Prerequisites for the Distributed-Workload Systems Papers

Dyson–Schwinger solvers, physics-informed networks, information geometry, and the Dirac spinor skeleton

Melvin Alvarez
HFThot Research Lab
contact@hfthot-lab.eu

October 18, 2026 (English edition)

This document collects, in one place and in English, the mathematical prerequisites for reading the laboratory's version-2 systems papers: the pysic-rs distributed-workload campaign (Dyson–Schwinger-type integral equations and distributed physics-informed network training), the polarway distributed-computation extension, the optimiz-rs stochastic-differential optimization section, and the ThotBook-AI information-geometry and Dirac spinor sections. Nothing here is new mathematics; each section states the precise facts the papers use, works the small examples in full, and points at the verification artifacts (notebooks and JSON certificates) that make the claims checkable. A reader knowing linear algebra, real analysis and basic measure theory can reconstruct every number quoted in the papers from this document plus the companion notebooks.

Contents

1 Notation and the computational substrate **2**

2 Linear Fredholm equations of the second kind: the Lippmann–Schwinger/Dyson prototype **2**
2.1 The equation and its discretisation . . . 2
2.2 The three verification identities . . . 3
2.3 Reader's checklist . . . 3

3 Self-consistent (gap) equations: Banach bounds and Picard convergence **3**

4 Physics-informed networks: loss structure, protocols, and the exact Fisher step **4**
4.1 The mechanics . . . 4
4.2 Why the last layer admits an exact solve . . . 4
4.3 Distribution and threading . . . 5

5 Information geometry: Fisher metric, natural gradient, invariance **5**
5.1 Fisher metric and its uniqueness . . . 5
5.2 Natural gradient, executed honestly . . . 5
5.3 Prior placement (Snoussi) . . . 6

6 The Dirac equation in 1+1 dimensions: the analytic skeleton **6**
6.1 Spinor algebra to dispersion . . . 6
6.2 Why this belongs in a systems programme . . . 7

7 Tensors, McKean–Vlasov and the distributed-primitives view 7

1 Notation and the computational substrate

All quantities are IEEE-754 double precision (float64) unless stated otherwise. "One task, one core" means each worker process runs with its BLAS library pinned to a single thread (`VECLIB_MAXIMUM_THREADS=1` here); the twelve-thread campaign host is a commodity computer with six physical cores and hyperthreading — a fact the batch experiments of the pysic-rs paper expose rather than assume (the measured speedups peak near six processes and degrade beyond). Distributed means *processes* on one host with fork-shared read-only memory: no interconnects, no network I/O in the measured region.

Three identities organize every claim about distribution:

1. **Streaming mean monoid.** (mean, n) composes associatively: the mean of two slab means with counts (n_1, n_2) is $(n_1\bar{x}_1 + n_2\bar{x}_2)/(n_1 + n_2)$. Consequently, for equal-size slabs the mean of slab gradients *is* the full-batch gradient up to floating-point summation order — no statistical approximation enters.

2. **Column decomposition.** An integral-equation sweep on energies $\{E_i\}_{i\le 96}$ is embarrassingly column-parallel: column i's discrete system is private to it. Per-column arithmetic is therefore identical whether run sequentially or distributed — the distributed result is exact, and only the *schedule* changes (the speedup is a scheduling property, not an accuracy property).

3. **Amdahl, heterogeneous host.** Speedup $S(p) = T_1/\max(T_{\text{seq}}/p,\ T_{\text{bottleneck}})$ degrades on a host whose twelve logical threads sit on six physical cores, and the measured campaign also feels memory-bandwidth contention: it peaks at $p = 4$ and degrades beyond, which the papers report rather than explain away.

2 Linear Fredholm equations of the second kind: the Lippmann–Schwinger/Dyson prototype

2.1 The equation and its discretisation

For a two-body potential V and the free resolvent $G_0(E) = \frac{1}{E' - q^2/2 + i\varepsilon}$, the half-off-shell Lippmann–Schwinger equation for the T matrix at energy E reads

$$t(p, p_0; E) = V(p, p_0) + \int_0^\infty V(p, q)\, \frac{q^2\, w_q\, \mathrm{d}q}{E - q^2/2 + i\varepsilon}\, t(q, p_0; E), \tag{1}$$

a Fredholm equation of the second kind, $t = v + Kt$ with kernel $K = (V\, G_0)$. The laboratory's campaign discretises (1) on an N-point trapezoidal grid with the standard $i\varepsilon$ reguliser ($\varepsilon = 0.05$) and an angle-averaged Yukawa kernel

$$\overline{V}(p, q) \;=\; \frac{1}{2pq} \ln \frac{\mu^2 + (p+q)^2}{\mu^2 + (p-q)^2}, \tag{2}$$

which is log-singular on the diagonal $p = q$ and drives the choice of quadrature care (the diagonal entries are handled by the limiting value $q \to p$; the trapezoidal rule distributes the

pole's imaginary part over a finite band, which the verification suite confirms is enough at the quoted tolerances).

Discretising (1) gives the linear system $(I - g\,VG_0)\,t = g\,v$ with coupling g multiplying the potential. Two solver families apply:

- **Direct (Gauss elimination),** cost $O(N^3)$: the campaign's $N = 8000$ single-core measurement (121.4 s, residual 5.9×10^{-14}) and its $N = 4000$ twin (15.4 s) realise the textbook factor 8 per grid doubling almost exactly.
- **Picard/Neumann iteration,** cost $O(N^2)$ per iteration, valid when the operator is contractive: the campaign fixes the coupling from the *measured* spectral radius, $g_{\text{ite}} = 0.8/\rho(VG_0)$ with $\rho = 25.159$ at unit coupling.

2.2 The three verification identities

Every solver run in the papers carries its own certificate, one of three kinds:

1. **Algebraic residual.** The computed $\hat{t}$ satisfies $\max_p |\hat{t} - (v + g\,K\hat{t})|_p$; a round-off-sized value (10^{-14}–10^{-13}) certifies that the linear system was solved exactly, independent of physics.
2. **Contraction and its cash value.** If $\rho(K) < 1$ then the Neumann series $\sum_n K^n v$ converges to the solution at ratio ρ *in any norm whose induced operator norm satisfies the bound.* The papers measure two estimates of ρ: the 2-norm by power iteration (0.7994) and the per-step decay of the max-component deviation of partial sums ($1/1.3011 = 0.7686$), and they report both with the norm distinction stated. Convergence of a bounded operator's spectral radius is the guarantee; the two measurements bracket it.
3. **Limits.** $T(0) = V$ (the bare potential) is the $g \to 0$ limit: the measured deviation collapses from 0.112 at $g = 10^{-2}$ to 8.4×10^{-5} at $g = 10^{-4}$. Grid convergence is the trapezoid's second order: the 8× ratio in the gap-equation ladder below is the measurable signature.

2.3 Reader's checklist

To reproduce Section 9, Part A of the pysic-rs paper: build (2) on grid $\{q_i\}$; assemble VG_0 (complex, $i\varepsilon$); take ρ by 50 power iterations (matvec cost only); set $g = 0.8/\rho$; run (i) one direct solve and (ii) the Picard loop $t_{n+1} = v + gKt_n$ with $t_0 = v$; print residuals, partial-sum deviations and the ratio column. The companion notebook `15_dyson_schwinger.ipynb` executes a reduced version ($N = 2000$) end-to-end.

3 Self-consistent (gap) equations: Banach bounds and Picard convergence

The nonlinear prototype is the rainbow-truncated gap equation

$$m^2(p) = m_0^2 + \lambda \int_0^\infty \frac{K(p,q)\; q^2\, w_q}{q^2 + m^2(q)}\, \mathrm{d}q, \tag{3}$$

thought of as a fixed-point problem $\Phi(m^2) = m^2$ on $[m_0^2, \infty)^N$. The only analytic fact the campaign needs is the **row-sum (Banach) bound**: if every row of the discrete Jacobian is

dominated by

$$\max_i \sum_j \Big|\frac{\partial \Phi_i}{\partial m_j^2}\Big| \;=\; \max_i \sum_j \Big|\lambda\, \frac{K_{ij}\, q_j^2 w_j}{(q_j^2+m_j^2)^2}\Big| \;\le\; C_{\text{rows}} < 1, \tag{4}$$

then Φ is a contraction on the whole admissible box and Picard iteration converges from anywhere in it. Because $(q^2+m^2)^2 \ge (q^2+m_0^2)^2$, the admissible numerator

$$C_{\text{rows}}(m_0^2) \;=\; \max_i \sum_j \Big|\lambda\, \frac{K_{ij}\, q_j^2 w_j}{(q_j^2+m_0^2)^2}\Big| \tag{5}$$

is computable *before* any iteration, and the campaign caps the coupling at $\lambda \le 0.7/C_{\text{rows}}$ — measured $C_{\text{rows}} = 0.3927$ at $\lambda = 1$, i.e. the cap never binds. The same bound is why the fixed point is reached in 15 iterations *independently of* N: the contraction factor is dimension-free.

The one variable that is *not* dimension-free is the trapezoid's $O(h^2)$: the ladder $|m_N^2(0) - m_{2N}^2(0)| = 5.4\times10^{-9} \to 1.4\times10^{-9} \to 3.3\times10^{-10}$ (factor ≈ 4 per doubling) is that signature, already at the 10^{-9} level where float64 is comfortable. The bare-mass limit $\lambda \to 0$ returns the input mass to 1.5×10^{-10} in two iterations — the second verification identity of the family.

What to compare when porting. The physically interesting scalar is the dynamical enhancement $m^2(0)/m_0^2 = 2.24$ and the profile monotonicity (the fixed point decays from 2.236 at the origin to 1.058 at the grid end, i.e. confinement-grade enhancement where the kernel is strong, near-bare where it is weak). These are outputs of the *discretised* problem: quoting them without the grid ladder and the fixed-point depth (15 iterations to 2.3×10^{-13}) is what the papers refuse to do.

4 Physics-informed networks: loss structure, protocols, and the exact Fisher step

4.1 The mechanics

A PINN represents the field by a multilayer perceptron $u_\theta \approx u$ and penalises the PDE residual computed through *autodiff* — here the laboratory's own `thotgrad`, whose forward pass carries first-derivative tangents exactly (one tape, no finite-difference contamination). For the Dirac benchmark (next section) the residual rows are

$$r_1 = \partial_x u - (E+m)\, v, \qquad r_2 = \partial_x v + (E-m)\, u, \tag{6}$$

plus enforce-at-origin boundary rows, and the loss is the mean of squares over collocation points.

The two-stage protocol (AdamW 3000, then L-BFGS 500) exists because of curvature: plain Adam is a *diagonal* preconditioner — it rescales coordinates independently, which is emphatically *not* the natural gradient (Section 5) — while L-BFGS accumulates (s, y) curvature pairs along the trajectory and thereby approximates the inverse metric *where the optimisation actually went*. The measured difference is three orders of magnitude in final loss.

4.2 Why the last layer admits an exact solve

Freeze the hidden layers. In the last layer's parameters θ_2 (the readout weights and biases, 130 numbers here), the network output is *affine*: $u = a^\top \theta_2 + b_\alpha$ with a the last hidden activations and b the constant-sheet contributions θ_2-free rows. Each residual row r_i of (6) is then affine in θ_2 with

a *known* coefficient row: the input-derivative of the last hidden layer $\mathrm{d}a/\mathrm{d}x = (1-a^2)\,\mathrm{d}a_{\mathrm{pre}}/\mathrm{d}x$ is computable in closed form by the same chain rule. The loss is therefore an *exact quadratic*

$$L(\theta_2) = \tfrac{1}{2}\,\|A\theta_2 + c\|_2^2, \qquad F = A^\top A, \quad \text{the empirical Fisher, exact, not sampled,} \tag{7}$$

and the Gauss–Newton/Fisher step is a least-squares solve $\theta_2^\star = -(A^\top A)^{-1}A^\top c$, costing single-digit milliseconds at these sizes. The measured effect on the protocol-trained network: a factor-level loss reduction from one solve. The subtlety worth internalising: for *hidden* layers the same $J^\top J$ is only an approximation (the chain is nonlinear), which is exactly why the campaign restricts the exact step to the readout — the structure, not the speed, is the result. (In the executed campaign the least-squares residual lands at 10^{-21}: the readout's fit is in span, not merely small.)

4.3 Distribution and threading

Two operational facts: the slab-aggregation monoid gives exactness (Section 1, identity 1); and the parallelism placement is measured, not assumed — inside one process, BLAS threading saturates at a 1.17× ceiling on this operation mix, while the same work sharded across processes accelerates 4.07×: the campaign's "one task, one core" is the measured optimum, not an apology. Reader's checklist: notebook `06_pinn_scaling.ipynb` runs the reduced protocol, the exact Gauss–Newton step, the slab sweep, and the Gaussian statistical-manifold demo of the next section.

5 Information geometry: Fisher metric, natural gradient, invariance

5.1 Fisher metric and its uniqueness

For a parametric family p_θ, the Fisher information matrix

$$F_{ij}(\theta) = \mathbb{E}_\theta[\partial_i \log p_\theta\, \partial_j \log p_\theta] = -\,\mathbb{E}_\theta[\partial_i \partial_j \log p_\theta] \tag{8}$$

(when differentiable twice) defines a Riemannian metric on the statistical manifold: it transforms as a tensor under reparameterisation and is positive semidefinite with degeneracies exactly at non-identified parameters. Uniqueness (Chentsov): up to monotone rescaling, *the Fisher metric is the only metric invariant under all randomised Markov morphisms.* The geometry is not a modelling choice; it is forced by what statistics may do to data.

5.2 Natural gradient, executed honestly

The natural gradient is $\widetilde{\nabla} L = F^{-1}\nabla L$ in any chart. Its invariance is a statement about *directions*: under reparameterisation $\theta' = \chi(\theta)$,

$$F'^{-1}\nabla' L \;=\; \mathsf{J}^{-\mathsf{T}} F^{-1} \nabla L, \qquad \mathsf{J} = \mathrm{d}\theta/\mathrm{d}\theta',$$

and the pushed-forward coordinate vector $\mathsf{J}\,(\mathsf{J}^{-\mathsf{T}}F^{-1}\nabla L)$ represents, in the new coordinates, the *same tangent vector* of the statistical manifold. What must be checked per family is the round-off behaviour of that push-forward: for the Gaussian family below the two charts share μ and differ by the scalar reparameterisation $\sigma = e^s$, for which $\mathsf{J}\,\mathsf{J}^{-\mathsf{T}}$ is exactly the identity on each axis — so the two representations agree *to floating-point round-off*, and the invariance

becomes an executable identity rather than an assertion. What is *not* invariant is the discrete Euler step: $\theta_+ = \theta - \eta F^{-1}\nabla L$ and its twin agree only to $O(\eta)$. The ThotBook paper's ladder executes this: direction invariance to 3.6×10^{-15} at every step; path discrepancy shrinking $1.19 \to 0.076 \to 0.014 \to 0.0053$ as η halves, while vanilla gradient descent's paths remain $O(1)$ apart because its *flow* differs between charts.

The Gaussian family example, worked. Fit $\mathcal{N}(\mu, \sigma^2)$ to fixed data ($n = 200$). The risk is $L = -\log\sigma + (\overline{y^2} - 2\mu\bar{y} + \mu^2)/2\sigma^2$. Fisher: $F = \mathrm{diag}(1/\sigma^2,\, 2/\sigma^2)$; in the $(\mu, s = \log\sigma)$ chart $F' = \mathrm{diag}(1/\sigma^2,\, 2)$ (the second component becomes *constant* — this is the natural-parameter convenience the exponential family buys). The chart-native natural-gradient step in (μ, s) is $\mu- = \eta\,\sigma^2\,\partial_\mu L$, $s- = \frac{\eta}{2}\sigma\,\partial_\sigma L$ with $\sigma = e^s$; pushed forward it equals the (μ, σ)-chart step's direction to round-off — the executed identity behind the papers' invariance claim.

5.3 Prior placement (Snoussi)

Given data x and a *model* $p(x|\theta)$, the Bayesian update needs a prior $\pi(\theta)$; selection among priors is a variational problem on the same manifold. In Snoussi's formulation the prior is placed by *projection*: a reference measure (often the invariant or Jeffreys measure ϖ) is transported into a restricted autoparallel submanifold $\mathcal{M}$ by minimising an information-type divergence, $\pi = \arg\min_{\mu\in\mathcal{M}} D(\mu\|\pi\varpi)$, giving the "dual" divergence $D_1 = D_0$ reversed — in the Gaussian-entropic case D_0/D_1 are the two KL symmetrisations, whose identity $D_1(p, q) = D_0(q, p)$ holds exceptionally for the inner product structure of the Gaussian family (verified in the paper's executed fragment). The practical reading: priors are geometric objects; changing charts does not change the belief state, only its description — the same invariance discipline as the natural gradient, one level up.

6 The Dirac equation in 1+1 dimensions: the analytic skeleton

6.1 Spinor algebra to dispersion

In 1+1 dimensions the Dirac equation is a two-component system: with $\alpha = \sigma_x$, $\beta = \sigma_z$,

$$i\partial_t\psi = \left(-i\,\alpha\,\partial_x + \beta\,m\right)\psi. \tag{9}$$

The stationary ansatz $\psi(x, t) = e^{-iEt}\,\varphi(x)$ and the plane-wave $\varphi = (u, v)^\top e^{ikx}$ give the coupled system

$$k\,v + m\,u = E\,u, \qquad k\,u - m\,v = E\,v, \tag{10}$$

with eigenspinor ratio $\phi_2/\phi_1 = k/(E + m)$ and dispersion $E^2 = k^2 + m^2$ (multiply the two equations). The benchmark values $E = 2$, $m = 1$ give $k = \sqrt{3}$ and ratio $\sqrt{3}/2$; the graded real-space target is $u(x) = \cos kx$, $v(x) = -\frac{k}{E+m}\sin kx$, which solves $\mathrm{d}u/\mathrm{d}x = (E + m)v$, $\mathrm{d}v/\mathrm{d}x = -(E-m)u$ exactly — the same pair of identities the PINN minimises in (6). The negative branch $E = -\sqrt{k^2 + m^2}$ is equally exact; upon quantisation it is the antiparticle sector. The papers' position is epistemic: the analytic spinor pair is the *grading target*, and every algebraic step above is verified independently (notebook `05_dirac_spinors.ipynb`: 16 Clifford anticommutators at round-off; plane-wave normalisation $u^\dagger u = 2E$, $\bar{u}u = 2m$; eigenspinor ratio to round-off; and a finite-difference Dirac Hamiltonian on a periodic grid whose spectrum reproduces $\pm\sqrt{\tilde{k}^2 + m^2}$ with the *discrete* momentum $\tilde{k} = \sin(\theta)/h$ of the central difference — the last distinction being exactly the kind of statement a validation discipline must make precise).

6.2 Why this belongs in a systems programme

Three reasons. First, the Dirac problem is the *stress test* with which the laboratory caught its own library defect (the norm-growth 3.6×10^{57} propagator): the closed form (10) supplies the invariant (unitarity) that a plausible-looking plot cannot. Second, its solution manifold is small enough to grade exactly — the PINN's rel-L^2 of 1.7×10^{-3} is against a *closed form*, not against a reference solver of unknown error. Third, the same dispersion argument transfers to every propagation workload the laboratory runs (Wheeler–DeWitt's "disappearance of time" aside).

7 Tensors, McKean–Vlasov and the distributed-primitives view

The optimiz-rs paper's new section optimizes SDE dynamics whose drift depends on the *law* of the solution — McKean–Vlasov diffusions $\mathrm{d}X_t = b(X_t, \mathrm{Law}(X_t))\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$ — via neural-network parameterisations in the spirit of *arXiv:1905.01205* (Zhang, Guo and Karniadakis: solving time-dependent stochastic PDEs by physics-informed networks in modal space). The prerequisite to keep in mind: for MV flows the objective is a functional of the distribution, so the network's "residual" is a *monte-carlo estimate of a mean-field term*, and the two-stage training discipline of Section 5 applies unchanged; the distributed slabs then shard the particle set, and the same monoid guarantees the gradient is the full-ensemble gradient — the pattern is one, the instances many. The polarway paper holds the corresponding theorem-level treatment for its reduction algebra; the ThotBook paper executes the Gaussian-family invariance; the pysic-rs paper measures the campaign; and this document states the common currency: exact aggregates, certified limits, measured schedules.

Verification artifacts index

- `15_dyson_schwinger.ipynb` (pysic-rs repo): LS/Dyson + gap equation + worker sweep, executed.
- `05_dirac_spinors.ipynb` (ThotBook-AI repo): Clifford, plane-wave spinors, stationary 1+1D solution, FD spectrum, executed.
- `06_pinn_scaling.ipynb` (ThotBook-AI repo): protocol replica, exact Fisher step, slab monoid, invariance demo, executed.
- `dse_results.json` / `pinn_case_results.json` / `ng_paths.json` + CSVs: the full campaign data behind every number quoted above.

Prerequisites edition — October 18, 2026