_{k-1}^2/p_{-,k-1}, & p_{+,k} &= 1 - \alpha_k - \beta_{k-1}^2/p_{+,k-1} \quad (2 \le k \le m). \end{aligned} \tag{64}$$

Suppose these coefficients are positive and the terminal coupling is feasible. Set $p_- = p_{-,m}$, $p_+ = p_{+,m}$. With A_-, A_+ from the main text, $(e_m^{(m)})^\top A_-^{-1} e_m^{(m)} = 1/p_-$ and $(e_m^{(m)})^\top A_+^{-1} e_m^{(m)} = 1/p_+$. For a trial first diagonal α and first tail coordinate one, the two energy expressions are $v^\top A_- v + 2\beta_m v_m + \alpha$ and $v^\top A_+ v - 2\beta_m v_m + 1 - \alpha$. Their minima over $v \in \mathbb{R}^m$ occur at $v = -\beta_m A_-^{-1} e_m^{(m)}$ and $v = \beta_m A_+^{-1} e_m^{(m)}$, leaving $\alpha - \beta_m^2/p_-$ and $1 - \alpha - \beta_m^2/p_+$, respectively. Both minima must be nonnegative. Thus

$$\ell = \beta_m^2/p_-, \qquad u = 1 - \beta_m^2/p_+, \qquad w_0 = u - \ell \ge 0.$$

The first added diagonal and updated coefficients are

$$\widehat{\alpha}_{m+1} = \ell + t_0 w_0, \qquad p_- \leftarrow t_0 w_0, \qquad p_+ \leftarrow (1-t_0) w_0, \tag{65}$$

while $\widehat{\beta}_m = \beta_m$. For the subsequent coordinates $j = 1, \dots, q-1$, use the current coefficients to set

$$\begin{aligned} \widehat{\beta}_{m+j}^2 &= \eta_j \frac{p_- p_+}{p_- + p_+}, & \widehat{\alpha}_{m+j+1} &= \eta_j \frac{p_+}{p_- + p_+} + t_j(1-\eta_j), \\ p_- &\leftarrow t_j(1-\eta_j), & p_+ &\leftarrow (1-t_j)(1-\eta_j). \end{aligned} \tag{66}$$

Fractions $t_j, \eta_j \in [0,1]$ keep both coefficients nonnegative. All new couplings are the nonnegative square roots. In the implementation, $t_0 = \sigma(\xi_1)$, $\eta_j = \sigma(\xi_{2j})$ and $t_j = \sigma(\xi_{2j+1})$, where $\sigma(t) = (1+e^{-t})^{-1}$, before the endpoint assignments below.

Why this chart contains the existence construction. For any feasible next diagonal α and coupling $b > 0$, keeping both energy matrices nonnegative gives

$$b^2/p_- \le \alpha \le 1 - b^2/p_+.$$

Consequently the inverse coordinates are

$$\eta = \frac{b^2(p_- + p_+)}{p_- p_+} \in [0,1], \qquad t = \frac{\alpha - b^2/p_-}{1-\eta} \in [0,1] \quad (\eta < 1). \tag{67}$$

The first diagonal similarly has $t_0 = (\alpha - \ell)/w_0$ when $w_0 > 0$. At $\eta = 1$ or $w_0 = 0$, the diagonal is uniquely fixed and both updated coefficients vanish. Such a coordinate permits no further positive coupling and ends the chain.

An irreducible positive semidefinite tridiagonal matrix has positive definite proper leading principal blocks: a null vector of such a block would extend to a null vector of the full matrix, and the next nonzero coupling forces its last coordinate, and recursively every coordinate, to vanish. Thus the two energy matrices of an irreducible admissible completion have positive coefficients in this recurrence until their final coordinate. Applying equation 67 successively recovers its fractions. Theorem 1 constructs such a Jacobi completion; if its cyclic support is smaller, the exact shorter chain suffices. This establishes inclusion in the closed chart, allowing termination. A finite sigmoid produces interior fractions; unsupplied boundary values are limits of that parameterization.

Reduced evaluation. For a completed matrix of dimension $N = m + q$, the implementation evaluates the equivalent continued fraction

$$d_N(s) = 1 + s\widehat{\alpha}_N, \qquad d_k(s) = 1 + s\widehat{\alpha}_k - \frac{s^2\widehat{\beta}_k^2}{d_{k+1}(s)}, \qquad \widehat{R}(s) = r_{0,X} + sc_X - \frac{s^2\rho}{d_1(s)}. \tag{68}$$

The compiled CPU implementation solves the same tridiagonal system. Zero logits initialize the unconstrained fractions at $1/2$; the resulting matrix still depends on the computed instance prefix.

D.2 Known zero-energy conditions

The minima in equation 22 are attained because the forms are nonnegative quadratics on a finite-dimensional affine space. Let w_* minimize a_X over $u_{0,h,X} + V_h$. For $s > 0$,

$$sd_\infty \le R_X(s) \le sd_\infty + a_0(w_*, w_*),$$

which proves $R_X(s) = sd_\infty + O(1)$. The same argument applies to the reduced energy and gives $\widehat{R}(s) = s\widehat{d}_\infty + O(1)$.

The two predicted minimum energies in equation 23 are

$$\widehat{d}_\infty = c_X - \rho(e_1^{(N)})^\top \widehat{J}^{-1} e_1^{(N)}, \qquad \widehat{e}_- = r_{0,X} - c_X - \rho(e_1^{(N)})^\top (I - \widehat{J})^{-1} e_1^{(N)}. \tag{69}$$

They are obtained by setting $t = 1$ in the two reduced energy forms and minimizing over z. The first also equals $\lim_{s\to\infty} \widehat{R}(s)/s$; the second is the predicted minimum complementary energy. Their nonnegativity follows from equation 20. On the coupled block, $\widehat{J}$ and $I - \widehat{J}$ are positive definite. Taking determinants gives

$$\det\left(\widehat{J} - \frac{\rho}{c_X} e_1^{(N)}(e_1^{(N)})^\top\right) = \det(\widehat{J})\frac{\widehat{d}_\infty}{c_X},$$
$$\det\left(I - \widehat{J} - \frac{\rho}{r_{0,X} - c_X} e_1^{(N)}(e_1^{(N)})^\top\right) = \det(I - \widehat{J})\frac{\widehat{e}_-}{r_{0,X} - c_X}.$$

Eliminating the first $N - 1$ coordinates of either energy matrix factors its determinant into positive leading coefficients and the final coefficient. Thus setting that final coefficient to zero is equivalent to setting the corresponding predicted minimum energy to zero, as stated in equation 24. Known zeros of the exact minimum energies supply the additional information used to select these cases. In the proof of Theorem 1, inverse-moment preservation also preserves the true values of both minimum energies.

In the conductivity benchmark, let S_0, S_1 indicate whether phase 0 or 1 connects the two electrodes in the assembled Q_1 graph. The occupied elements are $\eta_{X,e} < 1$ and $\eta_{X,e} > 0$, respectively. Elements sharing a node are connected, giving eight-neighbour contact on the square element array. If a phase has no component touching both electrodes, an admissible field can be constant on each component; that phase's minimum energy is zero. Thus $\neg S_1$ gives $d_\infty = 0$ and $\neg S_0$ gives $e_- = 0$ for this discrete problem.

With $j_\star = q - 1$, the implemented two-endpoint rule is

$$\eta_{j_\star} \leftarrow 1 \quad \text{if } \neg S_0 \wedge \neg S_1, \qquad t_{j_\star} \leftarrow \begin{cases} 0, & \neg S_1, \\ 1, & S_1 \wedge \neg S_0, \\ \sigma(\xi_{2j_\star+1}), & S_0 \wedge S_1. \end{cases} \tag{70}$$

Both zero slacks require $q \ge 2$. These assignments constrain the energy matrices; they do not place eigenvalues of $\widehat{J}$ at 0 or 1. The ablation without endpoint equality assignments uses the unmodified sigmoid fractions, while retaining the same network inputs. The reported conductivity timings include extraction of these indicators. No analogous inexpensive test is assumed for a general elliptic discretization.

E EXPERIMENTAL PROTOCOL

E.1 PDE BENCHMARK, DATA AND EVALUATION

Discretization and data. We use the Q1 conductivity problem with opposing electrodes at potentials zero and one and natural insulation on the transverse boundary. Responses are the minimized discrete energies, evaluated against independent references on the same mesh. The original training and validation sets contain 512 and 128 structures at 128^2, generated from isotropic spectral random fields with correlation lengths $(0.06, 0.06)$ and phase fraction $\phi \in [0.3, 0.7]$. The original training queries are $s \in \{1, 4, 9, 29, 99, 299, 999\}$.

Accuracy evaluation. The contrast comparison and in-domain curves use the same 40 held-out IID structures. The in-domain queries are $s \in \{1, 2, 4, 9, 19, 29, 99, 199, 299, 999\}$; $2, 19, 199$ are absent from training. Contrast extrapolation uses $s \in \{2999, 9999, 99999\}$. For each query, we compute $e = |\widehat{R} - R|/R$, average errors over two loading directions and seeds $11, 23, 37$ within each structure, and report the mean and sample standard deviation across structures. Deterministic methods average directions only. The reported standard deviation describes variation across structures.

Query timing. The cost comparison uses twenty fixed IID geometries with x loading at 4096^2. Prefixes and references are recomputed on each mesh; the MLP uses frozen seed-11 weights. Timed calls include instance-dependent assembly, numerical iterations, connectivity features, network evaluation and response reconstruction. Coefficient arrays are resident; file I/O, reusable reference-operator setup, model loading, label generation, training and solver calibration are outside these initialized query costs. Solver budgets are calibrated offline against the MLP's attained response error and frozen before timing; timed calls access no reference answers. Discrete iteration budgets yield close but unequal errors. Each geometry contributes the median of three complete calls, and reported means and sample standard deviations use the twenty medians. Measurements use an AMD Ryzen 9 9955HX, float64 numerical kernels and float32 networks, with one BLAS/network thread and one/eight FFT workers at 4096^2.

E.2 COMPUTED INFORMATION, MODELS AND TRAINING

Numerical inputs and model families. Prefix methods use $m = 24$ load-specific Lanczos steps. The 56-entry context contains phase fraction, loading direction, $c_X/r_{0,X}$, $\rho/r_{0,X}$, 24 diagonals, 23 internal couplings, the terminal coupling, logarithms of two prefix energy coefficients and two Q1 phase-connectivity indicators. Continuation heads output seven coordinates for a four-node tail, retaining J_m, β_m and using the same tail for every query. The MLP has three width-128 SiLU hidden layers; sequence Transformer and FNO encoders use width 48 and three layers. Ordinary Lanczos at $m = 24$ and $m = 28$ provides deterministic comparisons; the latter performs four additional PDE-space iterations.

Response and field controls. Free scalar-residual heads predict an unrestricted correction to R_m, using the same numerical context and six query features: $\log(1 + s)/\log 1000$, $s/(1 + s)$, L/U, $(U - L)/U$, $r_{0,X}/U$ and $(R_m - U)/U$, where L, U are numerical energy bounds. Direct field models instead receive the coefficient field, two coordinate channels, two electrode masks and the first two query features, without a Lanczos prefix. The field FNO has width 16, three Fourier blocks and eight modes; the field Transformer uses a 128^2 adaptive-pooling adapter, 8^2 patches, width 64, three layers and four attention heads.

Decoder ablation. Table 2 independently enables or removes the two phase-energy inequalities and the applicable known zero-energy equalities, yielding four combinations. All variants retain the same numerical inputs, MLP architecture, prefix, terminal coupling and four-node tridiagonal representation. The energy-only variant omits endpoint assignments. The endpoint-only variant enforces the known zero equalities without the energy inequalities. The unconstrained variant imposes neither: its four diagonals and three internal couplings are unrestricted real coordinates. Free coordinates use a fixed feasible initialization plus 0.05 times the network output; this affine parame-

terization imposes no bound on the learned coordinates. Connectivity indicators remain available to every variant.

Training. All four ablation variants use the same training structures, normalizer and seeds. Training uses AdamW with learning rate 10^{-3}, weight decay 10^{-4}, cosine decay to 2×10^{-5}. The loss is multiplied by 10^4 for optimization, with gradient clipping at 5×10^4; validation selection uses the corresponding unscaled weighted objective. The full-decoder checkpoints are reused, and the other three variants are fitted separately.