

Spectral Portfolio Theory: From SGD Weight Matrices to Wealth Dynamics

Anders G Frøseth*

8 March 2026

Abstract

We develop spectral portfolio theory by establishing a direct identification: neural network weight matrices trained on stochastic processes are portfolio allocation matrices, and their spectral structure encodes factor decompositions and wealth concentration patterns. The three forces governing stochastic gradient descent (SGD)—gradient signal, dimensional regularisation, and eigenvalue repulsion—translate directly into portfolio dynamics: smart money, survival constraint, and endogenous diversification. The spectral properties of SGD weight matrices transition from Marchenko–Pastur statistics (additive regime, short horizon) to inverse-Wishart via the free log-normal (multiplicative regime, long horizon), mirroring the transition from daily returns to long-run wealth compounding. We unify the cross-sectional wealth dynamics of [Bouchaud and Mézard \(2000\)](#), the within-portfolio dynamics of [Olsen et al. \(2025\)](#), and the scalar Fokker–Planck framework via a common spectral foundation. A central result is the Spectral Invariance Theorem: any isotropic perturbation to the portfolio objective preserves the singular-value distribution up to scale and shift, while anisotropic perturbations produce spectral distortion proportional to their cross-asset variance. We develop applications to portfolio design, wealth inequality measurement, tax policy, and neural network diagnostics. In the tax context, the invariance result recovers and generalises the neutrality conditions of [Frøseth \(2026b\)](#).

Contents

1	Introduction	1
2	Setup: Weight Matrices as Allocation Matrices	2
3	Portfolio Dynamics: The Three Forces of SGD	4
3.1	Gradient Signal: Smart Money	4
3.2	Dimensional Regularization: Survival Constraint	5
3.3	Eigenvalue Repulsion: Endogenous Diversification	5
4	The Stationary Portfolio: Bulk and Tail	6
5	Network Learning: Factor Complexity and Data Structure	7

*Independent Researcher. E-mail: indrefjorden@pm.me.

6	Timescale Regimes and Spectral Transitions	7
6.1	The Additive Regime: Short Horizon	7
6.2	The Multiplicative Regime: Long Horizon	8
6.3	The q -Transformation: Quantifying the Crossover	8
6.4	The Transition: Marchenko–Pastur to Inverse-Wishart	9
7	The Ergodicity Gap as Spectral Observable	10
8	Aggregation: From Matrix to Scalar via Itô’s Lemma	11
9	The Fokker–Planck Connection	13
10	The Bouchaud–Mézard Unification	14
11	Isotropic Perturbations and Spectral Invariance	15
11.1	The Isotropic Condition	15
11.2	Spectral Consequences	15
12	Anisotropic Perturbations and Spectral Distortion	17
13	The Spectral Invariance Conjecture	18
14	The Loss Function and Utility Correspondence	19
14.1	Maximum Likelihood Estimation	19
14.2	Score Matching and Denoising Score Matching	19
15	Applications	20
15.1	Portfolio Design and Factor Structure	20
15.2	Wealth Inequality and Power Laws	20
15.3	Tax Policy: Neutrality as Spectral Invariance	21
15.4	Neural Network Diagnostics	21
16	Testable Predictions	22
16.1	General Spectral Predictions	22
16.2	Tax-Specific Predictions	22
17	Open Questions and Future Work	22
18	Relation to Existing Literature	23
18.1	Random Matrix Theory in Finance	23
18.2	Deep Learning and Portfolio Optimization	23
18.3	Ergodicity Economics	23
18.4	SGD Spectral Theory	23
18.5	Classical Portfolio Theory	24
18.6	Wealth Dynamics and Power Laws	24
19	Conclusion	24

1 Introduction

Consider a feedforward neural network trained to learn a stochastic process—a drift function, a transition density, or a score function—from observed trajectories. At each layer, the weight matrix $W \in \mathbb{R}^{m \times n}$ evolves under stochastic gradient descent.

This paper takes the following identification as its foundation: the weight matrix at a given layer is a portfolio allocation matrix. Row i specifies how capital is distributed across n assets in state i , and columns index assets. Training the network on data from a stochastic process is equivalent to running an adaptive portfolio optimizer on that process.

The spectral theory of SGD dynamics developed by [Olsen et al. \(2025\)](#) reveals that this identification is not merely an analogy: the stationary portfolio structure, the factor decomposition, and the concentration–diversification tradeoff emerge as direct mathematical consequences. This paper develops the framework systematically, proving a spectral invariance theorem for isotropic perturbations and applying the results to portfolio design, wealth inequality, tax policy, and neural network diagnostics.

The main contributions are as follows. First, we establish that the three forces in the singular-value evolution equation have direct portfolio interpretations: gradient signal encodes smart money (return-seeking), dimensional regularisation captures an endogenous survival constraint, and eigenvalue repulsion implements endogenous diversification without explicit constraints. Second, we characterize the stationary spectral distribution as having a gamma-type bulk with power-law tail, giving rise to the core–satellite portfolio structure observed empirically in institutional wealth and household portfolio data. Third, we show that the spectral transition from additive (Marchenko–Pastur) to multiplicative (inverse-Wishart) regimes is governed by the free log-normal distribution and the matrix Kesten problem, establishing a precise mathematical duality between short-horizon and long-horizon portfolio dynamics. Fourth, we unify the cross-sectional wealth model of Bouchaud and Mézard, the within-portfolio model of Olsen et al., and the scalar Fokker–Planck framework via spectral decomposition. Fifth, we prove the Spectral Invariance Theorem: any isotropic perturbation to the portfolio objective preserves the spectral shape of the allocation matrix, while anisotropic perturbations produce distortion proportional to their cross-asset variance. Sixth, we develop applications to portfolio design, wealth inequality measurement, tax policy, and neural network diagnostics, demonstrating the breadth of the framework.

The paper is organized in five parts. Part I (Sections 2–5) establishes the foundations: the learning setup (Figure 1), the weight matrix–portfolio identification, the SVD decomposition into eigenportfolios, the three forces of SGD, the stationary spectral distribution, and factor complexity. Part II (Sections 6–9) develops the dynamics: timescale regimes governing the additive-to-multiplicative transition, the ergodicity gap as a spectral observable, the aggregation from matrix to scalar via Itô projection, and the Fokker–Planck connection. Part III (Sections 10–14) presents general results: the Bouchaud–Mézarid unification, the Spectral Invariance Theorem and its anisotropic counterpart, the Spectral Invariance Conjecture, and the loss–utility correspondence. Part IV (Section 15) develops applications to portfolio design, wealth

inequality, tax policy, and neural network diagnostics. Part V (Sections 16–19) collects testable predictions, open questions, literature context, and conclusions.

2 Setup: Weight Matrices as Allocation Matrices

The starting point is a stochastic process observed through discrete trajectories. The process follows a stochastic differential equation

$$dx = v(x) dt + \sqrt{2D} dB, \quad (1)$$

where $v(x)$ is the drift function and D the diffusion constant. A neural network is trained to learn either the drift $v(x)$, the transition density $p(x_{t+\Delta t} | x_t)$, or the score function $\nabla_x \log p(x)$ from observed trajectory data $\{x_0, x_{\Delta t}, x_{2\Delta t}, \dots\}$. The weight matrix W at each layer is updated by stochastic gradient descent on a loss function $\mathcal{L}(W)$ that is quadratic in the prediction residual (see Section 14 for details).

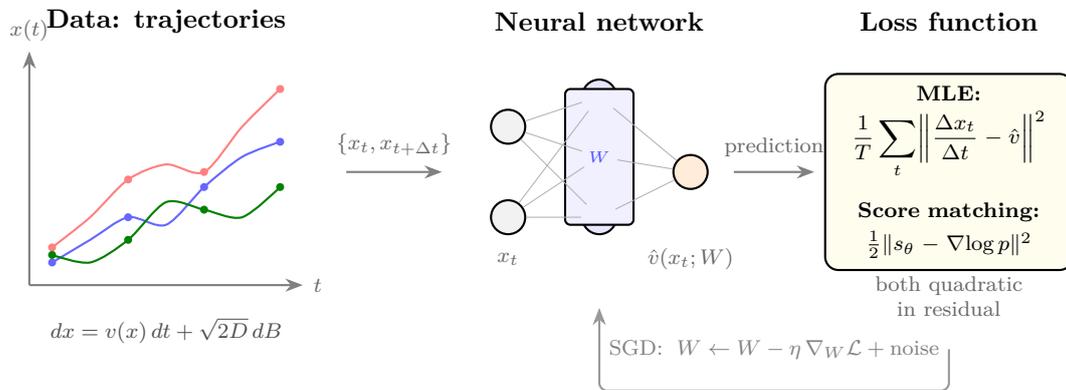


Figure 1: The learning setup. A stochastic process (1) generates trajectory data (left). A neural network with weight matrix W at each layer learns the drift function $\hat{v}(x; W)$ or the score function $s_\theta(x)$ (centre). The loss function is quadratic in the prediction residual (right), whether using maximum likelihood or score matching. SGD updates W iteratively, and the stationary spectral distribution of W encodes the structure of the learned process.

In Olsen et al. (2025), the weight matrix $W \in \mathbb{R}^{m \times n}$ evolves under the stochastic differential equation

$$dW = -\eta \frac{\partial \mathcal{L}}{\partial W} dt + \sqrt{2\eta D} d\mathcal{W}, \quad (2)$$

where η is the learning rate, D an effective diffusion constant, and $d\mathcal{W}$ is matrix-valued Wiener noise.

We relabel the dimensions: set $m = T$ (time periods or states of the world) and $n = N$ (number of assets). Then W_{ti} represents the portfolio weight on asset i in state t . The allocation matrix $W \in \mathbb{R}^{T \times N}$ evolves as investors rebalance in response to return signals, information noise, and constraints implicit in the learning dynamics.

The singular value decomposition $W = U\Sigma V^\top$ decomposes this allocation into principal components. The left singular vectors $u_k \in \mathbb{R}^T$ show when each factor is relevant (temporal patterns).

The right singular vectors $v_k \in \mathbb{R}^N$ show which assets compose each factor (eigenportfolios). The singular values σ_k measure the magnitude of allocation to each eigenportfolio.

This decomposition reveals the factor structure naturally: large singular values correspond to important factors (market-wide moves, systematic risks); small singular values correspond to idiosyncratic or second-order effects. The right singular vectors v_k define eigenportfolio compositions, and the left singular vectors u_k indicate when each factor dominates. Figure 2 illustrates this relabelling.

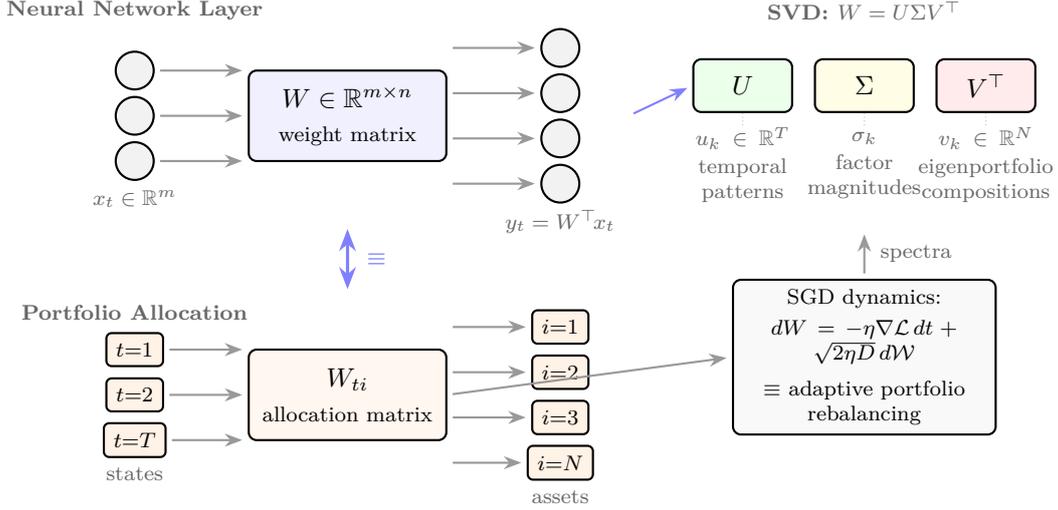


Figure 2: The neural network–portfolio identification. A single layer with weight matrix $W \in \mathbb{R}^{m \times n}$ maps input x_t to output $y_t = W^\top x_t$ (top left). Relabelling rows as states and columns as assets gives the allocation matrix W_{ti} (bottom left). The SVD decomposes W into temporal patterns u_k , factor magnitudes σ_k , and eigenportfolio compositions v_k (top right). SGD dynamics on W are equivalent to adaptive portfolio rebalancing (bottom right); the stationary spectral density of σ_k is determined by the signal-to-noise ratio $\beta_1/\eta D$.

3 Portfolio Dynamics: The Three Forces of SGD

By [Olsen et al. \(2025\)](#) (Theorem 3.1), the singular values evolve as

$$d\sigma_k = \left[-\eta u_k^\top (\nabla_W \mathcal{L}) v_k + \eta D \left(\frac{m-n+1}{2\sigma_k} - \sum_{j \neq k} \frac{\sigma_k}{\sigma_k^2 - \sigma_j^2} \right) \right] dt + \sqrt{2\eta D} d\beta_k. \quad (3)$$

This equation contains three distinct forces, each with a clear portfolio interpretation.

3.1 Gradient Signal: Smart Money

The term

$$-\eta u_k^\top (\nabla_W \mathcal{L}) v_k \quad (4)$$

is the projected return signal pushing capital toward high-performing eigenportfolios. This is the optimizer’s best estimate of where marginal utility of wealth is highest. In a mean–variance framework with loss $\mathcal{L} = -Wr + \frac{\gamma}{2} W^2$, this term becomes the excess return on factor k adjusted

for risk aversion. The negative sign reflects gradient descent: the optimizer increases weights on factors with positive marginal returns.

This force implements **smart money dynamics**: capital flows toward factors with superior risk-adjusted returns.

3.2 Dimensional Regularization: Survival Constraint

The term

$$\eta D \frac{m - n + 1}{2\sigma_k} \quad (5)$$

is a $1/\sigma_k$ restoring force that prevents any eigenportfolio loading from reaching zero. Small positions get pushed up. Economically, this is a minimum-position constraint arising *endogenously* from the noise structure. An investor observing returns with noise cannot rationally set any factor exposure exactly to zero, because zero exposure means zero information—creating an explore–exploit tension.

The prefactor $m - n + 1$ is the aspect ratio of the allocation matrix, encoding how many degrees of freedom remain after fitting the primary factors. More time periods (larger m) strengthen the survival constraint; more assets (larger n) weaken it.

This force implements **endogenous survival**: positions that would shrink to zero due to poor current performance are sustained by information and exploration incentives.

3.3 Eigenvalue Repulsion: Endogenous Diversification

The term

$$\eta D \sum_{j \neq k} \frac{\sigma_k}{\sigma_k^2 - \sigma_j^2} \quad (6)$$

prevents any two eigenportfolio loadings from becoming equal. In portfolio language, this is **endogenous diversification**. Even without an explicit diversification constraint, stochastic rebalancing naturally separates factor exposures. When two factors carry similar weight ($\sigma_k \approx \sigma_j$), the denominator diverges, creating an unstable repulsion that forces differentiation.

This is a new micro-foundation for diversification: it arises not from risk aversion (which is in the gradient term), but from the *noise structure of learning under partial information*. It holds regardless of the loss function or risk aversion parameter.

Remark (Connection to Merton). In the continuous-time portfolio problem, the optimal allocation involves the inverse covariance matrix Σ^{-1} (Merton, 1969, 1971). The eigenvalue repulsion ensures that the estimated covariance matrix remains well-conditioned—it is a regularisation of the Merton solution that emerges from the learning dynamics.

This paper	Nearest existing concept	Key distinction
Smart money (gradient signal)	Informed investor flows (Gruber, 1996)	Optimisation on available data, not asymmetric information; compatible with market efficiency
Survival constraint (dim. regularisation)	Safety-first / Kelly criterion (Roy, 1952)	Survival of information channels (factor exposures), not of the investor
Endogenous diversification (eigenvalue repulsion)	Mean-variance diversification (Markowitz, 1952)	Emerges from noise structure of learning; independent of risk aversion

Table 1: Terminology disambiguation. The portfolio labels proposed in this paper echo but differ from established usage. Each force in the singular-value SDE (3) has a nearest classical counterpart; the distinctions are summarised here.

Remark (Adaptive dynamics). Taken together, the three forces describe an explore–exploit–diversify triad: the gradient signal exploits current information, the survival constraint maintains exploratory positions, and eigenvalue repulsion enforces differentiation. This is reminiscent of the Adaptive Markets Hypothesis (Lo, 2004, 2017), in which market behaviour is driven by evolutionary dynamics among strategies. The spectral framework can be viewed as providing a mathematical instantiation of Lo’s qualitative programme.

4 The Stationary Portfolio: Bulk and Tail

By Olsen et al. (2025) (Theorem 3.2), the stationary distribution of singular values is

$$p_\sigma(\sigma) \propto \sigma^{m-n+1} e^{-(\beta_1/4\eta D)\sigma^2}, \quad (7)$$

where β_1 is a mean-field restoring-force constant.

This distribution exhibits a clear structure: a concentrated bulk followed by a power-law tail.

Definition 1 (Core–Satellite Portfolio). At stationarity, we partition the eigenportfolios into two groups:

- **Bulk:** most eigenportfolio loadings cluster in a concentrated region. This is the diversified *core*—broad market exposure spread across many factors.
- **Tail:** a few large σ_k values form a power-law tail (σ^{m-n+1} for large σ). These are concentrated *satellite* positions—large bets on specific factors.

This core–satellite structure is observed empirically in institutional portfolios and, crucially, in the Norwegian wealth register data of Frøseth (2026b,a,c).

The tail exponent $m - n + 1$ depends on the aspect ratio of the allocation matrix. For a market with N assets observed over T periods, the exponent is $T - N + 1$. More observation (larger T) fattens the tail (allows more concentration); more assets (larger N) thins it (forces diversification). This is testable against Norwegian portfolio data.

Definition 2 (Effective Spectral Rank). Define the *effective spectral rank* as

$$r_{\text{eff}} = \frac{\sum_k \sigma_k}{\max_k \sigma_k}, \quad (8)$$

the ratio of total allocation to the largest single factor exposure. This measures the complexity of the portfolio: $r_{\text{eff}} = 1$ means one factor dominates; $r_{\text{eff}} = N$ means all factors have equal weight. The stationary distribution predicts a specific value of r_{eff} as a function of the aspect ratio and the signal-to-noise ratio.

5 Network Learning: Factor Complexity and Data Structure

When a weight matrix W is trained on data from a stochastic process, it encodes a factor decomposition of that process. The singular values tell us the relative importance of each factor.

Consider three cases:

Case 1: Simple underlying structure. If the data is generated by a drift function with simple structure (mean-reverting, polynomial, Fourier basis), the trained weight matrix will have a small number of large singular values followed by a sharp drop. The bulk is thin, the tail is fat—a concentrated factor model.

Case 2: Complex or noisy data. If the data is noisy or has complex structure, the trained matrix will have a more gradual decay of singular values. The bulk is thick, the tail is thinner—a more diversified representation.

Case 3: Pareto-distributed data. The spectral complexity measure r_{eff} directly reflects the intrinsic dimensionality of the process. For Pareto-distributed data (as in wealth), the exponent of r_{eff} should match the Pareto exponent of the data.

This means the weight spectrum is not an artifact of the learning process; it is a faithful encoding of the underlying data structure. In deep networks with L layers, each layer performs a hierarchical factor decomposition: early layers capture primary factors, intermediate layers combine them into secondary factors, and the final layer produces the output. This is the fund-of-funds structure: not a single portfolio, but a hierarchy of portfolios.

6 Timescale Regimes and Spectral Transitions

The spectral portfolio framework must reconcile two well-established facts about asset dynamics that operate on different timescales.

6.1 The Additive Regime: Short Horizon

At timescales from minutes to days, asset returns are approximately additive: $r_t \approx \mu \delta t + \sigma \delta B_t$, with heavy tails but finite variance over short windows. This is the regime of the classical stylized facts (Cont, 2001; Bouchaud and Potters, 2003): fat-tailed return distributions, volatility

clustering, and approximately uncorrelated returns. The covariance matrix $\Sigma = \text{Cov}(r_t)$ has a Marchenko–Pastur bulk with a few spikes (market factor, sector factors), and random matrix theory cleaning (Laloux et al., 1999; Ledoit and Wolf, 2004) is effective precisely because the bulk is well described by random matrix universality.

This is also the regime where Olsen et al. (2025)’s spectral theory of SGD applies most directly: weight updates are small additive perturbations, the matrix-valued SDE is well approximated by Dyson Brownian motion, and the stationary spectral density follows a gamma-type distribution with power-law tails. Aarts et al. (2025) provide independent confirmation from the physics side, showing that the eigenvalue dynamics of SGD weight matrices are described by Dyson Brownian motion with a Coulomb gas stationary distribution whose variance scales as $\alpha/|\mathcal{B}|$ (learning rate over batch size).

6.2 The Multiplicative Regime: Long Horizon

At timescales from months to years, compounding dominates. Wealth dynamics are inherently multiplicative: $dW/W = \mu dt + \sigma dB_t$, and the natural coordinate is log-wealth $x = \ln W$, which converts multiplicative dynamics into an additive SDE with drift $v = \mu - \frac{1}{2}\sigma^2$ and diffusion $D = \sigma^2$. This is the regime of Frøseth (2026c), where the Fokker–Planck equation governs the wealth distribution and the Pareto tail emerges from drift–diffusion balance.

Tax effects and portfolio allocation decisions operate on this longer timescale. The Heston model (Heston, 1993) and other stochastic-volatility extensions capture both regimes: they remain multiplicative while capturing the volatility clustering persisting from the short-horizon regime.

6.3 The q -Transformation: Quantifying the Crossover

Bouchaud and Potters (2003) introduce a continuous parametrisation of the additive-to-multiplicative crossover. Define the price process

$$x(T) = x_0(1 + q(T)\xi(T))^{1/q(T)}, \quad q(T) \in [0, 1], \quad (9)$$

where $\xi(T)$ is the fundamental random variable (normalised return). At $q = 1$ the process is fully additive: $x = x_0(1 + \xi)$. As $q \rightarrow 0$, one recovers the multiplicative limit: $(1 + q\xi)^{1/q} \rightarrow e^\xi$, so $x = x_0 e^\xi$.

The index $q(T)$ is a quantitative measure of the additivity of the process at horizon T . Empirically, q decreases monotonically from near 1 at intraday timescales to near 0 at horizons of months to years. The crossover timescale T_c at which $q \approx 1/2$ is on the order of *months* for liquid equity markets.

The q -transformation also governs the skewness:

$$s(T) = -3q(T)\sigma(T), \quad (10)$$

so that skewness vanishes in the multiplicative limit ($q \rightarrow 0$) and is largest in the additive regime. This provides a measurable diagnostic: the observed skewness at a given horizon identifies where

the process sits in the crossover.

For the spectral portfolio framework, $q(T)$ determines which universality class governs the covariance spectrum at a given horizon. When $q \approx 1$ (additive, short horizon), the spectrum is Marchenko–Pastur with isolated signal spikes. When $q \approx 0$ (multiplicative, long horizon), the spectrum transitions toward the free log-normal/inverse-Wishart regime described next.

6.4 The Transition: Marchenko–Pastur to Inverse-Wishart

The passage from additive to multiplicative regime reflects not merely a change of resolution but a change in the dominant mechanism and, crucially, a change in the spectral universality class.

[Potters and Bouchaud \(2021\)](#) develop this transition explicitly in free probability language. The additive central limit theorem for free random matrices gives the Wigner semicircle (Marchenko–Pastur for rectangular matrices). The *multiplicative* central limit theorem gives the *free log-normal*, with S-transform

$$S_{\text{LN}}(t) = e^{-a/2-bt}, \quad (11)$$

where a controls the mean log-eigenvalue and b the variance. The free log-normal is stable under free products, as the semicircle is stable under free sums. For small a , the density is indistinguishable from a Wigner semicircle; as a grows, it develops the characteristic asymmetric shape of multiplicative dynamics.

Multiplicative Dyson Brownian Motion. The additive DBM governing eigenvalue dynamics in the short-horizon regime has a multiplicative counterpart ([Potters and Bouchaud, 2021](#)):

$$\frac{d\lambda_i}{dt} = \frac{a}{2} \lambda_i + \frac{b}{N} \sum_{j \neq i} \frac{\lambda_i \lambda_j}{\lambda_i - \lambda_j} + \sqrt{\frac{b}{N}} \lambda_i \xi_i, \quad (12)$$

where ξ_i are independent Langevin noise terms. Every term is proportional to λ_i : the drift, the repulsion, and the noise are all multiplicative. At arbitrary time t , the spectral density is the free log-normal with parameters ta and tb .

Matrix Kesten Problem. The scalar Kesten recursion $Z_{n+1} = Z_n(1 + \zeta_n)$, which generates Pareto tails in [Bouchaud and Mézard \(2000\)](#); [Frøseth \(2026c\)](#), has a matrix generalization. The corresponding Fokker–Planck equation is

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial U} [(1 + mU)P] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial U^2} [U^2 P], \quad (13)$$

with stationary distribution $P_{\text{eq}}(U) \propto U^{-1-\mu} e^{-2/(\sigma^2 U)}$, an inverse-gamma with power-law tail exponent

$$\mu = 1 + \frac{2\hat{m}}{\sigma^2}. \quad (14)$$

The matrix Kesten variable is an inverse-Wishart matrix, whose eigenvalue spectrum maps to Marchenko–Pastur under $\lambda \rightarrow 1/\lambda$. This establishes a precise duality: the additive regime is Marchenko–Pastur; the multiplicative regime is inverse-Wishart; the two are related by spectral

inversion.

Remark (Interpolating Family). The free log-normal provides a one-parameter family interpolating continuously between additive and multiplicative regimes. At small a (short horizon), the density resembles the Marchenko–Pastur semicircle. At large a (long horizon), the density develops a power-law tail controlled by $\mu = 1 + 2\hat{m}/\sigma^2$. The parameter a plays the role of effective horizon, growing with the number of compounding periods.

For the spectral portfolio framework:

1. The *within-layer* spectral structure of SGD weight matrices reflects the additive regime. The network is trained on short-horizon data increments; the relevant universality class is Marchenko–Pastur; the relevant dynamics are additive DBM.
2. The *cross-sectional* spectral structure of the population allocation matrix reflects the multiplicative regime. Agents compound returns over long horizons; the relevant universality class is the free log-normal/inverse-Wishart; the dynamics are multiplicative DBM.
3. The aggregation from matrix to scalar (Section 8) is the radial projection from matrix Fokker–Planck to scalar Fokker–Planck, mapping the spectral exponent μ of inverse-Wishart to the Pareto exponent $\alpha = 1 + v/D$ of the wealth distribution.
4. The scalar Kesten process of Frøseth (2026c) and the matrix Kesten problem are related by this same projection. The Pareto tail of the wealth distribution is the radial shadow of the inverse-Wishart tail of the allocation matrix.

Regime shifts manifest spectrally as changes in the effective parameter a of the free log-normal, from small a (Marchenko–Pastur bulk with isolated spikes) to large a (inverse-Wishart density with power-law tails). Market crashes or policy reforms can trigger cross-sectional regime shifts: assets in the additive regime may suddenly transition to multiplicative dynamics (fire sales, forced liquidation).

7 The Ergodicity Gap as Spectral Observable

A central insight from Peters (2019) and Frøseth (2026c) is the distinction between *ensemble average* and *time average*. The wealth distribution $\pi(x)$ represents the ensemble (cross-section at a fixed time). An individual trajectory $x(t)$ represents time evolution of a single agent.

In ergodic systems, the two averages coincide: an agent’s long-run average outcome equals the population average. In non-ergodic systems, they differ: the population distribution can have fat tails while individual trajectories concentrate.

Definition 3 (Ergodicity Gap). For a wealth process $x(t)$ with drift v and diffusion D , define the *ergodicity gap* as the difference between the ensemble and time-average growth rates:

$$\Delta g = g_{\text{ens}} - g_{\text{time}} = \left(v + \frac{1}{2}D\right) - \left(v - \frac{1}{2}D\right) = D. \quad (15)$$

Here $g_{\text{ens}} = \frac{d}{dt} \log \mathbb{E}[x]$ is the growth rate of the ensemble mean and $g_{\text{time}} = \mathbb{E}[\frac{d}{dt} \log x]$ is the expected time-average (geometric) growth rate.

In the scalar model of Frøseth (2026c), the ergodicity gap is simply the diffusion coefficient D . In the spectral framework, the Itô projection of Section 8 yields an effective scalar diffusion

$$D_{\text{eff}} = 2\eta D, \quad (16)$$

where η is the learning rate and D the data covariance scale. The radial projection collapses all mn noise directions onto one: each matrix entry contributes independently to $\|W\|_F^2$, but the quadratic variation of $x = \|W\|_F$ is $(1/x^2) \sum W_{ai}^2 \cdot 2\eta D = 2\eta D$. The ergodicity gap thus has a spectral interpretation:

$$\Delta g = D_{\text{eff}} = 2\eta D. \quad (17)$$

The gap depends on the noise level but not on matrix dimensions. However, the *effect* of the gap on the wealth distribution depends on the spectral structure: when the spectral tail is fat (small tail exponent), fewer factors carry most of the variance, individual trajectories are more volatile relative to their mean, and the distributional consequences of non-ergodicity are more severe. When the spectrum is flat (many comparable factors), the volatility is spread across dimensions and individual trajectories track the ensemble more closely.

An isotropic perturbation (Theorem 1) preserves the spectral shape, hence preserves D_{eff} and leaves the ergodicity gap unchanged. An anisotropic perturbation that compresses the spectral tail (e.g. differential taxation favouring diversified portfolios) reduces the effective number of dominant factors, lowers D_{eff} , and narrows the gap.

8 Aggregation: From Matrix to Scalar via Itô's Lemma

Define total portfolio value as $x = \|W\|_F = (\sum_{a,i} W_{ai}^2)^{1/2}$. By Itô's lemma applied to $f(W) = \|W\|_F$:

$$dx = \frac{1}{x} \sum_{a,i} W_{ai} dW_{ai} + \frac{1}{2x} \left(mn \cdot 2\eta D - \frac{1}{x^2} \sum_{a,i} W_{ai}^2 \cdot 2\eta D \right) dt. \quad (18)$$

The first term gives the drift contribution from the gradient signal projected onto the radial direction. The second term simplifies using $\sum W_{ai}^2 = x^2$:

$$dx = \frac{1}{x} \text{tr}(W^\top dW) + \frac{\eta D(mn - 1)}{x} dt. \quad (19)$$

In SVD coordinates, $\text{tr}(W^\top dW) = \sum_k \sigma_k d\sigma_k$ plus rotational terms. If the rotational terms average out (which they do under isotropic noise), the radial process becomes

$$\frac{dx}{x} = \mu_{\text{rad}} dt + \sigma_{\text{rad}} dB, \quad (20)$$

a geometric Brownian motion. This is exactly the scalar SDE of Frøseth (2026c). **The wealth Fokker–Planck is the radial projection of the matrix Fokker–Planck.**

The radial drift and diffusion are computed by substituting the multiplicative Dyson BM (12) into (19). Let $p = \min(m, n)$ be the number of singular values and write $x^2 = \sum_{k=1}^p \sigma_k^2$. The signal term contributes a radial drift from the restoring force in (7), and the Itô correction from (18) contributes $\eta D(mn - 1)/x$. The noise has quadratic variation $(1/x^2) \sum_{a,i} \|W_{ai}\|^2 \cdot 2\eta D = 2\eta D$, giving an effective scalar diffusion $D_{\text{eff}} = 2\eta D$ independent of matrix dimensions (the radial projection collapses all mn noise directions onto one). The Pareto exponent is

$$\alpha = 1 + \frac{v_{\text{radial}}}{D_{\text{eff}}}, \quad (21)$$

where v_{radial} depends on m , n , and $\beta_1/\eta D$.

The function f can be evaluated explicitly in two regimes.

Single-factor limit ($p = 1$, i.e. $n = 1$). The allocation matrix reduces to a vector, eigenvalue repulsion vanishes, and the spectral SDE collapses to a scalar Ornstein–Uhlenbeck process for σ^2 . The radial SDE (20) becomes

$$dx = \left(-\frac{\beta_1}{2}x + \frac{\eta D(m-1)}{x} \right) dt + \sqrt{2\eta D} dB.$$

At stationarity, the process for $y = x^2$ has drift $v_y = -\beta_1 y + 2\eta D \cdot m$ and diffusion $4\eta D \cdot y$. The stationary density is $P(y) \propto y^{-1-\mu} e^{-\text{const}/y}$ with

$$\mu|_{p=1} = 1 + \frac{(m-1)\beta_1}{2\eta D}, \quad (22)$$

which is exactly the scalar Kesten exponent (14) with $\hat{m} = (m-1)\beta_1/2$ and $\sigma^2 = 2\eta D$. The corresponding wealth Pareto exponent is $\alpha = \mu/2$ (since wealth scales as x , not x^2).

Large- N regime ($m, n \gg 1$, $m/n = \gamma$ fixed). The eigenvalue repulsion is of order p^2 but the repulsion terms cancel in the radial projection (they redistribute variance among singular values without changing $\sum \sigma_k^2$). The radial drift at stationarity is $v_{\text{radial}} = -\beta_1 x/2 + \eta D(mn - 1)/x$. Setting $v_{\text{radial}} = 0$ at the stationary mean $\bar{x}^2 = 2\eta D(mn - 1)/\beta_1$ and linearising, the Pareto exponent becomes

$$\alpha \approx 1 + \frac{(mn-1)\beta_1}{2mn\eta D} \approx 1 + \frac{\beta_1}{2\eta D}, \quad m, n \gg 1. \quad (23)$$

In the large- N limit the Pareto exponent depends on the signal-to-noise ratio $\beta_1/(2\eta D)$ alone, independent of matrix dimensions. This is the spectral counterpart of the Kesten result: strong signal (large β_1) produces a thin tail (large α), while strong noise (large ηD) produces a fat tail (small α).

For general (m, n) , the function f interpolates between the single-factor formula (22) and the large- N limit (23), and can be computed by numerical integration of the radial projection against the spectral density (7).

9 The Fokker–Planck Connection

Both the wealth distribution and the spectral density satisfy Fokker–Planck equations, revealing a deep structural parallel.

Wealth distribution (Frøseth, 2026c):

$$\frac{\partial \pi}{\partial t} = -\frac{\partial}{\partial x} [v(x) \pi] + D \frac{\partial^2 \pi}{\partial x^2}. \quad (24)$$

The stationary solution has Pareto tail with exponent $\alpha = 1 + v/D$.

Spectral density:

$$\frac{\partial \rho}{\partial \lambda} = \frac{\partial}{\partial \lambda} \left[(c(\lambda^* - \lambda) - \text{PV} \int \frac{\rho(\lambda')}{\lambda - \lambda'} d\lambda') \rho \right] + D_{\text{eff}} \frac{\partial^2 \rho}{\partial \lambda^2}. \quad (25)$$

The stationary solution has gamma-type density with power-law tail exponent $m - n + 1$.

The spectral equation has an extra nonlocal interaction term (the principal-value integral—the mean-field version of eigenvalue repulsion). Structurally, both are drift–diffusion PDEs with power-law stationary solutions.

If we consider a single-factor portfolio ($N = 1$, so the allocation matrix reduces to a vector), the repulsion term vanishes and the spectral FP reduces to the standard wealth FP. In this limit, the spectral exponent should reduce to the Pareto exponent. The scalar wealth model is the $N = 1$ projection of the matrix model. Figure 3 summarises the flow from the three forces through the spectral density to the wealth distribution.

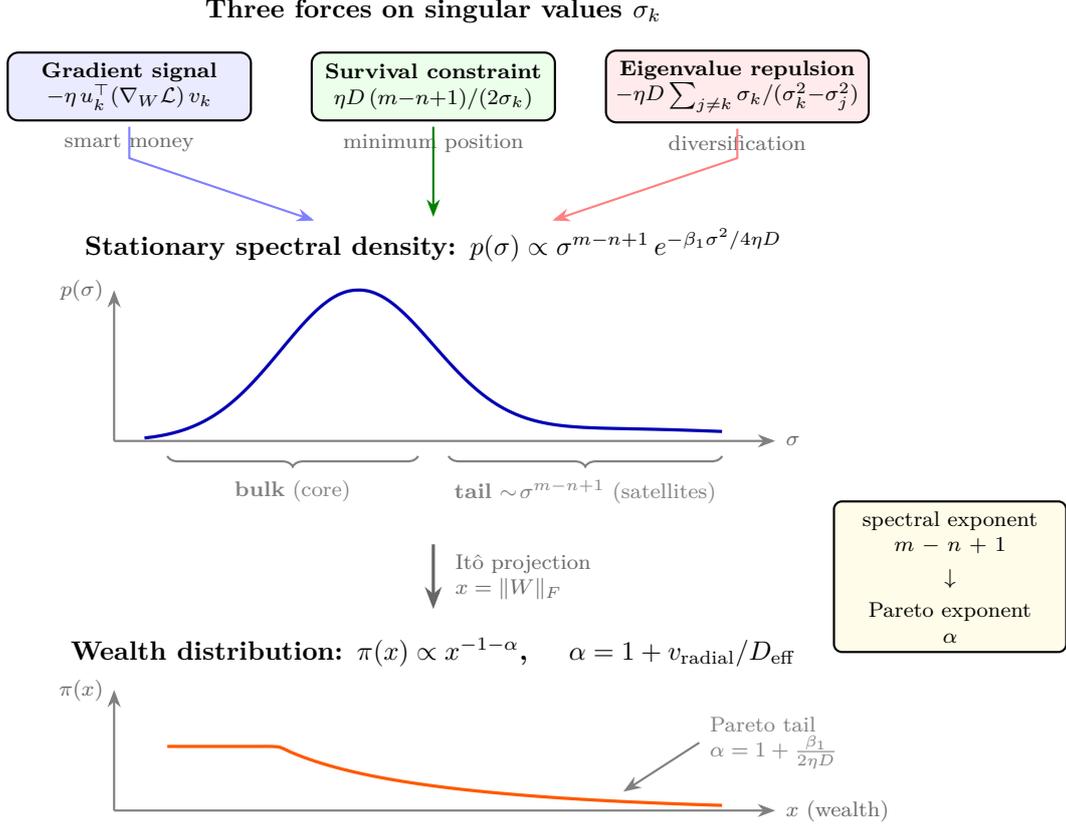


Figure 3: From forces to wealth distributions. The three forces in the singular-value SDE (3) — gradient signal, survival constraint, and eigenvalue repulsion — determine the stationary spectral density (middle). The gamma-type density has a concentrated bulk (core portfolio) and a power-law tail (satellite positions). The radial Itô projection (Section 8) maps the matrix-valued spectral density to a scalar wealth process $x = \|W\|_F$, whose stationary distribution exhibits a Pareto tail with exponent α determined by the signal-to-noise ratio $\beta_1/\eta D$.

10 The Bouchaud–Mézard Unification

Bouchaud and Mézard (2000) model wealth dynamics across a population of agents:

$$dw_i = \eta_i w_i dt + \sum_j J_{ij} (w_j - w_i) dt, \quad (26)$$

where η_i is multiplicative noise (investment returns) and J_{ij} describes wealth exchange (trade, lending, redistribution) between agents.

Stack N agents' wealths into a vector $w = (w_1, \dots, w_N)$. The exchange term is $Jw - \text{diag}(J\mathbf{1})w$, a matrix–vector product minus diagonal correction. The full dynamics:

$$dw = [\text{diag}(\eta) + J - \text{diag}(J\mathbf{1})] w dt. \quad (27)$$

The matrix $M = \text{diag}(\eta) + J - \text{diag}(J\mathbf{1})$ has eigenvalues whose distribution determines the steady-state wealth distribution. Bouchaud and Mézard (2000) show that the Pareto exponent is

determined by the balance between the noise spectrum (eigenvalues of $\text{diag}(\eta)$) and the exchange coupling J .

The exchange matrix J introduces **inter-agent coupling**—analogous to the eigenvalue repulsion in [Olsen et al. \(2025\)](#).

- Without coupling ($J = 0$ or no repulsion): wealth and singular values evolve independently, leading to condensation.
- With coupling: repulsive interaction prevents condensation and produces power-law tails.

This establishes a **unified framework** in which three descriptions of wealth-related dynamics share common spectral structure:

1. Bouchaud–Mézard describes the *cross-sectional* dynamics (wealth distribution across agents).
2. [Olsen et al. \(2025\)](#) describes the *within-portfolio* dynamics (allocations across assets).
3. The scalar Fokker–Planck framework of [Frøseth \(2026c\)](#) provides the *radial projection* linking spectral and Pareto exponents.

All three are instances of interacting particle systems with multiplicative noise, where the interaction term prevents condensation and produces power-law distributions. Any uniform perturbation (e.g. a tax, a fee, a regulatory cost) modifies the multiplicative noise term in all three descriptions simultaneously.

11 Isotropic Perturbations and Spectral Invariance

We now establish the central invariance result. Consider any perturbation to the portfolio objective that affects all assets uniformly—what we call an *isotropic perturbation*. Examples include a proportional wealth tax with uniform assessment, a flat management fee, a uniform transaction cost, or any regulatory levy that does not distinguish between asset classes.

11.1 The Isotropic Condition

An isotropic perturbation modifies the loss function as

$$\mathcal{L}_{\text{post}}(W) = k \cdot \mathcal{L}_{\text{pre}}(W) + \text{const}, \quad (28)$$

where $k > 0$ is a scalar that depends on the perturbation parameters but not on the direction in weight space. This is a uniform rescaling of the loss across all directions.

11.2 Spectral Consequences

This uniform rescaling produces three immediate consequences:

1. **Gradient Scaling.** The gradient rescales uniformly: $\nabla_W \mathcal{L}_{\text{post}} = k \nabla_W \mathcal{L}_{\text{pre}}$. All signal terms in the singular-value SDE scale by the same factor k .

2. **Repulsion Structure Preserved.** The eigenvalue repulsion term depends only on D and the geometry of the singular value space, not on the loss function. It is unaffected by the perturbation.

3. **Spectral Density Transformation.** The stationary singular-value density becomes

$$p_\sigma^{\text{post}}(\sigma) = p_\sigma^{\text{pre}}(\sigma/\sqrt{k})/\sqrt{k}, \quad (29)$$

a scale-shifted version of the pre-perturbation distribution. The shape (the exponent $m - n + 1$ in the power-law tail) is preserved.

Theorem 1 (Spectral Invariance). *An isotropic perturbation to the portfolio objective preserves the singular-value distribution of the allocation matrix up to a scale-and-shift transformation. The tail exponent, the eigenportfolio directions, and the effective spectral rank are all invariant.*

Proof sketch. Let $W \in \mathbb{R}^{T \times N}$ be the allocation matrix with singular-value decomposition $W = U\Sigma V^\top$, where $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_{\min(T,N)})$. The singular-value dynamics under SGD follow the multiplicative Dyson Brownian motion (12):

$$\frac{d\sigma_i}{dt} = \underbrace{g(\sigma_i)}_{\text{signal}} + \underbrace{\frac{b}{N} \sum_{j \neq i} \frac{\sigma_i \sigma_j}{\sigma_i - \sigma_j}}_{\text{repulsion}} + \underbrace{\sqrt{\frac{b}{N}} \sigma_i \xi_i}_{\text{noise}},$$

where $g(\sigma_i)$ encodes the gradient signal from the loss function.

Step 1: Gradient rescaling. Under the isotropic perturbation (28), $\nabla_W \mathcal{L}_{\text{post}} = k \nabla_W \mathcal{L}_{\text{pre}}$, so the signal term transforms as $g \rightarrow k g$. The repulsion term depends only on the spectral geometry of $W^\top W$ and is independent of the loss; the noise term is determined by the learning rate and data covariance. Both are unaffected.

Step 2: Rescaling. Define $\tilde{\sigma}_i = \sigma_i/\sqrt{k}$. By Itô's lemma,

$$\frac{d\tilde{\sigma}_i}{dt} = g(\tilde{\sigma}_i) + \frac{b}{N} \sum_{j \neq i} \frac{\tilde{\sigma}_i \tilde{\sigma}_j}{\tilde{\sigma}_i - \tilde{\sigma}_j} + \sqrt{\frac{b}{N}} \tilde{\sigma}_i \xi_i,$$

which is the original SDE. Hence $\tilde{\sigma}_i$ has the same stationary distribution as σ_i^{pre} .

Step 3: Spectral invariants. The stationary density transforms as $p_\sigma^{\text{post}}(\sigma) = p_\sigma^{\text{pre}}(\sigma/\sqrt{k})/\sqrt{k}$, confirming (5). The tail exponent $\mu = 1 + 2\hat{m}/\sigma^2$ depends on \hat{m} and σ^2 only through their ratio, which is scale-invariant. The eigenportfolio directions v_k (columns of V) are determined by the angular part of the SVD, which is unchanged by a scalar rescaling of the loss. The effective spectral rank $r_{\text{eff}} = (\sum \sigma_i)^2 / \sum \sigma_i^2$ is a ratio of homogeneous functions and is therefore scale-invariant. \square

Since the portfolio weights are defined as $w_i = W_{ti} / \sum_j W_{tj}$ (allocation proportion in state t),

scaling all of W by a constant k leaves the proportions unchanged:

$$w^{\text{post}} = w^{\text{pre}}. \quad (30)$$

The portfolio composition is invariant under any isotropic perturbation.

Corollary 1 (Tax Neutrality). *A proportional wealth tax satisfying conditions C1–C3 of Frøseth (2026b)—tax liability depends only on total wealth (C1), the rate structure is smooth (C2), and all asset classes are assessed uniformly (C3)—is an isotropic perturbation with $k = (1 - \tau_c)(1 - \tau_d)$. By Theorem 1, such a tax preserves the spectral structure of the allocation matrix. This recovers and generalises the neutrality result of Frøseth (2026b).*

Proof. Under a proportional wealth tax with capital-value rate τ_c and distribution rate τ_d , condition C3 (uniform assessment) ensures that the after-tax return on every asset i is

$$r_i^{\text{post}} = (1 - \tau_d) r_i^{\text{pre}} - \tau_c.$$

The portfolio loss function, which depends on returns through $\sum_i w_i r_i$, thus transforms as $\mathcal{L}_{\text{post}} = (1 - \tau_d) \mathcal{L}_{\text{pre}} + c(\tau_c)$, where $c(\tau_c)$ is a constant independent of W . This is the isotropic form (28) with $k = (1 - \tau_c)(1 - \tau_d)$. The result follows from Theorem 1. \square

12 Anisotropic Perturbations and Spectral Distortion

When a perturbation affects different assets differently, the loss landscape is modified *anisotropically*:

$$\frac{\partial \mathcal{L}}{\partial W_{ti}} \rightarrow \frac{\partial \mathcal{L}}{\partial W_{ti}} + \delta_i W_{ti}, \quad (31)$$

where δ_i is the asset-specific perturbation strength. This is the regime Olsen et al. (2025) identify (Proposition 6.17) as breaking their isotropic analysis and producing distorted spectra.

The consequences are:

- Eigenportfolio directions v_k rotate (portfolio tilt toward less-perturbed assets).
- The repulsion structure changes (some factors compressed, others stretched).
- The tail exponent becomes direction-dependent.

The portfolio weight distortion, to first order in the perturbation, is

$$\Delta w^* = -\frac{1}{\gamma} V^{-1} (\delta - \bar{\delta} \cdot \mathbf{1}), \quad (32)$$

where γ is risk aversion, V is the return covariance matrix, $\delta = (\delta_1, \dots, \delta_N)$ is the vector of perturbation strengths, and $\bar{\delta}$ is their mean.

Example 1 (Differential Taxation). In the Norwegian wealth tax system, $\delta_i = \tau_w(1 - \alpha_i)$ where α_i is the assessment fraction for asset class i . With $\alpha_{\text{housing}} = 0.25$, $\alpha_{\text{shares}} = 0.80$,

$\alpha_{\text{deposits}} = 1.00$, the cross-asset variance $\text{Var}(\alpha_i) \approx 0.07$ drives portfolio tilt toward housing. Post-2017 discount changes (which changed α_{shares} from 1.00 to 0.80 over several steps) should produce time-varying spectral distortions testable in SSB microdata.

Example 2 (Sector-Specific Regulation). A financial regulation imposing different capital requirements on different sectors (e.g. Basel risk weights) enters as $\delta_i = c_i$ where c_i is the capital charge per unit of exposure to sector i . The spectral distortion is proportional to $\text{Var}(c_1, \dots, c_N)$.

Example 3 (Differential Transaction Costs). Market microstructure costs that vary across asset classes (e.g. bid–ask spreads, brokerage fees) enter as $\delta_i = \kappa_i$ where κ_i is the effective cost of trading asset i . Illiquid assets (high κ_i) are underweighted relative to the frictionless optimum.

13 The Spectral Invariance Conjecture

Conjecture 1 (Spectral Invariance). *Consider N assets with returns observed over T periods. Let the allocation matrix $W \in \mathbb{R}^{T \times N}$ evolve under stochastic gradient descent on a portfolio objective $\mathcal{L}(W)$ with isotropic noise of intensity D .*

- (a) *The stationary singular-value distribution has tail exponent $T - N + 1$, and the aggregate wealth distribution formed by $x = \|W\|_F$ has Pareto tail exponent $\alpha = 1 + v/D$ where v is the mean drift of $\|W\|_F$.*
- (b) *Any isotropic perturbation to \mathcal{L} preserves the spectral shape: the singular-value distribution shifts in scale but does not change its tail exponent or eigenportfolio directions.*
- (c) *Any anisotropic perturbation to \mathcal{L} with asset-specific weights δ_i rotates the eigenportfolio directions and changes the tail exponent in a direction-dependent way. The magnitude of distortion is proportional to the cross-asset variance:*

$$\text{distortion} \propto \text{Var}(\delta_1, \dots, \delta_N). \tag{33}$$

Status. Parts (a) and (b) follow from combining [Olsen et al. \(2025\)](#)’s theorems with the isotropic rescaling argument of [Theorem 1](#). Part (c) requires extending [Olsen et al. \(2025\)](#)’s analysis to the anisotropic case; their [Proposition 6.17](#) is a starting point but does not give the full spectral distortion.

Corollary 2 (Tax Neutrality). *Setting $\delta_i = \tau_w(1 - \alpha_i)$, part (b) recovers the neutrality conditions C1–C3 of [Frøseth \(2026b\)](#) and part (c) quantifies the cost of violating condition C3.*

Interpretation. The Conjecture provides a unified criterion for evaluating any environmental perturbation—tax, regulation, transaction cost, or fee structure—in terms of its spectral signature. Part (a) connects spectral and Pareto exponents; part (b) characterises benign (isotropic) perturbations; part (c) quantifies distortion from differential treatment.

14 The Loss Function and Utility Correspondence

The loss function $\mathcal{L}(W)$ encodes the portfolio objective. Different loss functions (MLE, score matching, denoising score matching) encode different utility assumptions.

14.1 Maximum Likelihood Estimation

For maximum likelihood estimation of a drift function, the loss is

$$\mathcal{L}_{\text{MLE}}(W) = -\frac{1}{T} \sum_{t=1}^T \log p(x_{t+1} | x_t; W). \quad (34)$$

Under the SDE model $dx = \hat{v}(x; W) dt + \sqrt{2D} dB$, the transition density is approximately Gaussian:

$$\mathcal{L}_{\text{MLE}}(W) = \frac{1}{4DT\Delta t} \sum_t (x_{t+1} - x_t - \hat{v}(x_t; W) \Delta t)^2 + \text{const}. \quad (35)$$

This is a **quadratic loss in the residuals**—a mean-squared-error objective. Minimizing this is equivalent to maximizing expected utility in a quadratic (CRRA with $\gamma = 2$) utility framework.

14.2 Score Matching and Denoising Score Matching

Score matching (Hyvärinen, 2005) minimizes

$$\mathcal{L}_{\text{SM}}(\theta) = \mathbb{E}_{x \sim p_{\text{data}}} \left[\frac{1}{2} \|s_{\theta}(x) - \nabla_x \log p_{\text{data}}(x)\|^2 \right], \quad (36)$$

where s_{θ} is the parametric score function. At steady state, $v(x) = D \cdot s(x)$, so learning the score is equivalent to learning the drift.

Denoising score matching (DSM) adds Gaussian noise $\tilde{x} = x + \sigma\epsilon$ and regresses $s_{\theta}(\tilde{x})$ onto $-\epsilon/\sigma$. The noise structure is isotropic by construction, which is exactly the regime where Olsen et al. (2025)'s spectral theory applies cleanly.

Key observation: score matching loss is quadratic in the score residual, just as MLE loss is quadratic in the drift residual. The SGD dynamics on the network weights therefore have the same mathematical structure regardless of whether we use MLE or score matching.

Proposition 1 (Loss–Utility Correspondence). *Let $\mathcal{L}(W) = \frac{1}{T} \sum_t \|r_t(W)\|^2$ be any loss function that is quadratic in a residual vector r_t (encompassing MLE (35), score matching, and denoising score matching). If the residuals satisfy $\mathbb{E}[r_t r_t^{\top}] = \Sigma_r$ with Σ_r full rank, then:*

1. *The SGD gradient covariance takes the form $C = \frac{4}{T} \sum_t (r_t r_t^{\top}) \otimes (x_t x_t^{\top})$, which has the Kronecker structure required by Olsen et al. (2025).*
2. *The stationary spectral density of W falls in the universality class characterised by the free log-normal (11) with parameters determined by the signal-to-noise ratio $\|v_{\text{true}}\|^2/D$.*
3. *The spectral tail exponent encodes the Pareto exponent of the target distribution via (21).*

Proof sketch. For a linear network $\hat{v}(x; W) = Wx$, the gradient is $\nabla_W \mathcal{L} = -\frac{2}{T} \sum_t r_t x_t^\top$, and the gradient covariance factors as $C = 4 \Sigma_r \otimes \Sigma_x / T$. This Kronecker structure is the input assumption of [Olsen et al. \(2025\)](#), Theorem 3.2 (spectral characterisation of SGD stationary distributions). The resulting spectral density of W follows a gamma-type distribution with tail exponent $m - n + 1$, which maps to a Pareto exponent $\alpha = 1 + f(m, n, \beta_1 / \eta D)$ via (21).

For a nonlinear network, the gradient is $\nabla_W \mathcal{L} = -\frac{2}{T} \sum_t r_t (\partial \hat{v} / \partial W)^\top$. The Jacobian $\partial \hat{v} / \partial W$ introduces input-dependent curvature. However, at each layer, the effective gradient covariance retains the Kronecker structure $C_l \approx \Sigma_{r,l} \otimes \Sigma_{a,l}$, where $\Sigma_{a,l}$ is the covariance of layer- l activations. This per-layer factorisation is empirically well-established and is the basis for natural gradient methods (KFAC). Under this factorisation, the spectral analysis of [Olsen et al. \(2025\)](#) applies layer-by-layer. \square

Remark. The remaining gap for a fully rigorous result is the per-layer Kronecker approximation $C_l \approx \Sigma_{r,l} \otimes \Sigma_{a,l}$, which holds exactly for linear networks and approximately for networks with smooth activations in the large-width regime. Establishing this rigorously for finite-width networks with standard activations (ReLU, GELU) remains open.

15 Applications

The spectral portfolio framework applies to several domains. We develop four here, emphasising that the general results of Part III specialise differently in each context.

15.1 Portfolio Design and Factor Structure

The core–satellite structure of the stationary spectral distribution (Definition 1) provides a principled decomposition of any portfolio. The bulk of the singular-value distribution corresponds to broad, diversified market exposure; the power-law tail corresponds to concentrated factor bets. The effective spectral rank r_{eff} (Definition 2) measures portfolio complexity and is directly computable from allocation data.

For portfolio construction, the three forces (Section 3) provide new micro-foundations: the gradient signal identifies high-return factors, the survival constraint prevents premature abandonment of underperforming positions, and eigenvalue repulsion enforces diversification without explicit constraints. In deep networks with L layers, the hierarchical SVD decomposes the allocation into primary factors, combinations of factors, and final output—a fund-of-funds architecture.

15.2 Wealth Inequality and Power Laws

The aggregation result (Section 8) establishes that the Pareto exponent of the wealth distribution is determined by the spectral exponent of the allocation matrix via the radial projection $\alpha = 1 + f(m, n, \beta_1 / \eta D)$. The Bouchaud–Mézard unification (Section 10) shows that the same eigenvalue repulsion mechanism that prevents portfolio concentration also prevents wealth condensation. The matrix Kesten problem (Section 6.4) provides the explicit link: the inverse-

Wishart spectrum of the multiplicative regime maps to the Pareto tail of the wealth distribution under $\lambda \rightarrow 1/\lambda$.

These results connect the observable spectral structure of portfolio allocations to the shape of the wealth distribution, offering a new diagnostic for inequality: the spectral tail exponent is directly measurable from microdata without parametric assumptions about the wealth distribution itself.

15.3 Tax Policy: Neutrality as Spectral Invariance

The Spectral Invariance Theorem (Theorem 1) has a direct application to wealth tax design. The identification between weight matrices and portfolio allocations enables a dictionary between the domains:

SGD (Olsen et al.)	Portfolio	Tax Context
Weight matrix W	Allocation matrix	Wealth composition
Loss function $\mathcal{L}(W)$	Negative expected utility	Tax-adjusted returns
Learning rate η	Rebalancing speed	Adjustment speed to tax
Diffusion constant D	Information noise	Market microstructure
Singular values σ_k	Factor concentrations	Sector concentration
Repulsion term	Diversification pressure	Tendency away from concentration

Table 2: Dictionary linking SGD weight dynamics, portfolio optimization, and tax design.

Under conditions C1–C3 of Frøseth (2026b)—tax liability depends only on total wealth (C1), rates are smooth (C2), and all asset classes are assessed uniformly (C3)—the post-tax loss satisfies $\mathcal{L}_{\text{post}} = k \cdot \mathcal{L}_{\text{pre}} + \text{const}$ with $k = (1 - \tau_c)(1 - \tau_d)$, which is an isotropic perturbation. By Corollary 1, such a tax preserves the spectral structure.

When C3 is violated, the Norwegian system provides a concrete test case. With $\alpha_{\text{housing}} = 0.25$, $\alpha_{\text{shares}} = 0.80$, $\alpha_{\text{deposits}} = 1.00$, the cross-asset variance $\text{Var}(\alpha_i) \approx 0.07$ drives spectral distortion aligned with the assessment differentials. The post-2017 changes to α_{shares} should produce time-varying spectral shifts testable in wealth register data.

15.4 Neural Network Diagnostics

The spectral theory predicts that the singular-value distribution of trained weight matrices should reflect the complexity of the target distribution. For neural networks trained on stochastic processes—such as neural SDEs learning drift or score functions—the weight spectrum serves as a convergence diagnostic: deviation from the predicted spectral shape indicates incomplete learning.

Denosing score matching should produce weight matrices with cleaner spectral structure than MLE training, because DSM’s noise is isotropic by construction and thus falls squarely within the universality class of Olsen et al. (2025). This prediction is testable by comparing weight spectra across training objectives on the same data.

16 Testable Predictions

The framework yields predictions at three levels.

16.1 General Spectral Predictions

1. **Spectral tail exponent from portfolio data.** Compute the SVD of wealth portfolios (e.g. from Norwegian SSB microdata). The tail exponent of the singular-value distribution should be $T - N + 1$ where T is the observation window and N the number of asset classes.
2. **Cross-sectional variation by investor horizon.** Long-horizon investors (low effective D) should have fatter spectral tails (more concentrated factor exposures) than short-horizon investors. Testable by segmenting wealth-holders by holding period.
3. **Weight spectrum of trained networks.** After training a neural network on wealth or return data, the weight spectrum should reflect the Pareto exponent of the target distribution. Deviation signals incomplete convergence.
4. **DSM vs. MLE spectral comparison.** Denoising score matching should produce weight matrices with cleaner spectral structure (closer to [Olsen et al. \(2025\)](#)'s predictions) than MLE training, because DSM's noise is isotropic by construction.
5. **Computational cost vs. spectral complexity.** Training networks on data with high intrinsic dimensionality (fat spectral tails, large r_{eff}) requires more computation. Testable by comparing training time across datasets with known spectral structure.
6. **Finite-width effects.** [Olsen et al. \(2025\)](#)'s results hold in the large- r limit. The predicted spectral exponent should hold approximately for finite-width networks, with deviations decreasing as width increases.

16.2 Tax-Specific Predictions

7. **Spectral shift around Norwegian tax reforms.** The introduction and variation of the *verdsettelsesrabatt* (valuation discount) changed α_{shares} from 1.00 to 0.80 over several steps. The anisotropic perturbation should produce measurable rotation of eigenportfolio directions and direction-dependent tail distortion.
8. **Distortion proportional to $\text{Var}(\alpha_i)$.** The magnitude of portfolio distortion should be proportional to the cross-asset variance of assessment fractions. Norway's system ($\text{Var} \approx 0.07$) provides a quantitative benchmark.

17 Open Questions and Future Work

Several important questions remain open:

1. **Anisotropic extension.** Part (c) of Conjecture 1 requires extending [Olsen et al. \(2025\)](#)'s analysis to anisotropic noise. Their Proposition 6.17 begins this, but the full spectral distortion (eigenportfolio rotation, tail exponent change) remains to be derived.

2. **Finite-width corrections.** [Olsen et al. \(2025\)](#)'s results are asymptotic. What are the correction terms for finite-width networks? How large do m, n need to be for the predictions to hold?
3. **Budget constraints.** In practice, investors face constraints ($\sum_i w_i = 1$). The unconstrained Olsen SDE does not enforce this. How do constraints modify the spectral evolution?
4. **Nonlinear activations.** We assume linear output (network is a linear projection of hidden units). How do nonlinear final layers (softmax, sigmoid) change the spectral dynamics?
5. **Compute f for intermediate regimes.** The single-factor (22) and large- N (23) limits are now explicit. Extend to finite $p > 1$ with eigenvalue repulsion corrections, and verify numerically against simulated spectra.
6. **Policy optimization.** Given a target Pareto exponent (e.g., $\alpha = 2$ as observed in Norway), what perturbation structure minimizes spectral distortion while achieving the desired distributional shift? This inverse problem applies to tax design, regulation, and fee structures alike.

18 Relation to Existing Literature

We position this work relative to six major research areas.

18.1 Random Matrix Theory in Finance

[Laloux et al. \(1999\)](#) and [Ledoit and Wolf \(2004\)](#) apply RMT to estimate asset covariance matrices. Our contribution: RMT also describes the structure of portfolio weight matrices (not just return covariances), and this structure is directly observable in trained neural networks. The connection to Pareto wealth distributions is new.

18.2 Deep Learning and Portfolio Optimization

[Heaton et al. \(2017\)](#); [Zhang et al. \(2021\)](#); [Buehler et al. \(2019\)](#) develop neural network methods for portfolio optimization. Our framework provides a spectral lens on why these networks learn specific factor structures, and predicts their spectra from the data distribution.

18.3 Ergodicity Economics

[Peters \(2019\)](#); [Peters and Gell-Mann \(2016\)](#); [Berman et al. \(2020\)](#) distinguish ensemble and time averages in wealth dynamics. Our contribution: the spectral gap (Section 7) is a measurable proxy for non-ergodicity, and spectral-preserving perturbations minimize it.

18.4 SGD Spectral Theory

[Martin and Mahoney \(2021\)](#); [Pennington and Worah \(2017\)](#); [Gunasekar et al. \(2018\)](#) develop spectral theory of SGD in deep learning. [Olsen et al. \(2025\)](#) extend this to matrix-valued weight

evolution. Our contribution: we show this theory has direct portfolio and wealth-distribution interpretations.

18.5 Classical Portfolio Theory

Merton (1969, 1971) define optimal portfolio choice via mean–variance analysis. Cover (1991) develops universal portfolio theory. Our framework extends these: the spectral decomposition reveals factor structure emerging from learning data, and spectral invariance provides a new rationale for neutral policy design.

18.6 Wealth Dynamics and Power Laws

Drăgulescu and Yakovenko (2000); Yakovenko and Rosser (2009); Gabaix and Koijen (2022) study mechanisms generating power-law wealth distributions. Kesten (1973) develops the Kesten recursion, central to our multiplicative regime. Our contribution: we connect these scalar mechanisms to the matrix spectral theory via the matrix Kesten problem and the free log-normal interpolation.

19 Conclusion

This paper establishes spectral portfolio theory as a unified framework connecting random matrix theory, portfolio dynamics, and wealth distribution.

The core results are:

1. Neural network weight matrices are portfolio allocation matrices, and their spectral structure encodes factor decompositions faithful to the underlying data.
2. The three forces in the singular-value evolution—gradient signal, survival constraint, and eigenvalue repulsion—translate directly into portfolio economics: smart money, information-driven survival, and endogenous diversification.
3. The stationary spectral distribution exhibits a core–satellite structure (bulk plus power-law tail) that matches empirical wealth and institutional portfolio data.
4. The spectral transition from additive (Marchenko–Pastur) to multiplicative (inverse-Wishart) regimes, mediated by the free log-normal, explains how daily return statistics relate to long-run wealth concentration.
5. The Bouchaud–Mézard cross-sectional model, the Olsen within-portfolio model, and the scalar Fokker–Planck framework all emerge from the same underlying spectral dynamics, unified via common power-law mechanisms.
6. The Spectral Invariance Theorem establishes that isotropic perturbations preserve spectral shape, while anisotropic perturbations produce distortion proportional to their cross-asset variance. This general result specialises to tax neutrality, regulatory impact assessment, and transaction cost analysis.

The framework is testable: portfolio SVDs can be computed from microdata, spectral exponents can be compared against wealth Pareto exponents, and spectral shifts around policy reforms can be measured directly. Applications span portfolio construction, inequality measurement, tax design, and neural network diagnostics.

Future work should: (1) extend the theory to anisotropic noise; (2) derive finite-width corrections; (3) handle budget constraints and nonlinear activations; (4) explicitly compute the spectral-to-Pareto mapping; and (5) solve the inverse problem of perturbation design—choosing the structure that achieves a distributional target while minimizing spectral distortion.

Acknowledgements

The author acknowledges the use of Claude (Anthropic) for assistance with literature review, L^AT_EX typesetting, mathematical exposition, and editorial refinement, and Lemma (Axiomatic AI) for review and proof checking. All substantive arguments, economic reasoning, and conclusions are the author’s own.

References

- Gert Aarts, Biagio Lucini, and Chanju Park. Stochastic weight matrix dynamics during learning and Dyson Brownian motion. *Physical Review E*, 111:015303, 2025. doi: 10.1103/PhysRevE.111.015303.
- Yonatan Berman, Ole Peters, and Alexander Adamou. Far from the madding crowd: Collective wisdom in prediction markets. *arXiv preprint*, 2020.
- Jean-Philippe Bouchaud and Marc Mézard. Wealth condensation in a simple model of economy. *Physica A*, 282(3–4):536–545, 2000.
- Jean-Philippe Bouchaud and Marc Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press, 2nd edition, 2003. doi: 10.1017/CBO9780511753893.
- Hans Buehler, Lukas Gonon, Josef Teichmann, and Ben Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019. doi: 10.1080/14697688.2019.1571683.
- Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. doi: 10.1080/713665670.
- Thomas M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991. doi: 10.1111/j.1467-9965.1991.tb00002.x.
- Adrian Drăgulescu and Victor M. Yakovenko. Statistical mechanics of money. *European Physical Journal B*, 17(4):723–729, 2000.
- Anders G. Frøseth. Extensions to the wealth tax neutrality framework. 2026a. arXiv:[2603.05277](https://arxiv.org/abs/2603.05277) [physics.soc-ph].

- Anders G. Frøseth. Asset returns, portfolio choice, and proportional wealth taxation. 2026b. arXiv:[2603.05264](https://arxiv.org/abs/2603.05264) [physics.soc-ph].
- Anders G. Frøseth. Wealth taxation as a drift modification: A Fokker–Planck approach to tax neutrality. 2026c. arXiv:[2603.05283](https://arxiv.org/abs/2603.05283) [physics.soc-ph].
- Xavier Gabaix and Ralph S. J. Koijen. In search of the origins of financial fluctuations: The inelastic markets hypothesis. Working paper, Harvard University and University of Chicago, 2022.
- Martin J. Gruber. Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance*, 51(3):783–810, 1996. doi: 10.1111/j.1540-6261.1996.tb02707.x.
- Suriya Gunasekar, Jason Lee, Daniel Soudry, and Nathan Srebro. Implicit regularization in matrix sensing. In *Advances in Neural Information Processing Systems*, volume 31, 2018.
- J. B. Heaton, N. G. Polson, and J. H. Witte. Deep learning for finance: Deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12, 2017. doi: 10.1002/asmb.2209.
- Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993. doi: 10.1093/rfs/6.2.327.
- Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6:695–709, 2005.
- Harry Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248, 1973.
- Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Physical Review Letters*, 83(7):1467–1470, 1999. doi: 10.1103/PhysRevLett.83.1467.
- Olivier Ledoit and Michael Wolf. Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004. doi: 10.3905/jpm.2004.110.
- Andrew W. Lo. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *The Journal of Portfolio Management*, 30(5):15–29, 2004. doi: 10.3905/jpm.2004.442611.
- Andrew W. Lo. *Adaptive Markets: Financial Evolution at the Speed of Thought*. Princeton University Press, Princeton, 2017.
- Harry Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.
- Charles H. Martin and Michael W. Mahoney. Implicit self-regularization in deep neural networks: Evidence from random matrix theory and implications for training. In *Journal of Machine Learning Research*, volume 22, pages 1–73, 2021.

- Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, 51(3):247–257, 1969. doi: 10.2307/1926560.
- Robert C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, 1971. doi: 10.1016/0022-0531(71)90038-X.
- Thijs Olsen, Mona Fatehmanesh, James Lim, and Roger Grosse. From SGD to spectra: The matrix and spectral geometry of neural network training. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025. arXiv:2507.12709.
- Jeffrey Pennington and Pratik Worah. Nonlinear random matrix theory for deep learning. In *Advances in Neural Information Processing Systems*, volume 30, pages 2637–2646, 2017.
- Ole Peters. The ergodicity problem in economics. *Nature Physics*, 15:1216–1221, 2019.
- Ole Peters and Murray Gell-Mann. Evaluating gambles using dynamics. *Chaos*, 26:023103, 2016. doi: 10.1063/1.4940236.
- Marc Potters and Jean-Philippe Bouchaud. *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge University Press, 2021. doi: 10.1017/9781108768900.
- A. D. Roy. Safety first and the holding of assets. *Econometrica*, 20(3):431–449, 1952. doi: 10.2307/1907413.
- Victor M. Yakovenko and J. Barkley Rosser, Jr. Colloquium: Statistical mechanics of money, wealth, and income. *Reviews of Modern Physics*, 81(4):1703–1725, 2009.
- Zihao Zhang, Stefan Zohren, and Stephen Roberts. Deep learning for portfolio optimization. *The Journal of Financial Data Science*, 3(2):8–20, 2021. doi: 10.3905/jfds.2020.1.052.