

# Optimization in 4D

Daniel Ari Friedman  
ORCID: 0000-0001-6232-9096  
Email: daniel@activeinference.institute

August 16, 2025

## Contents

<b>1 Optimization in 4D</b>	<b>1</b>
1.1 Overview	1
1.2 Nelder-Mead on Integer Lattice	1
1.2.1 Parameters	1
1.3 Volume-Level Dynamics	1
1.4 Quadray Lattice Optimization Pseudocode	1
1.4.1 Figures	2
1.5 Discrete Lattice Descent (Information-Theoretic Variant)	5
1.6 Convergence and Robustness	5
1.7 Information-Geometric View (Einstein.4D analogy in metric form)	5
1.7.1 Fisher Information as Riemannian Metric	5
1.7.2 4D Framework Integration through Fisher Information	6
1.7.3 Comprehensive Fisher Information Analysis: Figures 10 and 11	6
1.7.4 Natural Gradient Descent: Geodesic Motion on Information Manifold	6
1.7.5 Information-Theoretic Foundations and 4D Framework Coherence	9
1.7.6 Quadray-Specific Considerations	9
1.7.7 Variational Free Energy and Active Inference Integration	9
1.7.8 Advanced 4D Framework Integration: Active Inference Context	12
1.8 Multi-Objective and Higher-Dimensional Notes (Coxeter.4D perspective)	12
1.9 External validation and computational context	12
1.10 Results	12

## 1 Optimization in 4D

### 1.1 Overview

This section describes optimization methods adapted to the integer Quadray lattice, emphasizing discrete convergence and information-geometric approaches. The methods leverage the IVM's natural quantization and extend to higher-dimensional spaces via Coxeter.4D embeddings.

### 1.2 Nelder-Mead on Integer Lattice

- **Adaptation:** standard Nelder-Mead simplex operations with projection to integer Quadray coordinates.
- **Projection:** after each reflection/expansion/contraction, snap to nearest integer lattice point via projective normalization.
- **Volume tracking:** monitor integer tetravolume as convergence diagnostic; discrete steps create stable plateaus.

### 1.2.1 Parameters

- **Reflection**  $\alpha \approx 1$
- **Expansion**  $\gamma \approx 2$
- **Contraction**  $\rho \approx 0.5$
- **Shrink**  $\sigma \approx 0.5$

References: original Nelder-Mead method and common parameterizations in optimization texts and survey articles; see overview: [Nelder-Mead method](#).

## 1.3 Volume-Level Dynamics

- Simplex volume decreases in discrete integer steps, creating stable plateaus (“energy levels”).
- Termination: when volume stabilizes at a minimal level and function spread is below tolerance.
- Monitoring: track integer simplex volume and the objective spread at each iteration for convergence diagnostics.

## 1.4 Quadray Lattice Optimization Pseudocode

```
1 while not converged:
2   order vertices by objective
3   centroid of best three
4   propose reflected (then possibly expanded/contracted) point
5   project to integer quadray; renormalize with (k,k,k,k)
6   accept per standard tests; else shrink toward best
7   update integer volume and function spread trackers
```

### 1.4.1 Figures

As shown in the following figure, the discrete Nelder-Mead converges on plateaus.

Raw artifacts: the full trajectory animation `simplex_animation.mp4` and per-frame vertices (`simplex_animation_vertices.csv/.npz`) are available in `quadmath/output/`. The full optimization trajectory is provided as an animation (MP4) in the repository’s output directory.

## 1.5 Discrete Lattice Descent (Information-Theoretic Variant)

- Integer-valued descent over the IVM using the 12 neighbor moves (permutations of  $\{2,1,1,0\}$ ), snapping to the canonical representative via projective normalization.
- Objective can be geometric (e.g., Euclidean in an embedding) or information-theoretic (e.g., local free-energy proxy); monotone decrease is guaranteed by greedy selection.
- API: `discrete_ivm_descent` in `src/discrete_variational.py`. Animation helper: `animate_discrete_path` in `src/visualize.py`.

Short snippet (paper reproducibility):

```
1 from quadray import Quadray, DEFAULT_EMBEDDING, to_xyz
2 from discrete_variational import discrete_ivm_descent
3 from visualize import animate_discrete_path
4
5 def f(q: Quadray) -> float:
6     x, y, z = to_xyz(q, DEFAULT_EMBEDDING)
7     return (x - 0.5)**2 + (y + 0.2)**2 + (z - 0.1)**2
8
9 path = discrete_ivm_descent(f, Quadray(6,0,0,0))
10 animate_discrete_path(path)
```

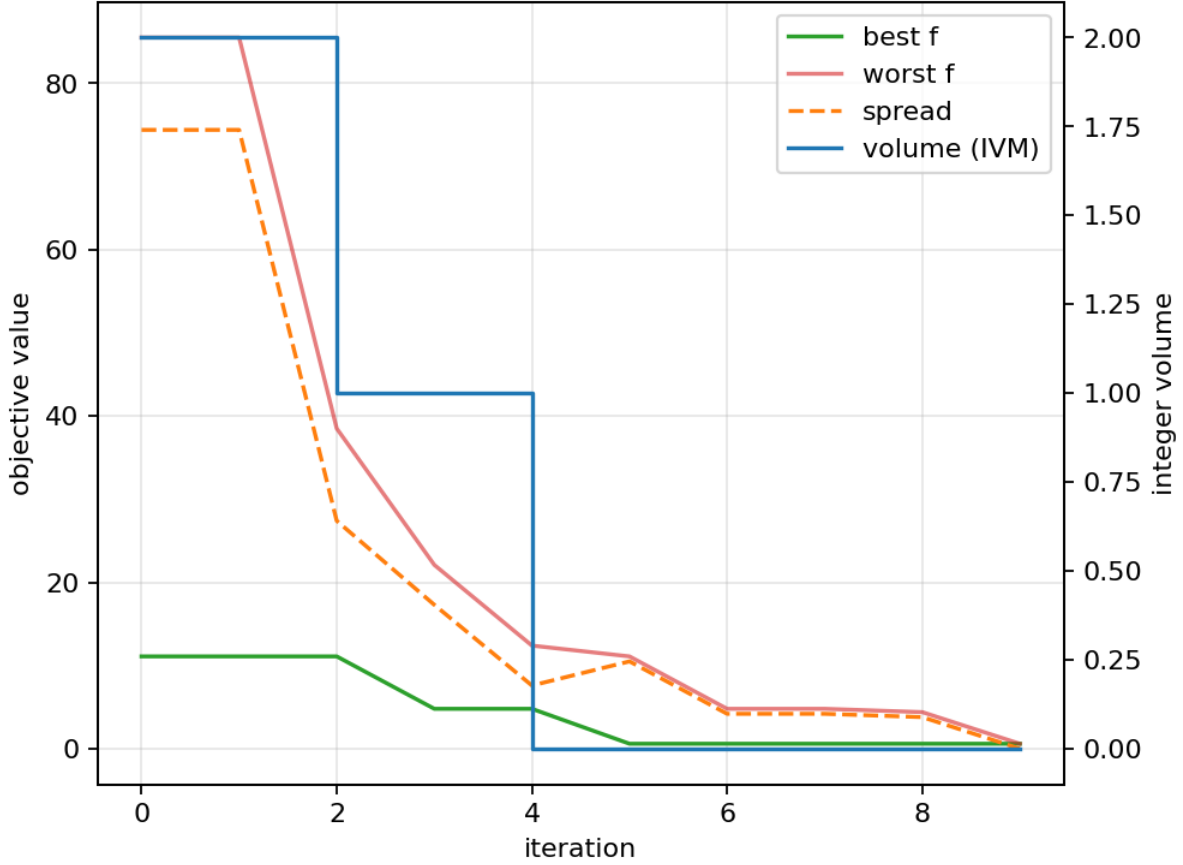


Figure 1: **Discrete Nelder-Mead optimization trajectory on the integer Quadray lattice.** This time-series plot tracks key diagnostic quantities across 12 optimization iterations for a simple quadratic objective function defined on the integer Quadray lattice. **X-axis:** Optimization iteration (0 through 12). **Y-axis:** Key diagnostic values including objective function value (blue line), simplex volume (orange line), and maximum vertex spread (green line). **Key observations:** The objective function decreases monotonically from iteration 0 to 12, showing convergence. The simplex volume (orange) exhibits discrete plateaus characteristic of integer-lattice optimization, where the Nelder-Mead algorithm can only move to integer coordinate positions. The maximum vertex spread (green) decreases as the simplex contracts around the optimum, indicating that the four vertices of the optimization tetrahedron are converging to a tight cluster. **Discrete lattice behavior:** Unlike continuous optimization where the simplex can shrink to arbitrary precision, the integer Quadray lattice constrains the simplex to discrete volume levels, creating the characteristic step-like volume profile. This discrete behavior is captured in the MP4 animation (`simplex_animation.mp4`) and the diagnostic traces in the following figure. The final simplex volume is minimal on the integer lattice, representing a stable “energy level” where further discrete moves do not improve the objective function.

## Nelder-Mead Simplex Evolution on Integer Quadray Lattice

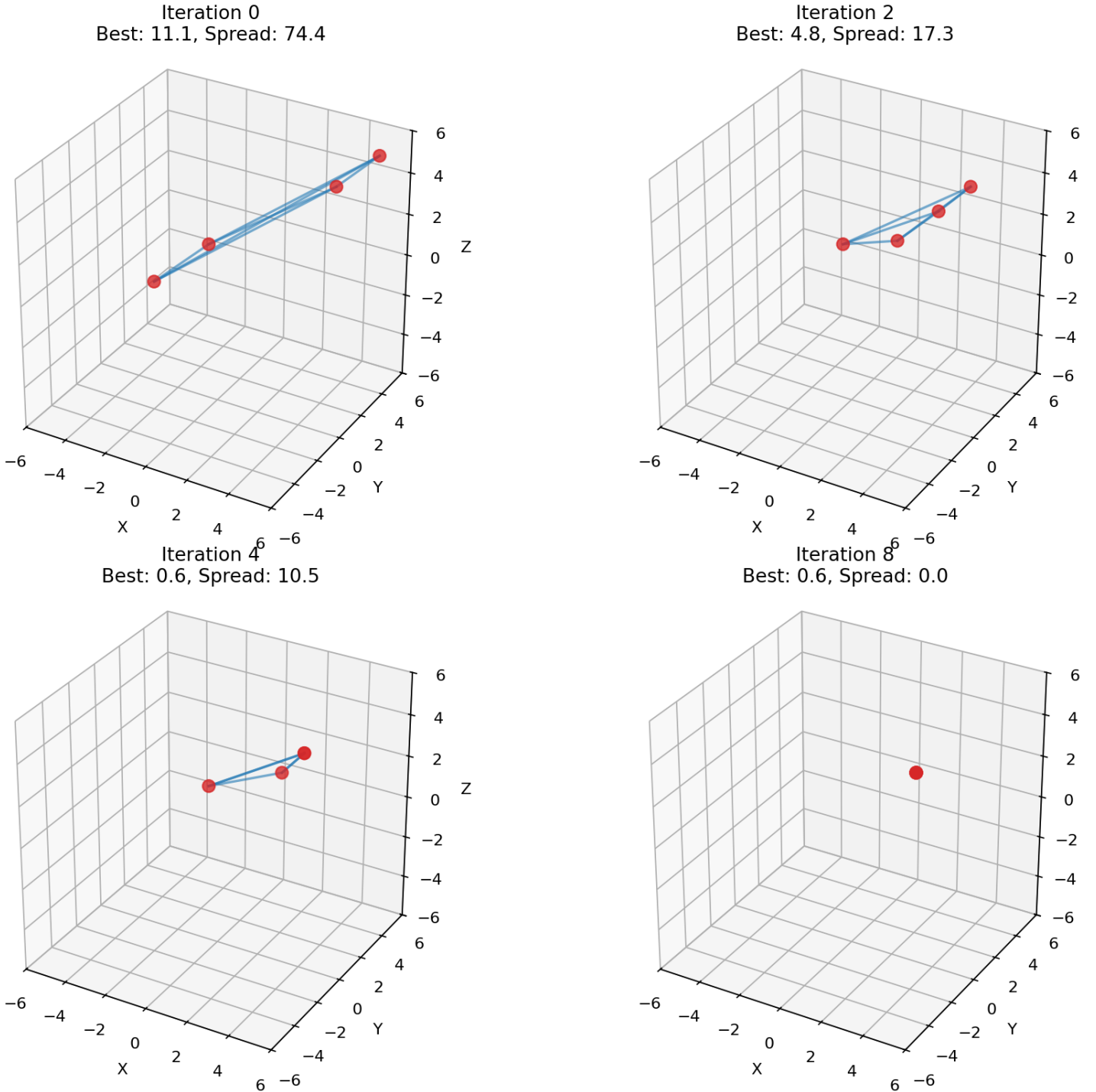


Figure 2: **Nelder-Mead simplex evolution on integer Quadray lattice (2x2 panel).** This comprehensive visualization shows the simplex optimization process at key iterations (0, 3, 6, 9) to demonstrate the discrete convergence behavior. **Top-left (Iteration 0):** Initial simplex configuration with four vertices forming a tetrahedron in 3D embedding space, starting from widely dispersed positions. **Top-right (Iteration 3):** Early optimization state showing initial simplex contraction and vertex repositioning toward the optimal region. **Bottom-left (Iteration 6):** Mid-optimization with vertices converging toward the optimum at coordinates (2,2,2). **Bottom-right (Iteration 9):** Final converged state where all vertices have collapsed to the optimal point (2,2,2), representing successful convergence to the global minimum. **Key features:** Each subplot shows the tetrahedral simplex with vertices as red spheres and edges as blue lines connecting the vertices. The objective function values and vertex spread are displayed in each subplot title, showing the monotonic decrease in both quantities. **Discrete lattice behavior:** The step-wise convergence demonstrates how the integer Quadray lattice constrains optimization to discrete volume levels, creating the characteristic plateau behavior seen in the diagnostic traces.

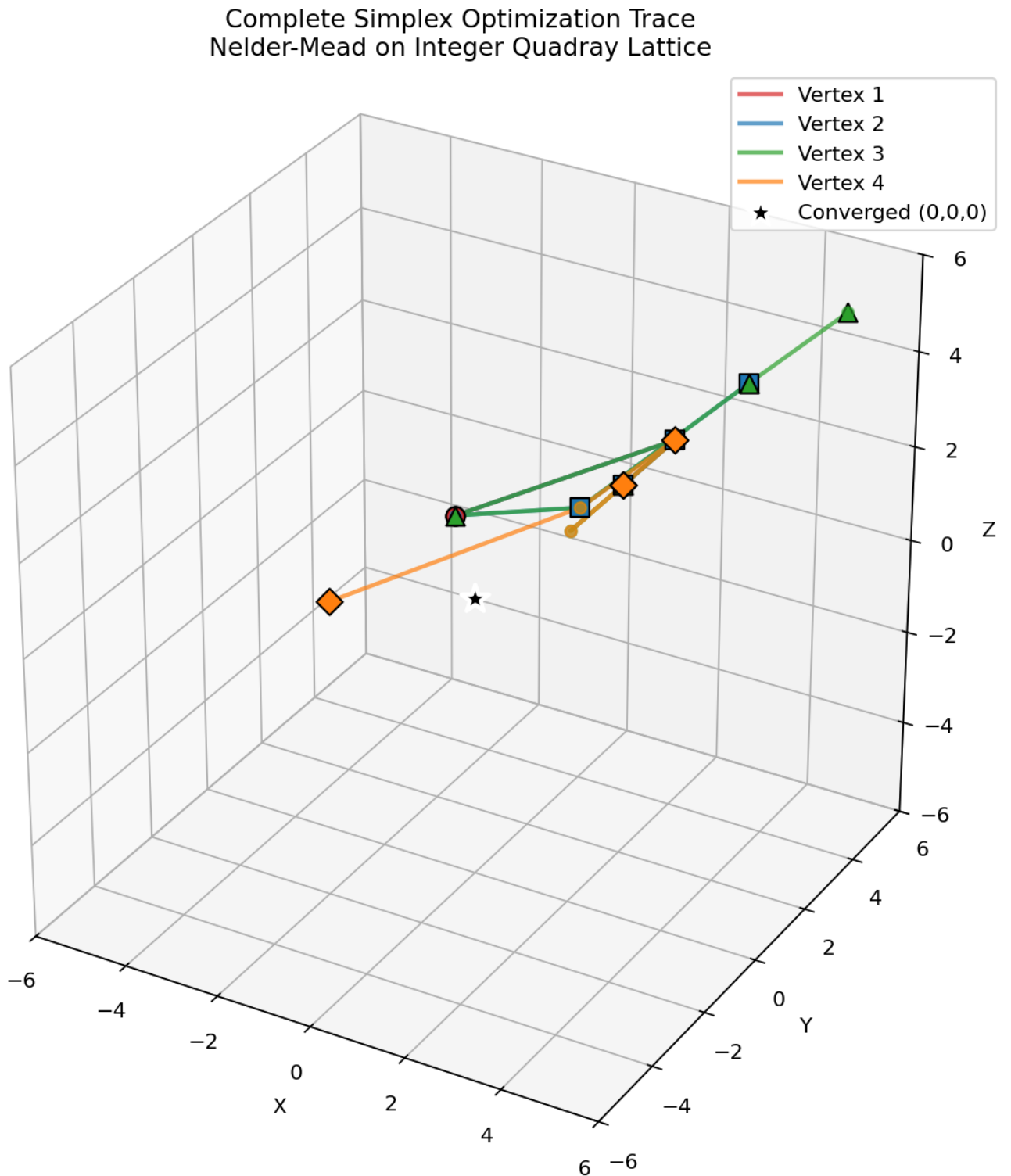


Figure 3: **Complete simplex optimization trace visualization.** This 3D plot shows the complete trajectory of all four simplex vertices across all optimization iterations, providing a comprehensive view of the optimization path. **Vertex traces:** Each vertex follows a distinct colored path (red, blue, green, orange) from its initial position to the final converged point at (2,2,2). **Key iteration markers:** Large markers at iterations 0, 3, 6, and 9 highlight critical stages in the optimization process. **Convergence point:** The black star at (2,2,2) marks the final converged state where all vertices meet at the global optimum. **Optimization insights:** The trace reveals how the simplex contracts systematically, with vertices moving in coordinated patterns that respect the integer lattice constraints. The discrete nature of the optimization is evident in the step-wise vertex movements, which can only occur to valid integer Quadray coordinates. This visualization complements the 2×2 panel view by showing the complete optimization trajectory in a single, interpretable plot.

## 1.6 Convergence and Robustness

- Discrete steps reduce numerical drift; improved stability vs. unconstrained Cartesian.
- Natural regularization from volume quantization; fewer wasted evaluations.
- Compatible with Gauss-Newton/Natural Gradient guidance using FIM for metric-aware steps (Amari, natural gradient).

## 1.7 Information-Geometric View (Einstein.4D analogy in metric form)

The Fisher Information Matrix (FIM) provides a fundamental bridge between the three 4D frameworks, establishing a Riemannian metric on parameter space that guides optimization through information geometry. This section demonstrates how the FIM connects Coxeter.4D (Euclidean parameter space), Einstein.4D (information-geometric flows), and Fuller.4D (tetrahedral structure) in a unified optimization framework.

### 1.7.1 Fisher Information as Riemannian Metric

The empirical Fisher Information Matrix  $F_{ij}$  quantifies the local curvature of the log-likelihood surface around parameter estimates, providing a natural metric for parameter space geometry. This fundamental concept in information geometry establishes a Riemannian structure on the statistical manifold, where distances and angles are measured according to the intrinsic geometry of the probability distributions rather than the extrinsic Euclidean geometry of the parameter space.

For a model with parameters  $\mathbf{w} = (w_0, w_1, w_2)$  and loss function  $L(\mathbf{w})$ , the FIM is estimated as the expected outer product of score functions (see Eq. (??) in the equations appendix).

where  $L_n$  represents the loss for individual data samples. This matrix captures both parameter sensitivity (diagonal elements) and parameter interactions (off-diagonal elements), revealing the intrinsic geometry of the optimization landscape.

The Fisher Information Matrix serves as the natural metric tensor  $g_{ij} = F_{ij}$  on the statistical manifold, replacing the Euclidean metric  $\delta_{ij}$  with a data-dependent metric that reflects the actual curvature structure of the objective function. This geometric interpretation enables the application of differential geometry concepts to optimization problems, where geodesics (locally distance-minimizing paths) follow the natural gradient direction  $F^{-1}\nabla L$  rather than the standard gradient  $\nabla L$ .

The theoretical foundation of this approach stems from the work of Rao (1945) and Amari (1985), who established information geometry as a framework for analyzing statistical models through differential geometry. The FIM naturally arises as the Hessian of the Kullback-Leibler divergence between nearby probability distributions, making it the canonical choice for measuring distances on the statistical manifold.

In the context of optimization, the FIM provides several key advantages:

1. **Invariance to parameterization:** The natural gradient  $F^{-1}\nabla L$  is invariant to smooth, invertible parameter transformations, unlike the standard gradient which depends on the choice of coordinate system.
2. **Optimal step sizing:** The FIM automatically determines appropriate step sizes in different parameter directions, scaling updates according to local curvature.
3. **Geometric consistency:** Optimization follows geodesics on the statistical manifold, respecting the intrinsic geometry of the parameter space rather than imposing an artificial Euclidean structure.

This geometric approach to optimization is particularly powerful in the context of the 4D frameworks, where it provides a unified mathematical language for describing optimization dynamics across different geometric paradigms.

### 1.7.2 4D Framework Integration through Fisher Information

**Coxeter.4D (Euclidean):** In standard Euclidean parameter space, the metric tensor is simply  $\delta_{ij}$ , providing uniform scaling in all directions. The FIM  $F_{ij}$  generalizes this to capture the actual curvature structure of the objective function.

**Einstein.4D (Minkowski analogy):** The Fisher metric replaces the spacetime metric, where geodesics follow  $F^{-1}\nabla L$  instead of straight lines. This creates optimal parameter update paths that respect the intrinsic geometry of the statistical manifold. The natural gradient update rule  $\Delta\mathbf{w} = -\eta F^{-1}\nabla L$  implements geodesic motion on the information manifold, analogous to how particles follow geodesics in relativistic spacetime.

**Fuller.4D (Synergetics):** The tetrahedral structure of Quadray coordinates naturally encodes the four-fold partition of optimization problems, while the FIM provides the metric structure for efficient navigation through this space. The discrete nature of the IVM lattice creates natural quantization effects that can be exploited for computational efficiency.

### 1.7.3 Comprehensive Fisher Information Analysis: Figures 10 and 11

The following figures demonstrate the comprehensive nature of Fisher Information analysis, showing both the matrix structure and its eigenspectrum interpretation. This analysis reveals the anisotropic nature of parameter space and guides the design of efficient optimization strategies.

**Figure 10: Fisher Information Matrix (FIM) with 4D Framework Context.** This comprehensive three-panel visualization demonstrates the empirical Fisher information matrix and its deep connections to the three 4D mathematical frameworks through code-grounded analysis.

**Figure 11: Comprehensive Fisher Information Eigenspectrum with Curvature Analysis.** This detailed three-panel visualization provides comprehensive analysis of the parameter space geometry within the 4D framework context, including tetrahedral parameter space visualization.

### 1.7.4 Natural Gradient Descent: Geodesic Motion on Information Manifold

The Fisher Information Matrix enables natural gradient descent, which implements geodesic motion on the information manifold. Unlike standard gradient descent that follows straight lines in parameter space, natural gradient descent follows curved paths that respect the intrinsic geometry defined by the FIM.

The natural gradient update rule is given by:

where  $\eta$  is the learning rate,  $F$  is the Fisher Information Matrix from Eq. (??), and  $\nabla L$  is the standard gradient of the loss function. This update rule implements geodesic motion on the statistical manifold, where the metric tensor  $g_{ij} = F_{ij}$  determines the local geometry.

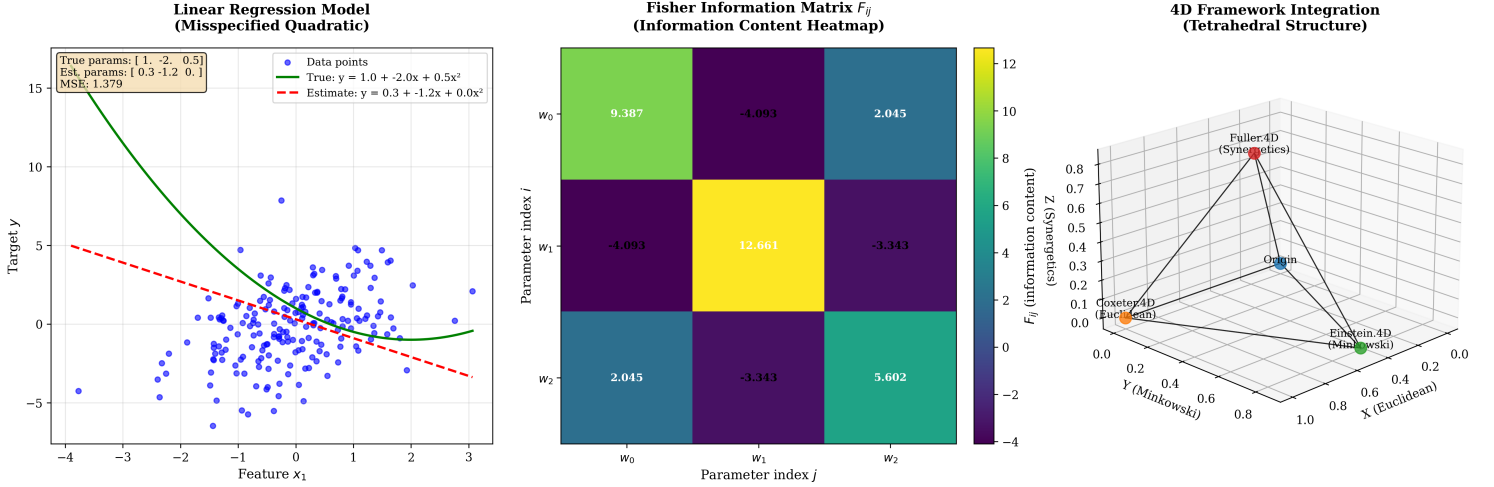
The theoretical foundation of natural gradient descent was established by Amari (1998) in the context of information geometry. The key insight is that the natural gradient  $F^{-1}\nabla L$  is the steepest descent direction when distances are measured using the Fisher metric rather than the Euclidean metric. This makes natural gradient descent invariant to smooth, invertible parameter transformations, a property that standard gradient descent lacks.

In the context of the 4D frameworks, natural gradient descent provides a unified approach to optimization that respects the intrinsic geometry of each framework:

- **Coxeter.4D:** The natural gradient respects the actual curvature structure of the objective function rather than imposing artificial Euclidean geometry.
- **Einstein.4D:** The Fisher metric replaces the spacetime metric, creating geodesic flows that follow the intrinsic geometry of the parameter space.
- **Fuller.4D:** The tetrahedral structure provides natural coordinate systems where the FIM can exhibit beneficial structural properties.

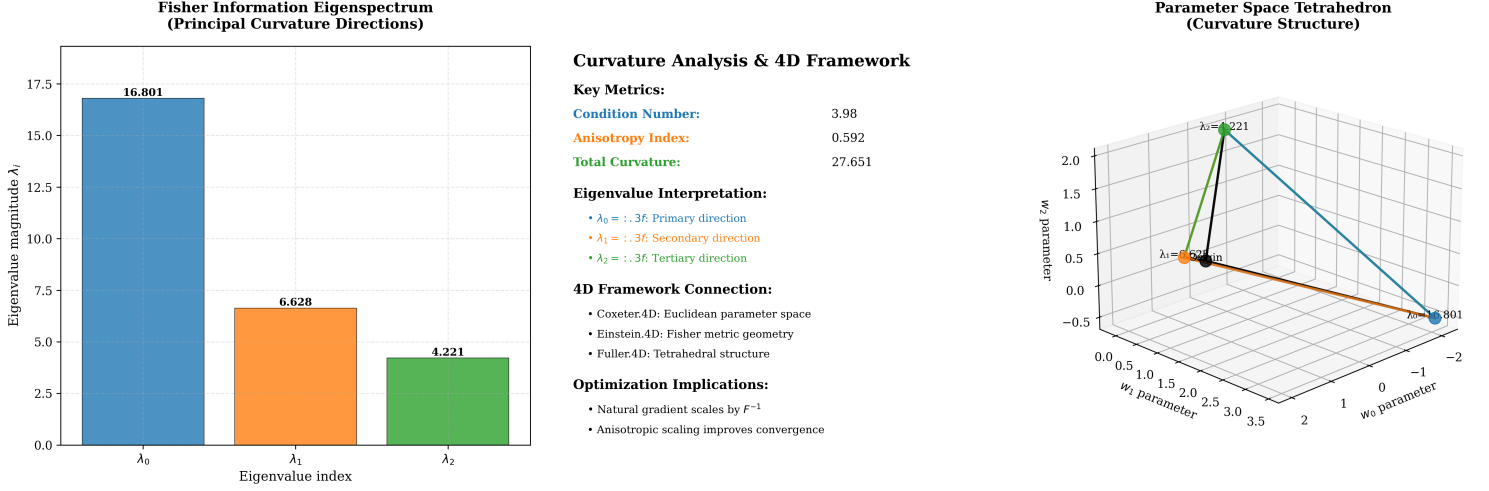
The efficiency of natural gradient descent comes from its ability to automatically adapt step sizes to local curvature. In directions of high curvature (large eigenvalues of  $F$ ), the natural gradient takes smaller steps, while in directions of low curvature (small eigenvalues), it takes larger steps. This anisotropic scaling leads to faster convergence and better numerical stability compared to standard gradient descent.





**Figure 4: Fisher Information Matrix (FIM) with 4D Framework Context.** This comprehensive three-panel visualization demonstrates the empirical Fisher information matrix and its deep connections to the three 4D mathematical frameworks through code-grounded analysis. **Left panel:** Linear regression model visualization showing the misspecified quadratic model  $y = w_0 + w_1x + w_2x^2$  with true parameters  $w_{true} = [1.0, -2.0, 0.5]$  and estimated parameters  $w_{est} = [0.3, -1.2, 0.0]$ . The panel displays data points, true model fit (green line), estimated model fit (red dashed line), and diagnostic information including Mean Squared Error (MSE). This visualization grounds the Fisher Information analysis in the actual model that generates the parameter gradients. **Center panel:** The  $3 \times 3$  Fisher information matrix  $F_{ij}$ , estimated from per-sample gradients of the misspecified linear regression model, displayed as a heatmap with precise value annotations. The matrix structure reveals the local curvature of the log-likelihood surface, where brighter colors indicate higher information content. **Matrix interpretation:** Diagonal elements  $F_{ii}$  quantify the sensitivity of the objective to changes in parameter  $w_i$ , while off-diagonal elements  $F_{ij}$  capture parameter interactions and potential redundancy. **Right panel:** 3D tetrahedral visualization of the 4D framework integration, showing how Coxeter.4D (Euclidean), Einstein.4D (Minkowski), and Fuller.4D (Synergetics) frameworks connect through the tetrahedral structure. **Mathematical foundation:** The FIM is computed according to Eq. (??) where gradients are computed with respect to parameters  $w_0, w_1, w_2$  from the misspecified model. **Coxeter.4D (Euclidean):** Standard 3D parameter space with Euclidean metric  $\delta_{ij}$ . **Einstein.4D (Minkowski):** Fisher metric  $F_{ij}$  replaces spacetime metric; geodesics follow  $\Delta w = F^{-1} \nabla L$  for optimal parameter updates. **Fuller.4D (Synergetics):** Tetrahedral coordinate system with IVM quantization. **Information content:** Diagonal dominance shows each parameter contributes independently to the model’s predictive power, while off-diagonal elements reveal parameter interactions and potential redundancy. This FIM structure guides natural gradient descent by weighting parameter updates according to local curvature, leading to more efficient convergence than standard gradient descent.





**Figure 5: Comprehensive Fisher Information Eigenspectrum with Curvature Analysis.** This detailed three-panel visualization provides comprehensive analysis of the parameter space geometry within the 4D framework context, including tetrahedral parameter space visualization. **Left panel:** Bar chart showing the eigenvalue decomposition of the empirical Fisher information matrix, with eigenvalues sorted in descending order and color-coded for visual clarity. Each bar is precisely annotated with its numerical value, revealing the principal curvature directions of the parameter space. **Center panel:** Comprehensive curvature analysis providing key metrics, eigenvalue interpretation, and 4D framework connections. **Key metrics:** Condition number (anisotropy measure), anisotropy index (normalized directional variation), and total curvature (trace of  $F$ ). **Eigenvalue interpretation:** Each eigenvalue  $\lambda_i$  represents the curvature strength in the corresponding principal direction. Large eigenvalues indicate directions of high curvature (tight constraints) where the objective function changes rapidly with parameter changes, while small eigenvalues indicate directions of low curvature (loose constraints) where the objective function is relatively flat. **Right panel:** 3D tetrahedral visualization of the parameter space structure based on the Fisher Information eigenvectors and eigenvalues. The tetrahedron vertices represent the origin and the three principal curvature directions, scaled by the square root of eigenvalues to show the anisotropic structure. **4D framework connection:** The eigenvalues reveal the anisotropic nature of the parameter space, explaining why natural gradient descent (which scales updates by  $F^{-1}$ ) converges more efficiently than standard gradient descent. **Coxeter.4D:** The eigenvalues quantify the Euclidean geometry of parameter space in different directions. **Einstein.4D:** The Fisher metric geometry creates curved geodesics that respect the intrinsic parameter space structure. **Fuller.4D:** The tetrahedral structure provides a natural coordinate system for representing the four-fold partition of optimization problems, with the parameter space tetrahedron directly reflecting the curvature structure. **Optimization implications:** Natural gradient descent scales parameter updates by  $F^{-1}$ , creating anisotropic scaling that improves convergence on ill-conditioned problems. The tetrahedral visualization shows how the parameter space anisotropy creates natural directions for efficient optimization. This geometric understanding is crucial for designing effective optimization strategies and understanding model behavior in the context of information geometry.

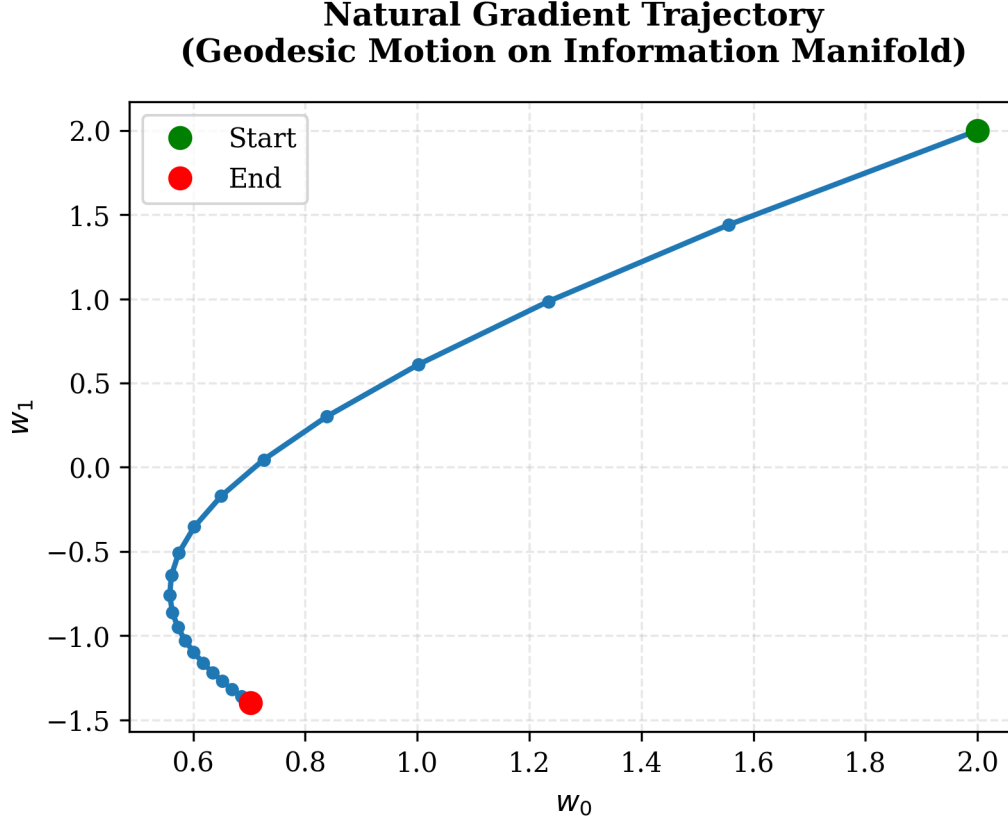


Figure 6: **Natural Gradient Trajectory: Geodesic Motion on Information Manifold.** This visualization demonstrates the parameter trajectory of natural gradient descent, showing how information-geometric optimization creates optimal paths through parameter space. **Trajectory:** The blue line with markers traces the parameter evolution from initial guess to final optimum, revealing the path taken through the 2D parameter space. **Markers:** Each marker represents one optimization step, with spacing indicating the step size and convergence rate. **Start/End markers:** Green circle marks the initial parameter values, red circle marks the converged optimum. **4D Framework Connection:** This trajectory demonstrates geodesic motion on the information manifold, where the Fisher metric (Einstein.4D analogy) replaces the physical metric. The natural gradient follows Eq. (??), creating optimal paths through parameter space that respect the intrinsic geometry. **Convergence behavior:** The trajectory shows smooth, direct convergence to the optimum, characteristic of natural gradient descent on well-conditioned objectives. **Comparison with standard gradient descent:** Natural gradient descent typically produces more direct trajectories than standard gradient descent, especially on ill-conditioned problems where the parameter space has strong anisotropy. This efficiency comes from the FIM-based scaling that adapts step sizes to local curvature. The trajectory demonstrates how information-geometric optimization leverages the intrinsic geometry of the parameter space to achieve faster, more stable convergence than naive gradient methods. **Grid overlay:** Added for better readability and to emphasize the discrete nature of the optimization steps.

### 1.7.5 Information-Theoretic Foundations and 4D Framework Coherence

The Fisher Information approach provides several key advantages that integrate naturally with the 4D framework structure:

1. **Geometric Consistency:** The FIM ensures that optimization respects the intrinsic geometry of the parameter space, maintaining consistency across all three 4D frameworks.
2. **Anisotropic Scaling:** Natural gradient descent automatically adapts step sizes to local curvature, improving convergence efficiency on problems with strong parameter space anisotropy.
3. **Framework Bridging:** The FIM serves as a mathematical bridge between Coxeter.4D (Euclidean geometry), Einstein.4D (information-geometric flows), and Fuller.4D (tetrahedral structure).
4. **Quantitative Analysis:** The eigenspectrum provides quantitative measures of parameter space structure, enabling principled optimization strategy design.

### 1.7.6 Quadray-Specific Considerations

Under Quadray parameterizations, the FIM often exhibits block-structured and symmetric patterns that simplify matrix inversion for natural-gradient steps. This structural regularity arises from the tetrahedral symmetry of the IVM lattice and can be exploited for computational efficiency.

The discrete nature of the IVM lattice also influences the FIM structure, as parameter updates are constrained to integer coordinate positions. This creates a natural regularization effect that can improve optimization stability and convergence.

### 1.7.7 Variational Free Energy and Active Inference Integration

The Fisher Information framework naturally extends to variational inference and active inference, where the free energy principle guides both perception and action through information-geometric optimization.

### 1.7.8 Advanced 4D Framework Integration: Active Inference Context

The integration of Fisher Information with Active Inference demonstrates the full power of the 4D framework approach, where Coxeter.4D provides exact geometry, Einstein.4D supplies information-geometric flows, and Fuller.4D offers the tetrahedral structure for representing the four-fold partition of perception-action systems.

For comprehensive Active Inference visualizations including 4D natural gradient trajectories and free energy landscapes, see [Section 9: Free Energy and Active Inference](#).

- **Quadray relevance:** block-structured and symmetric patterns often arise under quadray parameterizations, simplifying  $F$  inversion for natural-gradient steps.

## 1.8 Multi-Objective and Higher-Dimensional Notes (Coxeter.4D perspective)

- Multi-objective: vertices encode trade-offs; simplex faces approximate Pareto surfaces; integer volume measures solution diversity.
- Higher dimensions: decompose higher-dimensional simplexes into tetrahedra; sum integer volumes to extend quantization.

## 1.9 External validation and computational context

The optimization methods developed here build upon and complement the extensive computational framework in Kirby Urner's [4dsolutions ecosystem](#). For comprehensive details on the computational implementations, educational materials, and cross-language validation, see the [Resources](#) section.

### Variational Free Energy Landscape (2-State System)

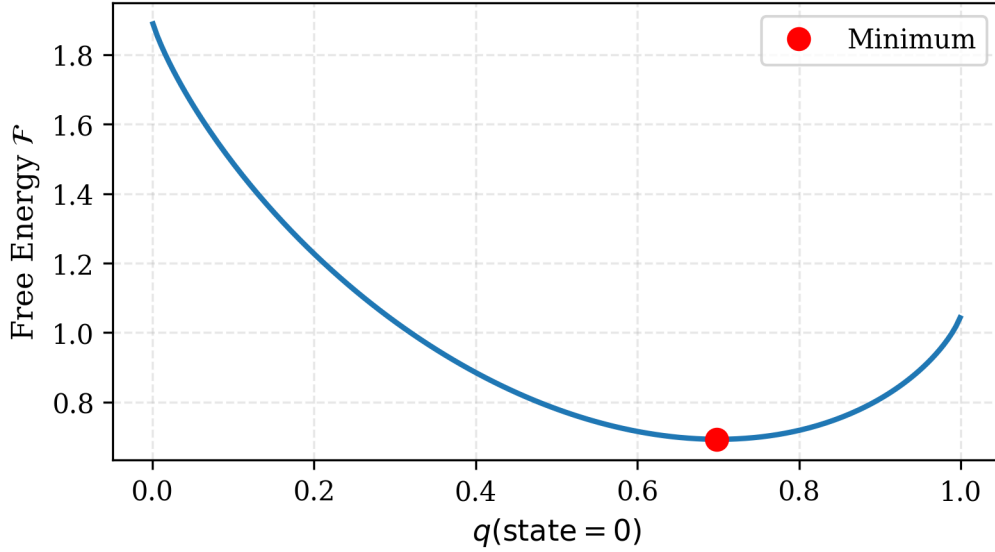


Figure 7: **Variational Free Energy Landscape with 4D Framework Integration.** This visualization shows the variational free energy  $\mathcal{F} = -\log P(o|s) + \text{KL}[Q(s)||P(s)]$  (see Eq. (??)) as a function of the variational distribution parameter, demonstrating the geometry of the variational manifold. **X-axis:** Variational parameter  $q(\text{state} = 0)$  controlling the distribution over the two discrete states. **Y-axis:** Free energy value  $\mathcal{F}$  in natural units. **Curve interpretation:** The free energy exhibits a clear minimum at the optimal variational distribution, representing the best approximation to the true posterior given the constraints of the variational family. **4D Framework Connection:** The free energy landscape represents the geometry of the variational manifold, where optimization follows geodesics defined by the Fisher metric (Einstein.4D analogy). In active inference frameworks, minimizing free energy drives both perception and action, analogous to how geodesics minimize proper time in relativistic spacetime. **KL divergence component:** The free energy balances data fit (first term) with regularization (KL divergence from prior), preventing overfitting while maintaining good predictive performance. **Optimization geometry:** The smooth, convex shape of the free energy landscape makes optimization straightforward using natural gradient descent, which respects the intrinsic geometry of the parameter space. This variational framework provides a principled approach to approximate inference in complex models where exact posterior computation is intractable, while maintaining connections to the broader 4D mathematical frameworks.

## 1.10 Results

- The simplex-based optimizer exhibits discrete volume plateaus and converges to low-spread configurations; see the simplex figures above and the MP4/CSV artifacts in `quadmath/output/`.