Online Resource 1

Mathematical and numerical material

A scale-covariant pre-solve screening algorithm for scalar finite-difference discretizations on positive nonuniform meshes

A. S. Krylov

Numerical Algorithms

Faculty of Computational Mathematics and Cybernetics,

Lomonosov Moscow State University, Moscow, Russia; kryl@cs.msu.ru

This PDF is Online Resource 1 and contains implementation-level benchmark definitions and secondary numerical tables supporting the main article. The definitions of the Mellin audit, the B, A, and M feature blocks, the $B + A + M$ workflow, and every theorem needed to interpret the main claims remain in the article. The optional derived contrast block C is defined here because it is a secondary representation ablation rather than part of the method. Online Resource 2 is the separately uploaded versioned ZIP archive containing the executable scripts and all row-level and aggregate data.

S1 Complete mesh-generator specification

Put $s_j = j/N$, $L = b - a$, and

$$q_c(s) = \frac{1}{2}\left(1 + \frac{\tanh(c(2s-1))}{\tanh c}\right), \qquad \mathcal{J}_\alpha(y)_j = y_j + \alpha \min(\Delta y_{j-1}, \Delta y_j)\sin(1.731j + 0.37),$$

at interior indices, with fixed endpoints and $\Delta y_j = y_{j+1} - y_j$. A piecewise grid is uniform on each listed subinterval with the stated number of intervals.

Euler family on $[1, e]$. With $t_a = 0$ and $t_b = 1$, the six grids are

1. uniform in x: $x_j = 1 + (e-1)s_j$;
2. uniform in $t = \log x$: $x_j = e^{s_j}$;
3. smoothly graded in t: $x_j = e^{q_{1.9}(s_j)}$;
4. piecewise uniform in t with normalized breakpoints $(0, 0.25, 0.72, 1)$ and counts $(\lfloor N/3\rfloor, \lfloor N/3\rfloor, N - 2\lfloor N/3\rfloor)$;
5. log-jittered: $t = \mathcal{J}_{0.33}(s)$ and $x = e^t$;
6. power graded: $x_j = 1 + (e-1)s_j^{2.4}$.

Reaction–diffusion family.
On $[a, b] = [0.1, 1]$, let $\delta_{\text{lay}} = \sqrt{\varepsilon}$ and $\tau = \min(0.24L, 2.2\delta_{\text{lay}} \log\max(N, 4))$. The six grids are uniform, logarithmic, two-sided Shishkin, smoothly graded $x = a + Lq_{2.6}(s)$, jittered Shishkin, and the one-sided negative control $x = a + Ls^{2.8}$. The Shishkin partition and counts are

$$(a, a+\tau, b-\tau, b), \qquad (\lfloor N/4\rfloor, N - 2\lfloor N/4\rfloor, \lfloor N/4\rfloor);$$

the jittered version applies $\mathcal{J}_{0.24}$.

Convection–diffusion family. Let $\delta_{\mathrm{lay}} = \varepsilon$ and $\tau = \min(0.48L, 2.2\delta_{\mathrm{lay}} \log\max(N, 4))$. The six generators are uniform, logarithmic, right-Shishkin, right-smooth, jittered right-Shishkin, and wrong-left. The Shishkin partition is $(a, b-\tau, b)$ with counts $(N - \lfloor N/3 \rfloor, \lfloor N/3 \rfloor)$; its jittered version applies $\mathcal{J}_{0.27}$. The last two analytic maps are

$$x = a + L\left[1 - \frac{e^{4.2(1-s)} - 1}{e^{4.2} - 1}\right], \qquad x = a + Ls^3.$$

Table S1: Benchmark inventory. Each one-dimensional candidate is evaluated at three tolerances; the 120 FNO packet problems are evaluated at four mesh sizes.

Block	Problems/cases	Decisions	Grouping variable
Three 1D families	1764 grids	5292	physical parameter level
Omitted-mode Euler shift	588 grids	1764	withheld Euler frequency
Barrier test	80 grids	–	frequency and mesh family
Tensor-product test	140 grids	–	modal pair and mesh family
Ordinary FNO proposals	480 per method	1440 per method	complete packet
Rare FNO tail test	360 grids	1080	complete tail group

S2 FNO architecture, data ranges, and training

The ordinary packet is parameterized by center c, integer carrier k, amplitude a, and width σ. Ordinary training draws use

$$c \in [0.22, 0.78], \quad k \in \{8, \dots, 18\}, \quad a \in [0.18, 0.42], \quad \sigma \in [0.06, 0.13].$$

The training and validation sets contain 768 and 160 independently drawn packets. The 60 interpolation tests use the same ranges; the 60 extrapolation tests use $c \in [0.10, 0.20] \cup [0.80, 0.90]$, $k \in \{20, \dots, 28\}$, and $\sigma \in [0.035, 0.06]$.

On 64 uniform t samples, the three input channels are

$$f/\|f\|_\infty, \qquad 2t - 1, \qquad \frac{\log_{10}\|f\|_\infty - \mu_{\log f}}{s_{\log f}}.$$

The real-valued FNO has width 12, two zero-padded Fourier layers, 12 retained modes, a scalar projection, and 7285 trainable parameters. Adam uses batches of 64, gradient norm clipping at 5, a cosine learning rate beginning at 2.5×10^{-3}, at most 220 epochs, and early stopping after 42 validation epochs without an improvement of 10^{-6}. The comparison MLP has one 96-unit tanh layer and 12544 parameters; it uses Adam with initial rate 2×10^{-3}, batch size 64, at most 300 iterations, and early stopping.

Both networks predict the smooth monitor $\rho_\star(t) = 1 + 4\exp[-(t-c)^2/(2(1.5\sigma)^2)]$. The fixed smoothing and validity projection used in the main-text stress test is

$$\widehat{g}_\theta = \mathrm{clip}\{\mathcal{S}(g_\theta - \overline{g}_\theta), -1.35, 1.35\}, \qquad \rho_\theta = e^{\widehat{g}_\theta}, \qquad \int_0^{t_i} \rho_\theta(s)\, ds = \frac{i}{N}\int_0^1 \rho_\theta(s)\, ds.$$

Consequently $x_i = e^{t_i}$ is positive and strictly increasing. The learned proposer is therefore separated from the deterministic audit, which is computed only after the finite-difference rows have been assembled.

Twenty-four rare scale-structured cases (3.1% of the ordinary training count) and eight validation cases are appended during training. Their ranges are $T \in [1.4, 2.4]$, $k \in \{8, \dots, 18\}$, $\beta \in [-1.25, 1.25]$, amplitude in $[0.12, 0.28]$, and phase in $[0, 2\pi]$. The external tail law uses

$T \in [3, 4.5]$, $k \in \{14, \ldots, 24\}$, and β within 0.08 of ± 1.4 or ± 1.8. Fifteen groups fit the tail screen, 15 disjoint groups calibrate it, and 60 new groups form the external test. The random seed is 20260805 throughout.

Table S2 uses the same data draws, optimizer, stopping rule, and external tests. The 8/12/16-mode alternatives were fixed before the tail comparison. The zero-shot row removes all 24 rare training and eight rare validation cases while retaining the ordinary data. Ordinary performance is stable, whereas zero-shot tail performance deteriorates sharply. The reported tail failure is therefore residual extrapolation after limited risk-informed training, not an artefact of selecting an anomalously weak FNO.

All complex-valued Mellin responses are evaluated in NumPy `complex128` arithmetic from the real logarithms $\delta_{jk} = \log(x_k/x_j)$. Conjugate modes are stored and evaluated explicitly; magnitudes agree to rounding, while the phase feature uses the principal argument and its circular distance to zero. The near-zero gate and all exponential-clipping counts are recorded in machine-readable diagnostics rather than silently discarded.

Table S2: Prespecified FNO controls. Error ratios compare the proposed mesh with a log-uniform mesh for the same problem and N.

Modes	Parameters	Rare train	Ordinary median	Ordinary Q0.9	Tail median	Tail Q0.9
8	4981	24 cases	0.284	0.505	1.681	2.021
12	7285	24 cases	0.306	0.560	1.097	1.296
16	9589	24 cases	0.409	0.782	1.789	3.056
12	7285	none (zero-shot)	0.307	0.579	2.360	11.785

Classifier fitting and tuning

All numeric variables are standardized inside each fitting fold and family is one-hot encoded with unseen levels ignored. Logistic regression uses balanced class weights, at most 3000 iterations, and $C \in \{0.1, 1, 10\}$. Histogram gradient boosting uses learning rate 0.07, minimum leaf size 20, maximum leaf count in $\{7, 15\}$, and ℓ_2 regularization in $\{0, 1\}$. Random forests use 300 trees, balanced-subsample weights, minimum leaf size in $\{2, 6\}$, and maximum feature count in $\{\sqrt{p}, 0.8p\}$. Hyperparameters maximize ROC AUC in a three-fold stratified group search on fitting groups only. The outer five-fold stratified group split, the subsequent 25% group calibration split, and every model use seed 20260805. The fixed resolution rule and two-grid score have no tuned shape parameters; only their safe-score thresholds are calibrated on the same groups and at the same nominal α.

S3 Secondary validation and complete ablations

Guarantee coverage in the designed benchmarks

Table S3 reports eligibility within the designed experiments, not an estimate of prevalence in applications. A case falling outside a stronger row remains eligible only for the lower diagnostic or calibrated branch. In particular, failure to certify is not rejection.

Table S3: Observed coverage of the guarantee hierarchy.

Level	Eligible cases	Outcome
Deterministic row audit	5292 main-corpus decisions	Response available for every assembled positive-mesh row.
Dilation-covariant Euler subclass	588 grids (1764 decisions)	Structural theorem applies to covariantly generated homogeneous rows.
Exact tensor modal identity	28 uniform-log cases of 140 tensor cases	Maximum discrepancy 1.1×10^{-14}.
Row-sum barrier experiment	80 of 80 test matrices	No bound violation; safe-case acceptance 55.1–60.0% at the three reported budgets.
Rare finite-spectrum FNO tail	360 of 360 assembled rows	Bound valid but too conservative for direct acceptance; calibrated branch used.
Five-point/nonmonotone or undeclared cases	No hard certificate	Diagnostic/reject/defer branch only.

Conditional overlap of the additive and Mellin blocks

The analysis below is label-free. Tolerance duplicates are removed, complete physical-parameter groups are withheld in five folds, and each transformed Mellin summary is predicted from the conventional block without tolerance and the full additive block. High conditional R^2 explains why flexible finite-sample classifiers obtain little or negative gain from adding M; it does not alter the covariance theorem or exact modal identity.

Table S4: Redundancy of corresponding A/M summaries. The last column is grouped out-of-fold R^2 for predicting M from $B + A$ without labels.

Summary	Spearman $\rho(A, M)$	Conditional R^2
Defect maximum	0.988	0.984
Defect Q0.9	0.998	0.995
Inverse-response Q0.9	0.996	0.983
Phase Q0.9	0.983	0.962
Roughness Q0.9	0.983	0.985

The five canonical correlations between the standardized blocks are 0.999, 0.990, 0.986, 0.979, and 0.912. Thus the overlap is multivariate rather than an artefact of one dominant summary, while a nonzero residual subspace remains.

Optional contrast block

For completeness, write $T(v) = \log_{10} \max(v, 10^{-14})$. The predeclared seven-dimensional secondary block is

$$\begin{aligned} C = (&T(M_{D_{\max}}) - T(A_{D_{\max}}), T(M_{D_{0.9}}) - T(A_{D_{0.9}}), T(M_{I_{0.9}}) - T(A_{I_{0.9}}), \\ &T(M_{R_{0.9}}) - T(A_{R_{0.9}}), M_{\Phi_{0.9}} - A_{\Phi_{0.9}}, |T(M_{D_{\max}}) - T(A_{D_{\max}})|, \\ &|T(M_{D_{0.9}}) - T(A_{D_{0.9}})|). \end{aligned} \tag{S1}$$

The first four entries are stabilized log-ratios. C is neither a third audit nor a guarantee. Its negative aggregate and ordinary-FNO ablations are reported to prevent selecting only the small favourable tail result.

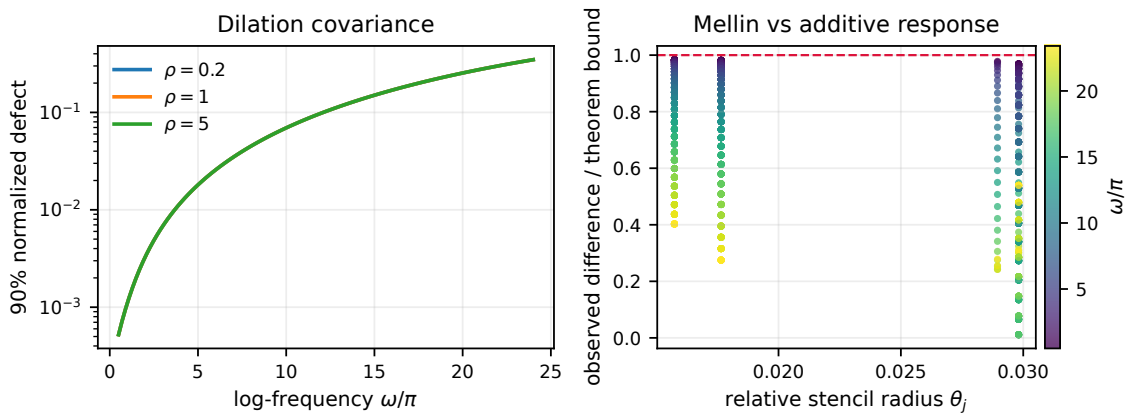

Figure S1: Multiplicative geometry. Left: invariance of the normalized Euler response under physical dilation. Right: observed Mellin–additive response difference divided by the proved second-relative-step bound.

Table S5: Joint-audit ablation. Increments compare $B+A+M$ with the two one-audit alternatives at independently calibrated operating points.

Model	UR, $B+A+M$	ΔUR vs. $B+A$	ΔUR vs. $B+M$	WR
Logistic	57.4	+2.5	+4.8	3.1
Gradient boosting	80.4	−1.4	−0.1	1.9
Random forest	68.3	−6.6	−4.9	1.1

Table S6: Explicit audit-contrast ablation. The interval is the paired problem-group bootstrap interval for the UR change from $B+A+M$ to $B+A+M+C$; every operating point is independently calibrated.

Model	UR, $B+A+M+C$	ΔUR	95% interval	WR
Logistic	54.6	−2.77	[−6.14, 0.41]	3.0
Gradient boosting	78.5	−1.92	[−3.69, −0.07]	1.6
Random forest	66.3	−1.99	[−5.02, 0.93]	1.1

Table S7: Family-level held-out gradient-boosting performance. Every feature set has its own nominal 3% safe-case calibration.

	B		$B+A$		$B+M$		$B+A+M$		$B+A+M+C$	
Family	UR	WR	UR	WR	UR	WR	UR	WR	UR	WR
Euler	75.5	1.7	83.8	1.3	82.5	1.1	82.4	1.1	79.0	1.2
Reaction–diffusion	68.8	1.0	73.6	2.4	72.4	1.7	71.2	2.2	71.6	1.9
Convection–diffusion	84.2	1.8	86.7	2.5	85.5	2.3	86.1	2.6	83.4	2.0

Table S8: Spectrum sensitivity for independently fitted and calibrated logistic screens.

	$B+M$		$B+A+M$		$B+A+M+C$	
Working exponent set	UR	WR	UR	WR	UR	WR
Misspecified $0.75z$	42.5	2.5	60.4	2.4	61.2	2.4
Nominal z	52.5	2.9	57.4	3.1	54.6	3.0
Robust band $\pm 25\%$	53.2	3.2	59.7	3.2	60.7	3.1
Robust band $\pm 50\%$	51.3	3.2	59.1	3.0	58.8	2.7

Table S9: Accuracy of learned log-coordinate mesh proposals. Ratios use the log-uniform error for the same problem and N.

Proposal	Median E_h/E_{unif}	Q0.9 E_h/E_{unif}	Median E_h
Log-uniform	1.000	1.000	3.09×10^{-2}
Local MLP	0.571	55.196	2.65×10^{-2}
FNO in $t = \log x$	0.306	0.560	1.08×10^{-2}
Reference monitor	0.238	0.475	8.05×10^{-3}

Table S10: Complete ordinary-FNO screening ablation after generator-specific grouped fitting, calibration, and testing at nominal $\alpha = 0.03$. The learned proposer is treated as an external stress test.

Screen	UR (%)	WR (%)
Logistic, B	82.1	1.5
Logistic, $B+A$	85.2	0.4
Logistic, $B+M$	85.7	1.1
Logistic, $B+A+M$	85.4	0.8
Logistic, $B+A+M+C$	84.5	0.4
Gradient boosting, B	73.8	0.4
Gradient boosting, $B+A$	75.3	0.8
Gradient boosting, $B+M$	66.4	0.4
Gradient boosting, $B+A+M$	73.1	1.1
Gradient boosting, $B+A+M+C$	71.2	1.1

Table S11: Complete rare FNO tail ablation. The M block includes the coefficient-weighted pre-solve score; every threshold uses disjoint tail calibration groups.

Screen	UR (%)	WR (%)
Logistic, B	86.5	4.9
Logistic, $B+A$	90.9	4.5
Logistic, $B+M$	90.7	2.1
Logistic, $B+A+M$	90.7	2.1
Logistic, $B+A+M+C$	91.6	2.1

Table S12: Rare FNO tail interaction. Paired group-bootstrap change caused by adding C to $B + A + M$.

Metric	Change (points)	95% lower	95% upper
Unsafe rejection	+0.96	+0.12	+1.81
Wrong rejection	0.00	0.00	0.00

Calibration-budget and timing diagnostics

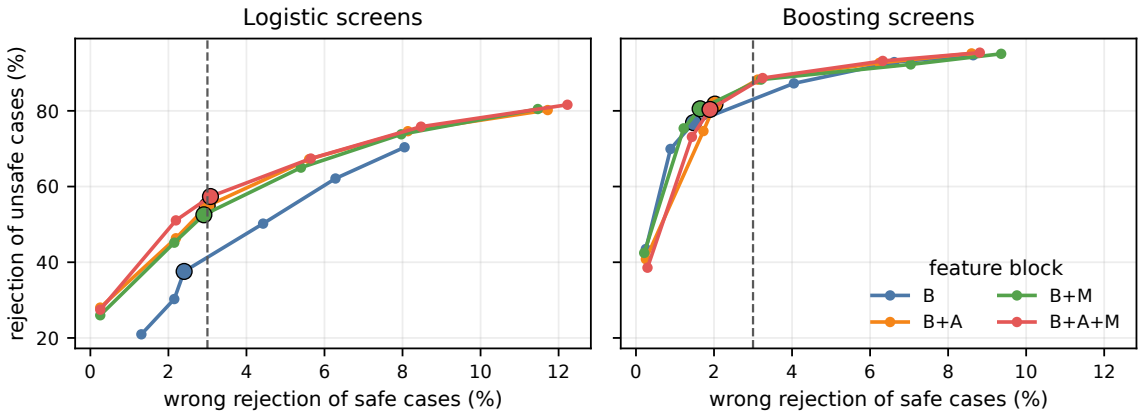

Figure S2: Realized held-out unsafe-rejection versus wrong-rejection tradeoff when the safe-class calibration budget varies over $\alpha \in \{0.01, 0.02, 0.03, 0.05, 0.075, 0.10\}$. Curves reuse the same fitted models, outer group folds, and safe calibration scores; only the conformal quantile changes. Large black-edged markers denote the predeclared $\alpha = 0.03$ operating point, and the dashed line marks 3% WR.

Table S13: Representative timing scaling from the reproducibility run. Entries are medians in milliseconds; ratios are medians of casewise ratios rather than ratios of the displayed medians. One-dimensional tolerance duplicates are removed (252 candidate discretizations per size); each two-dimensional size contains 20 mesh–mode cases. Absolute wall times depend on the Python, linear-algebra, and hardware environment.

Dimension	N	Audit (ms)	Solve (ms)	Two-grid (ms)	Audit/solve	Audit/two-grid
1D	24	0.796	0.215	1.986	3.825	0.459
1D	48	1.330	0.282	3.875	5.214	0.401
1D	96	2.424	0.432	7.377	6.486	0.378
1D	192	4.561	0.713	14.560	7.722	0.362
2D	16	0.563	0.291	—	1.900	—
2D	32	1.071	1.282	—	0.829	—
2D	64	2.286	6.147	—	0.362	—
2D	128	4.861	41.230	—	0.120	—

Continuous-band finite-net diagnostic

The five-point robust grid on $[0.75, 1.25]$ has fill distance $\eta_G = 0.0625$. Table S14 evaluates the additional derivative term in the finite-net uncertainty-band theorem of the main article on every resolution-qualified row–mode pair in the one-dimensional corpus. No exponential clipping is used in this calculation. The reported count is obtained by enumerating every interior row

and declared base mode in all 1764 one-dimensional candidates and retaining the 290247 pairs satisfying $|z_0|H_j \le 1$. The theorem itself remains valid outside it, but its computable majorant correctly becomes conservative.

Table S14: Finite-net certification diagnostic for the continuous $[0.75, 1.25]$ multiplier band. The correction is $(\eta_G M_{1,j} + \eta_G^2 \overline{M}_{2,j}/2)/s_j(z_0)$. The last column is the largest ratio of the maximum on an additional 65-point validation net to the certified bound B_j; the analytic theorem, not this denser net, supplies continuum coverage.

Population	Pairs	Median	q90	q99	Validation/bound max
All resolution-qualified	290247	0.0070	0.0112	0.0311	0.894
Euler	177254	0.0071	0.0119	0.0343	0.854
Convection–diffusion	49524	0.0054	0.0085	0.0170	0.870
Reaction–diffusion	63469	0.0058	0.0116	0.0257	0.894

S4 Clustered uncertainty and numerical-scope diagnostics

The screening calibration uses candidate scores after group-disjoint splitting. It is not a cluster-level conformal construction: the 4–5 calibration groups per fold do not provide the 33 independent group scores needed for a nontrivial 3% split-conformal threshold. The following tables therefore distinguish the candidate-score operating point from clustered held-out uncertainty.

Table S15: Fold-level calibration diagnostics for boosting with $B + M$. Ties are counted among safe calibration scores exactly at the strict rejection threshold. The complete table for every screen is machine-readable in Online Resource 2.

Fold	Fit gr.	Cal. gr.	Test gr.	Safe cal.	Distinct	$q_{0.03}$	Ties
0	12	5	4	474	468	0.8928	1
1	12	4	5	373	365	0.9253	1
2	12	5	4	526	503	0.8314	1
3	12	5	4	432	376	0.8668	1
4	12	5	4	729	670	0.8033	1

Table S16: Absolute outer-test counts and 95% percentile intervals from 10000 problem-group bootstrap replications. Each row has 2920 unsafe and 2372 safe decisions in 21 groups.

Screen	Rejected unsafe	UR (%)	95% interval	Rejected safe	WR (%)	95% interval
Logistic $B + M$	1534	52.5	42.4–62.3	69	2.9	0.3–7.5
Logistic $B + A + M$	1675	57.4	48.1–66.1	73	3.1	0.6–7.2
Boosting $B + M$	2352	80.5	73.3–86.1	39	1.6	0.5–3.4
Boosting $B + A + M$	2347	80.4	72.8–86.3	45	1.9	0.7–3.8
Two-grid, post-solve	2356	80.7	74.7–85.6	80	3.4	1.9–5.4

For boosting $B + M$, the macro-average of the 21 within-group safe rejection rates is 3.45%, and 9 of 21 groups contain at least one wrong rejection. These descriptive cluster summaries are not a replacement for a conformal group guarantee.

Table S17: Tolerance-resolved absolute results. Entries are rejected/total, followed by the corresponding percentage.

	Boosting $B + M$		Two-grid	
Tolerance	Unsafe	Safe	Unsafe	Safe
10^{-3}	1027/1207 (85.1%)	10/557 (1.8%)	1068/1207 (88.5%)	55/557 (9.9%)
3×10^{-3}	803/995 (80.7%)	12/769 (1.6%)	797/995 (80.1%)	22/769 (2.9%)
10^{-2}	522/718 (72.7%)	17/1046 (1.6%)	491/718 (68.4%)	3/1046 (0.3%)

Table S18: Sensitivity of nested logistic $B + M$ screening to the prespecified inverse-response regularizer η. The same group splits and nominal $\alpha = 0.03$ are used.

η	UR (%)	WR (%)
0.01	55.1	2.6
0.05	52.5	2.9
0.10	50.9	3.1

Among 387744 nominal node–exponent evaluations, real-exponent clipping was activated 736 times (0.19%), defect clipping at 10^{12} 1752 times (0.45%), and inverse-response clipping and the 10^{-30} discrete-response floor never activated. Across the three η values, the continuous near-zero phase gate activated in 449542 of 1163232 evaluations (38.6%); this is a deliberate phase-undefined rule rather than overflow protection.

Table S19: Applicability of the Mellin–additive comparison theorem. Screening strata use held-out boosting $B + M$ decisions; the theorem is not invoked in the second row.

Population	Candidates	Decisions	UR (%)	WR (%)
$\theta_{\max} < 1$	1736	5208	80.6	1.6
$\theta_{\max} \geq 1$	28	84	77.3	2.5

All 72702 three-point rows and 99.9615% of the 72702 five-point rows satisfy $\theta_j < 1$. Over applicable tested row–frequency pairs, the largest observed ratio of the two sides of the theorem is 0.999963; ratios are formed only for a strictly positive right-hand side.

Table S20: Constructive barrier ablation on the common 80-case certificate set. Every assembled barrier satisfies $\min_i (A_h w_h)_i \geq 1$.

Barrier	Median $\|w_h\|_\infty$	Median effectivity	Q0.9 effectivity
Row sum	0.010000	7.83	27.28
Analytic trial	0.009879	7.84	27.04
Five-function dictionary LP	0.009869	7.78	26.97

The tuned barriers improve this reaction-dominated test only marginally. Its conservatism therefore cannot be attributed solely to the parameter-free row-sum choice; the rare FNO tail remains a separate, substantially more conservative certificate regime.

Table S21: FNO robustness over five prespecified training seeds. Ratios compare the proposed mesh with the log-uniform mesh for the same problem and N. The final column is the fraction of rare-tail cases for which the FNO mesh is better.

Seed	Ordinary median	Ordinary Q0.9	Tail median	Tail Q0.9	Tail better (%)
20260805	0.306	0.560	1.097	1.296	26.7
20260806	0.323	0.603	1.227	1.581	9.7
20260807	0.341	0.576	1.526	2.755	1.9
20260808	0.325	0.527	1.871	2.447	4.7
20260809	0.331	0.544	0.994	1.398	52.2

For ordinary cases, a 10000-replicate paired bootstrap resampling complete packets gives median FNO-minus-comparator error-ratio differences of -0.195 (95% interval $[-0.353, -0.125]$) against the MLP, 0.044 ($[0.031, 0.059]$) against the reference monitor, and -0.694 ($[-0.717, -0.670]$) against the log-uniform mesh. The rare-tail spread in the table prevents a seed-robust superiority claim for that external law.

S5 Contents of Online Resource 2: code and data archive

Online Resource 2 is organized as a reproducibility deposit rather than as an unstructured PDF supplement. The directory `scripts/` contains the deterministic benchmark generator, the complete analysis driver, and the NumPy FNO experiment. The directory `data/` contains raw candidate tables, fold-level predictions, aggregate metrics, and machine-readable run summaries. Vector figures are stored in `figures/`. The root README states the commands, software versions, output files, and expected run order. No external observational data or pretrained neural-network weights are used.

The article source and bibliography are submission-source files and are not part of the research-data deposit. Archive structure, software versions, and execution instructions are stated in the README distributed with the code-and-data archive.