Research Article *Open Access*

Journal of Pioneering Artificial Intelligence Research

ISSN: 3069-0846

DOI: doi.org/10.63721/26JPAIR0126

From Eigen-Decomposition to Iterative Descent: Structural and Trajectory Paradigms in Artificial Intelligence

Huseyin Murat Cekirge

Department of Mechanical Engineering, The City College of New York (CUNY), New York, USA

Citation: Huseyin Murat Cekirge (2026) From Eigen-Decomposition to Iterative Descent: Structural and Trajectory Paradigms in Artificial Intelligence. J of Poin Artf Research 2(1), 1-13. WMJ-JPAIR-126

Abstract

Contemporary machine learning is overwhelmingly framed as an optimization problem. Gradient descent and its variants define learning as a trajectory in parameter space, governed by step sizes, penalty weights, and convergence criteria. In such formulations, the solution is not assumed it is approached. This work challenges that framing in the linear regime. When a linear system satisfies the compatibility condition A W = b, equilibrium is not produced by optimization. It is already encoded in the algebraic structure of the system. Both L_1 and L_2 objectives attain zero residual at the same structural solution. The minimum does not emerge from descent dynamics; it exists as a consequence of determinacy. Optimization, in this setting, is epistemic — it reveals a solution. Equilibrium is structural it precedes the algorithm. Regularization does not create equilibrium. It modifies the geometry surrounding an equilibrium that may already be present. It stabilizes degeneracy, suppresses ill-conditioning, and reshapes curvature. But when the system is structurally solvable, the equilibrium is not a product of penalty tuning. Principal Component Analysis exemplifies this distinction. Principal directions are obtained through eigen-decomposition of the covariance matrix. No learning rate governs their emergence. No penalty trajectory produces them. The spectral structure determines them outright. The prevailing narrative in machine learning equates learning with iterative convergence. This work separates the two. Trajectory-based convergence is necessary when structural determinacy is absent or obscured. But when determinacy is present, equilibrium-based inference supersedes optimization. The distinction is not computational, it is ontological, optimization describes how we search and determinacy describes what already is. Recognizing this separation clarifies a central issue: iteration is indispensable in underdetermined or nonlinear regimes, but in structurally determined systems it is often a method of access, not a mechanism of existence.

***Corresponding author:** Huseyin Murat Cekirge, Department of Mechanical Engineering, The City College of New York (CUNY), New York, USA. **E-mail:** hmcekirge@usa.net.

Submitted: 16.02.2026 **Accepted:** 20.02.2026 **Published:** 10.03.2026

Keywords: Structural Learning, Trajectory-Based Optimization, Equilibrium Inference, Eigen-Decomposition, Principal Component Analysis, Gradient Descent, Regularization, Elastic Net, L1 And L2 Penalties, Parameter Dependence, Structural Determinacy, Sigma-Regularization, Linear Inverse Problems, Learning Paradigms

Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
EB	Equilibrium-Based Inference
ED	Eigen-Decomposition
EN	Elastic Net
GD	Gradient Descent
IP	Inverse Problem
L1	L1 Norm Regularization
L2	L2 Norm Regularization
LR	Learning Rate
LS	Least Squares
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
RB	Regularization-Based Stabilization
SGD	Stochastic Gradient Descent
SPD	Symmetric Positive Definite
SVD	Singular Value Decomposition
TB	Trajectory-Based Learning
σ	Stabilizing Regularization Parameter

Introduction

Machine learning methodologies are commonly organized around iterative optimization. From early backpropagation algorithms to modern large-scale stochastic gradient procedures, model training is typically described as a sequence of parameter updates guided by gradient information [1-7]. In regularized regression, L_1 and L_2 penalties are introduced to promote sparsity and numerical conditioning, while Elastic Net combines both to address multicollinearity and grouping effects. These formulations depend explicitly on hyperparameters such as λ_1, λ_2, and learning rate η, which influence convergence behavior and final parameter configuration. This trajectory-based perspective has become dominant, particularly in deep learning and high-dimensional inference. Within this view, the solution is defined operationally by the path taken in parameter space. However, not all learning problems are structurally trajectory-defined. In classical linear analysis, solutions are characterized by equilibrium conditions rather than by update dynamics. For instance, Principal Component Analysis is derived from eigen-decomposition of the covariance matrix and does not require gradient descent to define its solution [8,9]. The principal components are determined by intrinsic structural properties of the data rather than by iterative adjustment or learning-rate control.

This distinction motivates a broader conceptual separation between parameter dependence and structural determinacy.

When a linear system satisfies $A W = b$ exactly, both L_1 and L_2 objectives attain zero residual at the same equilibrium point. Under such conditions, the minimum is determined by algebraic compatibility rather than by descent dynamics. The equilibrium is encoded in the system itself. Iterative optimization may serve as a numerical mechanism for approaching the solution, but it does not define the solution's existence.

Recent work on deterministic σ-regularized equilibrium inference further explores this structural viewpoint, demonstrating how stability can be ensured without reliance on trajectory-dependent training [10-18]. In this framework, regularization does not create minima through parameterized descent. Instead, it stabilizes degeneracies, controls ill-conditioning, and selects unique representatives among structurally equivalent equilibria. Regularization reshapes geometry; it does not manufacture equilibrium where algebraic compatibility already determines it.

The present study therefore advances a paradigm-level clarification. Iterative methods govern how solutions are approached; structural equilibrium governs whether solutions are already encoded in the system. Optimization describes the procedural route. Determinacy describes the structural fact. By separating these roles, optimization-based and equilibrium-based perspectives can be interpreted as complementary rather than contradictory, operating at distinct logical levels within learning theory [19].

L_2–L_1 Minimum Equivalence and the Geometry of Decision

Consider the residual

$r(W) = A\,W - b$ (1)

and define the objectives

$L_2(W) = \|A\,W - b\|_2^2$ (2)

$L_1(W) = \|A\,W - b\|_1$ (3)

Assume structural compatibility: there exists W^* such that

$A\,W^* = b.$ (4)

Then $r(W^*) = 0.$ (5)

Since both L_2 and L_1 are nonnegative and attain zero if and only if the residual vanishes, it follows that

$L_2(W^*) = 0$ (6)

$L_1(W^*) = 0.$ (7)

Therefore, W^* minimizes both objectives. If A has full column rank (rank n), the minimizer is unique. Under structural compatibility and full column rank, the compatibility solution is the unique minimizer for both L1 and L2; “

$\arg\min L_2(W) = \arg\min L_1(W).$ (8)

The coincidence follows directly from algebraic compatibility rather than from geometric similarity.

Geometric Interpretation

The L_2 objective defines a smooth quadratic valley with gradient

$\nabla L_2(W) = 2\,A^T (A\,W - b).$ (9)

At W^*, the gradient vanishes. The L_1 objective is convex but non-smooth. Its optimality condition is expressed through subgradients:

$0 \in A^T s,$ (10)

where $s_i \in \partial|r_i|$ and component-wise,

$$s_i \in \partial|r_i| = \begin{cases} \operatorname{sign}(r_i), & r_i \neq 0 \\ [-1, 1], & r_i = 0 \end{cases} \quad (11)$$

At equilibrium point W^*, where $r(W^*) = 0$ the choice of $s=0$ satisfies the optimality condition.

Hence both the smooth Gaussian model (L_2) and the non-smooth Laplace model (L_1) admit the same equilibrium under compatibility.

The distinction between quadratic curvature and piecewise-linear geometry becomes relevant only away from equilibrium.

Parameter Dependence versus Structural Determinacy

Modern optimization-based learning typically defines solutions through parameterized trajectories. In gradient-based methods,

$W(t+1) = W(t) - \eta\,\nabla L(W(t)),$ (12)

where η is the learning rate. Convergence depends on initialization, step size, and stopping criteria.

Regularized objectives introduce additional parameters:

$L(W) = \|A\,W - b\|_2^2 + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_2^2 .$ (13)

The penalty magnitudes λ_1 and λ_2 shape sparsity and conditioning. The resulting estimator depends not only on the data (A, b) but also on hyperparameter selection.

Elastic Net combines L_1 and L_2 penalties to balance sparsity and stability. Its behavior is governed by both penalty magnitudes and the chosen optimization trajectory.

Structural determinacy differs fundamentally. When $A\,W = b$ is solvable, the equilibrium is encoded in the algebraic structure. Iteration governs how the solution is approached; compatibility governs whether the solution already exists.

Regularization becomes structurally relevant only under degeneracy, incompatibility, or noise.

Role of Regularization

If A^TA is singular or ill-conditioned, the equilibrium may be non-unique or unstable. Introducing σ-regularization yields

$W\sigma = (A^TA + \sigma I)^{-1} A^T b.$ (14)

This shifts eigenvalues upward, ensuring positive definiteness and uniqueness.

When compatibility is exact and the solution is unique, regularization does not create a new minimum; it stabilizes or selects among solutions only when structural freedom exists.

Regularization parameters govern the bias–variance trade-off. Increasing penalty strength reduces flexibility and may induce underfitting. Decreasing penalty strength increases flexibility and may expose the model to overfitting in noisy regimes.

In compatible and uniquely solvable systems, equilibrium is independent of penalty magnitude.

Prevention of Overfitting

Overfitting arises when structural compatibility is absent or when the parameter space contains unnecessary degrees of freedom.

Equilibrium-based inference prevents unnecessary parameter chasing by resolving compatibility prior to introducing penalty-driven preferences.

Regularization may be chosen through a discrepancy principle:
$\|A W - b\| \leq \varepsilon$, (15)

where ε reflects noise magnitude. When $\varepsilon = 0$, equilibrium is independent of σ. Under incompatibility, σ is selected to enforce the prescribed tolerance.

Structural Conclusion

If the system is consistent:

- Both L_1 and L_2 attain zero residual.
- The equilibrium condition is identical.
- The difference between models is geometric, not algebraic.
- Compatibility defines equilibrium.
- Regularization stabilizes degeneracy.
- Gradient descent provides trajectory.

These roles operate at distinct conceptual levels.

Conceptual Diagram of Three Learning Regimes

Learning frameworks can be organized into three structurally distinct regimes:

1. Trajectory-Based Learning
   Solution defined by iterative dynamics.
2. Regularization-Based Stabilization
   Solution shaped by penalty-modified geometry.
3. Equilibrium-Based Inference
   Solution defined by structural compatibility $A W = b$.

The distinction is structural rather than oppositional. Iterative methods approximate equilibria; regularization stabilizes imperfections; equilibrium analysis identifies when the solution is already encoded.

To clarify how penalty terms modify the equilibrium, consider the unregularized least-squares solution
$W^* = (A^T A)^{-1} A^T b$, (16)

assuming $A^T A$ is invertible.

With quadratic regularization,
$W\sigma = (A^T A + \sigma I)^{-1} A^T b$. (17)
If $A^T A = Q \Lambda Q^T$ (18)

is the eigen-decomposition, then
$W^* = \Sigma\, (1/\lambda_i)\, q_i q_i^T A^T b$ (19)

And projection components,
$W\sigma = \Sigma\, (1/(\lambda_i + \sigma))\, q_i q_i^T A^T b$ (20)

projection components.

$$\frac{\lambda_i}{\lambda_i + \sigma} \qquad (21)$$

Each eigendirection is scaled by implying stronger shrinkage along directions associated with small eigenvalues. Thus, quadratic regularization induces continuous contraction toward the origin, particularly in ill-conditioned directions.

$0 \in A^T(AW - b) + \lambda\, \partial\|W\|_1$ (22)

For L_1 regularization, the optimality condition produces thresholding behavior: small coefficients may collapse to zero, leading to sparsity. The displacement is no longer uniform across eigendirections but depends on coordinate magnitude relative to λ.

Elastic Net combines these effects, blending eigenvalue-dependent shrinkage with sparsity-inducing thresholding.

In compatible and uniquely solvable systems, these modifications represent bias rather than structural necessity. As $\sigma \to 0$ (or $\lambda \to 0$), the regularized solution converges to the structural equilibrium W^*.

Determination of the Regularization Parameter σ

Consider the regularized solution

$$W\sigma = (A^T A + \sigma I)^{-1} A^T b. \qquad (23)$$

Let the eigen-decomposition of $A^T A$ be

$$A^T A = Q \Lambda Q^T, \qquad (24)$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, with $\lambda_i \geq 0$.
Then

$$W\sigma = \Sigma\, (1/(\lambda_i + \sigma))\, q_i q_i^T A^T b. \; (25)$$

Suppression of Small Eigenvalues

If λ_i is very small, the corresponding direction is ill-conditioned. Without regularization:

$1/\lambda_i$ becomes large → instability. (26)

With regularization:

$$1/(\lambda_i + \sigma) \qquad (27)$$

remains bounded. Thus σ acts as a lower spectral threshold. Directions with $\lambda_i \ll \sigma$ are effectively suppressed. This removes numerically insignificant components.

Discrepancy Principle

Let ε denote an acceptable global error tolerance.
Define σ such that

$$\|A W\sigma - b\|_2 \leq \varepsilon. \qquad (28)$$

This ensures that:

- The solution fits the data within the prescribed tolerance.
- Noise amplification is avoided.
- Overfitting is controlled structurally.

In the compatible regime ($\varepsilon = 0$), σ may be chosen minimal or zero. In the incompatible regime, σ is chosen to match the noise level.

Practical Determination

$$\|A W\sigma - b\|_2 = \varepsilon \qquad (29)$$

σ can be selected by enforcing or by monitoring the decay of eigenvalues and choosing σ near the numerical stability threshold. Thus, σ balances:

- Stability (spectral conditioning)
- Accuracy (residual tolerance)

Structural Interpretation

Regularization does not define equilibrium when compatibility holds. Instead, σ governs the admissible resolution of unstable spectral directions and enforces a global error bound.
Small σ → higher resolution, higher variance.
Large σ → stronger suppression, higher bias.

The parameter therefore controls spectral filtering rather than equilibrium existence. The regularization parameter σ may be determined through spectral thresholding and global discrepancy control, ensuring stability while respecting a prescribed residual bound. In compatible systems, σ regulates numerical conditioning; in incompatible systems, it governs admissible approximation error.

Elastic Net as a Combined Geometric Regime

Elastic Net minimizes

$$\|A W - b\|_2^2 + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_2^2. \qquad (30)$$

It combines two geometric effects:

- L_2 penalty → smooth spectral shrinkage
- L_1 penalty → sparsity via thresholding

Thus, Elastic Net operates simultaneously in:

1. Spectral contraction space (eigenvalue-dependent shrinkage)
2. Coordinate sparsity space (polyhedral selection)

Structural Interpretation

Elastic Net does not introduce a new equilibrium principle. It combines two regularization geometries to stabilize ill-posed or high-dimensional systems. When $A W = b$ is structurally compatible and uniquely solvable:
The residual-zero equilibrium remains primary.

Elastic Net then modifies that equilibrium through:
As $\lambda_1, \lambda_2 \to 0$,
• Continuous contraction (λ_2 effect)
• Variable elimination (λ_1 effect)

$$\lim_{\lambda_1 \to 0,\, \lambda_2 \to 0} W_{EN}(\lambda_1, \lambda_2) = W^*$$

$$W_{EN}(\lambda_1, \lambda_2) = \arg\min_W \left(\|AW - b\|_2^2 + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_2^2\right) \;, \quad (31)$$

where WEN denotes the elastic net estimator.
Thus, the combined penalty governs bias and stability, not existence of equilibrium.

Figure 1: Level sets of L_2, L_1, and $L_2 + L_1$ (Elastic Net) regularization norms in parameter space. The L_2 norm produces isotropic circular geometry, the L_1 norm produces polyhedral sparsity-inducing geometry, and Elastic Net blends smooth shrinkage with softened sparsity structure.

The Elastic Net constraint is
$\lambda_1 \|W\|_1 + \lambda_2 \|W\|_2^2 \leq t$ (32)

In 2D, let $W = (w_1, w_2)$. Then the boundary satisfies
$\lambda_1 (|w_1| + |w_2|) + \lambda_2 (w_1^2 + w_2^2) = t$. (33)

This equation tells you exactly how the shape should look. Elastic Net represents a combined regularization regime that blends spectral shrinkage and sparsity induction. By simultaneously penalizing $\|W\|_1$ and $\|W\|_2^2$, it stabilizes ill-conditioned directions while promoting variable selection. Within the present structural hierarchy, Elastic Net occupies an intermediate position: it does not redefine equilibrium under compatibility, but modulates solution geometry when degeneracy, noise, or high dimensionality is present.

Norm Geometry and Dual-Parameter Determination

Regularization is typically introduced as an algebraic penalty appended to a loss function. This view is operational but incomplete. Before discussing parameters, one must recognize that norms impose geometry. Regularization does not merely scale a term; it reshapes the energy landscape in which determination occurs.

The figures in this section are therefore not algorithmic diagrams. They are structural maps.

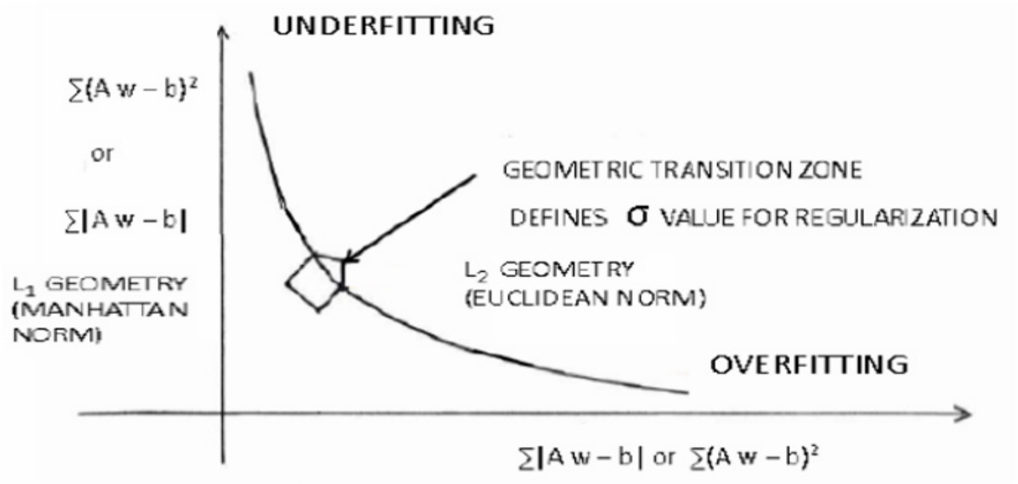

Figure 2: Norm-Dependent Regularization Geometry

The horizontal axis represents the intrinsic hybrid data misfit
$\sum|A\,w - b| + \sum(A\,w - b)^2$, (34)

while the vertical axis represents regularization magnitude measured by
$\sum|w|$ (L_1 norm) (35)
or
$\sum w^2$ (L_2 norm). (36)

The distinction is geometric and fundamental.

The L_2 norm induces smooth, isotropic contraction. Coefficients shrink continuously under increasing regularization. The geometry is curvature-dominated; equilibrium emerges through smooth spectral damping.

The L_1 norm, in contrast, induces axis-aligned sparsity. Small coefficients collapse abruptly toward zero. The geometry is corner-dominated; equilibrium emerges through directional elimination.

Thus, L_1 and L_2 are not variations of the same mechanism.

They are distinct determination principles:

- Curvature-based stabilization (L_2)
- Structure-selective elimination (L_1)

Even before blending them, the geometry already encodes two incompatible structural logics.

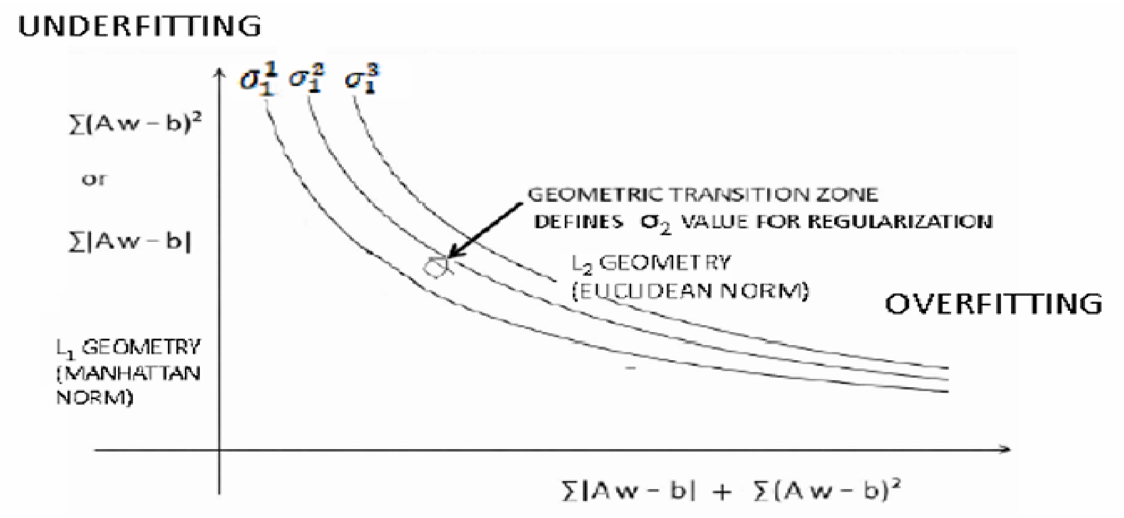

Figure 3: Structural (σ_1) and Curvature (σ_2) Meridian Families

In Figure 3, the axes represent intrinsic data energy and regularization magnitude as before. However, σ does not appear explicitly as a multiplier.

Instead, σ_1 and σ_2 generate meridian families through the deterministic mapping $w = w(\sigma)$.
Each meridian is a structural trajectory in energy space.

For small σ_1, the solution retains many active coefficients. Data fidelity dominates; variance is high. As σ_1 increases, coefficients are selectively eliminated. Structural magnitude decreases, and the slope of the meridian shifts. This slope change marks the geometric boundary between overfitting and structural compression. σ_1 is therefore not a scaling factor. It is a coordinate that generates structural elimination paths. σ_2 governs a different mechanism. Increasing σ_2 suppresses directions associated with small spectral components. Contraction becomes isotropic; instability diminishes. However, excessive σ_2 introduces bias through over-smoothing. Again, the transition is visible as a geometric slope change along the meridian family.

The two parameters act along orthogonal structural dimensions:
σ_1 → sparsity direction
σ_2 → curvature direction

Geometric Interpretation: Elastic Net as a Unified Regime

Elastic Net is often described as a weighted sum of L_1 and L_2 penalties. This algebraic description is secondary. Geometrically, Elastic Net is a unified regime in which sparsity and curvature coexist within a single energy surface. The horizontal axis measures intrinsic data fidelity. The vertical axes measure structural magnitude and curvature magnitude.
σ_1 and σ_2 do not scale the axes; they generate families of admissible geometric trajectories.

The equilibrium is not produced by regularization. The equilibrium exists as a structural possibility. Regularization determines how that equilibrium is stabilized. In this sense, Elastic Net is not a compromise. It is a geometric synthesis and the fact that this entire synthesis is captured in a single figure is not a simplification. It is evidence that the determination mechanism is fundamentally geometric.

Cekirge Extension: Deterministic Dual-Geometry Stabilization

The geometric interpretation above admits a further structural refinement within the Cekirge equilibrium framework. In conventional formulations, Elastic Net is treated as a weighted superposition of penalties. In the Cekirge extension, however, the dual-parameter structure is not interpreted as a compromise between norms, but as a deterministic stabilization surface embedded in equilibrium space.

The equilibrium is defined first through structural compatibility. The parameters σ_1 and σ_2 do not create this equilibrium. Instead, they govern how the equilibrium behaves under structural and spectral perturbations.

Within this framework:

- σ_1 regulates structural elimination along coordinate directions.
- σ_2 regulates spectral stabilization along eigen-directions.

These two mechanisms operate on orthogonal geometric axes. Their interaction generates a two-dimensional stabilization surface rather than a one-dimensional penalty path.

Unlike trajectory-based optimization, where the solution is approached through descent dynamics, the Cekirge extension assumes that equilibrium is algebraically encoded whenever compatibility holds. The role of σ-parameters is therefore not generative but selective and stabilizing.

In this sense, Elastic Net becomes a special case of a broader deterministic stabilization principle. The meridian families depicted in Figure 3 are not optimization trajectories; they are structural slices of a pre-defined equilibrium surface parameterized by σ_1 and σ_2.

Regularization, under this interpretation, does not define the minimum. It controls the admissible geometric resolution of that minimum in the presence of degeneracy, ill-conditioning, or noise.
Thus, the Cekirge extension reframes dual-penalty learning:

From parameter-weighted optimization
to deterministic dual-geometry stabilization.

The fact that the entire dual-parameter interaction can be represented within a single structural diagram is not merely illustrative. It reflects the underlying determinacy of the equilibrium structure itself.

Spectral Determination Surface

The equilibrium solution emerges from the geometric interaction between L_1-induced sparsity, L_2-induced smooth curvature, and the intrinsic principal directions of the data covariance operator. Rather than following iterative descent trajectories, the solution collapses onto a structurally stable intersection region defined by curvature balance and constraint geometry.

Regularization is commonly interpreted as shrinkage. Here it is reinterpreted as a determination multiplier acting directly on the curvature spectrum of the operator A^TA. The σ-parameter does not merely scale a penalty term; it shifts the equilibrium along a curvature-controlled trade-off surface between data fidelity and structural constraint.

In two-dimensional spectral space defined by eigenvalues (λ_1, λ_2), equilibrium is governed not only by data misfit but by the relative magnitude of σ with respect to the spectral spread. When $\sigma \ll \lambda_1$, the system remains weakly determined along the dominant eigen-direction. As σ approaches λ_1, a regime transition occurs. Beyond this critical region, the optimization landscape becomes curvature-dominated and structurally stabilized.

This transition produces a geometric bifurcation in the determination surface, visible as a ridge in spectral space. The ridge separates weak-determination and strong-determination regimes. Thus, σ does not merely penalize magnitude; it reshapes the spectral geometry of optimization.

We extend the classical regularization framework into a three-axis determination space. The axes represent:

- Hybrid data misfit
- L_1 structural magnitude
- L_2 curvature energy

Each point on the resulting surface is generated parametrically by (λ_1, λ_2), forming a two-parameter determination surface. Meridian curves corresponding to constant λ_1 or constant λ_2 sweep the surface and reveal curvature transitions where sparsity and smoothness balance.

The equilibrium does not lie on a simple convex contour. It emerges as a triple intersection of data, sparsity, and curvature surface. The optimal spectral configuration arises as a critical point of this three-dimensional geometry. This point is not imposed externally; it follows from intrinsic spectral structure.

The deterministic solution $w(\sigma_1)$ is obtained prior to the introduction of σ_2. The parameter σ_2 does not alter the structural mapping itself; it scales the quadratic curvature component in the post-solution energy evaluation. This separation distinguishes structural determination from curvature assessment.

Although Elastic Net is convex and solvable through optimization algorithms, exhaustive parameter exploration via iterative descent becomes computationally intensive in high-dimensional settings. The present dual-σ construction enables direct generation of the determination surface without repeated trajectory-based optimization.

In this formulation, regularization weights are not treated as external tuning knobs but as coordinates in spectral geometry. The transition between overfitting and underfitting appears as a slope change along meridian curves of the surface rather than as an artifact of descent dynamics.

Principal Component Analysis

The coincidence of L_1 and L_2 residual minima under compatibility illustrates that equilibrium is determined by algebraic structure. Principal Component Analysis provides a canonical instance of this principle. The principal directions are obtained through eigen-decomposition of the covariance matrix without reliance on iterative descent or penalty tuning. The solution is encoded in the structure of the data itself.

PCA as a Canonical Example of Structural Equilibrium

Let $X \in \mathbb{R}^{m\times n}$ denote centered data and define the covariance matrix

$$C = (1/m)\, X^TX. \qquad (37)$$

Principal Component Analysis determines directions v by solving the eigenvalue problem

$$C\, v = \lambda\, v. \qquad (38)$$

The principal components are defined by spectral structure rather than by iterative descent. No learning rate, penalty parameter, or trajectory-dependent tuning is required. The dominant directions emerge from algebraic compatibility conditions intrinsic to the covariance operator.

When equilibrium equations are explicitly available, the solution is encoded in the operator itself. Optimization trajectories may approximate the solution numerically, but they do not define it.

Spectral Structure

Let the spectral decomposition of C be
$C = Q \Lambda Q^T$, (39)

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$, and Q contains orthonormal eigenvectors.

The k-th principal component corresponds to λ_k.
The total variance captured equals λ_k.

Ordering eigenvalues determines dominance. No descent trajectory defines these directions.

Explicit Two-Dimensional Illustration

Consider a 2×2 covariance matrix

$$C = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \qquad (41)$$

Its eigenvalues satisfy
$\lambda^2 - (a + b)\lambda + (ab - c^2) = 0$. (41)

Thus,
$\lambda = (½) [(a + b) \pm ((a - b)^2 + 4c^2)^{(1/2)}]$. (42)

The principal directions are obtained directly from this closed-form expression. No learning rate, penalty parameter, or iterative refinement is needed to define the solution. The equilibrium is algebraically encoded.

Structural Bridge to Regression

Regression equilibrium satisfies
$A^TA\,W = A^T b$. (43)

PCA equilibrium satisfies
$A^TA\,v = \lambda\,v$. (44)
Both are governed by the same symmetric operator A^TA.

In regression, the spectrum determines conditioning and stability of W.
In PCA, the spectrum determines variance directions and dimensional structure.

The distinction lies in the right-hand side: regression incorporates a response vector, whereas PCA extracts intrinsic structure through eigen-decomposition.

In both cases, the defining relation is algebraic rather than trajectory-based.

Constrained L_2 Interpretation

Maximizing projected variance
$\max \lVert X v \rVert_2^2$ (45)

subject to
$\lVert v \rVert_2 = 1$ (46)

leads directly to
$X^TX\,v = \lambda\,v$. (47)

The Lagrange multiplier arises from stationarity, not from hyperparameter tuning. The solution is structurally determined.

Structural Hierarchy

PCA demonstrates:

- Compatibility condition → eigen-equilibrium
- Constraint → normalization
- No learning rate
- No trajectory
- No shrinkage tuning

The equilibrium is defined before any algorithm is applied. Algorithms such as power iteration or SVD compute what already exists algebraically. This places PCA within the regime of equilibrium-based inference.

A Numerical Example: Pca without Iteration

Consider the symmetric covariance matrix

$$C = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix} \qquad (48)$$

This matrix could arise from centered 2D data.

Step 1 — Eigenvalues

By considering the matrice in Equation (48), solve

$$\det(C - \lambda I) = 0. \qquad (49)$$

That is,

$$\det \begin{bmatrix} 4 - \lambda & 2 \\ 2 & 3 - \lambda \end{bmatrix} \qquad (50)$$

Compute determinant:

$(4 - \lambda)(3 - \lambda) - 4 = 0$ (51)

Expand:

$12 - 4\lambda - 3\lambda + \lambda^2 - 4 = 0$ (52)

$\lambda^2 - 7\lambda + 8 = 0$ (53)

Solve quadratic:

$\lambda = [7 \pm (49 - 32)^{\frac{1}{2}}] / 2$ (54)

$\lambda = [7 \pm 17^{\frac{1}{2}}] / 2$ (55)

Numerically:

$\lambda_1 \approx (7 + 4.123) / 2 \approx 5.561$ (56)

$\lambda_2 \approx (7 - 4.123) / 2 \approx 1.439$ (57)

So total variance = 7 (58)

Principal variance direction ≈ 5.561 (59)

Step 2 — Eigenvectors

Solve, for $\lambda_1 \approx 5.561$:

$(C - \lambda_1 I)\, v = 0$ (60)

Matrix becomes approximately

$$\begin{bmatrix} -1.561 & 2 \\ 2 & -2.561 \end{bmatrix} \qquad (61)$$

From first row:

$-1.561\, v_1 + 2\, v_2 = 0 \rightarrow v_2 \approx 0.780\, v_1$ (62)

Normalize:

$$\begin{bmatrix} 0.7882 \\ 0.615 \end{bmatrix} \qquad (63)$$

This is the first principal component. The principal direction is determined entirely by the spectral equation

$C\, v = \lambda\, v$ (64)

The equilibrium is algebraic.

Structural Interpretation

This simple example illustrates the core principle:

The dominant direction is encoded in the operator C itself.

Computation may use iterative methods numerically, but the solution is defined structurally.
This is the equilibrium-based regime.

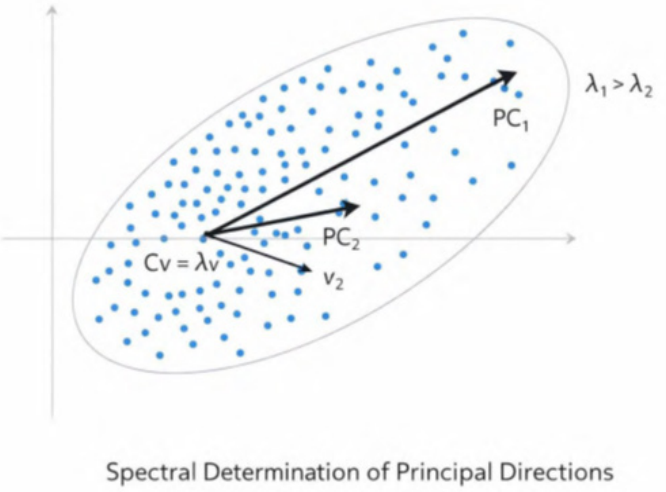

Figure 4: Principal Component Analysis identifies directions v satisfying C v = λ v, where C is the covariance matrix. The dominant principal component corresponds to the largest eigenvalue and aligns with the direction of maximum variance. The solution is determined by spectral structure rather than by iterative descent.

Relation of PCA to Structural Determinacy

Principal Component Analysis illustrates the structural regime identified in this work. While regression problems involve the normal equation

$A^TA\ W = A^T\ b,$ (65)

PCA solves the spectral equation

$A^TA\ v = \lambda\ v.$ (66)

In both cases, the same symmetric operator A^TA governs the solution. The distinction lies in the right-hand side: regression incorporates a response vector, whereas PCA extracts intrinsic structure through eigen-decomposition.

The key commonality is that the solution is encoded in the algebraic structure of the operator. The eigenvectors of A^TA are determined by spectral compatibility conditions and do not depend on learning rates, penalty magnitudes, or descent trajectories. Iterative algorithms may compute eigenvectors numerically, but they do not define them.

This parallels the L_2–L_1 coincidence established earlier. When the linear system $A\ W = b$ is structurally compatible, the residual vanishes and both Gaussian and Laplace residual models admit the same equilibrium. In PCA, the dominant directions similarly emerge from structural equilibrium rather than from path-dependent optimization.

Thus, PCA serves as a canonical example of equilibrium-based inference: the solution is determined by operator structure, while computation merely recovers what is already encoded.

Conclusion

This study examined the structural relationship between L_2 and L_1 residual formulations, regularization mechanisms, and spectral inference methods within a unified geometric framework.

It was shown that when the linear system $A\ W = b$ is structurally compatible, the residual vector vanishes and both Gaussian (L_2) and Laplace (L_1) objectives attain the same global minimum. In this regime, equilibrium is determined by algebraic compatibility rather than by descent dynamics or penalty magnitudes. The distinction between L_2 and L_1 lies in geometric smoothness and descent behavior, not in the location of the compatible equilibrium itself.

Regularization was analyzed as a stabilizing mechanism rather than a generator of equilibrium. Quadratic penalties modify the spectrum of A^TA and produce controlled spectral shrinkage. Absolute penalties induce sparsity through coordinate-wise thresholding. These mechanisms govern conditioning, uniqueness, and bias–variance trade-offs, particularly in ill-posed or incompatible systems. When compatibility is exact and uniqueness holds, regularization acts as stabilization—not as structural necessity.

Principal Component Analysis was presented as a canonical example of equilibrium-based inference. The eigenvalue equation $A^T\ A\ v = \lambda\ v$ demonstrates that certain solutions are encoded directly in operator structure and do not depend on trajectory-defined optimization. Iterative algorithms compute such solutions; they do not define them.

These results establish a structural classification of learning regimes:

- Trajectory-based methods define solutions through iterative convergence.
- Regularization-based methods stabilize and reshape solution geometry.
- Equilibrium-based inference identifies solutions already encoded in structural compatibility.

The distinction is structural, not stylistic. Iteration is procedural. Regularization is corrective. Equilibrium is foundational.

With the present work, this hierarchy is made explicit. Compatibility, stabilization, and trajectory are separated as distinct structural layers of learning theory. Optimization is thereby repositioned: it is not the origin of solution, but the mechanism of approximation when equilibrium is not directly resolved.
The central conclusion is clear:
Equilibrium may precede optimization.

When compatibility exists, the solution is already encoded in operator structure. Regularization governs stability. Iteration governs computation. Confusing these roles obscures the geometry of determination.

By restoring this separation, the framework provides a principled structural foundation for understanding overfitting, conditioning, spectral stability, and the geometry of linear learning systems. The present analysis concerns linear compatibility regimes; nonlinear systems may require additional structural considerations beyond the equilibrium classification developed here.

Author Contributions

Huseyin Murat Cekirge is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

1. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323: 533-536.
2. Bottou L (2010) Large-scale machine learning with stochastic gradient descent. Proceedings of COMPSTAT 177-186.
3. Tibshirani R (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58: 267-288.
4. Zou H, Hastie T (2005) Regularization and variable selection via the Elastic Net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67: 301-320.
5. Fu W (1998) Penalized regression: The bridge versus the lasso. Journal of Computational and Graphical Statistics 7: 397-416.
6. Golub G, Van Loan C (1983) Matrix Computations. Baltimore: Johns Hopkins University Press https://epubs.siam.org/doi/book/10.1137/1.9781421407944.
7. Golub T, Slonim D, Tamayo P, Huard C, Gaasenbeek M (1999) Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286: 513-536.
8. Jolliffe IT (2002) Principal Component Analysis. Springer https://royalsocietypublishing.org/rsta/article/374/2065/20150202/115142/Principal-component-analysis-a-review-and-recent.
9. Tanaka Yamawaki M, Ikura Y (2023) Principal Component Analysis and Randomness Test for Big Data Analysis. Springer https://link.springer.com/book/10.1007/978-981-19-3967-9.
10. Cekirge HM (2025) Tuning the training of neural networks by using the perturbation technique. American Journal of Artificial Intelligence 9: 107-109.
11. Cekirge HM (2025) An alternative way of determining biases and weights for the training of neural networks. American Journal of Artificial Intelligence 9: 129-132.
12. Cekirge HM (2025) Algebraic σ-based (Cekirge) model for deterministic and energy-efficient unsupervised machine learning. American Journal of Artificial Intelligence 9: 198-205.
13. Cekirge HM (2025) Cekirge's σ-based ANN model for deterministic, energy-efficient, scalable AI with large-matrix capability. American Journal of Artificial Intelligence 9: 206-216.
14. Cekirge HM (2025) Cekirge_Perturbation_Report_v4. Zenodo https://www.researchgate.net/publication/396921891_Cekirge_Method_Perturbation_Report_v4.
15. Cekirge HM (2025) Algebraic Cekirge method for deterministic and energy-efficient transformer language models. American Journal of Artificial Intelligence 9: 258-271.
16. Cekirge HM (2025) Deterministic σ-regularized benchmarking of the Cekirge model against GPT-transformer baseline. American Journal of Artificial Intelligence 9: 272-280.
17. Cekirge HM (2025) The Cekirge method for machine learning: A deterministic σ-regularized analytical solution for general minimum problems. American Journal of Artificial Intelligence, 9: 324-337.
18. Cekirge HM (2026) The Cekirge σ-method in AI: Analysis and broad applications. American Journal of Artificial Intelligence 10: 14-33.
19. Tabak D, Kuo BC (1971) Optimal Control by Mathematical Programming. Englewood Cliffs, NJ: Prentice-Hall 237.