Free Decompression with Algebraic Spectral Curves

Siavash Ameli
ICSI and Department of Statistics
University of California, Berkeley
sameli@berkeley.edu

Chris van der Heide
Dept. of Electrical and Electronic Engineering
University of Melbourne
chris.vdh@gmail.com

Liam Hodgkinson
School of Mathematics and Statistics
University of Melbourne
lhodgkinson@unimelb.edu.au

Michael W. Mahoney
ICSI, LBNL, and Department of Statistics
University of California, Berkeley
mmahoney@stat.berkeley.edu

Abstract

Tools from random matrix theory have become central to deep learning theory, using spectral information to provide mechanisms for modeling generalization, robustness, scaling, and failure modes. While often capable of modeling empirical behavior, practical computations are limited by matrix size, often imposing a restriction to models that are too small to be realistic. This motivates the inference of properties of larger models from the behavior of smaller ones. Free decompression (FD) is a recently proposed method for extrapolating spectral information across matrix sizes, but its utility is currently limited by strong assumptions that preclude its implementation on more realistic machine learning (ML) models. We use algebraic spectral curve theory to provide a general FD methodology for spectral densities whose Stieltjes transform satisfies an algebraic relation, a modeling assumption that is more likely to hold in practice. This recasts FD as an evolution along spectral curves which can be readily integrated. Our framework enables the expansion of spectral densities that have multiple or multi-modal bulks, that exist at multiple scales, and that contain atoms, all characteristic of real-world data and popular ML models. We demonstrate the efficacy of our framework on models of interest in modern ML, including Hessian and activation matrices associated with neural networks and large-scale diffusion models.

1 Introduction

At the core of scientific computing and much of modern machine learning (ML) lies the challenge of estimating the eigenvalues of high-dimensional Hermitian matrices. Such matrices, including kernels, Hessians, and graph representations, encode the intrinsic geometry and connectivity of the data and models built on them, rendering the pursuit of efficient spectral techniques a primary concern for both theory and practice. Studying eigenspectra has become a prominent approach to understanding performance and guiding training in deep learning [10, 22, 23, 24, 25, 40, 57]. In many cases, the spectra of such matrices have non-trivial structure, often containing spikes, multiple multi-modal bulks, and heavy-tails [16, 30]. Conventional algorithms to extract eigenvalue information from these matrices have required that the data are able to be stored in memory, scratch space, or can at least be accessed as an implicit operator (via matrix-vector products).

More recently, a new class of algorithms has emerged that is able to provide highly-accurate estimates of the eigenvalues (or summary functionals thereof [2]) of matrices, *even without implicit or explicit access to the full matrix*, i.e., of so-called *impalpable* matrices [1]. One such method, termed *Free Decompression* (FD), shows great promise as a tool for gaining access to the spectral distributions

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

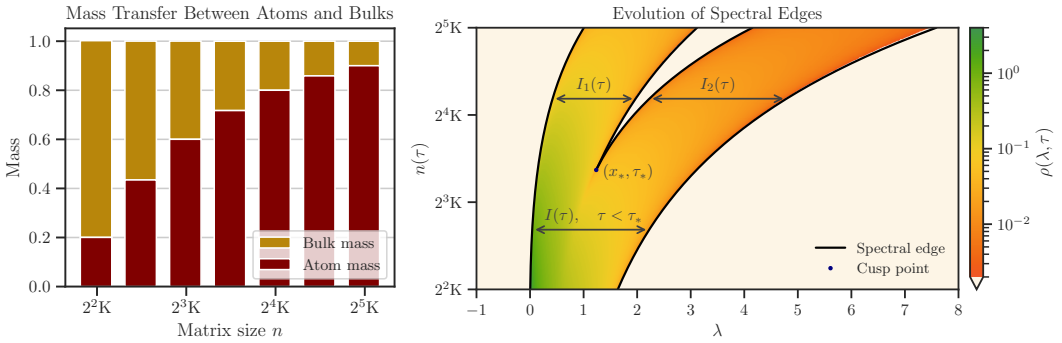

Figure 1: Evolution of atomic mass (left), density and spectral edges (right) with increasing matrix size $n(\tau)$ for the compound free Poisson law in Section 4. A complexity our method handles is that the bulk with support $I(\tau)$ splits at the cusp (x_*, τ_*) to reveal two evolving supports $I_1(\tau)$ and $I_2(\tau)$.

of such impalpable matrices. The central premise is that by appropriately sampling a small submatrix from the large impalpable matrix of interest, one can evolve a partial differential equation (PDE) in the Stieltjes transform of a spectral density in the decompression ratio to the desired matrix dimension. Despite demonstrating impressive accuracy when considering synthetic examples that naturally arise in random matrix theory (RMT) and free probability, the methods in [1] are highly restrictive, assuming the distribution of eigenvalues can be described by a single smooth connected bulk. This assumption is inevitably violated when challenged with more realistic examples, preventing widespread adoption of FD-based methods.

Instead, we develop a more broadly applicable FD methodology using algebraic spectral curve theory. We assume that the Stieltjes transform of the limiting spectral distribution of a class of random matrices is *algebraic*, i.e., it can be represented as a root of a (typically low-order) bivariate polynomial relation. All canonical random matrix ensembles satisfy this assumption (see Remark H.1), and with increasing polynomial degree, an extraordinarily broad class of spectral densities can be captured in this way. For example, Pennington et al. [40, 42] highlight that the Stieltjes transforms of certain Hessian approximations and neural network gram matrices are, respectively, solutions to cubic and quartic polynomial equations. Rao & Edelman [43] introduced the *polynomial method*, showing how algebraic representations can, in principle, simplify random-matrix computations because many operations and transforms preserve this structure. Conveniently, it turns out that FD does too, recasting decompression as evolution on a *spectral curve*. This fact can be exploited to circumvent many of the challenges faced by the approach taken in Ameli et al. [1], including the problem of analytic continuation, which can be handled geometrically on algebraic spectral curves. Unfortunately, to our knowledge, there has been no general numerical implementation of the polynomial method for empirical matrix computations to date, due to major practical challenges in fitting the polynomial and identifying its physical root. Indeed, Rao & Edelman [43, Sections 7 and 11] identified both as open problems. These issues must first be overcome to accomplish FD at scale.

Contributions. The central contribution of this work is to introduce a new general FD methodology and practical tools for spectral computations on large impalpable matrices. We develop the first general numerical implementation of the polynomial method for empirical spectra and apply it to obtain a state-of-the-art FD implementation. This enables us to directly trace several spectral quantities, including the edges of the distribution, locations of atoms and their weights, as well as spectral moments (see Figure 1). This is highly valuable in many applications: for positive-definite matrices, the ratio of the largest to the smallest eigenvalues constitutes the *condition number* $\mathrm{cond}(\mathbf{A})$ of the matrix $\mathbf{A}$, a critical component for understanding rates of convergence in optimization algorithms, for example. Similarly, it was demonstrated in Bonnaire et al. [10] that the edges of bulk regions of the spectrum play an important role in the training dynamics of diffusion models. In doing so, we

- develop an algebraic FD framework (AFD) that generalizes previous methods beyond quadratic Stieltjes transforms, extending previous work to accurately approximate multi-scale spectral distributions, which may be multi-modal and/or contain multiple bulk regions;
- derive equations to accurately compute the dynamics of important spectral quantities (atoms, moments, and edges), as well as track the mechanism of splitting of bulks, without the need to compute the full spectral distribution;

- demonstrate empirically that the method works on several classes of ML models; and
- deploy an implementation into the open source software project `freealg`[1].

The remainder of the paper is structured as follows. Section 2 outlines some background and related work, as well as tools from free probability. Section 3 describes algebraic free decompression (AFD), detailing FD in this context, and the numerical methods we use, including associated methods for atoms, spectral edges, and moments. Section 4 provides numerical examples demonstrating the utility of our method. We conclude and discuss limitations in Section 5. Background, proofs, additional information and experiments are provided in a detailed set of appendices.

2 Background on Free Decompression

Free decompression was recently introduced by Ameli et al. [1] as a method for gaining access to the spectral distribution of so-called *impalpable* matrices: cases where the user does not have complete access to the matrix, perhaps due to their sheer size, corruption, or data availability. The method relies upon tools from free probability to derive a PDE in the familiar Stieltjes transform, a standard tool in RMT that at this stage has been widely deployed in both ML theory and the development of practical algorithms.

We briefly recall some relevant notation. Let $\mathbf{A}$ denote a matrix of size $n \times n$, where n is assumed to be large, and let $m_{\mathbf{A}}$ denote the corresponding *empirical Stieltjes transform*, given by

$$m_{\mathbf{A}}(z) \coloneqq \frac{1}{n}\mathrm{tr}(\mathbf{A} - z\mathbf{I})^{-1} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i - z}, \tag{1}$$

where $z \in \mathbb{C} \setminus \{\lambda_i\}_{i=1}^{n}$ and λ_i is the i-th eigenvalue of $\mathbf{A}$. It is generally assumed that $\mathbf{A}$ is an element of a sequence $(\mathbf{A}_n)_{n=1}^{\infty}$ where each $\mathbf{A}_i$ is contained in $\mathbf{A}_j$ for $j > i$, $\mathbf{A}_n \in \mathbb{R}^{n \times n}$, and as $n \to \infty$, $m_{\mathbf{A}_n}(z) \to m(z)$ pointwise. Here, m is the *Stieltjes transform* of a limiting *spectral measure* ν in the sense that $m : \mathbb{C} \setminus \mathrm{supp}(\nu) \to \mathbb{C}$ satisfies $m(z) \coloneqq \int_{\mathbb{R}} \frac{1}{x-z}\,\nu(\mathrm{d}x)$. The R-transform $R(z)$ corresponding to m is the (possibly multi-valued) function satisfying $R(-m(z)) = z + \frac{1}{m(z)}$ for $z \in \mathbb{C} \setminus \mathrm{supp}(\nu)$.

To model the process of matrix subsampling, let σ_n denote a uniformly random permutation of $\{1, \ldots, n\}$ and for $1 \le k \le n$, let $\mathbf{A}_n^{(k)} = (\mathbf{A}_{\sigma_n(i)\sigma_n(j)})_{i,j=1}^{k}$. Then for any subsampling factor $\tau < 1$, $\mathbb{E} m_{\mathbf{A}_n^{([\tau n])}}(z) \to m_\tau(z)$ as $n \to \infty$. Ameli et al. [1] asserts that the map $m \mapsto m_\tau$ is equivalent to the map that operates on the R-transform by $R(z) \mapsto R(\tau z)$ via the Nica-Speicher Theorem [35]. However, this identification does not hold for arbitrary matrix sequences: it is justified when the matrix and the sampling projection are asymptotically free, with orthogonal invariance providing a sufficient, but not necessary, condition (see Appendix F.1 for a proof). [2] When $\tau < 1$, as is typical in free probability theory, this is called ***free compression***. However, this operation can be formally inverted by taking $\tau > 1$, moving from the Stieltjes transform of the submatrix to that of the full matrix; when well-defined, this inverse operation is ***free decompression***.

While a number of obvious strategies exist to implement free decompression, they each have significant drawbacks. The multi-valued nature of the R-transform, and the need to eventually recover m, make it difficult to work with directly. The approach taken by Ameli et al. [1] relies on numerically solving a PDE in m, where the major challenges faced are accurate curve-fitting of the approximated spectral distribution of the initial matrix, and analytic continuation of the Stieltjes transform across the x-axis. Reparameterization of this PDE results in an inviscid complex Burgers equation, and while this presents another potential approach, it does not seem straight-forward to implement in practice. While the formation of singularities and shocks can be expected for equations of this type, we have identified a family of distributions for which the flow exists for all time in Appendix F.2. For further discussion on these difficulties, we refer the reader to Appendix F.3.

[1] Available at `https://github.com/ameli/freealg`, with documentation at `https://ameli.github.io/freealg`.

[2] Orthogonal invariance arises naturally in some ML models, such as Gaussian Gram matrices of the form $\mathbf{Z}^\intercal \mathbf{\Sigma} \mathbf{Z}$ with iid Gaussian $\mathbf{Z}$. Related covariance approximations arise for certain high-dimensional kernel matrices [16], although these do not guarantee invariance in general. Our experiments suggest that FD can remain effective even beyond this sufficient setting.

3 Implementing the Polynomial Method

An especially wide array of operations in random matrix theory, including free additive and multiplicative convolution, can be distilled into operations acting on various *transforms* of the Stieltjes transform, including the R-transform, the S-transform, and the moment transform. We have seen that free decompression is defined in terms of the R-transform. ***These transforms are all related to each other by algebraic relations.*** This is inconvenient for performing direct numerical conversions between the transforms themselves [12, 37]. However, it is *very convenient* if the Stieltjes transform is already defined through an implicit algebraic relation, as deriving relations for each of the transforms and their inverses offers no further complexity. This algebraic viewpoint is central to the polynomial method introduced by Rao & Edelman [43]. At the heart of our approach is the following ansatz.

Assumption 1. The Stieltjes transform m is *algebraic*: there exists a nonzero bivariate polynomial $P : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ such that for any $z \in \mathbb{C} \setminus \mathcal{K}$ where $\mathcal{K}$ is a finite set of singularities,

$$P(z, m(z)) = 0. \tag{2}$$

We write d_z for the degree of this polynomial in z, and s for its degree in m. In addition to the quadratic examples previously handled by Ameli et al. [1], this unlocks, for example, models of Hessian matrices and neural network gram matrices [40, 42], which we demonstrate in Section 4. Under Assumption 1, Cauchy, R-, S- and moment transforms, as well as common algebraic operations and stochastic transformations, can also be realized as the roots of corresponding polynomials.

Despite its convenience, Rao & Edelman [43] highlight fundamental obstacles to a general numerical implementation of the polynomial method. While m is easily estimated, finding P is nontrivial. Polynomial roots are sensitive to their coefficients, and because valid P for algebraic Stieltjes transforms are rare among bivariate polynomials, naïve polynomial fits often radically fail. Furthermore, identifying the correct numerical solution for m from P alone is nearly impossible. Consequently, to our knowledge, no general numerical implementations of the polynomial method exist. In the following sections, we address these challenges and provide its first general numerical implementation.

Fitting the Bivariate Polynomial. To fit P, we first choose its degrees in z and m, denoted by d_z and s, respectively. These degrees can be selected post hoc using standard model-selection criteria [26, Section 6.1.3]; in practice, we increase them until the fitting error stabilizes while avoiding overparameterization. Our degree-sensitivity experiment in Table I.1 shows that choosing too low a degree can fail to capture the correct bulk topology, while modest over-specification is well tolerated. By exploiting the conjugate symmetry of the Stieltjes transform, we show that its irreducible defining polynomial must have real coefficients up to an overall scalar (Lemma D.1). Thus, we write

$$P(z, m) \coloneqq \sum_{i=0}^{d_z} \sum_{j=0}^{s} c_{ij} z^i m^j, \qquad c_{ij} \in \mathbb{R}. \tag{3}$$

To estimate the coefficients c_{ij}, we need samples of the Stieltjes transform. Unlike the density-fitting approach of Ameli et al. [1], which first fits the initial density and then computes its Stieltjes transform, we obtain these samples directly from the empirical Stieltjes transform (1), computed from the observed submatrix eigenvalues. We choose sampling points z on Bernstein ellipses of varying radii around each spectral bulk. For spectra spanning many orders of magnitude, we found that ordinary ellipses did not sample the bulks effectively, so we developed contours adapted to logarithmic scales. We further impose linear constraints on the coefficients of P using the Stieltjes transform's expansion at infinity, including its normalization $m(z) \sim -1/z$ and empirical spectral moments when available. These constraints enforce necessary conditions for P to admit a Stieltjes-transform branch. The fitting procedure is detailed in Appendix D, with pseudocode given in Algorithm D.1.

The Physical Branch. While algebraic manipulations of P are straightforward, a central numerical challenge is identifying, for each $z \in \mathbb{C}$, which root of $P(z, \cdot)$ is the physical Stieltjes value. The physical branch is characterized globally by the Herglotz property $\Im m(z) > 0$ for $z \in \mathbb{C}^+$ and by the normalization $m(z) = -z^{-1} + O(z^{-2})$ at infinity. However, these conditions do not by themselves provide a stable pointwise root-selection rule. In high-degree or multi-bulk examples, roots can come close, exchange ordering, or pass near branch points, and relying only on pointwise tests can easily lead to jumps between sheets.

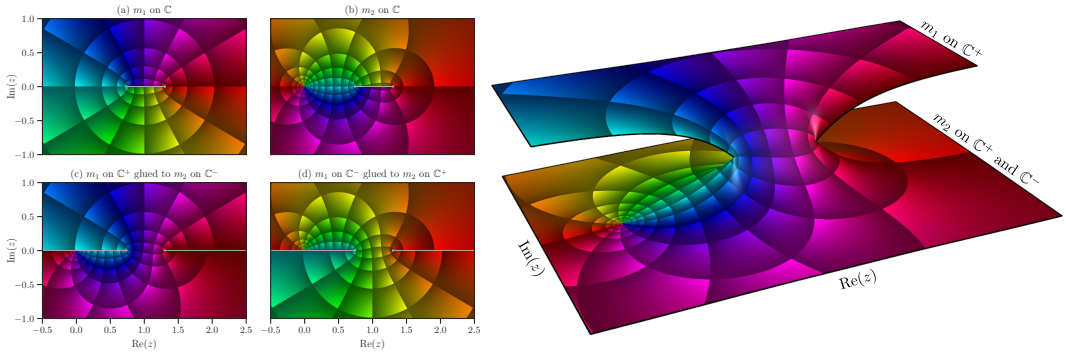

Figure 2: Visualization of multiple solutions to the algebraic relation (2) for the Marchenko–Pastur density with $\lambda = \frac{1}{50}$ (left), and the surface obtained by gluing the Stieltjes solution m_1 to the secondary branch m_2 (right); colors are interpreted as in Appendix A. Free decompression requires tracing families of continuous paths along this surface, from the top sheet down to the lower sheet.

Branch Selection via Continuation. We therefore select the physical branch by geometric continuation on the algebraic spectral curve $P(z, m) = 0$. Starting from an anchor point where the physical root can be identified reliably, either from the normalization at infinity or by comparison with the empirical Stieltjes transform, we continue the solution along a path in the z-plane. Along a smooth portion of the curve, implicit differentiation gives

$$\frac{\mathrm{d}m}{\mathrm{d}z} = -\frac{\partial_z P(z, m)}{\partial_m P(z, m)}.$$

This ordinary differential equation respects the local geometry of the sheet and provides a predictor for transporting the physical root from the anchor to the query point. The predictor is then corrected by Newton projection onto $P(z, m) = 0$, with step subdivision when the geometry becomes stiff. This continuation procedure is substantially more stable than independent pointwise root selection and allows the fitted polynomial to be used as a numerical Stieltjes transform even for the high-degree, multi-scale curves used in our experiments; details are given in Appendix E.1.

3.1 Free Decompression on Algebraic Spectral Curves

Since the R-transform that underpins free decompression satisfies an algebraic relation whenever Assumption 1 holds, the Stieltjes transform after free decompression also satisfies an algebraic relation. This is stated in our main Theorem, proved as Theorem F.4 in Appendix F.4.

Theorem 1. *Let m be a Stieltjes transform satisfying Assumption 1, $\tau \geq 1$ be a decompression ratio, and m_τ be the free decompression of m by the factor τ. Then every point $(z, m_\tau(z)) \in \mathbb{C}^2$ satisfies*

$$P\left(z + \frac{\tau - 1}{\tau m_\tau}, \tau m_\tau\right) = 0. \tag{4}$$

Especially noteworthy here is the lack of any further assumptions on the nature of m or on the explicit construction of the R-transform. Assumption 1 naturally allows free decompression to be viewed through a different lens, characterizing it as a family of rational reparameterizations of the spectral curve $P(z, m) = 0$. This characterization is convenient, as it circumvents the manual analytic continuation that plagued the PDE-based approach in Ameli et al. [1]. Indeed, for each fixed z away from the singular set, the Fundamental Theorem of Algebra guarantees that $P(z, m) = 0$ has s solutions in m, up to multiplicity. Viewing $m(z)$ as a multivalued algebraic function, these roots represent the sheets of the analytic continuation of the Stieltjes transform. The physical Stieltjes transform is one such sheet, and free decompression transports it along the spectral curve. This geometric viewpoint is developed in Appendix F.6; Figure 2 illustrates the sheet structure and the gluing of branches in the Marchenko–Pastur case.

Evolving the Physical Branch. The evolved algebraic relation in Theorem 1, and its polynomial form in Corollary F.5, gives all candidate roots of $m_\tau(z)$, but it does not identify the physical branch.

To select the physical root, we track the known physical sheet at $\tau = 1$ along the free-decompression flow. For a fixed query point z, we work in the original spectral-curve coordinates (ζ, y), where $P(\zeta, y) = 0$, and impose the characteristic relation

$$\zeta - (\tau - 1)y^{-1} - z = 0.$$

Equivalently, (ζ, y) is determined by the coupled system

$$F_1(\zeta, y) \coloneqq P(\zeta, y) = 0, \qquad F_2(\zeta, y; \tau, z) \coloneqq \zeta - (\tau - 1)y^{-1} - z = 0.$$

At $\tau = 1$, this system reduces to $\zeta = z$ and $y = m_0(z)$, where $m_0(z)$ is already known from the physical-branch selection procedure above. Differentiating the system with respect to τ, while keeping z fixed, gives the tangent equation

$$\begin{bmatrix} \partial_\zeta P(\zeta, y) & \partial_y P(\zeta, y) \\ 1 & (\tau - 1)y^{-2} \end{bmatrix} \begin{bmatrix} \dot{\zeta} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 \\ y^{-1} \end{bmatrix}.$$

This ordinary differential equation on the spectral curve provides the predictor direction for transporting the physical sheet from one decompression ratio to the next. The predicted point is then corrected by Newton projection onto the nonlinear system $(F_1, F_2) = (0, 0)$, with step subdivision when the local geometry becomes stiff. After correction, the decompressed Stieltjes value is recovered as $m_\tau(z) = \tau^{-1} y$. The complete geometric predictor–corrector procedure is given in Appendix F.6, while Algorithm 1 provides a high-level overview of the end-to-end density decompression workflow.

Algorithm 1: Algebraic free decompression from a random principal submatrix

Input : Eigenvalues $\{\lambda_i\}_{i=1}^{n_0}$ of an accessible random principal submatrix;
Grid of decompression ratios $1 = \tau_0 < \tau_1 < \cdots < \tau_T$;
Polynomial degrees (d_z, s) and optional submatrix spectral moments $\{\mu_p\}_{p=0}^{r}$;
Spectral grid $\{x_j\}_{j=1}^{N_x}$ and small imaginary offsets $0 < \delta_1 < \cdots < \delta_L$.

Output : Target density $\hat{\rho}_{\tau_T}$, optionally the path $\{\hat{\rho}_{\tau_\ell}\}_{\ell=0}^{T}$.

Construct contours Γ around the bulks. Sample $\hat{m}_1(z) \leftarrow \frac{1}{n_0} \sum_{i=1}^{n_0} (\lambda_i - z)^{-1}$ for $z \in \Gamma$.
Fit $\hat{P}(z, m) = 0$ to samples with a stabilized basis and moment constraints (Algorithm D.1).
foreach $z = x_j + \mathrm{i}\delta_k$ **do**
    Choose an anchor $z_a \in \mathbb{C}^+$ and the root of $\hat{P}(z_a, m) = 0$ nearest $\hat{m}_1(z_a)$; continue it geometrically along the path from z_a to z to obtain $m_1(z)$ (Algorithm E.1).
    Initialize $(\zeta, y) \leftarrow (z, m_1(z))$ and $m_{\tau_0} \leftarrow m_1(z)$.
    for $\ell = 1, \ldots, T$ **do**
        Evolve (ζ, y) from $\tau_{\ell-1}$ to τ_ℓ on $\hat{P}(\zeta, y) = 0$ with $z = \zeta - (\tau - 1)y^{-1}$, via tangent prediction, damped Newton correction, and adaptive subdivision (Algorithm F.1).
        Set $m_{\tau_\ell}(z) \leftarrow \tau_\ell^{-1} y$.
foreach (τ_ℓ, x_j) **do**
    Extrapolate $\{\frac{1}{\pi} \Im\, m_{\tau_\ell}(x_j + \mathrm{i}\delta_k)\}_{k=1}^{L}$ to $\delta = 0$, obtaining $\hat{\rho}_{\tau_\ell}(x_j)$ (Appendix E.2).
return $\hat{\rho}_{\tau_T}$, optionally $\{\hat{\rho}_{\tau_\ell}\}_{\ell=0}^{T}$

3.2 Evolution of Spectral Quantities

The algebraic framework does more than decompress the full spectral density: it also lets us evolve spectral quantities directly, without computing the density itself. These include atom weights, spectral moments, bulk edges, and gaps between bulks. We can track how these quantities change with matrix size and even identify where and when bulks split.

Atoms. An atom of a distribution at a point $x_\circ$ corresponds to a simple pole of the Stieltjes transform $m(z)$ at $z = x_\circ$, whose weight is given in terms of its residue $w = \lim_{z \to x_\circ} (x_\circ - z) m(z)$. Under Assumption 1, writing $P(z, m) = \sum_{i=0}^{s} a_i(z) m^i$, a necessary condition for the existence of a pole at finite real $x_\circ$ is that $x_\circ$ is a root of the leading coefficient, i.e., $a_s(x_\circ) = 0$. When $x_\circ$ is

a simple root of a_s, the residue gives $w(x_\circ) = a_{s-1}(x_\circ)/a_s'(x_\circ)$. Under free decompression, the atom's location stays fixed, and we can compute how this weight evolves with τ. This is depicted in Figure 1, showing the proportion of probability mass occupied by the atomic component as the matrix decompresses. The following proposition gives the precise dynamics, with proof in Appendix G.4.

Proposition 2. *Let $x_\circ \in \mathbb{R}$ be an atom of $m(z)$ with initial mass $w_\circ \in (0, 1]$. Under the free decompression flow, the atom location remains fixed at $x_\circ(\tau) = x_\circ$, while its weight evolves by $w(\tau) = 1 - \tau^{-1}(1 - w_\circ)$.*

Spectral Edges. We can also compute how the boundaries of spectral bulks evolve under free decompression, including when bulks split. This turns out to be a simpler task than accurately estimating the entire spectral density, and can help diagnose and correct poor density approximations. We assume that the spectral distribution consists of finitely many atoms and absolutely continuous bulk components. Writing the support of the latter as $I(\tau) = \bigcup_{j=1}^{k} I_j(\tau)$, where $I_j(\tau) = [a_j(\tau), b_j(\tau)]$, the spectral edges are the endpoints $\{a_j(\tau), b_j(\tau)\}_{j=1}^{k}$. Under Assumption 1, these points have a precise characterization in terms of the *branch points* of the decompressed spectral curve. At such points, different sheets coalesce, satisfying

$$P(\zeta, y) = 0, \qquad \text{and} \qquad y^2 \frac{\partial P}{\partial y}(\zeta, y) - (\tau - 1)\frac{\partial P}{\partial \zeta}(\zeta, y) = 0, \qquad \zeta \in \mathbb{R},$$

where y belongs to the physical branch, as derived in Proposition G.1 in Appendix G.1. To identify which candidates are spectral edges, we use continuation from known points on the physical branch, similar to the method described in Appendix F.6. At $\tau = 1$, we identify roots on the physical branch through anchored continuation, and for $\tau > 1$ we use the roots identified at the previous time-step. An algorithmic description of a procedure for finding the spectral edges is given in Algorithm G.1. When a spectral edge approaches zero, effects of finite matrix size can become significant for near-zero eigenvalues. We investigate this phenomenon in Appendix H, deriving a Tracy–Widom-type scaling regime from a finite-size spectral-curve ansatz.

Merging and Splitting Bulks. As distributions evolve under free decompression, a connected bulk may split into separate bulks. Viewed in reverse, this is the merger of neighboring bulks under free compression. We can identify where and when these transitions occur. Geometrically, regular transitions correspond to second-order critical points of the decompressed spectral projection, satisfying

$$P(\zeta_*, y_*) = 0, \qquad y_*^2\, \partial_y P - (\tau_* - 1)\, \partial_\zeta P = 0, \qquad \text{and}$$
$$y_* \Big((\partial_{\zeta\zeta} P)(\partial_y P)^2 - 2(\partial_{\zeta y} P)(\partial_\zeta P)(\partial_y P) + (\partial_{yy} P)(\partial_\zeta P)^2 \Big) + 2(\partial_\zeta P)^2 (\partial_y P) = 0.$$

These relations are derived in Proposition G.2 in Appendix G.2. We also derive equations for the evolution of bulk intervals and the gaps between them in Appendix G.3.

Moments. We can also evolve the spectral moments $\mu_n(\tau) = \int x^n \nu_\tau(\mathrm{d}x)$ directly from their initial values, without computing the density. Starting from the free-decompression PDE of Ameli et al. [1], we derive a triangular system of ODEs for these moments (Lemma G.6). Solving this system gives $\mu_n(\tau) = \sum_{k=0}^{n-1} \kappa_{n,k} \tau^k$, where the coefficients $\kappa_{n,k}$ depend only on the first n initial moments and are computed recursively in Proposition G.7. Notably, this evolution does not require Assumption 1. For algebraic distributions, Proposition C.2 also provides an efficient $\mathcal{O}(n^2)$ recurrence for computing the first n moments directly from P, without root selection.

4 Empirical Results

We now consider three examples that demonstrate the ability to decompress distributions of interest in ML. First, we consider a compound free Poisson law containing an atom and bulk regions that split under decompression. Then, we consider the Hessian of a two-layer neural-network autoencoder at initialization, related to the Pennington–Bahri model of limiting Hessian spectra at critical points of the MSE loss [40]. Finally, we consider the expected activation matrix of a random-features diffusion model. Further experiments with the free Lévy and Pennington–Bahri models are presented in Appendices I.2 and I.3.

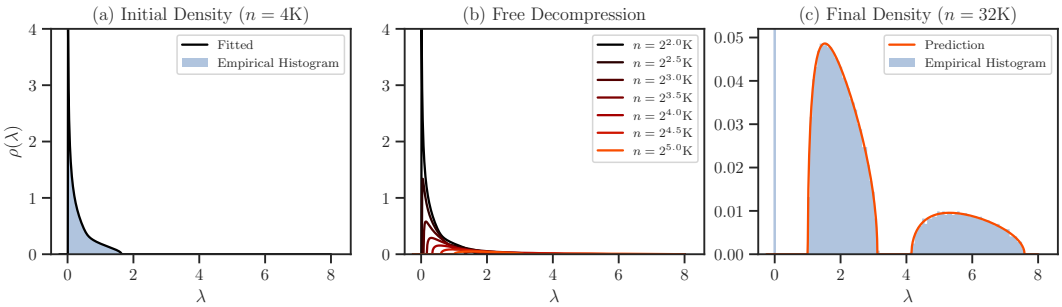

Figure 3: Example of free decompression for recovering the ESD of a $32{,}000 \times 32{,}000$ matrix, generated from the compound free Poisson law described in Section 4. (a) ESD of the sampled 4000×4000 submatrix (histogram) and the fitted density (solid curve). (b) Free decompression flow over n. (c) Final predicted ESD at size $n = 32{,}000$, compared with the ESD of the full matrix.

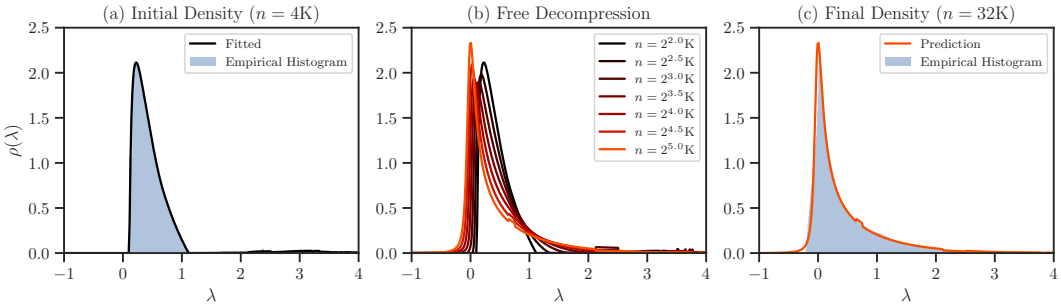

Figure 4: Example of free decompression for recovering the ESD of a $32{,}768 \times 32{,}768$ matrix, generated from the neural network Hessian described in Section 4. (a) ESD of the sampled 4096×4096 submatrix (histogram) and the fitted density (solid curve). (b) Free decompression flow from n_0 to n. (c) Final predicted ESD at size $n = 32{,}768$, compared with the ESD of the original matrix.

Compound Free Poisson. Our first example belongs to the family of compound free Poisson distributions (see Appendix I.1 for details). For aspect ratios $0 < \lambda \leq 1$, this law can be expressed as $\big((1-\lambda)\delta_0 + \lambda\nu\big) \boxtimes \mu_{\mathrm{MP}_1}$, where ν is the jump distribution. In our example, we take $\nu(\mathrm{d}x) = \frac{3}{4}\delta_2(\mathrm{d}x) + \frac{1}{4}\delta_{5.5}(\mathrm{d}x)$ and $\lambda = 0.1$. The resulting distribution contains an atom (spike) at zero and two absolutely continuous bulk regions. We generate a $32{,}000 \times 32{,}000$ matrix from this model and sample a 4000×4000 principal submatrix as initial data. At this smaller size, the fitted spectral density consists of a single bulk. Free decompression from this submatrix is depicted in Figure 3, where the bulk splits into two as the matrix size increases. This example clearly demonstrates both the atom dynamics and bulk-splitting behavior described in Section 3.2, as shown in Figure 1. We further demonstrate free decompression for the more general free Lévy law in Appendix I.2, where a single spectral bulk splits into three.

Neural Network Hessian. Our second example is the Hessian matrix of the ℓ^2-loss of a two-layer neural-network autoencoder with CIFAR-10 data, evaluated at initialization. This is related to the free-probabilistic Hessian approximation of Pennington & Bahri [40], who studied a similar but smaller autoencoder at critical points of the MSE loss. They analyzed the two components of the Hessian $\mathbf{H} = \mathbf{H}_0 + \mathbf{H}_1$ separately, where $\mathbf{H}_0$ is the squared-Jacobian component and $\mathbf{H}_1$ is the network Hessian term. We examine their theoretical model separately in Appendix I.3, while details of our network architecture, data, and Hessian construction are given in Appendix I.4.

For this experiment, the full $32{,}768 \times 32{,}768$ matrix was formed, and a 4096×4096 submatrix was sampled to form the initial empirical spectral density. After fitting the bivariate polynomial, we applied free decompression to recover the full-matrix spectral density, as shown in Figure 4. In Appendix I.4, we use the decompressed distributions to accurately track the normalized Hessian index (the fraction of negative eigenvalues) across matrix sizes. We remark that while the full Hessian was computed as a baseline for our experiments, subsamples of Hessian matrices too large to practically compute can be readily obtained by masking using standard automatic differentiation tools.

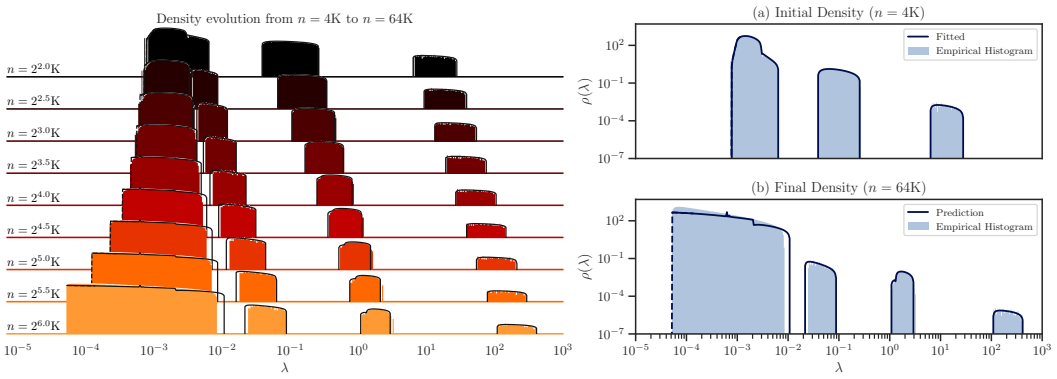

Figure 5: Free decompression for recovering the ESD of an $n \times n$ activation matrix $\mathbf{A}$ of a random-features diffusion model with $n = 64{,}000$ random feature weights. *Left:* ridge plot of the FD density evolution on log scale (solid lines), compared with empirical spectral distributions of sampled submatrices (colored histograms). *Right:* (a) ESD of the sampled $n_0 \times n_0$ matrix $\mathbf{A}_0$, with $n_0 = 4000$, together with the fitted density; and (b) final predicted ESD at $n = 64{,}000$, compared with the ESD of the original matrix $\mathbf{A}$. The y-axis is logarithmic in all panels; the leftmost narrow bulk is atom-like and stands *ten billion* times higher than the right bulk.

Diffusion Model. Our third example considers the expected activation matrix of a diffusion model studied in George et al. [18]; see Appendix I.5 for model and matrix details. Its large-scale limit, studied in Bonnaire et al. [10] in connection with memorization–generalization transitions, has a highly multi-scale spectrum with several bulks, including a narrow atom-like bulk away from zero, making this a stringent log-scale test case for branch selection and density recovery.

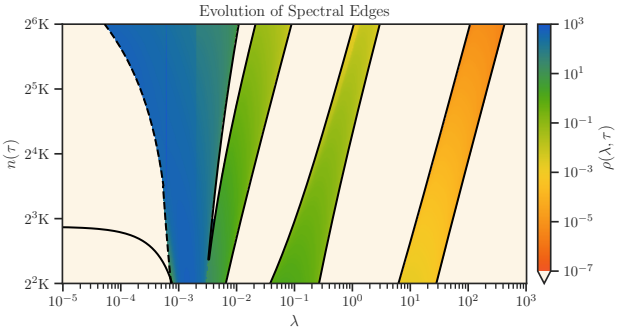

Figure 6: Evolution of the diffusion-model spectral density (colors, on a logarithmic scale), with independently evolved spectral edges (solid black curves). The edges closely follow the bulk boundaries and identify a cusp where the leftmost bulk splits. The dashed curve shows the finite-size estimate of the far-left edge, whose empirical evolution differs from the FD prediction near zero (Appendix H). The blue bulk exhibits atom-like behavior.

To perform free decompression, we formed a $64{,}000 \times 64{,}000$ matrix and sampled a 4000×4000 submatrix. The eigenvalues of the subsampled matrices were pooled to obtain a more accurate initial ESD. Figure 5 (left) shows ridge plots of the FD evolution up to 64K, overlaid with empirical spectra from subsamples at each scale. The right panels isolate the initial and final densities on the same vertical scale, showing that the final FD prediction matches the empirical spectrum across the full log-scale range.

Despite the enormous range in both eigenvalues and spectral density, a single fitted polynomial captures and evolves all the bulks. The narrow leftmost bulk is particularly interesting: although it resembles a smeared-out atom, we fit and evolve it as a continuous component alongside the other bulks, without modeling it separately as an atom. Remarkably, its center remains nearly fixed and its mass closely follows the atomic evolution law derived in Section 3.2, as quantified in Table I.3.

We also compute the spectral edges independently of the density evolution. In Figure 6, these edges (solid curves) closely follow the boundaries of the decompressed density (colored regions) and identify where the leftmost bulk splits. The left edges of the non-atomic bulks also matter for ML, as Bonnaire et al. [10] relates them to distinct dynamical timescales of generalization and memorization during diffusion-model training.

The far-left edge of the narrow atom-like bulk presents a different phenomenon. Its empirical location is the smallest eigenvalue of a finite matrix, which at microscopic scales near zero does not follow the macroscopic FD prediction. We investigated this discrepancy further and found a Tracy–Widom-type scaling regime at intermediate matrix sizes (Appendix H).

Table 1: Comparison of direct eigendecomposition with free decompression (FD) for the diffusion-model random-feature matrix. The Direct column reports the process time for computing the eigenvalues at each matrix size. In the FD column, the first term is the time to compute eigenvalues of the initial 2^2K-sized submatrix, and the second term is the FD evolution time. Distributional distances compare the empirical spectrum with the FD reconstruction, with W_1/L denoting the Wasserstein-1 distance on log-eigenvalues normalized by the log-width L of the plotted spectral range, and MMD denoting the Gaussian-kernel maximum mean discrepancy on log-eigenvalues. Moment errors are relative errors of the first moment and standard deviation. Log edge errors compare empirically detected spectral edges with FD-evolved edges on the log-λ axis, excluding the unstable left edge of the spike-like leftmost bulk. All reported errors are multiplied by 100 and shown as percentages.

Size	Process Time (sec)		Distributional Distance		Moments Rel. Error		Log Edges Error	
n	Direct	FD (ours)	W_1/L	MMD	$\Delta\mu_1/\mu_1$	$\Delta\sigma/\sigma$	Mean	Max
2^2K	32.2	$32.2 + 0.0$	0.01%	0.01%	0.01%	0.01%	0.1%	0.2%
2^3K	431.43	$32.2 + 2.3$	0.22%	0.08%	0.02%	0.03%	0.2%	0.5%
2^4K	3534.5	$32.2 + 2.9$	0.43%	0.12%	0.01%	0.00%	0.2%	0.2%
2^5K	27056.2	$32.2 + 4.3$	0.70%	0.15%	0.05%	0.08%	0.4%	0.8%
2^6K	192447.7	$32.2 + 6.8$	2.19%	0.53%	0.08%	0.09%	0.4%	1.2%

Table 2: Comparison of free decompression (FD) and stochastic Lanczos quadrature (SLQ) for the 64K diffusion-model matrix, on the same hardware. SLQ uses matrix–vector products with the full target, while FD uses only smaller principal submatrices. SLQ ($p \times \ell$) denotes p independent Rademacher probes with ℓ Lanczos steps each. Distributional errors use the same metrics as Table 1.

Method	Observed input	Time (sec)	Target matvecs	W_1/L	MMD
FD	4K submatrix	39.0	0	2.19%	0.53%
SLQ (8×60)	Full 64K target	39.1	480	5.79%	11.52%
SLQ (16×200)	Full 64K target	121.3	3200	2.69%	0.47%

We also compare FD with stochastic Lanczos quadrature (SLQ) on the same diffusion-model matrix (Table 2). The two methods operate under different information requirements: SLQ accesses the full 64K target through matrix–vector products, whereas FD observes only 4K principal submatrices. Even with this difference, FD achieves lower errors in both metrics at matched runtime. With over three times the runtime, SLQ achieves comparable accuracy. Moreover, FD recovers the entire spectral evolution across matrix sizes, including densities, edges, gaps, atoms, and bulk transitions, rather than only the spectrum of a fixed target.

5 Conclusions

We have developed a new method of free decompression that generalizes previous work by representing the Stieltjes transform of the spectral density through a bivariate polynomial of arbitrary degree. Under this algebraic ansatz, free decompression becomes an evolution on spectral curves, allowing us to track the physical branch, evolve the density, and derive equations for spectral quantities such as atoms, moments, and bulk edges. This enables FD to handle substantially more challenging spectra than previous approaches, including multi-modal, multi-bulk, and multi-scale distributions arising in realistic ML matrices.

Limitations. While the current method expands the range of matrices accessible to free decompression, some important regimes remain challenging. In particular, heavy-tailed spectral distributions remain elusive, as their unbounded support and possibly divergent moments lie beyond our current setting. Moreover, the underlying FD law is justified under asymptotic freeness between the matrix and the sampling projection, which need not hold for arbitrary ML matrices. For empirical spectra, the fitted polynomial serves as an algebraic surrogate, and obtaining a reliable fit and tracking its physical branch can remain numerically challenging. We hope these results provide a foundation for further development of methods for large-scale spectral inference.

Acknowledgments

LH is supported by the Australian Research Council through a Discovery Early Career Researcher Award (DE240100144). MWM acknowledges partial support from DARPA, DOE, NSF, and ONR. This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research's Applied Mathematics Competitive Portfolios program under Contract No. AC02-05CH11231.

Competing Interests

The authors declare no competing interests.

References

[1] Ameli, S., van der Heide, C., Hodgkinson, L., & Mahoney, M. (2025a). Spectral estimation with free decompression. In *Advances in Neural Information Processing Systems*, volume 38 (pp. 1888–1929).

[2] Ameli, S., van der Heide, C., Hodgkinson, L., Roosta, F., & Mahoney, M. W. (2025b). Determinant estimation under memory constraints and neural scaling laws. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research* (pp. 1367–1400).

[3] Anshelevich, M. (2008). Orthogonal polynomials with a resolvent-type generating function. *Transactions of the American Mathematical Society*, 360(8), 4125–4143.

[4] Benaych-Georges, F. (2005). Classical and free infinitely divisible distributions and random matrices. *The Annals of Probability*, 33(3), 1134–1170.

[5] Bercovici, H. & Voiculescu, D. (1993). Free convolution of measures with unbounded support. *Indiana University Mathematics Journal*, 42(3), 733–773.

[6] Biroli, G., Bonnaire, T., de Bortoli, V., & Mézard, M. (2024). Dynamical regimes of diffusion models. *Nature Communications*, 15(1), 9957.

[7] Blaizot, J.-P., Grela, J., Nowak, M., & Warchoł, P. (2015). Diffusion in the space of complex Hermitian matrices: microscopic properties of the averaged characteristic polynomial and the averaged inverse characteristic polynomial. *Acta Physica Polonica. B*, 46(9), 1801–1823.

[8] Blaizot, J.-P. & Nowak, M. A. (2010). Universal shocks in random matrix theory. *Phys. Rev. E*, 82(5), 051115.

[9] Blaizot, J.-P., Nowak, M. A., & Warchoł, P. (2013). Universal shocks in the Wishart random-matrix ensemble. *Phys. Rev. E*, 87(5), 052134.

[10] Bonnaire, T., Urfin, R., Biroli, G., & Mézard, M. (2025). Why diffusion models don't memorize: The role of implicit dynamical regularization in training. In *Advances in Neural Information Processing Systems*, volume 38 (pp. 141266–141286).

[11] Burda, Z., Janik, R. A., Jurkiewicz, J., Nowak, M. A., Papp, G., & Zahed, I. (2002). Free random Lévy matrices. *Phys. Rev. E*, 65(2), 021106.

[12] Chen, J. & Olver, S. (2026). Computing inverses of Stieltjes transforms of probability measures. *Mathematics of Computation*.

[13] Collins, B. & Śniady, P. (2006). Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3), 773–795.

[14] Couillet, R. & Liao, Z. (2022). *Random Matrix Methods for Machine Learning*. Cambridge University Press.

[15] DLMF (2026). NIST Digital Library of Mathematical Functions. https://dlmf.nist.gov/. Olver, F. W. J., Olde Daalhuis, A. B., Lozier, D. W., Schneider, B. I., Boisvert, R. F., Clark, C. W., Miller, B. R., Saunders, B. V., Cohl, H. S., and McClain, M. A., eds. Accessed 3 May 2026.

[16] El Karoui, N. (2010). The spectrum of kernel random matrices. *The Annals of Statistics*, 38(1), 1–50.

[17] Erdős, L., Krüger, T., & Schröder, D. (2020). Cusp universality for random matrices I: Local law and the complex Hermitian case. *Commun. Math. Phys.*, 378(2), 1203–1278.

[18] George, A. J., Veiga, R., & Macris, N. (2026). Denoising score matching with random features: Insights on diffusion models from precise learning curves. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research* (pp. 1405–1413).

[19] Gibbs, A. L. & Su, F. E. (2002). On choosing and bounding probability metrics. *International Statistical Review*, 70(3), 419–435.

[20] Guionnet, A. (2009). *Large random matrices: lectures on macroscopic asymptotics*. Springer-Verlag, Berlin.

[21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

[22] Hodgkinson, L., van der Heide, C., Roosta, F., & Mahoney, M. W. (2023a). Monotonicity and double descent in uncertainty estimation with Gaussian processes. In *Proceedings of the 40th International Conference on Machine Learning* (pp. 13085–13117).

[23] Hodgkinson, L., van der Heide, C., Salomone, R., Roosta, F., & Mahoney, M. W. (2023b). The interpolating information criterion for overparameterized models. *arXiv preprint arXiv:2307.07785v1*.

[24] Hodgkinson, L., van der Heide, C., Salomone, R., Roosta, F., & Mahoney, M. W. (2023c). A PAC-Bayesian perspective on the interpolating information criterion. In *NeurIPS 2023 Workshop on Mathematics of Modern Machine Learning*.

[25] Hodgkinson, L., Wang, Z., & Mahoney, M. W. (2025). Models of heavy-tailed mechanistic universality. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research* (pp. 23290–23329).

[26] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*, volume 103. Springer.

[27] Kesten, H. (1959). Symmetric random walks on groups. *Transactions of the American Mathematical Society*, 92(2), 336–354.

[28] Krizhevsky, A. (2009). *Learning Multiple Layers of Features from Tiny Images*. Technical report, University of Toronto.

[29] Liao, Z., Couillet, R., & Mahoney, M. W. (2021). A random matrix analysis of random Fourier features: beyond the Gaussian kernel, a precise phase transition, and the corresponding double descent. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(12), 124006.

[30] Martin, C. H. & Mahoney, M. W. (2021). Implicit self-regularization in deep neural networks: Evidence from random matrix theory and implications for learning. *Journal of Machine Learning Research*, 22(165), 1–73.

[31] Marčenko, V. A. & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4), 457–483.

[32] McKay, B. D. (1981). The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40, 203–216.

[33] Meixner, J. (1934). Orthogonale Polynomsysteme mit einer besonderen Gestalt der erzeugenden Funktion. *Journal of The London Mathematical Society*, s1-9(1), 6–13.

[34] Muskhelishvili, N. (1992). *Singular Integral Equations: Boundary Problems of Function Theory and Their Application to Mathematical Physics*. Dover Publications.

[35] Nica, A. & Speicher, R. (1996). On the multiplication of free N-tuples of noncommutative random variables. *Amer. J. Math.*, 118(4), 799–837.

[36] Nica, A. & Speicher, R. (2006). *Lectures on the Combinatorics of Free Probability*. London Mathematical Society Lecture Note Series. Cambridge University Press.

[37] Olver, S. & Nadakuditi, R. R. (2012). Numerical computation of convolutions in free probability theory. *arXiv preprint arXiv:1203.1958*.

[38] Panaretos, V. M. & Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and its Application*, 6(1), 405–431.

[39] Pastur, L. & Shcherbina, M. (2011). *Eigenvalue Distribution of Large Random Matrices*. Mathematical surveys and monographs. American Mathematical Society.

[40] Pennington, J. & Bahri, Y. (2017). Geometry of neural network loss surfaces via random matrix theory. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research* (pp. 2798–2806).

[41] Pennington, J., Schoenholz, S., & Ganguli, S. (2018). The emergence of spectral universality in deep networks. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research* (pp. 1924–1932).

[42] Pennington, J. & Worah, P. (2017). Nonlinear random matrix theory for deep learning. In *Advances in Neural Information Processing Systems*, volume 30: Curran Associates, Inc.

[43] Rao, N. R. & Edelman, A. (2008). The polynomial method for random matrices. *Foundations of Computational Mathematics*, 8(6), 649–702.

[44] Rudelson, M. & Vershynin, R. (2009). Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 62(12), 1707–1739.

[45] Saitoh, N. & Yoshida, H. (2001). The infinite divisibility and orthogonal polynomials with a constant recursion formula in free probability theory. *Probab. Math. Statist.*, 21, 159–170.

[46] Senouf, D., Caflisch, R., & Ercolani, N. (1996). Pole dynamics and oscillations for the complex Burgers equation in the small-dispersion limit. *Nonlinearity*, 9(6), 1671–1702.

[47] Shlyakhtenko, D. & Tao, T. (2022). Fractional free convolution powers. *Indiana University Mathematics Journal*, 71(6).

[48] Speicher, R. (2015). Free probability theory. In *The Oxford Handbook of Random Matrix Theory*. Oxford University Press.

[49] Stein, E. M. & Shakarchi, R. (2003). *Fourier Analysis: An Introduction*, volume 1 of *Princeton Lectures in Analysis*. Princeton, NJ: Princeton University Press.

[50] Sturmfels, B. (2002). *Solving Systems of Polynomial Equations*, volume 97 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society.

[51] Trefethen, L. N. (2019). *Approximation Theory and Approximation Practice*. Philadelphia, PA: Society for Industrial and Applied Mathematics, extended edition.

[52] Trefethen, L. N. (2023). Numerical analytic continuation. *Japan Journal of Industrial and Applied Mathematics*, 40(3), 1587–1636.

[53] Ventura, E., Achilli, B., Silvestri, G., Lucibello, C., & Ambrogioni, L. (2025). Manifolds, random matrices and spectral gaps: The geometric phases of generative diffusion. In *The Thirteenth International Conference on Learning Representations*.

[54] Voiculescu, D. (1991). Limit laws for random matrices and free products. *Inventiones mathematicae*, 104(1), 201–220.

[55] Wachter, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *The Annals of Probability*, 6(1), 1–18.

[56] Wegert, E. (2012). *Visual Complex Functions: An Introduction with Phase Portraits*. Springer Basel.

[57] Wei, A., Hu, W., & Steinhardt, J. (2022). More than a toy: Random matrix models predict how real-world neural representations generalize. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research* (pp. 23549–23588).

[58] Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3), 548–564.

Appendices

Contents

A Notation and Conventions **16**

B Background **17**

B.1 Stieltjes Transform and Boundary Values 17

B.2 Analytic Continuation Beyond the Physical Branch 18

C Algebraic Stieltjes Transforms **18**

C.1 Moments from the Algebraic Relation 19

D Fitting the Spectral Curve **21**

D.1 Constraints at Infinity from Moments 22

D.2 Constraints from Branch Points 23

D.3 Numerical Reconstruction from Samples 24

E Computing the Stieltjes Transform and Density **24**

E.1 Physical Root Selection for the Stieltjes Transform 25

E.2 Density Recovery and δ-Extrapolation 27

F Free Decompression on Algebraic Spectral Curves **29**

F.1 The Nica–Speicher Theorem 30

F.2 Infinite Decompressibility 33

F.3 The PDE Approach and Challenges 34

F.4 Formal Free Decompression along Spectral Curves 35

F.5 Polynomial Evolution under Free Decompression 35

F.6 Free Decompression via Geometric Continuation 37

G Evolution of Spectral Quantities **39**

G.1 Evolving Spectral Edges 39

G.2 Spectral Cusps and The Mechanism of Splitting Bulks 41

G.3 Dynamics of Gap and Support Intervals 44

G.4 Dynamics of Atoms 45

G.5 Evolution of Moments 46

H Finite-Size Correction to the Spectral Curve **48**

I Random Matrix Models **52**

I.1 Compound Free Poisson Law 52

I.2 Free Lévy Law 53

I.3 Pennington–Bahri Model 58

I.4 Hessian of Two-Layer Network 58
I.5 Details of the Diffusion Model Example 60

J Implementation and Reproducibility Guide **62**

Appendix A Notation and Conventions

Nomenclature. We use boldface lowercase letters for vectors, boldface upper case letters for matrices, and normal face letters for scalars, including the components of vectors and matrices. Table A.1 summarizes the main symbols and notations used throughout the paper, organized by context.

Table A.1: Common notations used throughout the manuscript.

Context	Symbol	Description
Matrices & Models	$\mathbf{A}, \mathbf{A}_n$	General $n \times n$ matrix (often symmetric or Hermitian)
	$\mathbf{A}_n^{(k)}$	$k \times k$ principal submatrix of $\mathbf{A}_n$
	$\mathbf{I}$	The identity matrix
	λ_i	The i-th eigenvalue of a matrix
	$\operatorname{trace}(\cdot)$	Standard trace of a matrix
	$\operatorname{tr}_n(\cdot)$	Normalized trace, $n^{-1}\operatorname{trace}(\cdot)$
	$\mathbf{Q}_n$	Haar-distributed random orthogonal matrix
	$\mathbf{H}, \mathbf{J}$	Hessian matrix (and subcomponents) and Jacobian matrix
	$\mathbf{\Sigma}$	Population covariance matrix
	$\mathbf{W}, \mathbf{X}, \mathbf{Z}$	Random matrices with i.i.d. entries (e.g., Wigner, Gaussian)
	$\mathbf{U}$	Expected diffusion activation matrix (random features)
	λ, c	Aspect ratios (p/n) or intensities/rates for random matrix models
Free Probability	ν	Spectral probability measure on $\mathbb{R}$, possibly including atoms
	ρ	Density of the absolutely continuous part of ν, $\rho = \mathrm{d}\nu_{\mathrm{ac}}/\mathrm{d}x$, $\mathbb{R} \to \mathbb{R}^+$
	I	Compact support of the density ρ, $I = \operatorname{supp}(\rho) \subsetneq \mathbb{R}$
	m	Stieltjes transform of measure, $m(z) = \int_{\mathbb{R}} \frac{\nu(\mathrm{d}y)}{y-z}$, $\mathbb{C} \setminus \operatorname{supp}(\nu) \to \mathbb{C}$
	ω	Functional inverse of the Stieltjes transform, $\omega(m(z)) = z$
	$\mathcal{H}\{\nu\}$	Hilbert transform of measure, $\mathcal{H}\{\nu\}(x) = \frac{1}{\pi}\,\mathrm{p.v.}\int_{\mathbb{R}} \frac{\nu(\mathrm{d}y)}{x-y}$, $\mathbb{R} \to \mathbb{R}$
	R	Voiculescu's R-transform, $R(z) = \omega(-z) - \frac{1}{z}$ with $\omega(m(z)) = z$
	S	S-transform, $S(z) = \frac{1+z}{z}\psi^{-1}(z)$, where $\psi(z) = \int_{\mathbb{R}} \frac{zy}{1-zy}\,\nu(\mathrm{d}y)$
	$\nu^{\boxplus s}$	s-fold free additive convolution power of ν
	$D_c\nu$	Dilation of measure ν by c, i.e., push-forward under $x \mapsto cx$
	$M(z)$	Truncated moment-generating function
	$\chi(z)$	Functional inverse of the truncated moment-generating function
	$r_k(\nu)$	k-th free cumulant of the measure ν
	$\mu_p, \mu_p(t)$	p-th moment of a probability measure (and its time evolution)
	$\boxplus$	Free additive convolution
	$\boxtimes$	Free multiplicative convolution
Algebraic Curve	$P(z, m)$	Algebraic equation for the Stieltjes transform, $P(z, m(z)) = 0$
	$\mathcal{C}$	Complex 1D spectral curve $\mathcal{C} = \{(z, m) \in \mathbb{C}^2 \mid P(z, m) = 0\}$
Free Decompression	n	Size of the target (large) matrix
	n_0	Size of the sampled (small) sub-matrix, $n_0 < n$
	τ, t	Decompression scale, $\tau = \frac{n}{n_0}$, with $t = \log(\tau)$
	ρ_0, m_0	Density and Stieltjes transform at $t = 0$
Notation	$\mathbb{C}^{\pm}$	Upper (lower) half complex plane
	$\Re, \Im$	Real and imaginary part of complex variable
	$\mathrm{p.v.}$	Cauchy principal value

Color code for plots. In our complex-plane plots, colors encode the values of a complex quantity using *domain coloring*: hue represents phase, while brightness represents magnitude; see the plain version in Figure A.1(b). For high-dynamic-range fields, we also use the tiled version shown in Figure A.1(a), in which phase and logarithmic magnitude are periodically reset, following the phase-portrait convention of Wegert [56, Section 2.5].

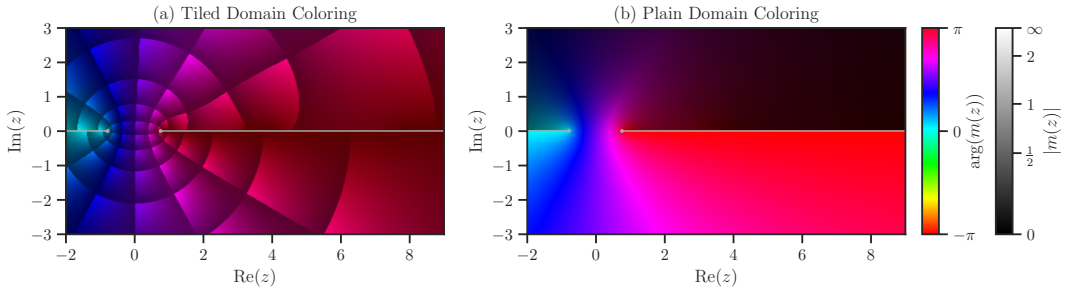

Figure A.1: Domain coloring of a Stieltjes-transform branch for the Free Lévy example in Figure I.3. Right: plain domain coloring, with hue encoding $\arg m(z)$ and brightness encoding $|m(z)|$. Left: tiled domain coloring, where phase and logarithmic magnitude are periodically reset to make high-dynamic-range structure more visible.

Appendix B Background

B.1 Stieltjes Transform and Boundary Values

Let ν be a probability measure on $\mathbb{R}$ with compact support $\operatorname{supp}(\nu) \subset \mathbb{R}$. Let ν_{ac} denote its absolutely continuous part with respect to Lebesgue measure, and let $\rho(x) := \nu_{\mathrm{ac}}(\mathrm{d}x)/\mathrm{d}x$ denote its density, given by the Radon–Nikodym derivative.

The *Stieltjes transform* of ν is the function $m : \mathbb{C} \setminus \operatorname{supp}(\nu) \to \mathbb{C}$ defined by

$$m(z) := \int_{\mathbb{R}} \frac{1}{x - z}\, \nu(\mathrm{d}x), \quad z \in \mathbb{C} \setminus \operatorname{supp}(\nu). \tag{B.1}$$

The Stieltjes transform is a *Herglotz* function: $\Im m(z) > 0$ for $z \in \mathbb{C}^+$. Here, $\mathbb{C}^+ := \{z \in \mathbb{C} \mid \Im z > 0\}$, and $\Im w$ denotes the imaginary part for $w \in \mathbb{C}$. Moreover, m satisfies the *normalization* at infinity: $m(z) = -z^{-1} + O(z^{-2})$ as $|z| \to \infty$.

The function m admits non-tangential boundary values on the real line, denoted by $m^{\pm}(x) := \lim_{y \to 0^{\pm}} m(x+iy)$ when the limits exist. If $\nu(\mathrm{d}x) = \rho(x)\,\mathrm{d}x$ in a neighborhood of x, the Sokhotski–Plemelj formula [34, p. 42] yields

$$m^{\pm}(x) = -\pi \mathcal{H}\{\nu\}(x) \pm i\pi\rho(x),$$

where $\mathcal{H}\{\nu\}(x) := \frac{1}{\pi}\,\mathrm{p.v.} \int_{\mathbb{R}} \frac{\nu(\mathrm{d}t)}{x-t}$ is the *Hilbert transform* of the full measure ν and p.v. denotes the Cauchy's principal value of a singular integral. The density can be recovered from boundary values by

$$\rho(x) = \pi^{-1} \Im m^{+}(x) = -\pi^{-1} \Im m^{-}(x), \tag{B.2}$$

which is the inversion of the Stieltjes transform.

The jump of m across the real axis is

$$m^{+}(x) - m^{-}(x) = 2i\pi\rho(x).$$

In particular, for $x \notin \operatorname{supp}(\nu)$, the boundary values coincide, and m is holomorphic across the real axis at x. In contrast, for $x \in \operatorname{supp}(\rho)$ the boundary values differ, which manifests as a branch cut in the boundary values of m across $\operatorname{supp}(\rho)$. An example of a branch cut is shown in Figure 2(a), illustrated by the white line segment in the center of the plot. In the remainder, we denote the support

of the absolutely continuous part of ν by $I \coloneqq \operatorname{supp}(\rho)$, and we will typically assume that it consists of finitely many disjoint compact intervals,

$$I = \bigcup_{j=1}^{k} I_j, \qquad I_j \coloneqq [a_j, b_j],$$

where $k \geq 1$, and $a_j < b_j < a_{j+1}$. We refer to k as the number of spectrum bulks, which is also the number of branch cuts of m. Also, any atomic component of ν produces isolated poles of m on the real axis rather than branch cuts, and will be treated separately when relevant.

B.2 Analytic Continuation Beyond the Physical Branch

Several applications, and in particular the method of characteristics used in free decompression [1], require evaluating the Stieltjes field $m(z)$ along trajectories in the complex z-plane that cross the cuts I. Since $I \subseteq \operatorname{supp}(\nu)$ and the physical Stieltjes transform $m(z)$ is holomorphic on $\mathbb{C} \setminus \operatorname{supp}(\nu)$, such evaluations require analytic continuation across I. It is therefore natural to regard m as belonging to a larger object: a multi-valued analytic function of z, whose single-valued realizations correspond to branches (or sheets) of a Riemann surface lying over the z-plane.

We denote the Stieltjes transform defined by (B.1) by m_1 and refer to it as the *physical branch*, characterized uniquely by the Herglotz property together with the normalization at infinity. Other branches, when they exist, are denoted by m_i, $i > 1$, and are referred to as *non-physical*.

Crossing a cut exchanges boundary traces of branches. Specifically, for x in the interior of an interval I_j, analytic continuation across I_j identifies the upper trace of one branch with the lower trace of another, and vice versa. Equivalently, two sheets are glued along I_j by matching boundary values with opposite orientations, so that a trajectory approaching I_j from $\mathbb{C}^+$ and crossing the cut may emerge on a different branch on the other side; see Figure 2 for an illustration.

In our setting, the physical branch m_1 is assumed known, either because ν is a model law or because it is estimated from an empirical spectrum and m_1 is computed by (B.1). Analytic continuation of an exact branch along a specified path is unique whenever it exists. However, continuation beyond the physical domain can be path-dependent. In particular, the phrase "the other branch" is generally ambiguous: continuations around different endpoints $\{a_j, b_j\}_{j=1}^{k}$ can lead to distinct sheets, and this ambiguity becomes more pronounced for multi-bulk spectra.

Numerically, recovering these continuations from approximate samples of the physical branch remains severely ill-posed. A convenient way to regularize this problem is to restrict attention to model classes in which the analytic continuation of m has only finitely many branches. This reduces the problem to a finite-sheeted branched covering of the z-plane. We adopt an algebraic model class, which we describe next.

Appendix C Algebraic Stieltjes Transforms

Our structural assumption is that the Stieltjes transform m is *algebraic*: there exists a nonzero bivariate polynomial P such that

$$P(z, m) = 0. \tag{C.1}$$

We write $s \coloneqq \deg_m(P)$ for the degree of P with respect to its second argument. For each z away from the branch locus, the algebraic equation $P(z, m) = 0$ admits s (not necessarily distinct) solutions in m, and the corresponding analytic branches are denoted by $\{m_i(z)\}_{i=1}^{s}$. In this notation, m_1 is the physical branch, singled out among the s branches by the Herglotz property together with the normalization at infinity, while the remaining branches are non-physical.

Branch points occur at those z_* for which two or more roots of P collide. In particular, there exists $m_* \in \mathbb{C}$ such that

$$P(z_*, m_*) = 0, \qquad \partial_m P(z_*, m_*) = 0. \tag{C.2}$$

Eliminating m_* yields the discriminant $\Delta(z) \coloneqq \operatorname{Disc}_m P(z, \cdot) \in \mathbb{R}[z]$, whose zeros provide candidates for the branch locus of the multi-valued function defined implicitly by (C.1). In Stieltjes settings with conjugation symmetry, the branch points z_* on the real line include the spectral edges

$\{a_j, b_j\}_{j=1}^k$ of the cut set I, though the algebraic model may admit additional branch points away from the real axis.

A *branch cut* I is a choice of curves connecting branch points such that each branch m_i becomes single-valued and holomorphic on $\mathbb{C}\setminus I$. Once I is fixed, each branch admits non-tangential boundary values on I, denoted by $m_i^{\pm}$. Across a cut component, analytic continuation can only permute branches. It is therefore convenient to collect branches into a vector,

$$\boldsymbol{m}(z) \coloneqq [m_1(z), m_2(z), \ldots, m_s(z)]^{\mathsf{T}}.$$

For $x \in I_j^{\circ}$, analytic continuation across I_j permutes the boundary traces, hence there exists a permutation matrix $\boldsymbol{\Pi}_j \in \{0,1\}^{s\times s}$ such that

$$\boldsymbol{m}^{+}(x) = \boldsymbol{\Pi}_j\, \boldsymbol{m}^{-}(x), \qquad x \in I_j^{\circ},$$

where $\boldsymbol{m}^{\pm}(x)$ denote non-tangential boundary values taken from $\mathbb{C}^{\pm}$. The collection $\{\boldsymbol{\Pi}_j\}_{j=1}^k$ encodes how the s sheets are glued along the real cuts and provides a finite description of the continuation rules needed to follow trajectories across I.

In Appendix D, we describe the construction of a finite-sheet continuation from samples of the physical branch m_1 by fitting (C.1). Before that, in Appendix C.1, we describe moment-based necessary conditions for the feasibility of such an algebraic representation.

C.1 Moments from the Algebraic Relation

One of the biggest challenges with the algebraic approach to performing free decompression is the identification of the root $m(z)$ in the algebraic relation that corresponds to the principal branch $m_1(z)$ of the Stieltjes transform we seek. This identification can be performed in one of two ways. The first is by homotopy, where one takes advantage of continuity and identifies the correct root at time $t+\delta t$ using knowledge of the correct root at time t. This approach performs well when the roots are distinct, but can fail in regimes where the roots become close or "cross over". The second approach involves the computation of *moments*; while this is less direct, it is arguably more concrete and provides a powerful check to ensure the correct solution is obtained.

Let μ_p denote the p-th moment of the probability measure ν

$$\mu_p \coloneqq \int_{\mathbb{R}} x^p \nu(\mathrm{d}x), \qquad p \in \mathbb{N}_0, \tag{C.3}$$

where $\mu_0 = 1$ corresponds to unit mass normalization. If m_1 denotes the Stieltjes transform of ν, then for sufficiently large $|z|$, m_1 satisfies the Laurent expansion

$$m_1(z) = \int_{\mathbb{R}} \frac{1}{x-z}\nu(\mathrm{d}x) = -\frac{1}{z}\int_{\mathbb{R}} \frac{1}{1-\frac{x}{z}}\nu(\mathrm{d}x) = -\frac{1}{z}\sum_{p=0}^{\infty}\frac{1}{z^p}\int_{\mathbb{R}} x^p\nu(\mathrm{d}x) = -\sum_{p=0}^{\infty}\frac{\mu_p}{z^{p+1}}, \tag{C.4}$$

providing a concrete characterisation of the Stieltjes transform in terms of its moments.

Suppose that a polynomial parameterized by an index set $\mathcal{A} \subset \mathbb{N}_0^2$, $P(z,m) = \sum_{(i,j)\in\mathcal{A}} c_{ij} z^i m^j$, is given. Our objective is to find a solution $m(z)$ satisfying $P(z, m(z)) = 0$ and an expansion of the form $m(z) = \sum_{k=0}^{\infty} \vartheta_k z^{-k-1}$ for all sufficiently large $|z|$. Let $w = 1/z$ and $S(w) = \sum_{k=0}^{\infty}\vartheta_k w^k$ so that $m(z) = wS(w)$. If $m(z)$ corresponds to the principal branch $m_1(z)$ with moments μ_k, then $\vartheta_0 = -1$ and $\mu_k = \vartheta_k/\vartheta_0 = -\vartheta_k$ for each $k = 1, 2, \ldots$.

Lemma C.1. *There exists a function $m(z)$ satisfying $P(z, m(z)) = 0$ and $m(z) = \sum_{k=0}^{\infty}\vartheta_k z^{-k-1}$ for sufficiently large $|z|$ only if ϑ_0 is a root of the polynomial*

$$L(\vartheta) = \sum_{\substack{(i,j)\in\mathcal{A}\\ i-j=e_{\max}}} c_{ij}\vartheta^j, \qquad e_{\max} = \max\{i-j : (i,j)\in\mathcal{A},\ c_{ij}\neq 0\}.$$

Conversely, if $L(\vartheta_0) = 0$ and $L'(\vartheta_0) \neq 0$, there exists a unique such function with prescribed coefficient ϑ_0 near infinity.

Proof. Setting $w = 1/z$ and $m(z) = wS(w)$, the equation $P(z, m(z)) = 0$ becomes

$$F(w, S) := \sum_{(i,j)\in\mathcal{A}} c_{ij} w^{e_{\max}-(i-j)} S^j = 0.$$

Since $F(0, S) = L(S)$, any such expansion must satisfy $L(\vartheta_0) = 0$. Conversely, if $L'(\vartheta_0) \neq 0$, the analytic implicit function theorem gives a unique analytic solution $S(w)$ near $w = 0$, with $S(0) = \vartheta_0$. Therefore, $m(z) = z^{-1}S(1/z)$ admits the required Laurent expansion. □

Example. As an example, consider the polynomial $P(z, m) = \lambda\sigma^2 zm^2 + (z - \sigma^2(1 - \lambda))m + 1$ for the Marchenko–Pastur distribution with scale σ^2 and aspect ratio λ. In this case, $e_{\max} = 0$ and $L(\vartheta) = 1 + \vartheta$, which does indeed have a single root at $\vartheta = -1$.

In many practical cases, a root of P will only *approximate* a Stieltjes transform, and so it may only be the case that $\vartheta_0 \approx -1$. Fortunately, ϑ_0 can be estimated by finding the closest root of $L(\vartheta)$ to -1. As long as the remaining coefficients ϑ_k can be recovered, the "moments" of the corresponding Stieltjes transform can be estimated by $\mu_k \approx \vartheta_k/\vartheta_0$. Recovery of ϑ_k is trickier, but can be achieved using Proposition C.2 under the condition that $L'(\vartheta_0) \neq 0$.

Proposition C.2. *Suppose $L(\vartheta_0) = 0$ and $L'(\vartheta_0) \neq 0$. Then, for $k \geq 1$, the coefficients ϑ_k can be computed using the recurrence relations*

$$d_{j,k} = \begin{cases} \vartheta_0^j & \text{if } j \geq 0,\ k = 0, \\ 0 & \text{if } j = 0,\ k \geq 1, \\ \sum_{l=0}^{k} \vartheta_l d_{j-1,k-l} & \text{otherwise,} \end{cases}$$

$$\tilde{d}_{j,k} = \begin{cases} 0 & \text{if } j = 0,\ k \geq 1, \\ \vartheta_0 \tilde{d}_{j-1,k} + \sum_{\ell=1}^{k-1} \vartheta_\ell d_{j-1,k-\ell} & \text{if } j \geq 1,\ k \geq 1, \end{cases}$$

where $d_{j,q} = 0$ for $q < 0$, and

$$\vartheta_k = -\frac{1}{L'(\vartheta_0)} \left(\sum_{\substack{(i,j)\in\mathcal{A} \\ i-j=e_{\max}}} c_{ij} \tilde{d}_{j,k} + \sum_{p<e_{\max}} \sum_{\substack{(i,j)\in\mathcal{A} \\ i-j=p}} c_{ij} d_{j,k-e_{\max}+p} \right),$$

with

$$L'(\vartheta) = \sum_{\substack{(i,j)\in\mathcal{A} \\ i-j=e_{\max}}} j c_{ij} \vartheta^{j-1}.$$

Proof. Let $w = 1/z$ and $S(w) = \sum_{k=0}^{\infty} \vartheta_k w^k$, so that $m(z) = wS(w)$. Let $d_{j,k}$ denote the coefficient of w^k in $S(w)^j$, so that

$$S(w)^j = \sum_{k=0}^{\infty} d_{j,k} w^k.$$

The Cauchy product immediately gives the stated recurrence for $d_{j,k}$.

To isolate the unknown coefficient ϑ_k, define

$$S_k(w) = \sum_{\ell=0}^{k-1} \vartheta_\ell w^\ell, \qquad k \geq 1.$$

Let $\tilde{d}_{j,k}$ denote the coefficient of w^k in $S_k(w)^j$. Since $S_k(w)$ and $S(w)$ agree through order $k - 1$, the Cauchy product gives, for $j, k \geq 1$,

$$\tilde{d}_{j,k} = \vartheta_0 \tilde{d}_{j-1,k} + \sum_{\ell=1}^{k-1} \vartheta_\ell d_{j-1,k-\ell}.$$

Moreover, $S(w) = S_k(w) + \vartheta_k w^k + O(w^{k+1})$, and hence

$$d_{j,k} = \tilde{d}_{j,k} + j\vartheta_0^{j-1}\vartheta_k.$$

Now, substituting $m(z) = wS(w)$ into $P(z, m(z)) = 0$ and multiplying by $w^{e_{\max}}$, we obtain

$$\sum_{(i,j)\in\mathcal{A}} c_{ij} w^{e_{\max}-(i-j)} S(w)^j = 0.$$

Equating the coefficients of w^k yields

$$\sum_{\substack{(i,j)\in\mathcal{A}\\ i-j=e_{\max}}} c_{ij} d_{j,k} + \sum_{p<e_{\max}} \sum_{\substack{(i,j)\in\mathcal{A}\\ i-j=p}} c_{ij} d_{j,k-e_{\max}+p} = 0,$$

where $d_{j,q} = 0$ for $q < 0$. Substituting the expression for $d_{j,k}$ and using

$$L'(\vartheta_0) = \sum_{\substack{(i,j)\in\mathcal{A}\\ i-j=e_{\max}}} j c_{ij} \vartheta_0^{j-1},$$

gives

$$\vartheta_k L'(\vartheta_0) + \sum_{\substack{(i,j)\in\mathcal{A}\\ i-j=e_{\max}}} c_{ij}\tilde{d}_{j,k} + \sum_{p<e_{\max}} \sum_{\substack{(i,j)\in\mathcal{A}\\ i-j=p}} c_{ij} d_{j,k-e_{\max}+p} = 0.$$

Since $L'(\vartheta_0) \neq 0$, solving for ϑ_k yields the stated recurrence. □

Appendix D Fitting the Spectral Curve

In this section, we describe a general-purpose procedure for reconstructing non-physical branches of an algebraic Stieltjes transform from pointwise evaluations of the physical branch m_1. The basic idea is to fit an implicit algebraic relation $P(z, m) = 0$ as in (C.1) from samples $\{(z_\ell, m_1(z_\ell))\}_{\ell=1}^n$. Once such a relation is available, the candidate branch values at a given $z \in \mathbb{C} \setminus I$ are obtained as the s roots of the univariate polynomial $m \mapsto P(z, m)$, and non-physical branches can be accessed and tracked by analytic continuation away from the branch locus.

Throughout, we assume that m is algebraic of degree s in the sense that there exists a nonzero bivariate polynomial $P \in \mathbb{C}[z, m]$, the ring of bivariate polynomials in (z, m) with complex coefficients, with $\deg_m P = s$ such that (C.1) holds for $m = m_1$ on $\mathbb{C} \setminus I$. Among all such relations, it is natural to seek a minimal one in the sense of having minimal degree in m, equivalently an irreducible P that does not contain spurious factors producing extraneous roots. Before turning to the numerical fitting procedure, we record structural constraints implied by the Stieltjes symmetry of m_1, starting with the fact that (C.1) admits a real-coefficient representative.

Lemma D.1. *Assume that $P \in \mathbb{C}[z, m]$ satisfies (C.1) for $m = m_1$ and is irreducible in $\mathbb{C}[z, m]$. Then there exists a nonzero scalar $\alpha \in \mathbb{C}$ such that $\alpha P \in \mathbb{R}[z, m]$. In particular, after an irrelevant rescaling of P, one may assume $P \in \mathbb{R}[z, m]$.*

Proof. Since the distribution ν is real-valued, the Stieltjes transform m_1 from (B.1) satisfies the Schwarz symmetry

$$m_1(\bar{z}) = \overline{m_1(z)}, \qquad z \in \mathbb{C} \setminus I,$$

where $\bar{z}$ denotes the complex conjugate of z. Define the conjugate-reflected polynomial $P^\sharp \in \mathbb{C}[z, m]$ by

$$P^\sharp(z, m) := \overline{P(\bar{z}, \bar{m})}.$$

Then, for every $z \in \mathbb{C} \setminus I$,

$$P^\sharp(z, m_1(z)) = \overline{P(\bar{z}, \overline{m_1(z)})} = \overline{P(\bar{z}, m_1(\bar{z}))} = 0,$$

so $P^\sharp$ also annihilates m_1 identically on $\mathbb{C} \setminus I$. By irreducibility, P and $P^\sharp$ must agree up to a nonzero scalar factor, hence there exists $\lambda \in \mathbb{C} \setminus \{0\}$ such that $P^\sharp = \lambda P$. Applying the $\sharp$ operation twice yields $P = (P^\sharp)^\sharp = \bar{\lambda}\lambda P$, so $|\lambda| = 1$. Choose $\alpha \in \mathbb{C} \setminus \{0\}$ such that $\alpha/\bar{\alpha} = \lambda$ and set $\tilde{P} := \alpha P$. Then $\tilde{P}^\sharp = \bar{\alpha} P^\sharp = \bar{\alpha}\lambda P = \alpha P = \tilde{P}$, which implies that $\tilde{P}$ has real coefficients, that is, $\tilde{P} \in \mathbb{R}[z, m]$. □

Remark D.1. The rescaling in Lemma D.1 does not affect the roots of $m \mapsto P(z,m)$ for any fixed z, and has no impact on the reconstructed branches. A numerical normalization fixing the scalar ambiguity is described later in this section. △

Having established that (C.1) admits a real-coefficient representative, we henceforth assume $P \in \mathbb{R}[z,m]$. To fit P from pointwise evaluations of m_1, we restrict P to a finite-dimensional model class by expanding it in a bivariate monomial basis. Concretely, we fix an index set $\mathcal{A} \subset \mathbb{N}_0^2$ and parameterize

$$P(z,m) := \sum_{(i,j)\in\mathcal{A}} c_{ij} z^i m^j, \qquad c_{ij} \in \mathbb{R}. \tag{D.1}$$

Let $N := |\mathcal{A}|$, and collect the coefficients into a vector $\boldsymbol{c} \in \mathbb{R}^N$ according to a fixed ordering of $\mathcal{A}$.

A common choice for $\mathcal{A}$ is a rectangular support,

$$\mathcal{A} := \{0,1,\dots,d_z\} \times \{0,1,\dots,s\}, \tag{D.2}$$

which corresponds to allowing degrees up to d_z in z and up to s in m. More generally, the flexibility of $\mathcal{A}$ allows us to incorporate structural constraints on (C.1) such as moment constraints at infinity as in Appendix D.1 or reduced supports motivated by asymptotic cancellations.

D.1 Constraints at Infinity from Moments

Assuming that a finite number of moments of the target probability measure ν are available, the Laurent expansion of m_1 at infinity induces linear constraints on the coefficients $\{c_{ij}\}_{(i,j)\in\mathcal{A}}$. Fix $r \in \mathbb{N}_0$ and define the moment vector $\boldsymbol{\mu} := (\mu_0, \mu_1, \dots, \mu_r) \in \mathbb{R}^{r+1}$.

Definition D.1. For $j \in \mathbb{N}_0$, define the j-fold discrete convolution

$$\boldsymbol{\mu}^{*j} := \underbrace{\boldsymbol{\mu} * \cdots * \boldsymbol{\mu}}_{j \text{ times}},$$

with the convention $\boldsymbol{\mu}^{*0} = [1]$. For $q \in \mathbb{N}_0$, denote by μ_q^{*j} the q-th entry of $\boldsymbol{\mu}^{*j}$, equivalently

$$\mu_q^{*j} := \sum_{\substack{p_1+\cdots+p_j=q \\ 0\le p_\ell \le r}} \prod_{\ell=1}^{j} \mu_{p_\ell},$$

with the conventions $\mu_0^{*0} = 1$, $\mu_q^{*0} = 0$ for $q \ge 1$, and $\mu_q^{*j} = 0$ whenever $q < 0$ or $q > jr$.

The following Proposition D.2 provides linear constraints on the coefficients using known moments. We note that Proposition D.2 is effectively a higher-order generalization of Lemma C.1.

Proposition D.2. *Let P be parameterized as in (D.1) for a fixed choice of index set $\mathcal{A} \subset \mathbb{N}_0^2$, and assume that (C.1) holds for $m = m_1$. Suppose that the moments $\{\mu_p\}_{p=0}^r$ are available, and let*

$$e_{\max} := \max\{i - j \ : \ (i,j) \in \mathcal{A},\ c_{ij} \neq 0\}. \tag{D.3}$$

Then the coefficients $\{c_{ij}\}_{(i,j)\in\mathcal{A}}$ satisfy the linear constraints

$$\sum_{(i,j)\in\mathcal{A}} (-1)^j\, c_{ij}\, \mu_{i-j-e_{\max}+l}^{*j} = 0, \qquad l = 0,1,\dots,r. \tag{D.4}$$

Proof. Using the identity $(x-z)^{-1} = -z^{-1}\sum_{p=0}^{r}(x/z)^p + O\big(z^{-r-2}\big)$ in (B.1) and applying (C.3), we obtain the truncated Laurent expansion

$$m_1(z) = -\sum_{p=0}^{r} \frac{\mu_p}{z^{p+1}} + O\big(z^{-r-2}\big), \qquad |z| \to \infty.$$

Set $\phi_r(z) := \sum_{p=0}^{r} \mu_p z^{-p}$, so that $m_1(z) = -z^{-1}\phi_r(z) + O(z^{-r-2})$. For each $j \in \mathbb{N}_0$,

$$m_1(z)^j = (-1)^j z^{-j} \phi_r(z)^j + O\big(z^{-(j+r+1)}\big), \qquad |z| \to \infty.$$

Expanding the multinomial $\phi_r(z)^j$ and collecting coefficients gives

$$\phi_r(z)^j = \sum_{q=0}^{jr} \mu_q^{*j} z^{-q},$$

where μ_q^{*j} is the q-th entry of the j-fold discrete convolution of $\boldsymbol{\mu}$, as stated in Definition D.1. Consequently, for each $(i,j) \in \mathcal{A}$,

$$z^i m_1(z)^j = (-1)^j \sum_{q=0}^{jr} \mu_q^{*j} z^{i-j-q} + O\big(z^{i-j-(r+1)}\big), \qquad |z| \to \infty.$$

Substituting this into (D.1) and canceling the coefficients of $z^{e_{\max}-l}$ for $l = 0, 1, \ldots, r$ in the truncated expansion of $P(z, m_1(z))$ yields (D.4). □

Remark D.2. In practice, while higher moments of an empirical distribution may be too noisy to estimate reliably, the leading constraint $l = 0$ in (D.4) is always available since $\mu_0 = 1$ holds unconditionally. This constraint enforces cancellation on the top $(i-j)$-edge of $\mathcal{A}$,

$$\sum_{\substack{(i,j)\in\mathcal{A} \\ i-j=e_{\max}}} (-1)^j c_{ij} = 0.$$

For a full rectangular support (D.2), the top $(i-j)$-edge consists of the single monomial $(i,j) = (d_z, 0)$, and the above condition reduces to $c_{d_z,0} = 0$. It is therefore natural to exclude this monomial from $\mathcal{A}$ a priori. More refined choices of $\mathcal{A}$, including slanted Newton polygon supports, are discussed later in connection with branch-point constraints. △

D.2 Constraints from Branch Points

We now exploit the prior knowledge of the branch-cut geometry $I = \bigcup_{j=1}^{k} [a_j, b_j]$. As discussed in Appendix C, the branch locus of the algebraic relation (C.1) is contained among the zeros of the discriminant $\Delta(z)$, obtained by eliminating m_* from (C.2). In the present setting, we assume that the spectral edges $\{a_j, b_j\}_{j=1}^{k}$ are among these branch points, which yields the nonlinear endpoint constraints

$$\Delta(a_j) = 0 \qquad \text{and} \qquad \Delta(b_j) = 0, \qquad j = 1, \ldots, k. \tag{D.5}$$

Since $\Delta \in \mathbb{R}[z]$ is a univariate polynomial, each condition in (D.5) implies that $(z - a_j)$ and $(z - b_j)$ divide Δ. If the endpoints $\{a_j, b_j\}_{j=1}^{k}$ are distinct, then with $\pi(z) \coloneqq \prod_{j=1}^{k}(z - a_j)(z - b_j)$ we obtain the divisibility condition $\pi \mid \Delta$. This divisibility viewpoint suggests an alternative enforcement strategy, for instance by parameterizing $\Delta(z) = \pi(z)\tilde{\Delta}(z)$.

The endpoint constraints (D.5) also imply a simple necessary richness condition on the polynomial model class. Recall $s = \deg_m(P)$ and write $P(z,m) = \sum_{j=0}^{s} a_j(z) m^j$ with $\deg_z(a_j) \le d_z$. Then

$$\deg_z(\Delta) \le (2s-1) d_z.$$

Indeed, up to a nonzero factor depending on the leading coefficient $a_s(z)$, one has $\Delta(z) = \mathrm{Res}_m(P, \partial_m P)$, and the resultant is the determinant of the $(2s-1) \times (2s-1)$ Sylvester matrix in the variable m. Since all Sylvester entries are coefficients of P and $\partial_m P$ and satisfy $\deg_z(P) = \deg_z(\partial_m P) \le d_z$, it follows that $\deg_z(\Delta) \le (2s-1)d_z$. See, for example, [50, Section 4.1].

From the fact that a nonzero polynomial with $2k$ distinct real zeros must have degree at least $2k$, this yields the necessary condition

$$(2s-1) d_z \ge 2k, \tag{D.6}$$

for P to accommodate $2k$ prescribed distinct endpoints. We emphasize that (D.6) is only a lower bound; the algebraic model may admit additional branch points away from $\{a_j, b_j\}_{j=1}^{k}$, contributing further zeros of Δ.

D.3 Numerical Reconstruction from Samples

We now describe the numerical procedure used to fit an implicit relation $P(z, m) = 0$ from samples of the physical branch m_1, and to recover non-physical branches by root finding and continuation. Throughout, we fix an index set $\mathcal{A} \subset \mathbb{N}_0^2$ and parameterize P by (D.1) with coefficient vector $\boldsymbol{c} \in \mathbb{R}^N$, $N := |\mathcal{A}|$.

A convenient choice of $\mathcal{A}$ is the rectangular set (D.2), where the sheet count s is set empirically, while d_z can be initialized using the lower bound (D.6) and increased as needed. In practice, we increase (s, d_z) until the fitting error stabilizes while avoiding over-parameterization.

Sampling. We choose sampling points $\{z_\ell\}_{\ell=1}^n \subset \mathbb{C} \setminus I$ along multiple contours enclosing each I_j and the overall I, but away from the branch locus, in particular separated from the endpoints $\{a_j, b_j\}_{j=1}^k$ by a small buffer. At each sample, we evaluate $\hat{m}_\ell := m_1(z_\ell)$. The details of our approach for evaluating m_1 from empirical measures are described in Appendix E. To preserve the Schwarz symmetry needed for real-coefficient structure as discussed in Lemma D.1, we sample in conjugate pairs: whenever z_ℓ is included, we also include $\bar{z}_\ell$, which yields $m_1(\bar{z}_\ell) = \overline{m_1(z_\ell)}$.

Linear system for the coefficients. Define the design matrix $\mathbf{A} \in \mathbb{C}^{n \times N}$ with entries

$$A_{\ell,(i,j)} := z_\ell^i \hat{m}_\ell^j, \qquad (i,j) \in \mathcal{A}, \quad \ell = 1, \dots, n, \tag{D.7}$$

where the column index (i, j) follows a fixed ordering of $\mathcal{A}$. Recall that $\boldsymbol{c} \in \mathbb{R}^N$ collects the coefficients c_{ij} in this same order. By construction, $(\mathbf{A}\boldsymbol{c})_\ell = P(z_\ell, \hat{m}_\ell)$. The samples satisfy $P(z_\ell, \hat{m}_\ell) \approx 0$, hence $\mathbf{A}\boldsymbol{c} \approx \mathbf{0}$. We compute $\boldsymbol{c}$ as a nontrivial approximate null vector via the augmented homogeneous least-squares problem

$$\hat{\boldsymbol{c}} := \operatorname*{argmin}_{\boldsymbol{c} \in \mathbb{R}^N} \|\mathbf{A}_{\mathbb{R}} \boldsymbol{c}\|_2, \qquad \text{subject to} \qquad \|\boldsymbol{c}\|_2 = 1, \tag{D.8}$$

where

$$\mathbf{A}_{\mathbb{R}} := \begin{bmatrix} \Re(\mathbf{A}) \\ \Im(\mathbf{A}) \end{bmatrix}. \tag{D.9}$$

The solution is given by the right singular vector corresponding to the smallest singular value of $\mathbf{A}_{\mathbb{R}}$. The scalar ambiguity of P (see Remark D.1) is fixed by the normalization $\|\boldsymbol{c}\|_2 = 1$ and a deterministic sign convention, for example by requiring the first nonzero entry of $\boldsymbol{c}$ to be positive.

Incorporating moment constraints. When moment constraints derived in Proposition D.2 are used, they can be written as a linear system $\mathbf{B}\boldsymbol{c} = \mathbf{0}$, where $\mathbf{B} \in \mathbb{R}^{(r+1) \times N}$ encodes the $r + 1$ enforced constraints $\{\mu_p\}_{p=0}^r$. To enforce these constraints, we restrict $\boldsymbol{c}$ to $\ker(\mathbf{B})$. Concretely, let $\mathbf{Q} \in \mathbb{R}^{N \times N_c}$ have orthonormal columns spanning $\ker(\mathbf{B})$, obtained for example from an SVD of $\mathbf{B}$. Then, any feasible coefficient vector can be written as $\boldsymbol{c} = \mathbf{Q}\boldsymbol{y}$. We then solve the reduced problem

$$\hat{\boldsymbol{y}} := \operatorname*{argmin}_{\boldsymbol{y} \in \mathbb{R}^{N_c}} \|\tilde{\mathbf{A}}_{\mathbb{R}} \boldsymbol{y}\|_2 \qquad \text{subject to} \qquad \|\boldsymbol{y}\|_2 = 1, \tag{D.10}$$

where

$$\tilde{\mathbf{A}}_{\mathbb{R}} := \begin{bmatrix} \Re(\mathbf{A}\mathbf{Q}) \\ \Im(\mathbf{A}\mathbf{Q}) \end{bmatrix},$$

followed by $\hat{\boldsymbol{c}} := \mathbf{Q}\hat{\boldsymbol{y}}$. In practice, enforcing only the leading constraint $\mu_0 = 1$ (that is, $r = 0$) is often beneficial even when higher-order moments are unreliable.

Algorithm D.1 summarizes the coefficient-estimation step (with optional moment constraints), producing a fitted polynomial $\hat{P}$.

Appendix E Computing the Stieltjes Transform and Density

This section describes the numerical procedure used to evaluate the Stieltjes transform and to recover the density from it. The first task is to evaluate the physical Stieltjes branch $m(z)$. When empirical

Algorithm D.1: Implicit fitting of an algebraic relation $P(z, m) = 0$ from samples of m_1

Input : Index set $\mathcal{A} \subset \mathbb{N}_0^2$ such as in (D.2) with $N := |\mathcal{A}|$;
Oracle for the empirical Stieltjes transform $\hat{m}_1(z)$ using submatrix eigenvalues;
Number of sampling points n (even);
Optional moments $\{\mu_p\}_{p=0}^{r}$ (to enforce Proposition D.2).

Output : Coefficient vector $\hat{\boldsymbol{c}} \in \mathbb{R}^N$ defining $\hat{P}(z, m) = \sum_{(i,j)\in\mathcal{A}} \hat{c}_{ij} z^i m^j$;
Fitting residual error $e := \|\mathbf{A}\hat{\boldsymbol{c}}\|_2 / \|\mathbf{A}\|_F$ for model selection.

// Sample the physical branch away from the cut, in conjugate pairs

Initialize set of points $\mathcal{Z} \leftarrow \{\}$ and function values $\mathcal{M} \leftarrow \{\}$.

Generate $n/2$ points $\{z_\ell\}_{\ell=1}^{n/2} \subset \mathbb{C}^+$ on arbitrary contours enclosing I, excluding ϵ-neighborhood from I.

Evaluate $\hat{m}_\ell \leftarrow \hat{m}_1(z_\ell)$.

Append $(z_\ell, \hat{m}_\ell)$ and $(\bar{z}_\ell, \overline{\hat{m}_\ell})$ to $(\mathcal{Z}, \mathcal{M})$. *// Total of n samples with conjugate points*

// Build the design matrix (D.7) from samples

Form $\mathbf{A} \in \mathbb{C}^{n\times N}$ with columns indexed by $(i, j) \in \mathcal{A}$:

$$A_{\ell,(i,j)} \leftarrow z_\ell^i \hat{m}_\ell^j, \qquad (z_\ell, \hat{m}_\ell) \in (\mathcal{Z}, \mathcal{M}).$$

Form $\mathbf{A}_\mathbb{R} \leftarrow \begin{bmatrix} \Re(\mathbf{A}) \\ \Im(\mathbf{A}) \end{bmatrix}$ as in (D.9) *// Form the real-valued design matrix*

// Optional: enforce linear moment constraints $\mathbf{B}\boldsymbol{c} = \mathbf{0}$ from Proposition D.2

if *moments $\{\mu_p\}_{p=0}^{r}$ are provided* **then**

  Build $\mathbf{B} \in \mathbb{R}^{(r+1)\times N}$ from (D.4)

  Compute an orthonormal basis $\mathbf{Q} \in \mathbb{R}^{N\times N_c}$ for $\ker(\mathbf{B})$ (e.g. using SVD)
  // $N_c = N - \operatorname{rank}(\mathbf{B})$

  Set $\mathbf{A}_\mathbb{R} \leftarrow \mathbf{A}_\mathbb{R}\mathbf{Q}$ and interpret unknowns as $\boldsymbol{y} \in \mathbb{R}^{N_c}$

// Solve (D.8), or (D.10) with moment constraints

$d \leftarrow \text{size}(\mathbf{A}_\mathbb{R}, 2)$ *// Number of columns*

$\mathbf{U}\boldsymbol{\Sigma}\mathbf{V} \leftarrow \mathbf{A}_\mathbb{R}$ *// Thin SVD where $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_d)$ and $\sigma_1 \geq \cdots \geq \sigma_d$*

$\hat{\boldsymbol{v}} \leftarrow V_{[:,d]}$ *// Right singular vector for the smallest singular value σ_d*

if *moment constraints were enforced* **then** $\hat{\boldsymbol{c}} \leftarrow \mathbf{Q}\hat{\boldsymbol{v}}$ **else** $\hat{\boldsymbol{c}} \leftarrow \hat{\boldsymbol{v}}$

$\hat{\boldsymbol{c}} \leftarrow \hat{\boldsymbol{c}}/\|\hat{\boldsymbol{c}}\|_2$ *// Normalize, optionally, fix a deterministic sign convention*

return $\hat{\boldsymbol{c}}$ *and residual error* $e := \|\mathbf{A}\hat{\boldsymbol{c}}\|_2 / \|\mathbf{A}\|_F$.

eigenvalues are available, this can be done directly by summing the Cauchy kernel. However, after fitting an algebraic spectral curve $P(z, m) = 0$, evaluating $m(z)$ becomes a *root-selection* problem: for each query point z, one must identify the *physical* root among all algebraic roots of $P(z, \cdot)$. This process is described in Appendix E.1. The second task, discussed in Appendix E.2, is the inverse Stieltjes problem: once $m(z)$ has been evaluated near the real axis, we recover the density ρ from its boundary values.

E.1 Physical Root Selection for the Stieltjes Transform

Recall from (B.1) that the Stieltjes transform of a spectral measure ν is

$$m(z) = \int_\mathbb{R} \frac{1}{x - z}\, \nu(\mathrm{d}x), \quad z \in \mathbb{C} \setminus \operatorname{supp}(\nu). \tag{E.1}$$

For empirical eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$, this quantity is computed directly as

$$m_n(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i - z}.$$

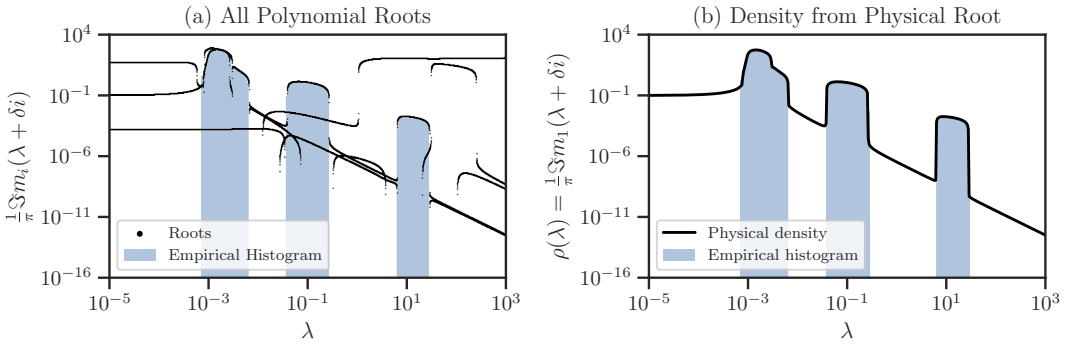

Figure E.1: Physical branch selection for the fitted algebraic Stieltjes transform in the diffusion-model example at the initial size $n = 2^2$K. Panel (a) shows all candidate densities $\pi^{-1}\Im m_i(\lambda + i\delta)$ with $\delta = 10^{-6}$, obtained from the seven roots m_i of $P(\lambda + i\delta, m) = 0$, where $\deg_m(P) = 7$. These unordered candidates include both physical and non-physical sheets and illustrate the difficulty of pointwise root selection in a multi-bulk, log-scale spectrum. Panel (b) shows that the continuation procedure successfully selects the physical branch, producing a coherent density curve that matches the empirical histogram on the bulks. The remaining sloped baseline in the spectral gaps is a finite-δ artifact and is addressed in Appendix E.2.

The more difficult problem arises when we evaluate the Stieltjes transform through a fitted algebraic relation $P(z, m) = 0$. For each fixed query point $z \in \mathbb{C} \setminus I$, the polynomial $m \mapsto P(z, m)$ has $s = \deg_m(P)$ roots. Only one of these roots is the physical Stieltjes value, while the remaining roots belong to non-physical sheets of the algebraic curve. Thus, evaluating $m(z)$ from P is a nontrivial branch-selection problem, especially for high-degree curves and multi-bulk spectra, where roots may approach one another, exchange ordering, or pass near branch points.

Figure E.1 illustrates this issue for the diffusion model example; see Appendix I.5. In the left panel, we evaluate all roots of $m \mapsto P(z, m)$ at $z = \lambda + i\delta$, for a small fixed $\delta > 0$ and a broad range of λ, and plot the candidate values $\pi^{-1}\Im m_i(\lambda + i\delta)$, $i = 1, \ldots, s$. Since the roots are unordered from one value of λ to the next, the result is an unordered scatter of algebraic candidates rather than a connected curve. At sufficiently fine resolution, the neighboring points visually trace several candidate branches. Only one of these traces corresponds to the physical density,

$$\rho(\lambda) \approx \frac{1}{\pi}\Im m(\lambda + i\delta),$$

namely the trace that agrees with the empirical histogram shown in the background. The right panel shows the result after continuously selecting the physical branch. Thus, the purpose of the procedure below is to pass from the unordered scatter of algebraic roots in the left panel to the coherent physical curve in the right panel.

The physical branch is characterized by the Herglotz property and by its normalization at infinity,

$$\Im m(z) > 0 \quad \text{for } z \in \mathbb{C}^+, \qquad m(z) \sim -\frac{1}{z} \quad \text{as } |z| \to \infty.$$

However, these conditions alone are not always sufficient for stable pointwise root selection. Instead, we treat the fitted relation as a spectral curve

$$\mathcal{C} = \left\{ (z, m) \in \mathbb{C}^2 \mid P(z, m) = 0 \right\},$$

and track the physical sheet geometrically on this curve from an anchor point to the desired target point.

The anchor point z_{a} is chosen so that the physical root m_{a} can be identified reliably. One option is to place z_{a} far from the spectrum, where the normalization $m(z) \sim -1/z$ selects the correct algebraic root. Another option is to place z_{a} closer to the real axis and select the algebraic root closest to the empirical value $m_n(z_{\mathrm{a}})$ from (E.1). The latter approach is often more stable for complicated algebraic curves.

Starting from $(z_{\mathrm{a}}, m_{\mathrm{a}}) \in \mathcal{C}$, we choose a continuous path $\gamma : [0,1] \to \mathbb{C}^+$ with $\gamma(0) = z_{\mathrm{a}}$ and $\gamma(1) = z$, and lift this path to the spectral curve by seeking $m(s)$ such that

$$P(\gamma(s), m(s)) = 0, \qquad m(0) = m_{\mathrm{a}}.$$

Root ordering alone is unreliable along such paths: near branch points, two sheets may come close, exchange order, or locally appear to cross. To avoid jumping between sheets, we follow the local geometry of the curve using a predictor–corrector procedure. Differentiating $P(\gamma(s), m(s)) = 0$ along the path gives

$$\frac{\mathrm{d}m}{\mathrm{d}s} = -\frac{\partial_z P(\gamma(s), m(s))}{\partial_m P(\gamma(s), m(s))}\, \gamma'(s).$$

Equivalently, away from branch points, the local tangent satisfies

$$\frac{\mathrm{d}m}{\mathrm{d}z} = -\frac{\partial_z P(z, m)}{\partial_m P(z, m)}.$$

This tangent relation provides a predictor for the next point z_{new} along the path on the same sheet. The predicted value is then corrected by enforcing

$$P(z_{\mathrm{new}}, m_{\mathrm{new}}) = 0,$$

which amounts to projecting the predicted point back onto the spectral curve. This correction is performed by Newton iteration, viewing $m_{\mathrm{new}} \in \mathbb{C}$ as a two-dimensional real variable. The selected root at z_{new} is the algebraic root closest to this corrected value. If no root is sufficiently close, or if the local geometry is too stiff, the continuation step is subdivided and repeated with smaller increments until the root closeness criterion is satisfied. Repeating this predictor–corrector continuation along the path yields a continuous selection of the physical branch.

This continuation-based, or homotopic, root selection is what allows the fitted polynomial to be used as a numerical Stieltjes transform. The resulting physical-branch selection procedure is summarized in Algorithm E.1.

The continuation step presented here is not merely a numerical convenience. Without it, direct pointwise root selection is unstable even for relatively simple algebraic curves, because root orderings can change from one query point to the next. This instability becomes much more severe for high-degree, multi-bulk, or log-scale spectra, where several sheets may come close over many orders of magnitude. The diffusion-model example in Figure E.1 is deliberately chosen as a stringent stress test of this type: the fitted curve has degree $\deg_m(P) = 7$, the density spans many decades in both λ and $\rho(\lambda)$, the geometry is highly stiff near the spectral edges, and the support contains several strongly unbalanced bulks, with the leftmost bulk orders of magnitude taller than the others. Nevertheless, the geometric continuation procedure reliably resolves the physical branch and produces a coherent density from the fitted algebraic curve.

With the physical root $m(z)$ selected consistently over a query grid $z = \lambda + i\delta$, the density $\rho(\lambda)$ can then be recovered from its boundary values using the inverse Stieltjes procedure described next.

E.2 Density Recovery and δ-Extrapolation

Once the physical branch $m(z)$ has been selected, the density is recovered from the boundary values of the Stieltjes transform. Recall the Plemelj formula from (B.2):

$$\rho(\lambda) = \frac{1}{\pi} \lim_{\delta \to 0^+} \Im m(\lambda + i\delta).$$

In numerical computations, however, one cannot evaluate exactly on the real axis at $\delta = 0$. Instead, one evaluates m at a small positive offset δ, giving the regularized density

$$\rho_\delta(\lambda) \coloneqq \frac{1}{\pi} \Im m(\lambda + i\delta).$$

The parameter δ plays the role of a smoothing scale: larger values of δ improve numerical stability but smear sharp features of the density, while smaller values give sharper reconstructions at the cost of increased sensitivity near branch points and spectral edges.

Algorithm E.1: Physical branch selection for an algebraic Stieltjes transform

Input : Polynomial $P \in \mathbb{R}[z, m]$ (see Algorithm D.1);
Query points $\{z_j\}_{j=1}^{N_z} \subset \mathbb{C}^+$;
Anchor rule for selecting m_{a} at z_{a};
Newton tolerance $\epsilon > 0$.

Output : Physical Stieltjes values $m(z_j)$ satisfying $P(z_j, m(z_j)) = 0$.

for $j = 1, \ldots, N_z$ **do**
  Choose an anchor point z_{a} and a path $\gamma : [0, 1] \to \mathbb{C}^+$ with $\gamma(0) = z_{\mathrm{a}}, \gamma(1) = z_j$.
  Select the physical root m_{a} at z_{a} using either $m(z) \sim -z^{-1}$ or the empirical value $m_n(z_{\mathrm{a}})$.
  Initialize $m^{(0)} \leftarrow m_{\mathrm{a}}$.
  // Lift the path γ to the spectral curve $P(z, m) = 0$
  for *successive points $z_{\mathrm{old}}, z_{\mathrm{new}}$ along γ* **do**
    Compute the tangent predictor

$$\dot{m} = -\frac{\partial_z P(z_{\mathrm{old}}, m)}{\partial_m P(z_{\mathrm{old}}, m)} \dot{z}.$$

    Predict m_{pred} at z_{new}.
    Correct m_{pred} by Newton iteration on

$$P(z_{\mathrm{new}}, m) = 0.$$

    if *Newton correction is accepted* **then**
      Set $m \leftarrow m_{\mathrm{new}}$.
    else
      Subdivide the path segment and repeat. *// Avoid sheet jumping*
  Store $m(z_j) \leftarrow m$.
return $\{m(z_j)\}_{j=1}^{N_z}$.

This smoothing effect has a precise interpretation. Indeed,

$$\rho_\delta(\lambda) = \frac{1}{\pi} \Im m(\lambda + i\delta) = \int_{\mathbb{R}} \frac{1}{\pi} \frac{\delta}{(\lambda - \xi)^2 + \delta^2} \nu(\mathrm{d}\xi) = (P_\delta * \nu)(\lambda),$$

where

$$P_\delta(\lambda) \coloneqq \frac{1}{\pi} \frac{\delta}{\lambda^2 + \delta^2}$$

is the Poisson kernel in the upper half-plane [49]. Thus, evaluating the inverse Stieltjes formula at a single finite δ amounts to convolving ν with P_δ, which introduces a finite-δ blur and a nonzero baseline outside the support.

This effect is especially visible for spectra plotted on logarithmic scales. In spectral gaps, where the limiting density should vanish, the finite-δ Poisson tail creates a nonzero baseline. Away from the support, this baseline behaves like $P_\delta(\lambda) \sim \delta/(\pi\lambda^2)$, which appears as a line of slope -2 on a log–log plot. We refer to this artifact as the Poisson δ-floor. Decreasing δ lowers this floor and makes the bulks stand out more sharply, but it also makes branch tracking and root selection more sensitive. The left panel of Figure E.2 illustrates this issue for the diffusion-model example. In the right panel, the baseline is eliminated by the extrapolation procedure described below, and the bulk edges are sharply resolved.

To remove this artifact, we evaluate the physical branch on a ladder of offsets

$$\delta_0 < \delta_1 < \cdots < \delta_L$$

and extrapolate the sequence $\rho_{\delta_i}(\lambda)$ to $\delta = 0$ at each fixed λ. A simple approach is to use a low-degree polynomial fit in δ. In practice, we use a Chebyshev basis for this extrapolation, which is

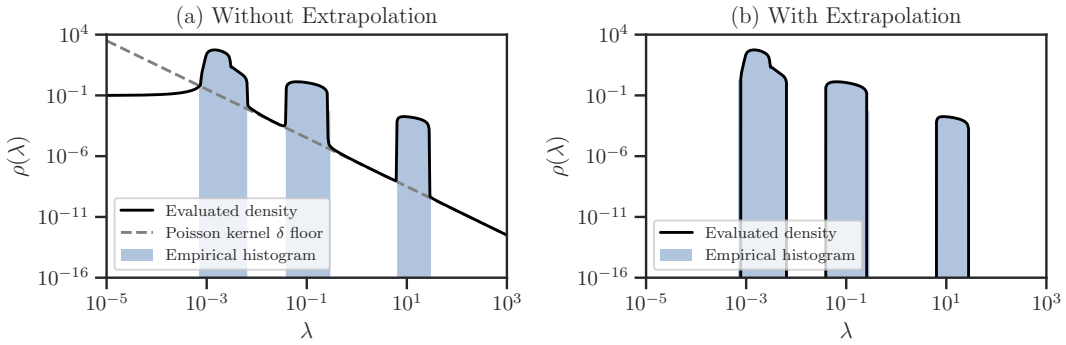

Figure E.2: Inverse Stieltjes recovery for the diffusion-model spectrum at the initial size $n = 2^2\mathrm{K}$. Panel (a) shows the direct finite-δ density $\rho_\delta(\lambda) = \pi^{-1}\Im m(\lambda + i\delta)$, which agrees with the empirical histogram on the bulks but exhibits a Poisson δ-floor in spectral gaps. Panel (b) shows the density after extrapolating the finite-δ evaluations to $\delta = 0$, removing the artificial gap baseline while preserving the bulk densities.

more stable than a monomial basis when several δ-levels are used [51]. Concretely, after mapping the offset ladder $\{\delta_i\}_{i=0}^{L}$ to points $\{\eta_i\}_{i=0}^{L} \subset [-1, 1]$, we approximate

$$\rho_{\delta_i}(\lambda) \approx \sum_{k=0}^{q} c_k(\lambda) T_k(\eta_i),$$

where T_k are Chebyshev polynomials. The extrapolated density is then obtained by evaluating this fitted polynomial at the value of η corresponding to $\delta = 0$. This approach is particularly useful for log-scale spectra, such as the diffusion-model example shown in Figure E.2. In this example, we used four δ-levels and a quadratic Chebyshev extrapolation. This was sufficient to reduce the artificial Poisson floor below 10^{-20} across the spectral gaps, without additional regularization.

For absolutely continuous measures with densities represented on a uniform linear grid, one can alternatively exploit the convolution identity $\rho_\delta = P_\delta * \rho$. In that case, the collection of finite-δ evaluations can be viewed as a regularized deconvolution problem, which can be solved efficiently in Fourier space. Indeed, on a uniform linear grid, the convolution relation gives

$$\mathcal{F}\{\rho_\delta\}(\omega) = \mathcal{F}\{P_\delta\}(\omega)\,\mathcal{F}\{\rho\}(\omega) = e^{-\delta|\omega|}\mathcal{F}\{\rho\}(\omega),$$

where $\mathcal{F}$ is the Fourier transform. Thus, if ρ_δ were known exactly on a uniform grid, the formal inverse would be

$$\rho = \mathcal{F}^{-1}\left[e^{\delta|\omega|}\mathcal{F}\{\rho_\delta\}(\omega)\right].$$

The factor $e^{\delta|\omega|}$ inverts the Poisson smoothing, but it also amplifies high-frequency numerical noise. Hence, the Fourier recovery must be regularized, for instance by damping or truncating high-frequency modes.

The δ-extrapolation is similarly important for densities with wide log–log scale spectra, such as the diffusion-model example shown in Figure E.2. Without it, the finite-δ Poisson floor obscures spectral gaps and makes narrow bulks appear artificially connected on a log–log plot. After extrapolation, the density retains the same bulk-scale agreement with the empirical histogram, while the gap baseline is pushed below the plotted range. This allows the multi-bulk structure and spectral edges to be resolved cleanly, even in this stringent log-scale test case.

Appendix F Free Decompression on Algebraic Spectral Curves

In this section, we develop the mathematical and computational foundation of free decompression from an algebraic Stieltjes relation. We begin in Appendix F.1 with the Nica–Speicher theorem, which identifies the R-transform scaling that underlies free compression and motivates its inverse operation. We then define infinite decompressibility in Appendix F.2 and show its equivalence with free infinite divisibility.

The next issue is computational. Although free decompression leads naturally to a nonlinear PDE for the Stieltjes transform, Appendix F.3 explains why solving this PDE directly is numerically unstable in realistic examples, especially when analytic continuation across branch cuts is required. The main purpose of the remainder of the section is to show that this difficulty can be avoided when the initial Stieltjes transform is represented by an algebraic relation $P(z, m) = 0$. In Appendix F.4, free decompression is lifted to a rational transformation of the associated spectral curve. In Appendix F.5, this transformation is converted into an evolved polynomial relation whose roots give all algebraic candidates at later times.

The remaining challenge is to select the physical Stieltjes branch from these candidates. We address this in Appendix F.6 by tracking the physical sheet geometrically along the free-decompression flow. This leads to a predictor–corrector method on the spectral curve, formulated as a 2×2 Newton system in the natural curve coordinates. This viewpoint is the basis of the practical free-decompression algorithm used in the numerical experiments.

F.1 The Nica–Speicher Theorem

Our exposition begins with a complete proof of the R-transform relationship which underpins the free decompression procedure, also known as the *Nica–Speicher Theorem*. Although our conditions appear to be less general than required in practice, our emphasis here is on establishing a rigorous set of assumptions upon which free decompression can be guaranteed to hold.

To do so, it is important to define several important objects pertaining to free probability theory. First, we begin with the definition of "limits" of matrices in the noncommutative algebraic sense.

Definition F.1 (Free Algebra Limit)**.** Let tr_n denote the normalized trace, given by $\mathrm{tr}_n(\mathbf{A}) = \frac{1}{n}\mathrm{tr}(\mathbf{A})$. A sequence of matrices $\mathbf{A}_n$ has a *limit* a if for every integer k, $\mathrm{tr}_n(\mathbf{A}_n^k)$ converges as $n \to \infty$ to a value, denoted $\mathrm{tr}_\infty(a^k)$. More generally, a family of sequences of matrices

$$(\{\mathbf{A}_{1,n}\}_{n=1}^\infty, \ldots, \{\mathbf{A}_{k,n}\}_{n=1}^\infty)$$

has limits $(a_1, \ldots, a_k)$ if for any (noncommutative) polynomial combination $p(x_1, \ldots, x_k)$, $\mathrm{tr}_n(p(\mathbf{A}_{1,n}, \ldots, \mathbf{A}_{k,n}))$ converges to a value, denoted $\mathrm{tr}_\infty(p(a_1, \ldots, a_k))$.

Next, we consider the critical definition of asymptotic freeness, first introduced in Voiculescu [54].

Definition F.2 (Asymptotically Free)**.** A family of sequences of matrices

$$(\{\mathbf{A}_{1,n}\}_{n=1}^\infty, \ldots, \{\mathbf{A}_{k,n}\}_{n=1}^\infty)$$

is *asymptotically free* from another family of sequences of matrices

$$(\{\mathbf{B}_{1,n}\}_{n=1}^\infty, \ldots, \{\mathbf{B}_{l,n}\}_{n=1}^\infty)$$

if the two families have a joint limit $(a_1, \ldots, a_k, b_1, \ldots, b_l)$ and $(a_1, \ldots, a_k)$, $(b_1, \ldots, b_l)$ are *freely independent* in the following sense: for any integer $m \geq 1$ and any "alternating" sequence of elements $x_1, x_2, \ldots, x_m$ such that x_{2i+1} is a noncommutative polynomial combination of $(a_1, \ldots, a_k)$ and x_{2i} is a noncommutative polynomial combination of $(b_1, \ldots, b_l)$, if $\mathrm{tr}_\infty(x_i) = 0$ for all i, then $\mathrm{tr}_\infty(x_1 x_2 \cdots x_m) = 0$.

Finally, we will require another important transform, similar to the R-transform.

Definition F.3 (S-Transform)**.** Let ν be a probability measure and let $\mu_n = \int x^n \mathrm{d}\nu(x)$ denote its moments for each integer n. The *truncated moment-generating function* of ν is given by

$$M(z) = \sum_{n=1}^{\infty} \mu_n z^n.$$

If $\chi(z)$ is the functional inverse of M so that $M(\chi(z)) = z$, then the S-transform of ν is given by

$$S(z) = \frac{z+1}{z}\chi(z).$$

The significance of the S-transform is its behavior for multiplicative free convolution. If a and b are freely independent, then the spectral distribution of $b^{1/2}ab^{1/2}$ has S-transform S_{ab} satisfying [14, Lemma 2.19]

$$S_{ab}(z) = S_a(z)S_b(z),$$

where S_a and S_b are the S-transforms of a and b respectively. To relate the R- and S-transforms, we have the following lemma.

Lemma F.1. *Letting $R(z)$ and $S(z)$ denote the R- and S-transforms of a probability measure, if $y = zS(z)$, then $z = yR(y)$.*

Proof. From Couillet & Liao [14, Definition 5], if $m(z)$ is the Stieltjes transform, then

$$m\left(\frac{z+1}{zS(z)}\right) = -zS(z).$$

Substituting $y = zS(z)$ gives

$$m\left(\frac{z+1}{y}\right) = -y.$$

But since $m(R(y) + y^{-1}) = -y$ from Couillet & Liao [14, Definition 5], $z/y = R(y)$ and so $z = yR(y)$. □

Now we may prove the Nica–Speicher Theorem [35].

Theorem F.2 (Nica–Speicher)**.** *Let $\{\mathbf{A}_n\}_{n=1}^{\infty}$ be a sequence of $n \times n$ deterministic real symmetric matrices whose empirical spectral distribution converges weakly as $n \to \infty$ to a compactly supported probability measure with R-transform $R(z)$ and S-transform $S(z)$. Consider a sequence of integers $\{k_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} \frac{k_n}{n} = \tau \in (0,1)$. For each n, let $\mathbf{Q}_n$ be a Haar-distributed random orthogonal matrix of size $n \times n$ and define $\tilde{\mathbf{A}}_n = \mathbf{Q}_n^{\intercal}\mathbf{A}_n\mathbf{Q}_n$. Let σ_n be a random permutation of $\{1,\dots,n\}$, independent of $\mathbf{Q}_n$, and let $\mathbf{B}_n = (\tilde{\mathbf{A}}_{\sigma_n(i)\sigma_n(j)})_{i,j=1}^{k_n}$. Then as $n \to \infty$, the empirical spectral distribution of $\mathbf{B}_n$ converges weakly almost surely to a limiting probability measure whose R- and S-transforms $R_\tau(z)$ and $S_\tau(z)$ are given by*

$$R_\tau(z) = R(\tau z), \qquad S_\tau(z) = S(\tau z).$$

Proof. First, choose $M > 0$ such that the limiting measure is supported in $(-M, M)$. Weak convergence implies that clipping the eigenvalues of $\mathbf{A}_n$ to $[-M, M]$ changes only $o(n)$ eigenvalues. This changes $\mathbf{A}_n$, and hence the corresponding compressed matrix $\mathbf{B}_n$, by rank $o(n)$. By the rank inequality, the empirical distribution functions of the original and clipped compressed matrices differ by at most $o(n)/k_n = o(1)$. It therefore suffices to prove the result for the clipped sequence, for which $\|\mathbf{A}_n\| \le M$. We retain the same notation for the resulting matrices.

Fix an integer n. For a permutation σ of $\{1,\dots,n\}$ into itself, we may define the permutation matrix $\mathbf{\Pi}_\sigma$ such that $(\mathbf{\Pi}_\sigma)_{ij} = 1$ if $i = \sigma(j)$ and 0 otherwise. In this way, $\mathbf{B}_n$ becomes the top-left $k_n \times k_n$ block of

$$\tilde{\tilde{\mathbf{A}}}_n = \mathbf{\Pi}_{\sigma_n}^{\intercal}\tilde{\mathbf{A}}_n\mathbf{\Pi}_{\sigma_n} = \mathbf{\Pi}_{\sigma_n}^{\intercal}\mathbf{Q}_n^{\intercal}\mathbf{A}_n\mathbf{Q}_n\mathbf{\Pi}_{\sigma_n}.$$

Since $\tilde{\mathbf{Q}}_n = \mathbf{Q}_n\mathbf{\Pi}_{\sigma_n}$ is also an orthogonal matrix and the Haar measure is invariant under permutations, $\tilde{\mathbf{Q}}_n \overset{\mathcal{D}}{=} \mathbf{Q}_n$ and so $\tilde{\tilde{\mathbf{A}}}_n \overset{\mathcal{D}}{=} \tilde{\mathbf{A}}_n$. Consequently, we can assume without loss of generality that the permutations σ_n are the identity and consider

$$\tilde{\mathbf{B}}_n = \mathbf{P}_n\tilde{\mathbf{A}}_n\mathbf{P}_n,$$

where $\mathbf{P}_n$ is a diagonal matrix with $(\mathbf{P}_n)_{ii} = 1$ if $i \le k_n$ and 0 otherwise.

Because $\mathbf{Q}_n$ is orthogonal, the conjugation $\tilde{\mathbf{A}}_n = \mathbf{Q}_n^{\intercal}\mathbf{A}_n\mathbf{Q}_n$ leaves the eigenvalues invariant. Therefore, the empirical spectral distribution of $\tilde{\mathbf{A}}_n$ is identical to that of $\mathbf{A}_n$ for all n, and converges weakly to the same compactly supported probability measure ν, which is uniquely characterised by its R-transform $R(z)$. Furthermore, $\mathbf{P}_n$ clearly has an empirical spectral distribution converging weakly to a Bernoulli distribution:

$$\lim_{n\to\infty} \nu_{\mathbf{P}_n} = \nu_p = (1-\tau)\delta_0 + \tau\delta_1.$$

Let $\mathrm{tr}_n(\mathbf{A}) = \frac{1}{n}\mathrm{tr}(\mathbf{A})$ denote the normalized trace. We note that $\lim_{n\to\infty} \mathrm{tr}_n(\mathbf{P}_n) = \lim_{n\to\infty} \frac{k_n}{n} = \tau$. By Haar invariance, we may assume that $\mathbf{A}_n$ is diagonal with its eigenvalues in increasing order.

The uniform bound and weak convergence imply convergence of all moments. Moreover, if F_ν^{-1} denotes the quantile function of ν, then

$$\lim_{n\to\infty} \operatorname{tr}_n(\mathbf{A}_n^r \mathbf{P}_n) = \int_0^\tau \left(F_\nu^{-1}(s)\right)^r \mathrm{d}s, \qquad r \in \{0, 1, 2, \dots\}.$$

Since $\mathbf{A}_n$ and $\mathbf{P}_n$ commute and $\mathbf{P}_n^2 = \mathbf{P}_n$, their joint moments therefore converge to those of a pair (a, p). The orthogonal asymptotic freeness theorem [13, Theorem 5.2] then gives almost-sure convergence of $(\mathbf{A}_n, \mathbf{P}_n, \mathbf{Q}_n, \mathbf{Q}_n^*)$ in joint $*$-moments to (a, p, u, u^*), where u is a Haar unitary free from (a, p). To show that $\{\tilde{\mathbf{A}}_n\}_{n=1}^\infty$ and $\{\mathbf{P}_n\}_{n=1}^\infty$ are asymptotically free, let $p_i(u^*au)$ and $q_i(p)$ be centered polynomials such that $\operatorname{tr}_\infty(p_i) = \operatorname{tr}_\infty(q_i) = 0$. Note that $p_i(u^*au) = u^* p_i(a) u$ since u^*u is the identity. Now since $\operatorname{tr}_\infty(xy) = \operatorname{tr}_\infty(yx)$,

$$\begin{aligned}\operatorname{tr}_\infty(p_1(u^*au)q_1(p)p_2(u^*au)q_2(p)\cdots) &= \operatorname{tr}_\infty(u^*p_1(a)uq_1(p)u^*p_2(a)uq_2(p)\cdots) \\ &= \operatorname{tr}_\infty(p_1(a)uq_1(p)u^*p_2(a)uq_2(p)\cdots u^*) = 0,\end{aligned}$$

since it is an alternating sequence of elements belonging to $\{a, p\}$ and $\{u, u^*\}$. Since p_i and q_i were arbitrary, $\tilde{a} = u^*au$ and p are freely independent.

To identify the distribution of $\tilde{b} = p\tilde{a}p$, we note that $p^{1/2} = p$ and so the S-transform of $\tilde{b}$, denoted $S_{\tilde{b}}$ satisfies $S_{\tilde{b}} = S_{\tilde{a}} S_p = S_a S_p$. Therefore, we will need to find the S-transform of p.

To do so, note that $m_n = \int x^n \nu_p(\mathrm{d}x) = \tau$ for every integer $n \geq 1$. Therefore, the truncated moment-generating function $M_p(z) = \sum_{n=1}^\infty \tau z^n = \frac{\tau z}{1-z}$. The inverse is given by $\chi_p(z) = \frac{z}{z+\tau}$ and so

$$S_p(z) = \frac{z+1}{z}\chi_p(z) = \frac{z+1}{z+\tau}.$$

Consequently,

$$S_{\tilde{b}}(z) = \frac{z+1}{z+\tau} S_a(z). \tag{F.1}$$

Now observe that $\tilde{\mathbf{B}}_n$ is a block matrix, with the top-left $k_n \times k_n$ block comprised of $\mathbf{B}_n$, and every other elements is zero. Consequently, $\operatorname{tr}(\mathbf{B}_n^k) = \operatorname{tr}(\tilde{\mathbf{B}}_n^k)$ for every integer $k \geq 1$, and so $\operatorname{tr}_n(\tilde{\mathbf{B}}_n^k) = \frac{1}{n}\operatorname{tr}(\mathbf{B}_n^k) = \frac{k_n}{n}\operatorname{tr}_{k_n}(\mathbf{B}_n^k)$. The limit of $\tilde{\mathbf{B}}_n$ is $\tilde{b} = p\tilde{a}p$ and so $\mathbf{B}_n$ has a limit b satisfying

$$\operatorname{tr}_\infty(b^k) = \frac{1}{\tau}\operatorname{tr}_\infty(\tilde{b}^k).$$

Therefore,

$$M_{\tilde{b}}(z) = \sum_{k=1}^\infty \operatorname{tr}_\infty(\tilde{b}^k) z^k = \tau \sum_{k=1}^\infty \operatorname{tr}_\infty(b^k) z^k = \tau M_b(z).$$

For the inverses, since $z = M_{\tilde{b}}(\chi_{\tilde{b}}(z)) = \tau M_b(\chi_{\tilde{b}}(z))$, we find that

$$\chi_{\tilde{b}}(z) = \chi_b\left(\frac{z}{\tau}\right).$$

Now observe that

$$\begin{aligned} S_{\tilde{b}}(z) &= \frac{z+1}{z}\chi_b\left(\frac{z}{\tau}\right) \\ S_b\left(\frac{z}{\tau}\right) &= \frac{\frac{z}{\tau}+1}{\frac{z}{\tau}}\chi_b\left(\frac{z}{\tau}\right) = \frac{z+\tau}{z}\chi_b\left(\frac{z}{\tau}\right), \end{aligned}$$

and so by equating the two through χ_b,

$$S_{\tilde{b}}(z) = \frac{z+1}{z+\tau} S_b\left(\frac{z}{\tau}\right).$$

Equating with (F.1) gives $S_b(z) = S_a(\tau z)$, completing the first part of the result. Now we need to express this relationship in terms of the R-transform. To do so, recall from Lemma F.1 that if $y = zS_b(z)$, then $z = yR_b(y)$. Since $S_b(z) = S_a(\tau z)$, $y = zS_b(z) = zS_a(\tau z)$ implies that

$$\tau y = (\tau z) S_a(\tau z).$$

This immediately implies that

$$\tau z = (\tau y) R_a(\tau y).$$

But since $y = zS_b(z)$ also implies that $z = yR_b(y)$, there is

$$\tau y R_b(y) = \tau y R_a(\tau y).$$

Therefore, finally, $R_b(y) = R_a(\tau y) = R(\tau y)$. The uniform norm bound turns the almost-sure moment convergence into weak convergence, and the preceding rank argument transfers this conclusion to the original matrices, completing the proof. □

F.2 Infinite Decompressibility

The Nica–Speicher Theorem guarantees that the distribution obtained under free compression is well-defined through the R-transform by taking $\tau < 1$. However, it does not imply that the inverse operation, taking $\tau > 1$, is generally well-defined. Indeed, there are probability measures where free decompression need not be valid. In this section, we define the class of probability measures for which free decompression can be applied *without bound* on τ. This motivates the following definition.

Definition F.4 (Infinitely Decompressible Measure)**.** A probability measure ν with R-transform R is ***infinitely decompressible*** if for every $\tau \geq 1$, there exists a probability measure ν_τ with R-transform satisfying $R_{\nu_\tau}(z) = R(\tau z)$.

Since the R transform admits the series expansion

$$R_{\nu_\tau}(w) = \sum_{k \geq 1} r_k(\nu_\tau) w^{k-1},$$

where $r_k(\nu_\tau)$ is the k-th ***free cumulant*** of ν_τ, the Nica–Speicher Theorem implies

$$r_k(\nu_\tau) = \tau^{k-1} r_k(\nu), \qquad k \geq 1.$$

Let $D_c\nu$ denote the *dilation* of a measure ν, defined as the law of cX when $X \sim \nu$. This is also written as $c_*\nu$, as this operation is the *push-forward* of ν induced by the map $x \mapsto cx$. Also, denote by $\nu^{\boxplus s}$ the s-fold *free additive convolution power*. The R-transform of dilation and convolution power operations satisfies the identities

$$R_{D_c\nu}(w) = c\, R_\nu(cw), \tag{F.2a}$$

$$R_{\nu^{\boxplus s}}(w) = s\, R_\nu(w), \tag{F.2b}$$

or, in terms of cumulants:

$$r_n(D_c\nu) = c^n r_n(\nu), \tag{F.3a}$$

$$r_n\big(\nu^{\boxplus s}\big) = s r_n(\nu). \tag{F.3b}$$

From the above, it is immediate that dilation and power operations commute. The fractional powers s in the above relations are due to Nica & Speicher [35, Proposition 1.2] (see also [47, Proposition 1.2]).

Definition F.5. The measure $\nu \in \mathcal{P}(\mathbb{R})$ is ***freely infinitely divisible*** (FID, also called $\boxplus$-infinitely divisible), if for any $n \in \mathbb{N}$, there exists $\nu_n \in \mathcal{P}(\mathbb{R})$ such that

$$\nu = \underbrace{\nu_n \boxplus \cdots \boxplus \nu_n}_{n \text{ times}} = \nu_n^{\boxplus n}.$$

A necessary and sufficient condition for a measure to be FID is that its R transform should have a holomorphic extension defined on $\mathbb{C}^-$ with values in $\mathbb{C}^- \cup \mathbb{R}$ [5, Theorem 5.10]. The condition that ν is FID is closely related to free decompression through the following proposition.

Proposition F.3 (Infinitely Decompressible $\iff$ FID)**.** *A probability law ν is infinitely decompressible if and only if it is freely infinitely divisible.*

Proof. (*Sufficiency*) If ν is FID, then for each $\tau \geq 1$, the dilation $D_\tau \nu$ is also FID, hence the free convolution power $\nu_\tau \coloneqq (D_\tau \nu)^{\boxplus \tau^{-1}}$ exists. Using (F.2a) and (F.2b), we obtain

$$R_{\nu_\tau}(w) = R_{(D_\tau \nu)^{\boxplus \tau^{-1}}}(w) = \tau^{-1} R_{D_\tau \nu}(w) = R_\nu(\tau w),$$

so the free decompression flow exists.

(*Necessity*) Assume ν is infinitely decompressible. Fix $n \in \mathbb{N}$ and set $\tau = n$, yielding a measure ν_n such that $R_{\nu_n}(w) = R_\nu(nw)$. Let $\sigma_n \coloneqq D_{n^{-1}} \nu_n$. Then $R_{\sigma_n}(w) = n^{-1} R_{\nu_n}(n^{-1} w) = n^{-1} R_\nu(w)$, hence $R_{\sigma_n^{\boxplus n}}(w) = R_\nu(w)$ and thus $\nu = \sigma_n^{\boxplus n}$. Since this holds for all n, ν is FID. □

We also note that any FID distribution can be approximated by compound free Poisson distributions [48, p. 117] given in Appendix I.1.

F.3 The PDE Approach and Challenges

Let ν be a probability measure and let $\{\nu_\tau\}_{\tau \geq 1}$ be a family of probability measures obtained by free decompression such that for each τ, denoting by R_τ the R-transform of ν_τ, $R_\tau(z) = R(\tau z)$, where R is the R-transform of ν. For each $\tau \geq 1$, let m_τ denote the Stieltjes transform corresponding to ν_τ. Ameli et al. [1] proved that $m(t, z) \coloneqq m_{e^t}(z)$ satisfies the PDE

$$\frac{\partial m(t,z)}{\partial t} = -m(t,z) + m(t,z)^{-1} \frac{\partial m(t,z)}{\partial z}. \tag{F.4}$$

In the following exposition, the time change $\tau = e^t$ will be used repeatedly with reference to the flow (F.4). While free decompression is always achievable in principle via (F.4), *in practice*, it is **extraordinarily difficult** to perform for several reasons. The primary difficulty lies in the ill-posedness of (F.4), which can be seen in two ways.

The first is through the relationship between (F.4) and the Burgers equation: if $\tau = e^t - 1$ and $u(\tau, z) = -e^{-t}/m(\tau, z)$, then u satisfies the inviscid complex Burgers equation $\partial_\tau u + u \partial_z u = 0$. This equation is well-studied in the setting where m is real-valued, including in related problems in RMT [7, 8, 9], and a number of techniques have been developed to obtain accurate approximations of solutions. Arguably the most important consideration is the development of *shocks*, where after some time t, multiple solutions emerge. These shocks present fatal instabilities when solving the PDE directly. To deal with this issue, one often introduces regularization through the viscous complex Burgers equation $\partial_\tau u + u \partial_z u = \xi \partial_{zz} u$, taking ξ to be small [46]. This equation has a unique explicit solution, called the *Cole-Hopf solution*, and taking $\xi \to 0^+$ yields the *entropy solution* to the inviscid Burgers equation. In the real-valued case, the entropy solution is given by the Hopf-Lax formula. However, the current setting requires that m is complex valued, and methods developed for the real-valued setting [46] cannot be readily deployed. When performing free decompression, shocks inevitably emerge, and while the Cole-Hopf solution can be computed, it need not converge as $\xi \to 0^+$ to the solution where $m(\tau, z)$ remains a valid Stieltjes transform. To our knowledge, no examination of the complex inviscid Burgers equation has presented a method to reliably estimate this particular solution.

The second is encountered when applying the method of characteristics to solve (F.4). This is the approach ultimately taken in Ameli et al. [1], but it is subject to several limitations. The characteristic curves for (F.4) are straightforward to obtain, but at the same time where shocks occur in the Burgers equation formulation, the characteristic curves cross a branch cut in the complex plane. To extend beyond this cut requires *analytic continuation* of the Stieltjes transform m. Numerical analytic continuation is notoriously ill-posed, with worst-case error from naïve approaches growing doubly exponentially fast in t, or exponentially fast in the original matrix size [52]. In reality, free decompression exhibits worst-case error that grows no more than polynomially in the matrix size [1, Proposition 3], and Ameli et al. [1] presents an approach using polynomial approximation with error rates often growing only *logarithmically* in the size of the original matrix, with the run-time complexity being independent of t. At these rates, free decompression becomes impressively practical. Unfortunately, the approach is mathematically brittle, highly sensitive to the accuracy of the curve-fit used to obtain the initial data, and imposes assumptions on the form of the spectral density that rarely hold outside of the simplest synthetic cases. Our primary contribution is to identify an

approach that maintains this high level of performance under less restrictive assumptions, rendering free decompression practical for more realistic examples.

F.4 Formal Free Decompression along Spectral Curves

Now, recall that the functional relation for the R-transform in terms of the Stieltjes transform is given by Couillet & Liao [14, Definition 5] as

$$R(-m(z)) = z + \frac{1}{m(z)}.$$

According to the Nica–Speicher Theorem, the operation of free decompression (inverting the operation of free compression) coincides with the map $\nu \mapsto \nu_\tau$ where the R-transforms R and R_τ of ν and ν_τ, respectively, satisfies

$$R_\tau(z) = R(\tau z), \qquad \tau > 1.$$

If m is defined to satisfy the formal algebraic relation $P(z, m) = 0$ for some polynomial P, then, formally, we can track the behavior of the Stieltjes transform of ν_τ under free decompression. In doing so, we can track solutions to free decompression while avoiding any potential issues with m and/or R being multivalued. This is captured in Theorem F.4.

Theorem F.4. *For a polynomial $P(z, m)$, define the following spectral curves:*

$$\begin{aligned}
\mathcal{C}_m &= \{(z, m) \in \mathbb{C}^2 \mid P(z, m) = 0\}, && \textit{(graph of } m\textit{)} \\
\mathcal{C}_R &= \{(-m, z + 1/m) \mid (z, m) \in \mathcal{C}_m,\ m \neq 0\}, && \textit{(R transform of } m\textit{)} \\
\mathcal{C}_{R_\tau} &= \{(w, r) \in \mathbb{C}^2 \mid (\tau w, r) \in \mathcal{C}_R\}, && \textit{(R transform of } m_\tau\textit{)} \\
\mathcal{C}_{m_\tau} &= \{(z, \omega) \in \mathbb{C}^2 \mid (-\omega, z + 1/\omega) \in \mathcal{C}_{R_\tau}, \omega \neq 0\}. && \textit{(graph of } m_\tau\textit{)}
\end{aligned}$$

Then

$$\mathcal{C}_{m_\tau} \subseteq \mathcal{C}_{P_\tau} \coloneqq \left\{ (z, \omega) \in \mathbb{C}^2 \,\middle|\, \omega \neq 0 \text{ and } P\left(z + \frac{\tau - 1}{\tau\omega}, \tau\omega\right) = 0 \right\}.$$

Proof. We begin by unpacking $\mathcal{C}_{m_\tau}$. Note that $(z, \omega) \in \mathcal{C}_{m_\tau}$ if and only if, with $\omega \neq 0$, $(-\omega, z + 1/\omega) \in \mathcal{C}_{R_\tau}$. This implies that $(-\tau\omega, z + 1/\omega) \in \mathcal{C}_R$, which in turn implies that there exists $(z_0, m) \in \mathcal{C}_m$ such that

$$-m = -\tau\omega \quad \text{and} \quad z_0 + \frac{1}{m} = z + \frac{1}{\omega}.$$

Consequently,

$$m = \tau\omega \quad \text{and} \quad z_0 = z + \frac{\tau - 1}{\tau\omega}.$$

Since $(z_0, m) \in \mathcal{C}_m$ implies that $P(z_0, m) = 0$, this implies that there exists z_0 such that

$$P(z_0, \tau\omega) = 0 \quad \text{and} \quad z_0 = z + \frac{\tau - 1}{\tau\omega}.$$

Consequently, any $(z, \omega) \in \mathcal{C}_{m_\tau}$ must also satisfy $(z, \omega) \in \mathcal{C}_{P_\tau}$. □

Theorem F.4 immediately implies Theorem 1, as we identify m_τ to be the free decompression of m by the factor τ.

F.5 Polynomial Evolution under Free Decompression

In this section, we use Theorem F.4 to derive a bivariate polynomial Q_τ such that the Stieltjes transform m_τ under free decompression by a decompression ratio τ satisfies $Q_\tau(z, m_\tau(z)) = 0$ for almost all z. Since this polynomial can be readily obtained from P and the roots obtained using standard methods, it provides a fast method of constructing candidate solutions for m_τ. The polynomial is constructed in Corollary F.5.

Corollary F.5. *Suppose that $m_0 : \mathbb{C}^+ \to \mathbb{C}^+$ satisfies an algebraic relation of the form $P(z, m_0(z)) = 0$ for $z \in \mathbb{C}^+$ where*

$$P(z,m) = \sum_{j=0}^{d_z} \sum_{k=0}^{s} c_{jk} z^j m^k.$$

Then for $\mathcal{C}_{m_\tau}$ defined in Theorem F.4,

$$\mathcal{C}_{m_\tau} \subseteq \left\{ (z,m) \in \mathbb{C}^2 \,\middle|\, Q_\tau(z,m) = 0 \right\},$$

where $Q_\tau(z,m)$ is the polynomial given by

$$Q_\tau(z,m) \coloneqq \sum_{k=0}^{d_z} \sum_{l=0}^{d_z+s} c_{kl}^{(\tau)} z^k m^l,$$

and

$$c_{pq}^{(\tau)} \coloneqq \sum_{j=p}^{d_z} c_{j,q-d_z+j-p} \tau^{q-d_z+j-p} \binom{j}{p} (1-\tau^{-1})^{j-p},$$

where $c_{ij} = 0$ if $j \notin \{0, \dots, s\}$.

Proof. From Theorem F.4, we have the inclusion

$$\mathcal{C}_{m_\tau} \subseteq \left\{ (z,m) \in \mathbb{C}^2 \,\middle|\, P\left(z + \frac{\tau-1}{\tau m}, \tau m\right) = 0 \right\}.$$

In particular, for $(z,m) \in \mathcal{C}_{m_\tau}$, $m \neq 0$ and

$$m^{d_z} P\left(z + \frac{\tau-1}{\tau m}, \tau m\right) = 0.$$

We will show that the left hand side of this expression is the polynomial $Q_\tau(z,m)$. First note that

$$P\left(z + \frac{\tau-1}{\tau m}, \tau m\right) = \sum_{j=0}^{d_z} \sum_{k=0}^{s} c_{jk} \left(z + \frac{\tau-1}{\tau m}\right)^j (\tau m)^k.$$

Expanding the binomial gives

$$\left(z + \frac{\tau-1}{\tau m}\right)^j = \sum_{i=0}^{j} \binom{j}{i} z^{j-i} \left(\frac{\tau-1}{\tau m}\right)^i = \sum_{i=0}^{j} \binom{j}{i} z^{j-i} (\tau-1)^i \tau^{-i} m^{-i}.$$

Therefore,

$$m^{d_z} P\left(z + \frac{\tau-1}{\tau m}, \tau m\right) = \sum_{j=0}^{d_z} \sum_{k=0}^{s} \sum_{i=0}^{j} c_{jk} \binom{j}{i} (\tau-1)^i \tau^{k-i} z^{j-i} m^{k-i+d_z}.$$

Now we introduce new summation indices in place of i, k by letting $p = j - i$ and $q = k - i + d_z$. Then $i = j - p$ and $k = q + i - d_z = q - d_z + j - p$ and $p \in \{0, \dots, d_z\}$, $q \in \{0, \dots, d_z + s\}$. Using these indices, since $j \geq p$ and $\binom{j}{i} = \binom{j}{j-i}$,

$$m^{d_z} P\left(z + \frac{\tau-1}{\tau m}, \tau m\right) = \sum_{p=0}^{d_z} \sum_{q=0}^{d_z+s} \sum_{j=p}^{d_z} c_{j,q-d_z+j-p} \binom{j}{p} (\tau-1)^{j-p} \tau^{q-d_z} z^p m^q,$$

where we have adopted the convention that $c_{j,k} = 0$ if $k \notin \{0, \dots, s\}$. Since $\tau^{q-d_z}(\tau-1)^{j-p} = \tau^{q-d_z+j-p}(1-\tau^{-1})^{j-p}$. Hence

$$m^{d_z} P\left(z + \frac{\tau-1}{\tau m}, \tau m\right) = \sum_{p=0}^{d_z} \sum_{q=0}^{d_z+s} c_{pq}^{(\tau)} z^p m^q,$$

as required. □

As stated, Q_τ need not have the same degree in m as the initial polynomial P. However, we recall that polynomial relations of this form are not unique. We reduce Q_τ by dividing out monomial factors introduced when clearing denominators, and fix the remaining scalar ambiguity by normalizing a nonzero coefficient. We refer to the resulting polynomial as P_τ. For common examples (e.g., Marchenko-Pastur or semicircle laws), P_τ often has the same degree as P, due to Lemma F.6 below.

Lemma F.6. *If the index set $\mathcal{A}$ for P is upper-triangular, then* $\deg_m(P_\tau) \le \deg_m(P)$.

Proof. Note that it will suffice to show that $P(z + (1-\tau^{-1})m^{-1}, \tau m)$ is a bivariate polynomial in z and m, as this will imply that m^{d_z} can be factorized from Q_τ. If $\mathcal{A}$ is upper-triangular, then every surviving monomial $z^j m^k$ in the expansion for P satisfies $k \ge j$. But this implies that

$$\begin{aligned} P\left(z + \frac{\tau-1}{\tau m}, \tau m\right) &= \sum_{j=0}^{d_z} \sum_{k=0}^{s-j} c_{j,j+k} \left(z + \frac{\tau-1}{\tau m}\right)^j (\tau m)^{j+k} \\ &= \sum_{j=0}^{d_z} \sum_{k=0}^{s-j} c_{j,j+k} (\tau z m + \tau - 1)^j (\tau m)^k, \end{aligned}$$

which has only positive powers of z and m. □

For a bivariate polynomial of the form $P(z,m) = q(z)m^2 - p(z)m + 1$, the condition in Lemma F.6 suggests that $\deg_z(q) \le 2$ and $\deg_z(p) \le 1$ implies that P_τ does not have a higher degree than P. Many ensemble models exhibit bivariate polynomial relations of this form, including [58] semicircle law (free Gaussian); [31] (free Poisson); [27] and [32] (free binomial); [55] (free Jacobi); and the general free-Meixner family [3, 45] (free analogs of the classical [33] laws). The Q and P for these ensemble models can be found in [1, Table B.3].

However, P need not be upper-triangular in general; for example, the deformed Wigner ensemble [39, Sections 2.2 and 18.3] is not upper-triangular. This increase in degree exacerbates the most significant difficulty with performing free decompression using Corollary F.5: identifying the correct root of P_τ is extraordinarily challenging. For this reason, we will require a more refined approach by tracing along spectral curves to identify the correct root.

F.6 Free Decompression via Geometric Continuation

The previous subsection shows that, once the initial algebraic curve $P(z,m) = 0$ is given, free decompression produces an evolved algebraic relation whose roots provide all candidate values of the transformed Stieltjes function at time τ. However, this polynomial evolution alone does not identify the physical branch. This is illustrated in Figure F.1(a), where the roots of the evolved polynomial form a cloud of candidate densities, only one of which agrees with the empirical spectrum.

This root-selection problem is fundamentally different from the one in Appendix E.1. There, for the initial curve, one may choose an anchor point directly in the z-plane and continue along that curve. After free decompression, by contrast, the evolved polynomial at a fixed $\tau > 1$ provides no new anchor identifying the physical sheet. Instead, the physical branch must be continued from the known physical branch at $\tau = 1$ along the free-decompression flow itself.

We now formulate this continuation geometrically. We write $m_\tau(z)$ for the Stieltjes transform after decompression ratio τ, equivalently at time $t = \log\tau$. In particular, m_0 denotes the initial Stieltjes transform at $t = 0$, or $\tau = 1$, and should not be confused with the indexed algebraic roots m_i used earlier. Let the initial fitted spectral curve be

$$\mathcal{C} = \left\{(\zeta, y) \in \mathbb{C}^2 \mid P(\zeta, y) = 0\right\}.$$

Here, $y = m_0(\zeta)$ denotes a branch of the initial algebraic Stieltjes transform. Under free decompression by ratio $\tau \ge 1$, the method of characteristics gives the change of variables

$$z = \zeta - \frac{\tau - 1}{y}, \qquad w = \frac{y}{\tau},$$

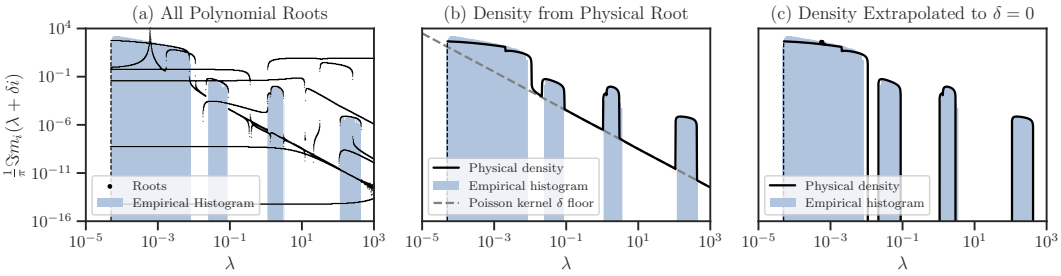

Figure F.1: Free decompression on the fitted algebraic spectral curve for the diffusion-model example at the final size $n = 64{,}000$, evolved from the initial $n = 4000$ submatrix. Panel (a) shows all candidate densities $\pi^{-1}\Im m_i(\lambda + i\delta)$ with $\delta = 10^{-6}$, obtained from the seven roots m_i of the evolved polynomial relation with $\deg_m(P) = 7$, at the final decompression time. These unordered candidates form a cloud of physical and non-physical sheets. Panel (b) shows the result after tracking the physical branch along the free-decompression flow, yielding a coherent density curve that matches the empirical histogram on the bulks, but still exhibits the finite-δ Poisson floor in the spectral gaps. Panel (c) shows the final density after δ-extrapolation to $\delta = 0$ using four δ-levels and a quadratic Chebyshev fit, which removes the artificial gap baseline while preserving the bulk densities.

where $w = m_\tau(z)$ is the decompressed Stieltjes value. Equivalently,

$$\zeta = z + \frac{\tau - 1}{\tau w}, \qquad y = \tau w.$$

Substituting this inverse map into $P(\zeta, y) = 0$ gives the evolved algebraic relation

$$P\left(z + \frac{\tau - 1}{\tau w}, \tau w\right) = 0.$$

This is precisely the relation obtained in Theorem F.4. After clearing denominators, it gives the evolved polynomial whose roots provide all algebraic candidates for $w = m_\tau(z)$, but not the physical one.

To track the physical root, we work instead in the original curve coordinates $(\zeta, y) \in \mathcal{C}$. For a fixed target point z, define

$$F_1(\zeta, y) \coloneqq P(\zeta, y),$$
$$F_2(\zeta, y; \tau, z) \coloneqq \zeta - \frac{\tau - 1}{y} - z.$$

Thus, a point $(\zeta, y) \in \mathcal{C}$ maps to the target z at decompression ratio τ precisely when

$$F_1(\zeta, y) = 0, \qquad F_2(\zeta, y; \tau, z) = 0.$$

At $\tau = 1$, this system reduces to $\zeta = z$ and $y = m_0(z)$, so the initial physical point is known from the branch-selection procedure in Appendix E.1. The task is then to continue this solution in τ, while keeping z fixed.

In practice, evolution along the spectral curve can pass through regions where sheets come close, roots change order, or the geometry becomes stiff. To resolve this, we evolve the above system using a predictor–corrector approach, where the predictor uses the local tangent along the constrained curve. Namely, differentiating $F_1 = 0$ with respect to τ, while keeping z fixed, gives

$$\partial_\zeta P(\zeta, y)\dot{\zeta} + \partial_y P(\zeta, y)\dot{y} = 0.$$

Similarly, differentiating $F_2 = 0$, namely

$$\zeta - \frac{\tau - 1}{y} - z = 0,$$

gives

$$\dot{\zeta} - y^{-1} + (\tau - 1)y^{-2}\dot{y} = 0,$$

or equivalently,

$$\dot{\zeta} + (\tau - 1)y^{-2}\dot{y} = y^{-1}.$$

Combining these two relations gives the 2×2 tangent system

$$\begin{bmatrix} \partial_\zeta P(\zeta, y) & \partial_y P(\zeta, y) \\ 1 & (\tau - 1)y^{-2} \end{bmatrix} \begin{bmatrix} \dot{\zeta} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 \\ y^{-1} \end{bmatrix}.$$

This tangent system provides the predictor direction for advancing the physical point from one value of τ to the next. The predicted point is then corrected by Newton iteration on the nonlinear system $(F_1, F_2) = (0, 0)$. After correction, the decompressed Stieltjes value is recovered as

$$m_\tau(z) = w = \frac{y}{\tau}.$$

The resulting geometric continuation procedure is summarized in Algorithm F.1. In the implementation, the Newton correction is damped and the τ-steps are adaptively subdivided when the local geometry becomes stiff; these safeguards prevent jumps between nearby sheets without changing the underlying geometric continuation described above.

The effect of this procedure is shown in Figure F.1. The evolved polynomial alone produces the unordered candidate cloud in panel (a), while the geometric continuation in τ selects the physical sheet shown in panel (b). This is a stringent test case: the density is log-scale, multi-bulk, and highly unbalanced, with nearby sheets interacting over several orders of magnitude. Nevertheless, the continuation remains on the physical branch and produces a coherent density matching the empirical bulks; the remaining finite-δ floor is then removed by the extrapolation step in panel (c).

Appendix G Evolution of Spectral Quantities

Free decompression primarily evolves the full density $\rho(t, x)$, or equivalently its Stieltjes transform $m(t, z)$. In many applications, however, one only needs derived spectral quantities such as edge locations, cusps (merging/splitting points), gap widths, atoms' weights and locations, or spectral moments. This section shows that these quantities can often be evolved directly from the spectral curve, without reconstructing the density pointwise.

G.1 Evolving Spectral Edges

The *spectral edges* of a density are of particular interest; for instance, [10] shows that the locations of bulk edges play an important role in the dynamics of diffusion models during training. Informally, if the support of the density $\rho(t, x)$ is a union of k disjoint intervals

$$I(t) \coloneqq \bigcup_{j=1}^{k} I_j(t), \qquad I_j(t) \coloneqq [a_j(t), b_j(t)],$$

then the edges are the endpoints $\{a_j(t), b_j(t)\}_{j=1}^{k}$, where $\rho(t, x)$ vanishes outside $I(t)$ and is positive inside the interior of each $I_j(t)$.

When the Stieltjes transform $m(t, z)$ is algebraic, the notion of edge admits a precise characterization in terms of the *branch points* of the spectral curve. Concretely, suppose m is a branch $m = m_1(z)$ of the algebraic relation $P(z, m) = 0$, and define the spectral curve

$$\mathcal{C} \coloneqq \left\{ (\zeta, y) \in \mathbb{C}^2 \mid P(\zeta, y) = 0 \right\}.$$

The curve $\mathcal{C}$ defines a branched covering of the ζ-plane, with local sheets given by the roots $y = y_i(\zeta)$ of $P(\zeta, y) = 0$. A *branch point* is a point $(\zeta_*, y_*) \in \mathcal{C}$ where two (or more) local sheets coalesce; equivalently,

$$P(\zeta_*, y_*) = 0, \qquad \text{and} \qquad \partial_y P(\zeta_*, y_*) = 0.$$

Branch points determine how sheets glue across branch cuts.

Among branch points, *spectral edges* are distinguished by the *physical sheet*. A real branch point $x_* \in \mathbb{R}$ is a spectral edge if the associated branch cut connects the physical sheet to a non-physical sheet, so that crossing the cut switches between m_1 and a different branch. In contrast, other branch

Algorithm F.1: Free decompression by geometric continuation on a spectral curve

Input : Polynomial $P \in \mathbb{R}[z, m]$ such that $P(z, m_0(z)) = 0$ (see Algorithm D.1);
Query grid $\{z_j\}_{j=1}^{N_z} \subset \mathbb{C}^+$;
Decompression ratios $1 = \tau_0 < \tau_1 < \cdots < \tau_T$;
Oracle for the initial physical branch $m_0(z)$ on $P(z, m) = 0$ (see Algorithm E.1);
Newton tolerance $\epsilon > 0$.

Output : Physical Stieltjes values $m_{\tau_\ell}(z_j)$ for all requested τ_ℓ and z_j.

// Initialize the physical sheet at $\tau_0 = 1$
for $j = 1, \ldots, N_z$ **do**
  $y_j^{(0)} \leftarrow m_0(z_j)$ *// Select physical root using Appendix E.1*
  $\zeta_j^{(0)} \leftarrow z_j$ *// At $\tau = 1$, the map is identity in z*
  $m_{\tau_0}(z_j) \leftarrow y_j^{(0)}$

// Continue the physical sheet in decompression ratio τ
for $\ell = 1, \ldots, T$ **do**
  for $j = 1, \ldots, N_z$ **do**
    Set $(\zeta, y) \leftarrow (\zeta_j^{(\ell-1)}, y_j^{(\ell-1)})$.
    // Predict along the tangent of the constrained system
    Solve

$$\begin{bmatrix} \partial_\zeta P(\zeta, y) & \partial_y P(\zeta, y) \\ 1 & (\tau_{\ell-1} - 1) y^{-2} \end{bmatrix} \begin{bmatrix} \dot{\zeta} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 \\ y^{-1} \end{bmatrix}.$$

    $\Delta\tau \leftarrow \tau_\ell - \tau_{\ell-1}$.
    $\zeta_{\text{pred}} \leftarrow \zeta + \Delta\tau\, \dot{\zeta}$.
    $y_{\text{pred}} \leftarrow y + \Delta\tau\, \dot{y}$.
    // Correct by projecting back to the decompression system
    Starting from $(\zeta_{\text{pred}}, y_{\text{pred}})$, apply damped Newton to solve

$$P(\zeta, y) = 0, \qquad \zeta - (\tau_\ell - 1) y^{-1} - z_j = 0.$$

    if *Newton converges to (ζ_*, y_*) within tolerance ϵ* **then**
      $\zeta_j^{(\ell)} \leftarrow \zeta_*, \quad y_j^{(\ell)} \leftarrow y_*$.
      $m_{\tau_\ell}(z_j) \leftarrow \tau_\ell^{-1} y_*$. *// Recover evolved Stieltjes value*
    else
      Subdivide $[\tau_{\ell-1}, \tau_\ell]$ and repeat the predictor–corrector step. *// Avoid sheet jumping in stiff regions*

return $\{m_{\tau_\ell}(z_j)\}_{\ell,j}$.

points whose cuts glue two non-physical sheets do not bound the support of the density (even if they are real branch points). In multi-bulk situations, this distinction is essential: not every real branch point corresponds to an endpoint of the support.

With this notion in mind, we first characterize the evolution of branch points under free decompression, and then interpret which of the resulting real branch points correspond to spectral edges via their sheet connectivity to the physical branch.

Proposition G.1 (Evolving Branch Point under Free Decompression)**.** *Let $(z_*, m_*) \in \mathbb{C}^2$ be a branch point of $m(t, z)$ with $m_* := m(t, z_*)$. Define the pullback $(\zeta_*, y_*) = \phi_t^{-1}(z_*, m_*)$, and*

assume that it is finite with $y_* \neq 0$. Then, $(\zeta_*, y_*) \in \mathcal{C}$ satisfies

$$P(\zeta_*, y_*) = 0, \tag{G.1a}$$

$$y_*^2\, \partial_y P(\zeta_*, y_*) - (\tau - 1)\, \partial_\zeta P(\zeta_*, y_*) = 0, \tag{G.1b}$$

followed by the pushforward $z_* = \zeta_* - (\tau - 1)y_*^{-1}$. In particular, real solutions $z_* \in \mathbb{R}$ of (G.1) include the evolving spectral edges of the real density.

Proof. The spectral edges of the real density at time t correspond to real branch points of the map $(\zeta, y) \mapsto z = \zeta - (\tau - 1)y^{-1}$ restricted to the spectral curve $\mathcal{C}$.

Along the spectral curve $\mathcal{C}$, implicit differentiation of $P(\zeta, y) = 0$ gives

$$\mathrm{d}P(\zeta, y) = \frac{\partial P(\zeta, y)}{\partial \zeta}\mathrm{d}\zeta + \frac{\partial P(\zeta, y)}{\partial y}\mathrm{d}y.$$

On the spectral curve, $P(\zeta, y) = 0$, hence $\mathrm{d}P(\zeta, y) = 0$. Whenever $\partial_y P(\zeta, y) \neq 0$, it holds

$$\frac{\mathrm{d}y}{\mathrm{d}\zeta} = -\frac{\partial_\zeta P(\zeta, y)}{\partial_y P(\zeta, y)}.$$

On the other hand, differentiating the projection $\pi \circ \phi_t(\zeta, y) = z(\zeta, y) = \zeta - (\tau - 1)y^{-1}$ along $\mathcal{C}$ yields

$$\frac{\mathrm{d}z}{\mathrm{d}\zeta} = 1 + (\tau - 1)y^{-2}\frac{\mathrm{d}y}{\mathrm{d}\zeta} = 1 - (\tau - 1)y^{-2}\frac{\partial_\zeta P(\zeta, y)}{\partial_y P(\zeta, y)}.$$

The critical points of the projection $z = \pi \circ \phi_t$ are characterized by $\mathrm{d}z/\mathrm{d}\zeta = 0$ together with $P(\zeta, y) = 0$, which is equivalent to (G.1). □

Since spectral edges are a subset of branch points, Proposition G.1 applies in particular to the evolution of the edges $\{a_j(t), b_j(t)\}_{j=1}^k$. For each fixed $t \geq 0$, the algebraic system (G.1) may admit a set of solutions

$$\mathcal{B}(t) := \left\{ z_*^{(j)}(t) \right\}_{j=1}^{p},$$

where typically $p \geq k$. The real subset $\mathcal{B}(t) \cap \mathbb{R}$ provides candidates for spectral edges; one then selects those candidates whose cuts connect the physical sheet to a non-physical sheet. Operationally, this can be checked by whether the physical branch exhibits a nontrivial jump across the real axis at that point (equivalently, by the Stieltjes inversion formula): a candidate $x_* \in \mathbb{R}$ corresponds to an edge of the density precisely when the boundary values $m_1(t, x_* + i0^+)$ and $m_1(t, x_* + i0^-)$ switch sheets across the associated cut. Algorithm G.1 summarizes the resulting procedure.

G.2 Spectral Cusps and The Mechanism of Splitting Bulks

Under free decompression, a bulk may split into two when a gap opens in its interior. Viewed in reverse, this is the merger of two neighboring bulks under free compression. We describe this transition through the geometry of the spectral curve, focusing on the branch points and their *sheet connectivity* (monodromy).

Let $I_j(t) := [a_j(t), b_j(t)]$ denote the j-th bulk before the split. Suppose that at time $t = t_* > 0$, a gap begins to open at a point x_* in its interior. We consider a regular transition, whose geometry is described below. Let $x_-(t)$ and $x_+(t)$ denote the left and right endpoints of the new gap. At the splitting time,

$$x_-(t_*) = x_+(t_*) = x_*.$$

For $t > t_*$ sufficiently close to t_*, we have $x_-(t) < x_+(t)$, and the density vanishes in the gap $(x_-(t), x_+(t))$. We then relabel the child bulks, increasing the indices of the subsequent bulks by one and simultaneously setting

$$(a_j, b_j, a_{j+1}, b_{j+1}) \leftarrow (a_j, x_-, x_+, b_j).$$

The two resulting bulks are $I_j(t) := [a_j(t), b_j(t)]$ and $I_{j+1}(t) := [a_{j+1}(t), b_{j+1}(t)]$, with $I_j(t) \cap I_{j+1}(t) = \emptyset$ just after the split. The new inner *spectral edges*, now labeled $b_j(t)$ and $a_{j+1}(t)$, continue to evolve according to the branch-point equations in Proposition G.1.

Algorithm G.1: Evolution of spectral edges under algebraic free decompression

Input : Polynomial $P \in \mathbb{R}[z, m]$ with $\deg_m(P) = s$ such that $P(z, m_0(z)) = 0$ (see Algorithm D.1);
Time grid $0 = t_0 < t_1 < \cdots < t_T$;
Initial support $I(0) = \bigcup_{j=1}^{k_0} [a_j(0), b_j(0)]$ (optional, for initialization);
Oracle for the physical branch $m_0(z)$ on $\mathbb{C} \setminus I$ (see Algorithm E.1); *// Only needed if $I(0)$ is provided.*
Tolerance $\epsilon > 0$.

Output : Edge trajectories $\{a_j(t_\ell), b_j(t_\ell)\}$ over the requested time grid;
Active bulk count $k(t_\ell)$ (number of connected components).

// Define the edge system from Proposition G.1

Set $\tau \leftarrow e^t$ and define the functions

$$G_1(\zeta, y; t) := P(\zeta, y)$$
$$G_2(\zeta, y; t) := y^2 \partial_y P(\zeta, y) - (\tau - 1)\partial_\zeta P(\zeta, y).$$

// Initialization at $t_0 = 0$

if *initial edges $I(0)$ are provided* **then**
  Initialize guesses $\{(\zeta_0^{(j)}, y_0^{(j)})\}_{j=1}^{2k_0}$ as follows: *// Any reasonable edge-local seeding works*
    ▷ Sample $2k_0$ points ζ, one near each endpoint
    ▷ Choose Herglotz-consistent root $y_0^{(j)} \leftarrow m_0(\zeta_0^{(j)})$.
else
  Obtain initial guesses $\{(\zeta_0^{(j)}, y_0^{(j)})\}_{j=1}^{2k_0}$ by solving $(G_1, G_2) = (0, 0)$ at t_0 using multiple seeds.

// Track edge preimages (ζ, y) by warm-started Newton over time

$k(t_0) \leftarrow k_0$
for $\ell = 1, \ldots, T$ **do**
  $\tau \leftarrow e^{t_\ell}$
  $k(t_\ell) \leftarrow k(t_{\ell-1})$
  foreach $j = 1, \cdots, 2k(t_{\ell-1})$ **do**
    Use previous solution $(\zeta_{\ell-1}^{(j)}, y_{\ell-1}^{(j)})$ as initial guess.
    Apply damped Newton to solve $(G_1(\zeta, y; t_\ell), G_2(\zeta, y; t_\ell)) = (0, 0)$ for $(\zeta_\ell^{(j)}, y_\ell^{(j)})$.
    $z_\ell^{(j)} \leftarrow \zeta_\ell^{(j)} - (\tau - 1)/y_\ell^{(j)}$ *// Pushforward to the z-plane*
    if $|\Im z_\ell^{(j)}| < \epsilon$ **then** Store $x_\ell^{(j)} \leftarrow \Re z_\ell^{(j)}$ **else** mark failure.
  // Detect newly opened gaps within the active bulks
  foreach $I_j(t_\ell) = [a_j(t_\ell), b_j(t_\ell)]$ **do**
    Search additional real solutions (ζ_-, y_-), (ζ_+, y_+) of $(G_1, G_2) = (0, 0)$ whose pushforwards $x_\pm = \zeta_\pm - (\tau - 1)/y_\pm$ satisfy $x_\pm \in I_j(t_\ell)$ and $x_- < x_+$
    if *such (x_-, x_+) exists and is a physical edge pair* **then**
      // Replace the parent bulk by the two split child bulks
      $\{I_r\}_{r=1}^{k(t_\ell)+1} \leftarrow \{I_1, \ldots, I_{j-1}, [a_j, x_-], [x_+, b_j], I_{j+1}, \ldots, I_{k(t_\ell)}\}$
      $k(t_\ell) \leftarrow k(t_\ell) + 1$ *// Update number of bulks*
      Store new preimages (ζ_-, y_-) and (ζ_+, y_+)

return *Spectral edges* $\left\{\{a_j(t_\ell), b_j(t_\ell)\}_{j=1}^{k(t_\ell)}\right\}_{\ell=0}^{T}$ *and active bulk count* $\{k(t_\ell)\}_{\ell=0}^{T}$;

This splitting is a phase transition in sheet gluing relative to the physical branch. The corresponding branch points of the deformed spectral projection, described by Proposition G.1, form a nonreal conjugate pair just before t_*. They locally glue non-physical branches in their respective half-planes,

where the physical Stieltjes transform remains holomorphic. At t_*, they meet on the real axis. Just after t_*, they connect the physical sheet to non-physical sheets and become the two inner spectral edges. In this sense, the branch points are present on both sides of the transition, but they bound a gap in the physical density only after the split.

At the splitting time t_*, the density develops a *cusp* (a cubic-root singularity at a sharp local minimum), associated with the Pearcey universality class in complex Wigner-type models [17]. Geometrically, the colliding branch points correspond to a higher-order critical point of the projection of the spectral curve. The following proposition gives algebraic conditions for this cusp point on $\mathcal{C}$.

Proposition G.2 (Cusp point at Splitting Bulks)**.** *Suppose a bulk splits at a regular spectral cusp at time $t_* > 0$, with $x_* = b_j(t_*) = a_{j+1}(t_*) \in \mathbb{R}$. Let $m_* := m(t_*, x_*)$ and define the pullback $(\zeta_*, y_*) := \phi_{t_*}^{-1}(x_*, m_*)$. Assume that $y_* \neq 0$ and $\partial_y P(\zeta_*, y_*) \neq 0$. Then, $(\zeta_*, y_*) \in \mathcal{C}$ is a higher-order critical point of the projection $z(\zeta, y) = \zeta - (\tau_* - 1)y^{-1}$ restricted to $\mathcal{C}$, and hence satisfies*

$$P(\zeta_*, y_*) = 0 \tag{G.2a}$$

$$y_*^2\, \partial_y P - (\tau_* - 1)\, \partial_\zeta P = 0, \tag{G.2b}$$

$$y_* \Big((\partial_{\zeta\zeta} P)(\partial_y P)^2 - 2(\partial_{\zeta y} P)(\partial_\zeta P)(\partial_y P) + (\partial_{yy} P)(\partial_\zeta P)^2 \Big) + 2(\partial_\zeta P)^2 (\partial_y P) = 0, \tag{G.2c}$$

where all derivatives are evaluated at (ζ_*, y_*), and $\tau_* := e^{t_*}$. Moreover, the cusp location is the pushforward $x_* = \zeta_* - (\tau_* - 1)y_*^{-1}$.

Proof. Consider the restriction of the map

$$z(\zeta, y) = \pi \circ \phi_t(\zeta, y) = \zeta - (\tau - 1)y^{-1}$$

to the spectral curve $\mathcal{C} = \{(\zeta, y) \in \mathbb{C}^2 \mid P(\zeta, y) = 0\}$. Near $(\zeta_*, y_*) \in \mathcal{C}$, let $y = y(\zeta)$ be a local parameterization of $\mathcal{C}$. A regular *cusp* at time t_* corresponds to a critical point of multiplicity two of the projection $\zeta \mapsto z(\zeta, y(\zeta)) = z(\zeta)$ along $\mathcal{C}$, i.e.,

$$\frac{\mathrm{d}z(\zeta_*)}{\mathrm{d}\zeta} = 0, \qquad \frac{\mathrm{d}^2 z(\zeta_*)}{\mathrm{d}\zeta^2} = 0, \qquad \frac{\mathrm{d}^3 z(\zeta_*)}{\mathrm{d}\zeta^3} \neq 0,$$

together with $P(\zeta_*, y_*) = 0$.

The identity $P(\zeta, y(\zeta)) = 0$ along $\mathcal{C}$ implies $\mathrm{d}P/\mathrm{d}\zeta = 0$ and $\mathrm{d}^2P/\mathrm{d}\zeta^2 = 0$ at (ζ_*, y_*). Implicit differentiation of $P(\zeta, y(\zeta)) = 0$ along $\mathcal{C}$ gives

$$\begin{aligned}
\frac{\mathrm{d}P}{\mathrm{d}\zeta} &= \frac{\partial P}{\partial \zeta} + y' \frac{\partial P}{\partial y}, \\
\frac{\mathrm{d}^2 P}{\mathrm{d}\zeta^2} &= \frac{\partial^2 P}{\partial \zeta^2} + 2y' \frac{\partial^2 P}{\partial \zeta \partial y} + (y')^2 \frac{\partial^2 P}{\partial y^2} + y'' \frac{\partial P}{\partial y},
\end{aligned}$$

where $y' := \mathrm{d}y/\mathrm{d}\zeta$ and $y'' := \mathrm{d}^2 y/\mathrm{d}\zeta^2$. Solving $\mathrm{d}P/\mathrm{d}\zeta = 0$ and $\mathrm{d}^2P/\mathrm{d}\zeta^2 = 0$ for y' and y'' yields

$$y' = -\frac{\partial_\zeta P}{\partial_y P}, \tag{G.3a}$$

$$y'' = -\frac{\partial_{\zeta\zeta} P + 2y' \partial_{\zeta y} P + (y')^2 \partial_{yy} P}{\partial_y P}, \tag{G.3b}$$

whenever $\partial_y P(\zeta, y) \neq 0$.

Next, differentiating $z(\zeta) = \zeta - (\tau - 1)y(\zeta)^{-1}$ along $\mathcal{C}$ yields

$$\frac{\mathrm{d}z}{\mathrm{d}\zeta} = 1 + (\tau - 1)y^{-2} y', \tag{G.4a}$$

$$\frac{\mathrm{d}^2 z}{\mathrm{d}\zeta^2} = (\tau - 1)\left(-2y^{-3}(y')^2 + y^{-2} y'' \right). \tag{G.4b}$$

Imposing $\mathrm{d}z/\mathrm{d}\zeta = 0$ and using y' from (G.3) gives the second equation in (G.2),

$$y^2 \partial_y P - (\tau - 1)\partial_\zeta P = 0.$$

Next, imposing $\mathrm{d}^2 z/\mathrm{d}\zeta^2 = 0$, substitute y' and y'' from (G.3), and clear denominators by multiplying by $y^3(\partial_y P)^3/(\tau - 1)$. After simplification, this yields the third equation in (G.2),

$$y\left(\partial_{\zeta\zeta} P(\partial_y P)^2 - 2(\partial_{\zeta y} P)(\partial_\zeta P)(\partial_y P) + (\partial_{yy} P)(\partial_\zeta P)^2\right) + 2(\partial_\zeta P)^2(\partial_y P) = 0,$$

where all partial derivatives are evaluated at (ζ_*, y_*). Finally, the cusp location in the physical coordinate is the pushforward $x_* = z(\zeta_*, y_*) = \zeta_* - (\tau_* - 1)y_*^{-1}$ at time $t_* = \log(\tau_*)$. □

Practically, (G.2) can be used to solve for candidate triples (ζ_*, y_*, τ_*) and hence recover both the splitting time $t_* = \log(\tau_*)$ and the split location x_*, after checking the stated nondegeneracy conditions and that the cusp lies on the physical branch. This provides a principled and more accurate alternative compared to monitoring gap opening on a discrete time grid. Once t_* is identified for an isolated split, one can treat the immediate post-split configuration as having one more bulk, and interpret subsequent real branch points through the updated sheet gluing relative to the physical branch.

G.3 Dynamics of Gap and Support Intervals

Beyond tracking the density itself, it is often useful to characterize whether a bulk interval widens or narrows, and whether a gap between two bulks shrinks or expands. Define the bulk widths s_j and gap widths g_j by

$$s_j(t) \coloneqq |I_j(t)| = b_j(t) - a_j(t), \qquad g_j(t) \coloneqq a_{j+1}(t) - b_j(t).$$

The dynamics of these intervals can be obtained from the relative velocity of edges. We derive an expression for the velocity of an edge, $x_*'(t)$, in terms of the geometry of the spectral curve, and then specialize at $t = 0$, where it admits a particularly simple interpretation in terms of the Hilbert transform.

Proposition G.3 (Edge Velocity under Free Decompression)**.** *Let $x_*(t) \in \mathbb{R}$ be a regular spectral edge of the density, and let $(\zeta_*(t), y_*(t))$ denote a corresponding preimage on the spectral curve $\mathcal{C} = \{(\zeta, y) \in \mathbb{C}^2 \mid P(\zeta, y) = 0\}$ under the free decompression map $z = \zeta - (\tau - 1)y^{-1}$ with $\tau = e^t$. Suppose (ζ_*, y_*) is a simple critical point of the projection (so it is not a cusp point), hence it satisfies*

$$P(\zeta_*, y_*) = 0, \tag{G.5a}$$

$$F(\zeta_*, y_*, \tau) = y_*^2 \partial_y P(\zeta_*, y_*) - (\tau - 1)\partial_\zeta P(\zeta_*, y_*) = 0, \tag{G.5b}$$

and the edge location is $x_*(t) = \zeta_*(t) - (\tau - 1)y_*(t)^{-1}$. Then, $x_*(t)$ is differentiable and its velocity is

$$x_*'(t) = \zeta_*'(t) - \tau\, y_*(t)^{-1} + (\tau - 1)\, y_*'(t)\, y_*(t)^{-2}, \tag{G.6}$$

where $(\zeta_*'(t), y_*'(t))$ is obtained by solving the linear system

$$\begin{bmatrix} \partial_\zeta P & \partial_y P \\ \partial_\zeta F & \partial_y F \end{bmatrix}_{(\zeta_*, y_*, \tau)} \begin{bmatrix} \zeta_*' \\ y_*' \end{bmatrix} = -\tau \begin{bmatrix} 0 \\ \partial_\tau F \end{bmatrix}_{(\zeta_*, y_*, \tau)}, \tag{G.7}$$

with

$$\begin{aligned} \partial_\tau F &= -\partial_\zeta P(\zeta, y), \\ \partial_\zeta F &= y^2 \partial_{\zeta y} P - (\tau - 1)\partial_{\zeta\zeta} P, \\ \partial_y F &= 2y\partial_y P + y^2 \partial_{yy} P - (\tau - 1)\partial_{\zeta y} P. \end{aligned}$$

Proof. We differentiate the defining relations (G.5) along the solution curve $t \mapsto (\zeta_*(t), y_*(t))$. From $P(\zeta_*, y_*) = 0$, we obtain

$$\partial_\zeta P(\zeta_*, y_*)\, \zeta_*' + \partial_y P(\zeta_*, y_*)\, y_*' = 0.$$

From $F(\zeta_*, y_*, \tau) = 0$ and $\tau'(t) = \tau(t)$, we obtain

$$\partial_\zeta F(\zeta_*, y_*, \tau)\, \zeta_*' + \partial_y F(\zeta_*, y_*, \tau)\, y_*' + \partial_\tau F(\zeta_*, y_*, \tau)\, \tau'(t) = 0,$$

which yields the linear system (G.7). Finally, differentiating $x_*(t) = \zeta_*(t) - (\tau(t) - 1) y_*(t)^{-1}$ gives (G.6). □

Corollary G.4 (Initial edge velocity)**.** *Let $x_*(0)$ be a regular spectral edge at $t = 0$, and let*

$$H(t, x) := \mathcal{H}\{\nu(t)\}(x) = \text{p.v.}\, \frac{1}{\pi} \int_{\mathbb{R}} \frac{\nu(t, \mathrm{d}y)}{x - y},$$

be the Hilbert transform of the full spectral measure $\nu(t)$ (including any atoms). Then,

$$x_*'(0) = \frac{1}{\pi H(0, x_*)}. \tag{G.8}$$

Proof. At $t = 0$, we have $\tau = 1$ and $x_*(0) = \zeta_*(0)$. Also $y_*(0) = m(0, x_*)$. Using the Plemelj formula, $m(0, x_*) = -\pi H(0, x_*) + i\pi\rho(0, x_*)$, but on the spectral edge $\rho(0, x_*) = 0$, so $y_*(0) = -\pi H(0, x_*)$. Also, from Proposition G.3, we have

$$x_*'(0) = \zeta_*'(0) - \tau(0)\, y_*(0)^{-1}.$$

Moreover, at $t = 0$ the criticality condition in (G.5) reduces to $\partial_y P(\zeta_*(0), y_*(0)) = 0$. If $x_*(0)$ is a regular edge (not a cusp), then $\partial_\zeta P(\zeta_*(0), y_*(0)) \neq 0$, and differentiating $P(\zeta_*(t), y_*(t)) = 0$ at $t = 0$ implies $\zeta_*'(0) = 0$. Hence $x_*'(0) = -y_*(0)^{-1} = 1/(\pi H(0, x_*))$. □

The dynamics of bulk widths $s_j(t)$ and gap widths $g_j(t)$ follows immediately. At $t = 0$, Corollary G.4 yields the explicit identities

$$s_j'(0) = \frac{1}{\pi H(0, b_j(0))} - \frac{1}{\pi H(0, a_j(0))}, \tag{G.9a}$$

$$g_j'(0) = \frac{1}{\pi H(0, a_{j+1}(0))} - \frac{1}{\pi H(0, b_j(0))}. \tag{G.9b}$$

The sign of the Hilbert transform at inner edges depends on the global arrangement of mass across bulks and atoms, and the identities above determine the initial change in each gap width. In contrast, if the full measure $\nu(0)$, including any atoms, is supported on $[a_1(0), b_k(0)]$, the signs at the two outermost bulk edges are fixed: the integral defining H gives $H(0, a_1(0)) < 0$ and $H(0, b_k(0)) > 0$, implying $a_1'(0) < 0$ and $b_k'(0) > 0$. Thus, under this assumption, the span $b_k(t) - a_1(t)$ is increasing at $t = 0$.

G.4 Dynamics of Atoms

An atom of ν corresponds to a simple pole of the Stieltjes transform $m(z)$ at $z = x_\circ$, with weight

$$w = -\operatorname*{Res}_{z = x_\circ} m(z) = \lim_{z \to x_\circ} (x_\circ - z) m(z).$$

When m is algebraic and satisfies $P(z, m) = 0$, where $P(z, m) := \sum_{j=0}^{s} a_j(z) m^j$, a necessary condition for a pole at a finite real point $x_\circ$ is that $x_\circ$ is a root of the leading coefficient, i.e., $a_s(x_\circ) = 0$. Conversely, a real root of a_s corresponds to an atom on the physical branch provided that m has a simple pole at $x_\circ$ (i.e., the singularity is not canceled and is not of higher order). In that case, if $x_\circ$ is a simple root of a_s, coefficient matching yields

$$w = \frac{a_{s-1}(x_\circ)}{a_s'(x_\circ)}.$$

Proposition G.5. *Let $x_\circ \in \mathbb{R}$ be an atom of $m(z)$ with initial mass $w_\circ \in (0, 1]$. Under the free decompression flow, the atom location remains fixed $x_\circ(t) = x_\circ$, while its weight evolves as*

$$w(t) = 1 - (1 - w_\circ) e^{-t}. \tag{G.10}$$

Proof. Let $x_\circ(t)$ denote the (a priori time-dependent) location of the atom with weight $w(t)$. The Laurent expansion of $m(t,z)$ near $z = x_\circ(t)$ is

$$m(t,z) = -\frac{w(t)}{z - x_\circ(t)} + O(1).$$

Differentiating gives

$$\partial_t m(t,z) = -\frac{w'(t)}{z - x_\circ(t)} - \frac{w(t)\, x_\circ'(t)}{(z - x_\circ(t))^2} + O(1),$$

and

$$\partial_z m(t,z) = \frac{w(t)}{(z - x_\circ(t))^2} + O(1).$$

Moreover, the leading term of m^{-1} near $z = x_\circ(t)$ is

$$m(t,z)^{-1} = -\frac{z - x_\circ(t)}{w(t)} + O\Big((z - x_\circ(t))^2\Big).$$

Substituting these expansions into the PDE $\partial_t m = -m + m^{-1}\partial_z m$ yields

$$-\frac{w'(t)}{z - x_\circ(t)} - \frac{w(t)\, x_\circ'(t)}{(z - x_\circ(t))^2} + O(1) = \frac{w(t)}{z - x_\circ(t)} - \frac{1}{z - x_\circ(t)} + O(1). \tag{G.11}$$

The right-hand side has no second-order pole $(z - x_\circ(t))^{-2}$, hence the coefficient of $(z - x_\circ(t))^{-2}$ on the left-hand side must vanish, implying $w(t)\, x_\circ'(t) = 0$. If $w(t) > 0$, then $x_\circ'(t) = 0$, and therefore $x_\circ(t) = x_\circ(0)$ is constant.

Finally, equating the coefficients of the first-order poles in (G.11) gives the ODE $w'(t) = 1 - w(t)$, which is solved by (G.10) with the initial condition $w(0) = w_\circ$. □

Remark G.1. The conclusion of Proposition G.5 is local: it does not depend on whether $x_\circ$ lies inside, outside, or at the boundary of the absolutely-continuous support of $\nu(0)$. Moreover, if $\nu(0)$ has finitely many atoms $\{(x_\circ^{(i)}, w_\circ^{(i)})\}_{i=1}^{n_a}$ at distinct locations $x_\circ^{(i)}$, then each location $x_\circ^{(i)}$ remains fixed and each weight evolves independently according to (G.10). △

Remark G.2. In backward flow with $t < 0$ (free *compression*) and $0 < w_\circ < 1$, there is a finite time $t_\circ := \log(1 - w_\circ) < 0$ at which the atom weight vanishes, $w(t_\circ) = 0$. For $t < t_\circ$, the pole (atom) ansatz is no longer valid, reflecting that the atom has disappeared.

In particular, the atom dynamics are forward deterministic while the atom is present ($w_\circ > 0$). However, disappearance of the atom does not imply loss of information from the full compressed measure. For a known compression ratio $c \in (0,1)$, the original and compressed R-transforms satisfy $R_c(w) = R_0(cw)$, hence $R_0(w) = R_c(w/c)$. Consequently, decompression from the exact compressed measure remains unique whenever it exists.

A concrete example is provided by rank-deficient matrices. Suppose $\mathbf{A}$ is an $n \times n$ symmetric matrix with nullity $0 < q < n$, so its empirical spectral measure has an atom of weight $w_\circ = q/n$ at the origin. Then, after a Haar-distributed orthogonal change of basis, a principal submatrix $\mathbf{A}_\circ$ of size $n_\circ$ is almost surely full rank when $n_\circ \le n - q$, and hence its empirical spectral measure has no atom at zero. △

G.5 Evolution of Moments

We now turn to computing the moments of the evolving measures ν_τ under free decompression starting from $\nu \equiv \nu_1$. For $n \ge 0$, define the time-dependent raw moment

$$\mu_n^{(\tau)} := \int x^n \mathrm{d}\nu_\tau(x).$$

These moments can be evolved without reconstructing ν_τ: the free decompression PDE from Ameli et al. [1, Proposition 1] induces a closed triangular system of ODEs for $\mu_n(t) = \mu_n^{\exp(t)}$, allowing each moment to be computed directly from lower-order moments. Throughout this section, we let $m(t,z)$ denote the Stieltjes transform of ν_{e^t} and assume that $m(t,z) \neq 0$ along the free decompression flow.

Lemma G.6 (ODEs for Moments)**.** *The evolution of the moment $\mu_n(t)$ is governed by the system of the triangularly coupled ODEs,*

$$\left(\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}+1\right)\sum_{i=0}^{n}\mu_i(t)\mu_{n-i}(t)=(n+1)\mu_n(t),\qquad n\in\mathbb{N},\tag{G.12}$$

and, in particular, $\mu_0(t)=1$ and $\mu_1(t)=\mu_1(0)$ for all $t>0$.

Proof. Since by the hypothesis $m(t,z)\neq 0$, we can rearrange the PDE from Ameli et al. [1, Proposition 1] to

$$\left(\frac{1}{2}\frac{\partial}{\partial t}+1\right)m(t,z)^2=\frac{\partial}{\partial z}m(t,z).\tag{G.13}$$

At $|z|>\max(\lambda)$ and under mild regularity conditions, we can expand a convergent Laurent series of $m(t,z)$ by

$$m(t,z)=\int\frac{\mathrm{d}\sigma(\lambda)}{\lambda-z}=-\int\left(\sum_{n=0}^{\infty}\frac{\lambda^n}{z^{n+1}}\right)\mathrm{d}\sigma(\lambda)=-\sum_{n=0}^{\infty}\frac{\mu_n(t)}{z^{n+1}}.$$

The Laurent series of $m(t,z)^2$ becomes

$$m(t,z)^2=\sum_{n=0}^{\infty}\left(\sum_{i=0}^{n}\mu_i(t)\mu_{n-i}(t)\right)\frac{1}{z^{n+2}}.$$

Also,

$$\frac{\partial}{\partial z}m(t,z)=\sum_{n=0}^{\infty}(n+1)\frac{\mu_n(t)}{z^{n+2}}.\tag{G.14}$$

Substituting the Laurent series into (G.13) and rearranging for the powers of $z^{-(n+1)}$, we obtain (G.12). For the first few terms, we find that $\mu_0'=1-\mu_0$, and since $\mu_0(0)=1$, $\mu_0(t)=1$ for all $t>0$. Similarly, we find $\mu_1'(t)=0$ and so $\mu_1(t)$ is also constant. □

The triangular structure of Lemma G.6 suggests solving the moment equations recursively. In particular, the lower-order moments determine the forcing term in the ODE for $\mu_n(t)$, and this forcing is a finite linear combination of powers of e^t. We therefore seek $\mu_n(t)$ in the form

$$\mu_n(t)=\sum_{k=0}^{n-1}\kappa_{n,k}e^{kt},\qquad\text{or equivalently}\qquad\mu_n^{(\tau)}=\sum_{k=0}^{n-1}\kappa_{n,k}\tau^k,$$

where the coefficients $\kappa_{n,k}$ are determined recursively from the initial moments. The following proposition gives the resulting recurrence.

Proposition G.7 (Moment Recurrence)**.** *For any $t\geq 0$ and $n\geq 1$, the n-th moment $\mu_n(t)$ satisfies $\mu_n(t)=\sum_{k=0}^{n-1}\kappa_{n,k}e^{kt}$ where $\kappa_{1,0}=\mu_1(0)$,*

$$\kappa_{n,k}=\begin{cases}\dfrac{\frac{1}{2}k+1}{n-k-1}\displaystyle\sum_{i=1}^{n-1}\sum_{j=0}^{k}\kappa_{i,j}\kappa_{n-i,k-j}, & 0\leq k\leq n-2,\\[2ex] \mu_n(0)-\displaystyle\sum_{j=0}^{n-2}\kappa_{n,j} & k=n-1,\end{cases}\tag{G.15}$$

and where $\kappa_{n,k}=0$ if $k\geq n$.

Proof. Starting from (G.12),

$$\left(\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}+1\right)\left[\sum_{i=1}^{n-1}\mu_i(t)\mu_{n-i}(t)+2\mu_n(t)\right]=(n+1)\mu_n(t).$$

Substituting $\mu_n(t) = \sum_{k=0}^{n-1} \kappa_{n,k} e^{kt}$ gives

$$\begin{aligned}(n+1)\mu_n(t) &= \left(\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} + 1\right)\left[\sum_{i=1}^{n-1}\mu_i(t)\mu_{n-i}(t) + 2\sum_{k=0}^{n-1}\kappa_{nk}e^{kt}\right] \\ &= \left(\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} + 1\right)\sum_{i=1}^{n-1}\mu_i(t)\mu_{n-i}(t) + \sum_{k=0}^{n-1}\kappa_{nk}\,(k+2)\,e^{kt},\end{aligned}$$

and hence

$$\left(\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} + 1\right)\sum_{i=1}^{n-1}\mu_i(t)\mu_{n-i}(t) = \sum_{k=0}^{n-2}(n-k-1)\,\kappa_{nk}e^{kt} \tag{G.16}$$

Rearranging terms gives

$$\begin{aligned}\sum_{i=1}^{n-1}\mu_i(t)\mu_{n-i}(t) &= \sum_{i=1}^{n-1}\left(\sum_{j=0}^{i-1}\kappa_{ij}e^{jt}\right)\left(\sum_{j=0}^{n-i-1}\kappa_{n-i,j}e^{jt}\right) \\ &= \sum_{i=1}^{n-1}\sum_{k=0}^{n-2}\sum_{j=0}^{k}\kappa_{ij}\kappa_{n-i,k-j}e^{kt}. \qquad \text{(G.17)}\end{aligned}$$

Substituting (G.17) into (G.16) gives

$$\sum_{k=0}^{n-2}\left(\frac{1}{2}k+1\right)\sum_{i=1}^{n-1}\sum_{j=0}^{k}\kappa_{ij}\kappa_{n-i,k-j}e^{kt} = \sum_{k=0}^{n-2}(n-k-1)\,\kappa_{nk}e^{kt},$$

and (G.15) follows by equating terms for $0 \le k \le n-2$. Only the expression for $\kappa_{n,n-1}$ remains, which follows from $\sum_{j=0}^{n-1}\kappa_{n,j} = \mu_n(0)$. □

Appendix H Finite-Size Correction to the Spectral Curve

It is critical to note that the discussion so far has focused on cases where the spectrum of a finite size matrix is approximated by that of a matrix with "equivalent" spectral distribution, but is of infinite size, and therefore has a continuous spectrum. This is typically fine for most applications, but it can fail in cases where a positive-definite matrix is considered with a nonzero minimal eigenvalue, yet the limiting spectrum possesses a left edge at zero. In these cases, especially to estimate the condition number of matrix, accounting for finite size effects becomes essential. This is intractable in the general cases considered here, so we shall instead develop an effective heuristic based on a special class of random matrix distributions. We note that the discussion below is *strictly informal*, as a robust proof is well beyond the scope of this paper.

Let $\mathbf{A}$ be an $n \times n$ real symmetric matrix and assume that the probability measure over the elements of $\mathbf{A}$ satisfies

$$p(\mathbf{A}) \propto \exp\left(-\frac{n}{2}\mathrm{tr}V(\mathbf{A})\right)\prod_{i,j=1}^{n}\mathrm{d}\mathbf{A}_{ij},$$

where V is an arbitrary polynomial. Our objective is to estimate $\bar{m}_n(z) = \mathbb{E}m_n(z)$ where $m_n(z) = \frac{1}{n}\mathrm{tr}G(z)$ and $\mathbf{G}(z) = (\mathbf{A} - z\mathbf{I})^{-1}$ is the resolvent for $z \in \mathbb{C} \setminus \mathbb{R}$. Consider Stein's Lemma: for any matrix-valued smooth function $f = (f_{ij})$ that grow subexponentially,

$$\int \sum_{i\le j}\frac{\partial}{\partial \mathbf{A}_{ij}}(f_{ij}(\mathbf{A})e^{-\frac{n}{2}\mathrm{tr}V(\mathbf{A})})\mathrm{d}\mathbf{A} = 0,$$

and so

$$\mathbb{E}\left[\sum_{i\le j}\frac{\partial f_{ij}(\mathbf{A})}{\partial \mathbf{A}_{ij}}\right] = \frac{n}{2}\mathbb{E}\left[\sum_{i\le j}f_{ij}(\mathbf{A})\frac{\partial}{\partial \mathbf{A}_{ij}}\mathrm{tr}V(\mathbf{A})\right] = \frac{n}{2}\mathbb{E}[\mathrm{tr}(f(\mathbf{A})V'(\mathbf{A}))].$$

Here, we want to take $f(\mathbf{A}) = \mathbf{G}(z)$. Note that since V is a polynomial, $Q(x, z) = (V'(z) - V'(x))/(z - x)$ is a bivariate polynomial. Then

$$\mathbf{G}(z)V'(\mathbf{A}) = (\mathbf{A} - z\mathbf{I})^{-1}V'(\mathbf{A}) = V'(z)(\mathbf{A} - z\mathbf{I})^{-1} + Q(\mathbf{A}, z) = V'(z)\mathbf{G}(z) + Q(\mathbf{A}, z).$$

Substituting this into the right-hand side of Stein's Lemma yields

$$\frac{n}{2}\mathbb{E}\mathrm{tr}(\mathbf{G}(z)V'(\mathbf{A})) = \frac{n}{2}(V'(z)\mathbb{E}\mathrm{tr}\mathbf{G}(z) + \mathbb{E}\mathrm{tr}Q(\mathbf{A}, z)).$$

By letting $m_n(z) = \frac{1}{n}\mathrm{tr}\mathbf{G}(z)$ and defining the polynomial expectation $Q_n(z) = \mathbb{E}[\frac{1}{n}\mathrm{tr}Q(\mathbf{A}, z)]$, the right-hand side simplifies:

$$\frac{n}{2}\mathbb{E}\mathrm{tr}(\mathbf{G}(z)V'(\mathbf{A})) = \frac{n^2}{2}(V'(z)\bar{m}_n(z) + Q_n(z)).$$

To evaluate the left-hand side, we apply the matrix chain rule $\mathrm{d}\mathbf{G} = -\mathbf{G}(\mathrm{d}\mathbf{A})\mathbf{G}$ to find the derivative

$$\frac{\partial \mathbf{G}_{ij}}{\partial \mathbf{A}_{ij}} = \begin{cases} -(\mathbf{G}_{ii}\mathbf{G}_{jj} + \mathbf{G}_{ij}^2) & \text{for } i \neq j, \\ -\mathbf{G}_{ii}^2 & \text{for } i = j. \end{cases}$$

Consequently,

$$\sum_{i \leq j} \frac{\partial \mathbf{G}_{ij}}{\partial \mathbf{A}_{ij}} = -\frac{1}{2}((\mathrm{tr}\mathbf{G}(z))^2 + \mathrm{tr}(\mathbf{G}(z))^2).$$

Since $\mathrm{tr}(\mathbf{G}(z)^2) = \frac{\mathrm{d}}{\mathrm{d}z}\mathrm{tr}\mathbf{G}(z) = nm_n'(z)$,

$$\sum_{i \leq j} \frac{\partial \mathbf{G}_{ij}}{\partial \mathbf{A}_{ij}} = -\frac{n^2}{2}\mathbb{E}[m_n(z)^2] - \frac{n}{2}\mathbb{E}[m_n'(z)].$$

Therefore,

$$\mathbb{E}[m_n(z)^2] + V'(z)\bar{m}_n(z) + Q_n(z) + \frac{1}{n}\mathbb{E}[m_n'(z)] = 0.$$

Here, we assert that $m_n(z) \approx m(z) + \mathcal{O}(1/n)$, and so $\mathrm{Var}(m_n(z)) = \mathcal{O}(n^{-2})$. To prove this rigorously is quite challenging, and we refer the reader to [20, Section 8]. Consequently, keeping only first-order terms,

$$\bar{m}_n(z)^2 + V'(z)\bar{m}_n(z) + Q_n(z) + \frac{1}{n}\bar{m}_n'(z) \approx 0. \tag{H.1}$$

Remark H.1. Taking $n \to \infty$ in (H.1) reveals that this large class of random matrix ensembles always exhibits a polynomial relation in the limiting Stieltjes transform m and its argument z that is quadratic in m. △

Motivated by (H.1), we now make the additional heuristic approximation of replacing the expected resolvent $\bar{m}_n$ by the finite-size resolvent m_n, while retaining the leading derivative correction. Observing that the first three terms comprise a bivariate polynomial in z and m, we approximate this polynomial near a point of interest m_* by its second-order expansion:

$$P_2(z, m) = P(z, m_*) + \partial_m P(z, m_*)(m - m_*) + \tfrac{1}{2}\partial_m^2 P(z, m_*)(m - m_*)^2.$$

This leads to our **correction ansatz**:

$$\boxed{P_2(z, m_n(z)) + \frac{1}{n}m_n'(z) = 0} \tag{H.2}$$

Now we consider how this affects the location of branch points.

Proposition H.1. *Let $(z_*, m_*) \in \mathbb{C}^2$ be a branch point of $m(t, z)$ with $m_* := m(t, z_*)$. Define the pullback $(\zeta_*, y_*) = \phi_t^{-1}(z_*, m_*)$. Then, evaluated along the integral curve $y(\zeta)$ satisfying the initial differential equation (H.2), the point (ζ_*, y_*) satisfies*

$$y_*^2 - n(\tau - 1)P_2(\zeta_*, y_*) = 0,$$

followed by the pushforward $z_* = \zeta_* - (\tau - 1)y_*^{-1}$.

Proof. From the differential equation $P_2(\zeta, y(\zeta)) + \frac{1}{n}y'(\zeta) = 0$, we immediately have that

$$\frac{\mathrm{d}y}{\mathrm{d}\zeta} = -nP_2(\zeta, y).$$

Substituting this into the criticality condition $\frac{\mathrm{d}z}{\mathrm{d}\zeta} = 0$ gives

$$\frac{\mathrm{d}z}{\mathrm{d}\zeta} = 1 + (\tau - 1)y^{-2}(-nP_2(\zeta, y)) = 0,$$

from whence the result follows. □

A few remarks about Proposition H.1: assuming that y_* remains finite as $n \to \infty$ and $P \approx P_2$, the condition $y_*^2 - n(\tau - 1)P_2(\zeta_*, y_*) = 0$ implies that

$$\lim_{n\to\infty} P_2(\zeta_*, y_*) = 0.$$

Furthermore, since $P_2(\zeta, y) = -\frac{1}{n}y'$, we find that $y_*^2 + (\tau - 1)y_*' = 0$, and so $y_*' = -\frac{1}{\tau-1}y_*^2$. Applying this to the derivative of (H.2) gives

$$0 = \partial_\zeta P_2 + (\partial_y P_2)y' + \frac{1}{n}y'' = \partial_\zeta P_2 + (\partial_y P_2)\left(-\frac{y_*^2}{\tau - 1}\right) + \frac{1}{n}y_*'' = 0.$$

Therefore, assuming additionally that $n^{-1}y_*'' \to 0$ as $n \to \infty$, we find that

$$y_*^2\partial_y P_2 - (\tau - 1)\partial_\zeta P_2 = 0.$$

Consequently, under these assumptions, we recover both of the original branch point conditions in the limit as $n \to \infty$. Most significantly, however, is that we can show that the finite-size correction suggests a Tracy–Widom-type scaling of $\mathcal{O}(n^{-2/3})$ in the left-edge. Performing a complete analysis in this respect is well beyond the scope of this paper. Instead, we can shortcut the proof by assuming a particular form of $m_n(z_n)$ where each z_n is the branch point of the finite-size correction m_n. In particular, if (z_*, m_*) is the limiting branch point as $n \to \infty$, we may assert that

$$m_n(z_n) \approx m_* + b_n u(a_n^{-1}(z_n - z_*)),$$

for sequences a_n, b_n and a function u.

Given that m_n has n poles located at the eigenvalues, and we should expect that a_n is the appropriate rescaling of the spectral gap as $n \to \infty$, u should contain at least two poles: one for the eigenvalue near the branch point (e.g., the left edge), and another for the next closest eigenvalue. This assumption is sufficient to uniquely determine the limiting behavior, and we will in fact find that u has infinitely many poles to capture *all* of the eigenvalues.

Proposition H.2. *Assume that for each n, the finite-order resolvent m_n satisfies (H.2) and that (z_*, m_*) is a branch point of P in the sense that $P(z_*, m_*) = 0$ and $\partial_m P(z_*, m_*) = 0$. Assume that $\partial_z P(z_*, m_*) \neq 0$ and $\partial_m^2 P(z_*, m_*) \neq 0$. Consider a sequence $(z_n, m_n(z_n))$ defined in terms of coefficients $a_n, b_n > 0$, a parameter ξ, and a function $u_n(\xi)$ by*

$$z_n = z_* + a_n\xi, \qquad m_n(z_n) = m_* + b_n u_n(\xi),$$

where $a_n, b_n \to 0$ as $n \to \infty$. Assume that $u_n(\xi)$ converges locally uniformly away from its poles to a meromorphic limit $u(\xi)$ with at least two poles. Then

$$a_n \asymp n^{-2/3}, \qquad b_n \asymp n^{-1/3}.$$

Proof. Let $x_n = a_n\xi$ and $y_n = b_n u_n(\xi)$ denote the local scaling variables such that $z_n = z_* + x_n$ and $m_n(z_n) = m_* + y_n$. Expanding P around z_* gives

$$P(z_n, m_*) = P(z_*, m_*) + \partial_z P(z_*, m_*)x_n + \mathcal{O}(x_n^2), \tag{H.3a}$$
$$\partial_m P(z_n, m_*) = \partial_m P(z_*, m_*) + \mathcal{O}(x_n), \tag{H.3b}$$
$$\partial_m^2 P(z_n, m_*) = \partial_m^2 P(z_*, m_*) + \mathcal{O}(x_n), \tag{H.3c}$$

Since (z_*, m_*) is a branch point, $P(z_*, m_*) = \partial_m P(z_*, m_*) = 0$. Now define the non-zero constants $A = \frac{1}{2}\partial_m^2 P(z_*, m_*)$ and $B = \partial_z P(z_*, m_*)$. Then substituting (H.3) into the definition of P_2 gives

$$P_2(z_n, m_n(z_n)) = Ay_n^2 + Bx_n + E_n,$$

where $E_n = o(x_n) + \mathcal{O}(x_n y_n) + o(y_n^2)$. But now

$$m_n'(z_n) = \frac{\mathrm{d}(m_* + b_n u_n(\xi))}{\mathrm{d}(z_* + a_n \xi)} = \frac{b_n}{a_n} u_n'(\xi),$$

and so (H.2) implies that

$$Ab_n^2 u_n(\xi)^2 + Ba_n \xi + \frac{b_n}{na_n} u_n'(\xi) + E_n = 0. \tag{H.4}$$

Now let $\theta_n = \max\{|A|b_n^2, |B|a_n, |b_n/(na_n)|\}$. Dividing (H.4) by θ_n gives

$$\alpha_n u_n(\xi)^2 + \beta_n \xi + \gamma_n u_n'(\xi) + \frac{E_n}{\theta_n} = 0,$$

where $\alpha_n = Ab_n^2/\theta_n$, $\beta_n = Ba_n/\theta_n$, and $\gamma_n = b_n/(na_n\theta_n)$. By construction, $\alpha_n, \beta_n, \gamma_n \in [-1, 1]$, and for any n, at least one of these coefficients has an absolute value of 1. By the Bolzano–Weierstrass Theorem, passing to a subsequence if necessary, these coefficients converge to finite limits α, β, γ, respectively, with at least one limit being strictly non-zero. Since E_n is strictly bounded by terms that vanish faster than θ_n, the remainder vanishes in the limit. Therefore, taking the limit as $n \to \infty$,

$$\alpha u(\xi)^2 + \beta \xi + \gamma u'(\xi) = 0. \tag{H.5}$$

If $\gamma = 0$, then $\alpha u^2 + \beta \xi = 0$. This implies that $u(\xi) = \pm\sqrt{-\frac{\beta}{\alpha}\xi}$, which has a branch cut and is without poles, contradicting the hypotheses[3]. If $\alpha = 0$ then $\beta\xi + \gamma u'(\xi) = 0$. Integrating this reveals that $u(\xi) = -\frac{\beta}{2\gamma}\xi^2 + C$ for some constant C which has no poles. If $\beta = 0$, the equation becomes $\alpha u^2 + \gamma u' = 0$. Solving this differential equation yields $u(\xi) = \frac{\gamma}{\alpha(\xi - C)}$ for some constant $C > 0$. This is meromorphic, but possesses only one pole[4]. Therefore, α, β, γ must all be non-zero. For this to occur, it must be the case that

$$b_n^2 \asymp a_n \asymp \frac{b_n}{na_n}.$$

Now, since $b_n^2 \asymp a_n$, this implies that $b_n^2 \asymp \frac{b_n}{nb_n^2} = \frac{1}{nb_n}$ and so $b_n^3 \asymp n^{-1}$. Immediately this implies that $b_n \asymp n^{-1/3}$ and $a_n \asymp n^{-2/3}$. It remains only to identify that the limiting u satisfies the condition of being meromorphic with at least two poles. Note that the limiting equation (H.5) is a Riccati equation, and so by applying the transformation

$$u(\xi) = \frac{\gamma}{\alpha} \cdot \frac{v'(\xi)}{v(\xi)},$$

and then the linear rescaling $w = -(\alpha\beta\gamma^{-2})^{1/3}\xi$, the differential equation of $v(\psi)$ becomes the Airy equation $v''(w) - wv(w) = 0$. The solutions for v are linear combinations of the Airy functions Ai and Bi in w [15, Section 9.2(i)]. For the relevant real solution, v is a nontrivial real linear combination of these functions and has infinitely many real zeros, which are simple [15, Section 9.9]. Therefore, $u(\xi)$ is meromorphic with infinitely many poles on the real axis, and in particular satisfies the required condition. □

The finite-size correction becomes particularly relevant in our diffusion model example, as free decompression predicts that the left-edge should go to zero, even though the matrix in question is

[3] This represents the regime where the finite-order correction is removed entirely, and so only describes the asymptotic macroscopic square-root behavior near the edge of the spectral density.

[4] This corresponds to the regime where the spectral gap has been stretched to the point where only one pole at the location of the left edge remains.

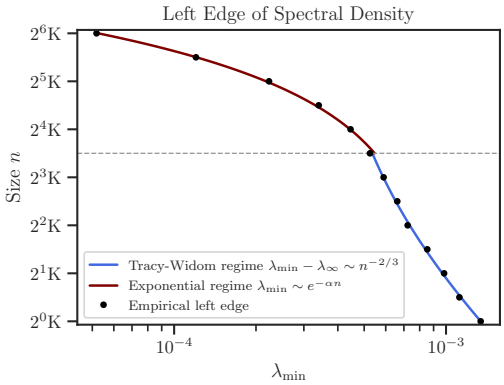

Figure H.1: Simulated (black) smallest eigenvalue $\lambda_{\min}$ of the diffusion model matrix example, with Tracy–Widom (blue) predicted decay at rate $\asymp n^{-2/3}$ and exponential (red) predicted decay at rate $\asymp e^{-\alpha n}$ for fitted α.

strictly positive-definite. Fortunately, the decay rate in the intermediate regime appears to be well-captured by the Tracy–Widom-type scaling behavior predicted in Proposition H.2; see Figure H.1. Unfortunately, as n grows larger, the left-edge begins to obey an exponential regime instead. This regime is not captured by the finite-size correction. Instead, over the larger matrix sizes considered here, the observed behavior is well described by an exponential fit. For related probabilistic bounds on the smallest singular values of random matrices with independent subgaussian entries, see Rudelson & Vershynin [44]. We do not expect that simple corrections to the free decompression operation will adequately capture this behavior at these size scales.

Appendix I Random Matrix Models

This section describes the random matrix models and empirical datasets used in Section 4. We begin with the analytically tractable compound free Poisson and free Lévy laws in Appendices I.1 and I.2, whose Stieltjes transforms satisfy explicit algebraic relations and therefore provide controlled benchmarks for free decompression. We then discuss the Pennington–Bahri Hessian model in Appendix I.3, which is a neural-network-inspired special case of the algebraic free Lévy family. Finally, we describe the empirical Hessian and diffusion-model matrices in Appendices I.4 and I.5, which serve as large-scale data-driven examples.

I.1 Compound Free Poisson Law

The compound free Poisson law [36, p. 218] is a convenient multi-bulk testbed whose Stieltjes transform satisfies an explicit algebraic relation. It may be viewed as a natural generalization of the Marchenko–Pastur law, with the usual Marchenko–Pastur case recovered as the simplest one-parameter member of the family.

This distribution admits a natural random-matrix realization [48, Section 22.5.4]. Let $\mathbf{Z} \in \mathbb{R}^{p\times n}$, with $p/n \to \lambda \in (0,\infty)$, have iid centered entries with variance 1. Suppose $\mathbf{\Sigma}$ is deterministic positive semidefinite with empirical spectral distribution converging weakly to a compactly supported probability measure ν. Then the companion sample covariance matrix

$$\underline{\mathbf{S}} := \frac{1}{n}\mathbf{X}^\intercal\mathbf{X}, \qquad \mathbf{X} := \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{Z}, \tag{I.1}$$

has a limiting eigenvalue distribution given by the compound free Poisson law with rate λ and jump distribution ν. When $\lambda < 1$, the companion matrix $\underline{\mathbf{S}}$ has at least $n - p$ zero eigenvalues. The limiting law has mass $\max\{1 - \lambda(1 - \nu(\{0\})), 0\}$ at the origin, which reduces to $\max\{1 - \lambda, 0\}$ when $\nu(\{0\}) = 0$.

The limiting eigenvalue distribution above can be characterized through its R-transform described below. Let ν denote the jump distribution of the compound free Poisson law. For the closed-form

algebraic examples used here, we take ν to be a finite-atom probability measure,

$$\nu(\mathrm{d}x) \coloneqq \sum_{i=1}^{r} w_i \delta_{t_i}(\mathrm{d}x), \tag{I.2}$$

where δ_{t_i} denotes a unit point mass at $x = t_i$, the atoms are distinct, $w_i > 0$, and $\sum_{i=1}^{r} w_i = 1$. The R-transform is then

$$R(w) = \lambda \int_{\mathbb{R}} \frac{x}{1 - xw}\, \nu(\mathrm{d}x) = \lambda \sum_{i=1}^{r} w_i \frac{t_i}{1 - t_i w}, \tag{I.3}$$

where $\lambda > 0$ denotes the intensity, or rate. Let $m(z)$ be the Stieltjes transform of the law. The standard relation between the Stieltjes transform and the R-transform reads; see, e.g., [36, Theorem 12.7],

$$z = R(w) + \frac{1}{w}, \qquad w = -m(z), \tag{I.4}$$

where the physical branch of m is selected by the Herglotz condition $\Im m(z) > 0$ for $z \in \mathbb{C}^+$.

When ν is finitely supported as in (I.2), multiplying (I.4) by the common denominator $m \prod_{i=1}^{r}(1 + t_i m)$ yields an explicit algebraic constraint

$$P(z, m) = 0, \tag{I.5}$$

where P has degree $r + 1$ in m. In the special case $r = 1$, this relation reduces to the familiar quadratic equation for the Marchenko–Pastur Stieltjes transform.

Example I.1. For the two-atom case $r = 2$ used in our experiments, P is cubic:

$$P(z, m) \coloneqq a_3(z)m^3 + a_2(z)m^2 + a_1(z)m + a_0(z) = 0, \tag{I.6}$$

with coefficients

$$\begin{aligned} a_3(z) &= z\, t_1 t_2, \\ a_2(z) &= z(t_1 + t_2) + (1 - \lambda) t_1 t_2, \\ a_1(z) &= z + (t_1 + t_2) - \lambda(w_1 t_1 + w_2 t_2), \\ a_0(z) &= 1, \end{aligned} \tag{I.7}$$

where $w_2 \coloneqq 1 - w_1$. Hence, $m(z)$ is a three-sheet algebraic function, providing a controlled benchmark for continuation and evolution in multi-bulk settings. △

For the numerical example using the compound free Poisson law, we use $\lambda = 0.1$, $(t_1, t_2) = (2, 5\frac{1}{2})$, and $(w_1, w_2) = (0.75, 0.25)$. This produces a cubic algebraic Stieltjes transform, so in the numerical fit we use $(\deg_m, \deg_z) = (3, 1)$ and impose only the zeroth-moment normalization constraint.

The cubic compound-free-Poisson example also provides a useful visualization of the sheet structure underlying algebraic Stieltjes transforms. Since (I.6) is cubic, each point z away from branch points has three algebraic roots, denoted m_1, m_2, m_3, shown in the first row of Figure I.1. The physical Stieltjes transform is the branch selected by the Herglotz property and the normalization at infinity; here, this branch is m_1. The remaining roots are nonphysical on $\mathbb{C}^+$, but they are not spurious: they encode the analytic continuation of the physical branch after crossing real branch cuts. For the physical branch, these cuts coincide with the support intervals of the density on the real axis. The lower panels of Figure I.1 show the half-plane gluings that continue the physical branch through these cuts, and Figure I.2 shows the same continuation geometry as a three-sheeted algebraic surface. In this surface view, the cuts are not endpoints of the Stieltjes transform, but seams where the physical sheet is attached to nonphysical sheets. Thus, crossing a cut does not terminate the analytic continuation; it moves the continuation from one sheet of the spectral curve to another. This is the main geometric advantage of working with the full algebraic relation $P(z, m) = 0$: the physical Stieltjes transform is only one sheet of a larger connected algebraic object, while the other sheets record its analytic continuations.

I.2 Free Lévy Law

The compound free Poisson law can be generalized by adding a drift and a semicircular, or free Gaussian, component. This yields the free Lévy law

$$\nu_{\mathrm{FL}} = \nu_{\delta_a} \boxplus \nu_{\mathrm{SC}_{\sigma^2}} \boxplus \nu_{\mathrm{CFP}(\lambda,\nu)},$$

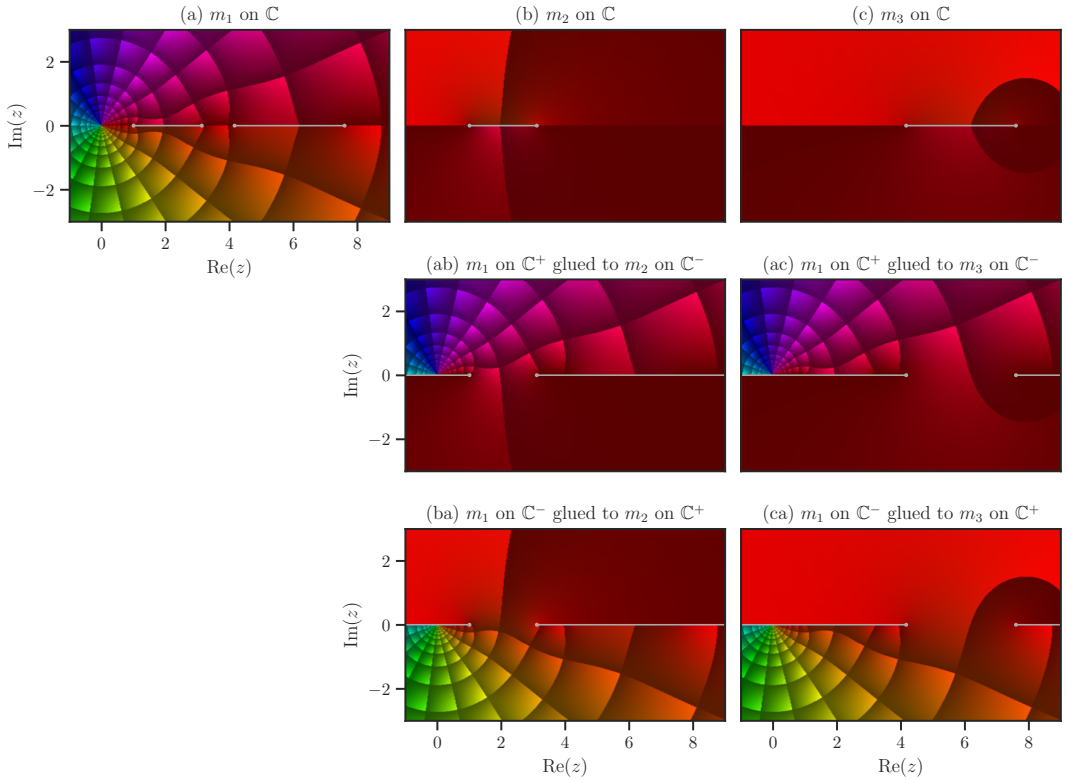

Figure I.1: Branches and analytic continuations of the cubic compound free Poisson Stieltjes transform in (I.6). The first row shows the three algebraic branches m_1, m_2, m_3 on the z-plane, with branch cuts marked on the real axis. The physical branch is m_1. The lower panels show the four possible half-plane gluings involving the physical sheet: for example, (ab) denotes m_1 on $\mathbb{C}^+$ glued to m_2 on $\mathbb{C}^-$, while (ba) denotes the opposite gluing. Across a shared cut, the individual branches are discontinuous, but the glued half-planes form a continuous analytic continuation. Colors encode the complex value of the branch using the domain-coloring convention of Appendix A.

where δ_a denotes the deterministic shift $X \mapsto X + a$, SC_{σ^2} is Wigner's semicircle law with variance σ^2, and $\mathrm{CFP}(\lambda, \nu)$ is the compound free Poisson law with rate λ and jump distribution ν, as in Appendix I.1. In full generality, the free Lévy–Khintchine representation characterizes freely infinitely divisible distributions: when ν is a Lévy measure, the corresponding law is freely infinitely divisible, and conversely every freely infinitely divisible law admits such a representation [36, Theorem 13.8]; see also Benaych-Georges [4] for random-matrix constructions associated with freely infinitely divisible laws. In this work, we restrict to the finite-atom case (I.2), which preserves an algebraic Stieltjes transform while still producing a broad family of multi-bulk distributions.

The free Lévy law also admits a natural random-matrix realization by combining the drift, Gaussian, and jump components; see Benaych-Georges [4], Burda et al. [11] for related random-matrix realizations of free Lévy and freely infinitely divisible laws. Let $\mathbf{W}$ be an $n \times n$ Wigner matrix with independent centered entries of variance $1/n$, up to symmetry, and let $\underline{\mathbf{S}}$ be the companion sample covariance matrix in (I.1), whose limiting law is $\mathrm{CFP}(\lambda, \nu)$. Assuming $\mathbf{W}$ and $\underline{\mathbf{S}}$ are independent, the additive model

$$\mathbf{A} \coloneqq a\mathbf{I} + \sigma\mathbf{W} + \underline{\mathbf{S}}$$

has an empirical spectral distribution converging to ν_{FL} under standard assumptions.

The R-transform of ν_{FL} is the sum of the R-transforms of its free additive components:

$$R_{\mathrm{FL}}(w) = R_{\delta_a}(w) + R_{\mathrm{SC}_{\sigma^2}}(w) + R_{\mathrm{CFP}(\lambda,\nu)}(w).$$

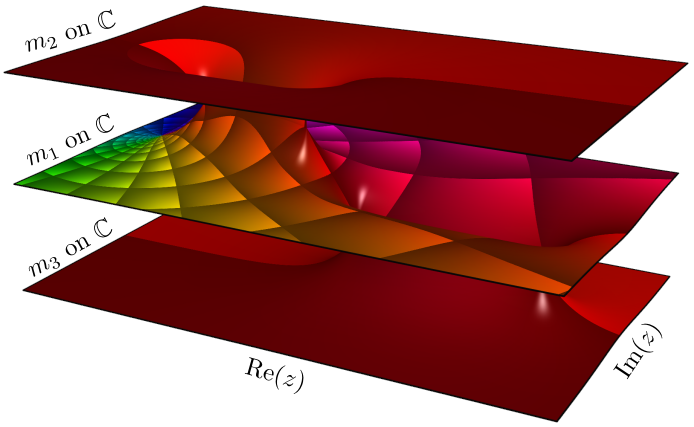

Figure I.2: Three-sheet visualization of the cubic compound free Poisson spectral curve from Figure I.1. The middle sheet represents the physical branch m_1, while the upper and lower sheets represent the non-physical branches m_2 and m_3. The sheets are vertically separated for visualization and are brought together along shared branch cuts to illustrate analytic continuation between branches. This geometric view explains why representing the Stieltjes transform as one sheet of a larger algebraic curve is useful: crossing a branch cut does not terminate the function, but moves the continuation to another sheet.

Here, $R_{\delta_a}(w) = a$, $R_{\mathrm{SC}_{\sigma^2}}(w) = \sigma^2 w$, and, for the finite-atom jump distribution (I.2), $R_{\mathrm{CFP}(\lambda,\nu)}$ is given by (I.3). Hence,

$$R_{\mathrm{FL}}(w) = a + \sigma^2 w + \lambda \sum_{i=1}^{r} w_i \frac{t_i}{1 - t_i w}, \tag{I.8}$$

which is a rational function.

Using the relation between the R-transform and the Stieltjes transform in (I.4), with $w = -m(z)$, we obtain the self-consistent equation

$$-\frac{1}{m(z)} = z - a + \sigma^2 m(z) - \lambda \sum_{i=1}^{r} w_i \frac{t_i}{1 + t_i m(z)}.$$

Multiplying by the common denominator $m \prod_{i=1}^{r}(1 + t_i m)$ yields the implicit algebraic constraint $P(z, m) = 0$, where

$$P(z,m) \coloneqq \Big(\sigma^2 m^2 + (z-a)m + 1\Big) \prod_{i=1}^{r}(1 + t_i m) - \lambda m \sum_{i=1}^{r} w_i t_i \prod_{j \neq i}(1 + t_j m). \tag{I.9}$$

For $\sigma = 0$, the polynomial (I.9) has degree $r + 1$ in m and reduces to the compound free Poisson relation, up to the drift shift a. When $\sigma > 0$, the additional semicircular term increases the degree to $r + 2$.

Example I.2. For the two-atom case $r = 2$ and $\sigma > 0$, (I.9) becomes a quartic polynomial in m,

$$P(z,m) = a_4(z)m^4 + a_3(z)m^3 + a_2(z)m^2 + a_1(z)m + a_0(z), \tag{I.10}$$

with coefficients

$$\begin{aligned}
a_4(z) &= \sigma^2 t_1 t_2, \\
a_3(z) &= (z-a)t_1 t_2 + \sigma^2(t_1 + t_2), \\
a_2(z) &= (z-a)(t_1 + t_2) + (1-\lambda)t_1 t_2 + \sigma^2, \\
a_1(z) &= (z-a) + (t_1 + t_2) - \lambda(w_1 t_1 + w_2 t_2), \\
a_0(z) &= 1.
\end{aligned} \tag{I.11}$$

In particular, for $\sigma = 0$, the leading coefficient a_4 vanishes and (I.10) reduces to the cubic compound free Poisson law with drift. △

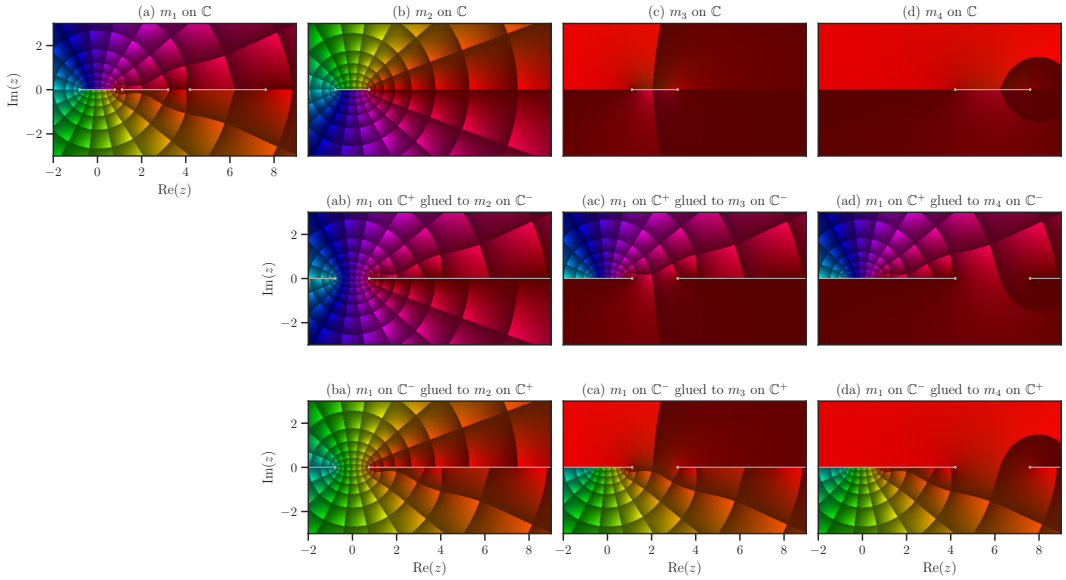

Figure I.3: Branches and analytic continuations of the quartic free Lévy Stieltjes transform. The first row shows the four algebraic branches of $P(z, m) = 0$ on the z-plane, with branch cuts marked on the real axis. The lower panels show half-plane gluings between the physical sheet and the non-physical sheets; for example, (ad) denotes the upper half-plane of the physical sheet glued to the lower half-plane of the branch shown in panel (d), while (da) denotes the opposite gluing. Across a shared cut, the separate branches are discontinuous, but the glued half-planes form a continuous analytic continuation. Colors encode the complex value of the branch using the domain-coloring convention of Appendix A.

For the numerical example below, we use $\lambda = 0.1$, $(t_1, t_2) = (2, 5\frac{1}{2})$, $(w_1, w_2) = (0.75, 0.25)$, $a = 0$, and $\sigma = 0.4$. This produces a quartic algebraic Stieltjes transform, so in the numerical fit we use $(\deg_m, \deg_z) = (4, 1)$ and impose only the zeroth-moment normalization constraint. Figure I.3 visualizes the four sheets and their analytic continuations. As in the compound free Poisson example, the individual branches are discontinuous across their cuts, but the appropriate half-plane gluings form continuous analytic continuations on the full spectral curve.

We next use this law as a controlled free-decompression benchmark. A matrix realization of size $n = 2^4\mathrm{K}$ is generated from the free Lévy law above, and a $2^2\mathrm{K}$-sized principal submatrix is used as the initial compressed matrix. The algebraic curve is fitted from the initial empirical spectrum and then evolved by free decompression to the full size.

Figure I.4 shows the resulting density evolution. Panel (a) compares the initial empirical histogram with the density recovered from the fitted algebraic curve. Panel (b) shows the free-decompression flow through intermediate sizes, and panel (c) compares the final FD prediction at $n = 2^4\mathrm{K}$ with the empirical spectrum of the full matrix. The initial density has one bulk, while the decompressed density splits into three bulks. The two right bulks arise from the compound free Poisson component, and the smallest rightmost bulk is resolved despite having much smaller mass and height than the dominant component.

The same example also illustrates the direct evolution of spectral edges. Rather than locating edges from the reconstructed density at each time, we evolve the branch-point conditions directly as described in Appendix G.1. Figure I.5 shows the edge trajectories from $n = 2^2\mathrm{K}$ to $n = 2^4\mathrm{K}$. The density field computed by free decompression is shown in the background, while the black curves show the independently evolved spectral edges. Starting from one initial bulk, two cusp events create four additional edges, producing three bulk intervals at the final size. The agreement between the black edge curves and the boundary of the nonzero-density region confirms that the edge evolution and density evolution are consistent.

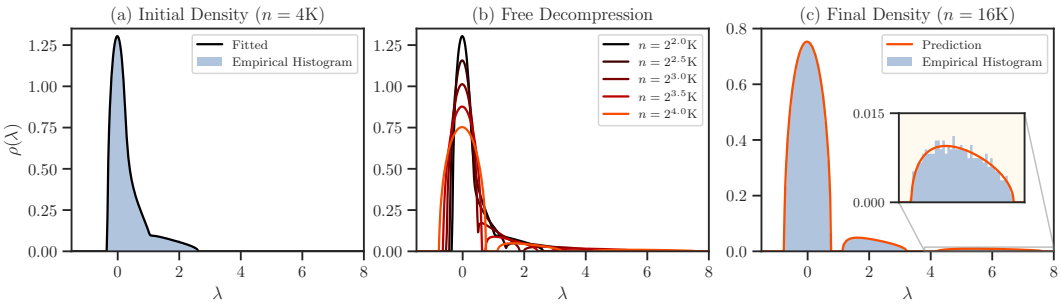

Figure I.4: Free decompression of the free Lévy benchmark. A matrix realization of size $n = 2^4$K is generated from the quartic free Lévy law, and a 2^2K-sized principal submatrix is used as the initial compressed matrix. Panel (a) shows the initial empirical histogram together with the density recovered from the fitted algebraic curve. Panel (b) shows the FD evolution through intermediate sizes. Panel (c) compares the final FD prediction with the empirical histogram of the full matrix. The initially single-bulk density splits into three bulks, including a small rightmost component that is accurately recovered by the evolved spectral curve.

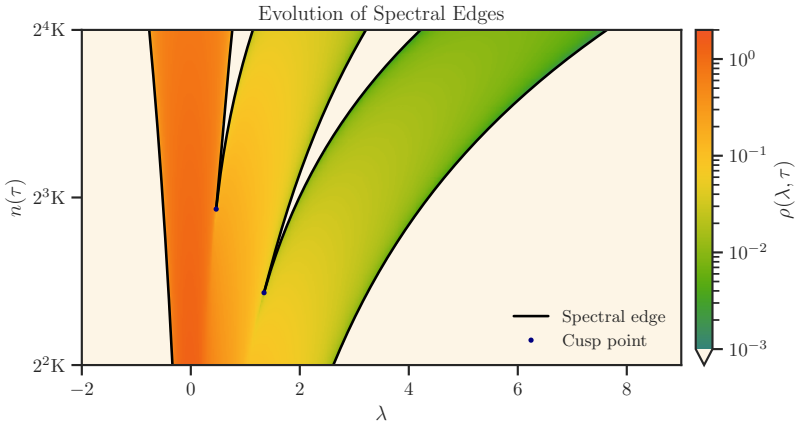

Figure I.5: Evolution of spectral edges for the free Lévy benchmark. The vertical axis is the matrix size $n(\tau)$, and the horizontal axis is the spectral coordinate λ. Black curves show spectral edges evolved directly from the branch-point equations in Appendix G.1, while the background color shows the density $\rho(\lambda, \tau)$ computed by free decompression on a fine time grid. The initial spectrum has one bulk; two cusp events create new pairs of edges, leading to three bulk intervals at the final size. The evolved edges align with the boundary of the nonzero-density region, showing consistency between direct edge evolution and full density evolution.

Sensitivity to Polynomial Degree. The free Lévy example provides a controlled setting for studying degree selection, since its Stieltjes transform has a known minimal bidegree $(d_z, s) = (1, 4)$. Keeping the initial submatrix and numerical settings fixed, we repeat the experiment with different polynomial degrees, as reported in Table I.1. We find that the dominant failure mode is structural under-specification. The pairs $(1, 2)$ and $(1, 3)$ fail to reproduce the correct three-bulk topology. This agrees with the necessary degree condition $(2s - 1)d_z \geq 2k$ for k bulk intervals, derived in Appendix D.2. Below this bound, the polynomial lacks sufficient degree to represent the required spectral cuts.

In contrast, all tested models from the minimal $(1, 4)$ through the larger $(4, 4)$ recover the three-bulk structure accurately. The small, non-monotone variations in error are consistent with numerical fluctuations from fitting and continuation, rather than a systematic penalty from modest over-specification. This is expected since larger polynomial classes still contain the minimal curve. Thus, the exact minimal degree need not be known in advance. We can begin with the smallest structurally admissible pair and verify stability across nearby larger orders. Excessively high degrees may nevertheless introduce conditioning and root-tracking difficulties.

Table I.1: Sensitivity to polynomial degrees for the free Lévy benchmark. All fits use the same 2^2K initial submatrix and are decompressed to 2^4K, varying only (d_z, s). Mass errors are measured before density normalization, while KS and W_1/L compare normalized recovered densities with the exact target density.

		Mass Error		Distributional Distance	
(d_z, s)	Specification	Initial	Target	KS	W_1/L
$(1, 2)$	Under	70.798%	85.301%	90.048%	28.493%
$(1, 3)$	Under	0.008%	0.336%	8.777%	1.634%
$(1, 4)$	Minimal exact	0.004%	0.023%	0.061%	0.009%
$(1, 5)$	Over	0.001%	0.745%	0.712%	0.196%
$(1, 6)$	Over	0.002%	0.127%	0.087%	0.023%
$(2, 4)$	Over	0.015%	0.029%	0.052%	0.015%
$(4, 4)$	Over	0.003%	0.162%	0.176%	0.029%

I.3 Pennington–Bahri Model

The Pennington–Bahri model [40] is a random-matrix model for Hessian matrices of single-layer neural networks with ReLU activations under simplifying assumptions. In this model, the network is assumed to be trained to mean-squared error training loss $n_2\varepsilon$ over m data points, where n_2 is the number of data labels. Writing $\mathbf{J}$ for the network Jacobian, the Hessian decomposes into two additive components

$$\mathbf{H} = \mathbf{H}_0 + \mathbf{H}_1,$$

where $\mathbf{H}_0 = \mathbf{J}\mathbf{J}^\intercal/m$ is the Jacobian component and $\mathbf{H}_1$ is the network Hessian term. Under the simplifying assumption that the entries of $\mathbf{J}$ and $\mathbf{H}_1$ are independent Gaussian variables, $\mathbf{H}_0$ and $\mathbf{H}_1$ have Wishart and Wigner limiting distributions, respectively.

Consequently, the limiting Hessian distribution can be written as an additive free convolution

$$\nu_H = \nu_{H_0} \boxplus \nu_{H_1}.$$

Equivalently, its R-transform is the sum of the R-transforms of the two components,

$$R_H(w) = R_{H_0}(w) + R_{H_1}(w) = \frac{1}{1 - \lambda w} + 2\varepsilon w,$$

where λ is the aspect ratio of the Jacobian. This places the Pennington–Bahri law within the finite-atom free Lévy family of Appendix I.2: it corresponds to zero drift, one compound-free-Poisson support point, and a nonzero semicircular component. In particular, its Stieltjes transform satisfies a cubic algebraic relation.

To perform decompression of this model, we set $\varepsilon = 0.1$ and $\lambda = \frac{3}{4}$, and sampled a 6000×6000 matrix from the Pennington–Bahri law. From this matrix, a 2000×2000 submatrix was uniformly sampled, and free decompression was applied to recover an estimate of the spectral distribution of the original matrix. Figure I.6 shows the result and should be compared with Figure 1 of Pennington & Bahri [40]. This example demonstrates the potential of free decompression as a tool for estimating spectral properties of large Hessian models from smaller submatrices that can be explicitly formed or diagonalized.

I.4 Hessian of Two-Layer Network

Suppose that $\mathbf{X} \in \mathbb{R}^{n\times p}$ is a matrix of n inputs in $\mathbb{R}^p$. Let $f(\boldsymbol{w}, \mathbf{X}) \in \mathbb{R}^{n\times m}$ denote the network output for inputs $\mathbf{X}$ with weights $\boldsymbol{w} = [w_i]_{i=1}^N$. To investigate the spectral behavior of Hessian matrices for f at initialization, we let $\boldsymbol{w}^*$ denote a target collection of weights comprised of i.i.d elements generated from a normal distribution $\mathcal{N}(0, \sigma^2)$. The choice of i.i.d elements here is to mimic standard initialization schemes, including the popular He initialization [21]. Let $\mathbf{E} \in \mathbb{R}^{n\times m}$ denote a collection of residual errors, where $E_{ij} \sim \mathcal{N}(0, 2\epsilon)$ are also independently generated for some $\epsilon > 0$. The Hessian matrix to be considered is given by $\mathbf{H} = [H_{ij}]_{i,j=1}^N$ where

$$H_{ij} = \frac{1}{2n} \frac{\partial^2}{\partial w_i \partial w_j} \|f(\boldsymbol{w}, \mathbf{X}) - \mathbf{Y}\|_F^2, \qquad \mathbf{Y} = f(\boldsymbol{w}^*, \mathbf{X}) + \mathbf{E}.$$

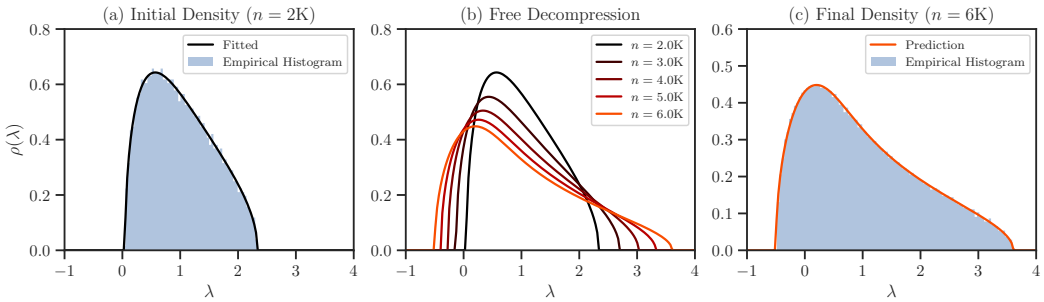

Figure I.6: Free decompression for the Pennington–Bahri Hessian model. A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with $n = 6000$ is generated from the Pennington–Bahri law with $\lambda = 3/4$ and $\varepsilon = 1/10$, and a submatrix of size $n_0 = 2000$ is used as the initial compressed matrix. Panel (a) shows the empirical spectrum of the submatrix together with the fitted density. Panel (b) shows the free-decompression flow from n_0 to n. Panel (c) compares the final FD prediction with the empirical spectrum of the full matrix.

Recall that $\mathbf{H}$ is symmetric and therefore has real eigenvalues. However, $\mathbf{H}$ is not necessarily positive-definite, and indeed, often possesses negative eigenvalues corresponding to directions of descent. The fraction of negative eigenvalues is called the *normalized index* and is a prominent object of study in random matrix analyses of deep learning models [40]:

$$\text{index} = \frac{\#\{\lambda_i(\mathbf{H}) < 0\}}{N}, \qquad \mathbf{H} \in \mathbb{R}^{N\times N}.$$

At scale, when the Hessian becomes so large that it cannot be stored in memory, the index becomes a prime target to be computed using free decompression.

As a concrete example, we consider a two-layer feedforward neural network with a ReLU activation function where

$$f(\mathbf{W}_1, \mathbf{W}_2, \mathbf{X}) = \phi(\mathbf{X}\mathbf{W}_1)\mathbf{W}_2, \qquad \mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{p\times p},$$

where $\phi(x) = \max\{0, x\}$ is the ReLU activation function applied elementwise. Our input data $\mathbf{X}$ is constructed from the CIFAR-10 dataset [28]. Let $\mathbf{X}_{\text{raw}} \in \mathbb{R}^{60000\times 1024}$ represent the total batch of 60,000 training images from CIFAR-10 of size 32×32 rendered in grayscale and flattened. To control the dimensionality, we truncate this matrix to its $n \times p$ upper-left corner, called $\tilde{\mathbf{X}}$. In our tests, we use $n = 512$ and $p = 128$. This matrix is centered by:

$$\mathbf{X}_c = \tilde{\mathbf{X}} - \frac{1}{p}\tilde{\mathbf{X}}\mathbf{1}\mathbf{1}^\intercal,$$

where $\mathbf{1}$ is a p-dimensional vector of ones. The empirical covariance matrix of this centered data is given by $\mathbf{\Sigma} = \frac{1}{p-1}\mathbf{X}_c\mathbf{X}_c^\intercal$. To ensure the data is isotropic as per Pennington et al. [41], we apply Zero Component Analysis (ZCA) whitening. Using the eigendecomposition $\mathbf{\Sigma} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\intercal$, we construct the ZCA whitening matrix

$$\mathbf{Y}_{\text{ZCA}} = \mathbf{V}\mathbf{\Lambda}^{-1/2}\mathbf{V}^\intercal,$$

where the inverse square root of the eigenvalue matrix $\mathbf{\Lambda}$ is bounded below for numerical stability. The final network input is the whitened data $\mathbf{X} = \mathbf{Y}_{\text{ZCA}}\mathbf{X}_c$. Finally, we choose $\sigma^2 = 1/p$, and $\epsilon = 0.1$.

Figure I.7 shows the normalized Hessian index estimated by free decompression across matrix sizes, together with the empirical index computed from direct eigendecomposition at the available sizes.

Table I.2 complements this comparison by reporting runtime and scalar accuracy metrics. Since the FD output is a continuous spectral measure, whereas the empirical spectrum is discrete, we use distributional metrics that compare probability measures directly. In particular, the normalized Wasserstein distance W_1/L provides a scale-free discrepancy that accounts for the geometry of the spectral axis and does not require an additional KDE or binning convention [19, 38]. The Hessian's index error remains small across all sizes, while the direct eigendecomposition time increases rapidly with $O(n^3)$. In contrast, free decompression only requires the eigendecomposition of the initial submatrix (at size $n_0 = 2^{12}$), followed by a comparatively inexpensive evolution step.

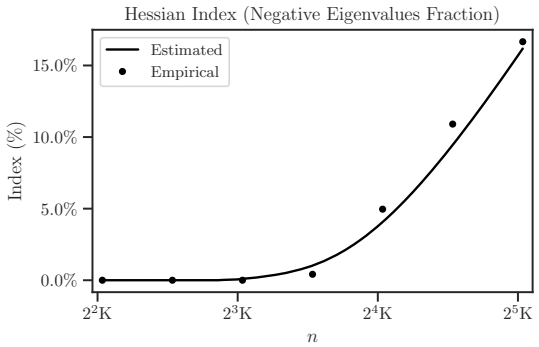

Figure I.7: Normalized Hessian index across matrix sizes. The curve shows the index estimated by free decompression, while markers show the empirical index computed directly from empirical eigenvalues at the available sizes.

Table I.2: Comparison of direct empirical eigenvalues versus free decompression (FD) for the Hessian example. The Direct column reports the process time for computing the eigenvalues of the Hessian at each size. In the FD column, the first term is the time to compute eigenvalues of the initial 2^{12}-sized submatrix, and the second term is the free-decompression evolution time. The index error is the absolute error in the normalized Hessian index, namely the fraction of negative eigenvalues. KS denotes the Kolmogorov–Smirnov distance between the empirical spectrum and samples drawn from the FD reconstruction. The normalized Wasserstein distance W_1/L is the Wasserstein-1 distance divided by the empirical spectral width $L = \lambda_{\max} - \lambda_{\min}$ at each size. The last two columns report relative errors of the first moment μ_1 and standard deviation σ. All accuracy metrics are multiplied by 100 and reported as percentages.

Size	Process Time (sec)		Accuracy Metrics			Moments Rel. Error	
n	Direct	FD (ours)	Index Error	KS	W_1/L	$\Delta\mu_1/\mu_1$	$\Delta\sigma/\sigma$
$2^{12.0}$	0.9	$0.9 + 0.0$	0.0%	0.4%	0.0%	0.3%	0.3%
$2^{12.5}$	91.9	$0.9 + 4.7$	0.0%	1.2%	0.3%	3.5%	6.1%
$2^{13.0}$	271.9	$0.9 + 5.4$	0.1%	3.0%	0.7%	8.5%	0.9%
$2^{13.5}$	759.4	$0.9 + 7.8$	0.6%	2.9%	0.7%	9.0%	0.1%
$2^{14.0}$	2154.4	$0.9 + 11.1$	0.8%	2.4%	0.5%	6.8%	2.0%
$2^{14.5}$	4755.3	$0.9 + 15.7$	1.3%	2.1%	0.3%	2.1%	4.9%
$2^{15.0}$	14023.5	$0.9 + 22.2$	0.4%	2.8%	0.3%	3.5%	13.0%

For this empirical Hessian example, we fit a polynomial with $(\deg_m, \deg_z) = (8, 4)$ and impose moment constraints up to degree two.

I.5 Details of the Diffusion Model Example

The diffusion model that we consider is based on a random features neural network model, which is by now well studied [29] and, in the diffusion-model setting, was introduced in George et al. [18]. A key advantage of this model is that, despite its heavy simplification, it still exhibits memorization–generalization phase transitions that are characteristic of full diffusion models with modern architectures [6, 53]. In this simplified setting, the spectral distribution of the expected activation matrix is tied to the transition between early training, generalization, and memorization behavior [10]. Our goal is to demonstrate that this spectral distribution, even when highly multi-scale and multi-bulk, can be accurately estimated from a sub-sampled model.

To this end, a $64{,}000 \times 64{,}000$ matrix was generated by sampling $m = 800$ i.i.d data-points $\boldsymbol{x}^j$, with each $\boldsymbol{x}^j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ for $d = 100$. The random features were generated by sampling $64{,}000$ i.i.d

Table I.3: Atom-like behavior of the leftmost bulk in the diffusion-model spectrum. The empirical mass is the mass of the leftmost bulk detected from the histogram, while the atom-law column reports the prediction $w(n) = 1 - (1 - w_0)\, n_0/n$ from Proposition G.5, with $w_0 = 0.8$ and $n_0 = 2^2\mathrm{K}$. The center-of-mass column reports the empirical center of the same bulk, and its error is measured relative to the reference center after the leftmost component has separated from the neighboring bulk.

Size	Left Bulk Mass			Left Bulk Center of Mass	
n	Empirical	Atom Law (ours)	Error	Empirical	Error
$2^2\mathrm{K}$	80.0%	80.0%	0.0%	1.87×10^{-3}	3.7%
$2^3\mathrm{K}$	88.8%	90.0%	1.3%	1.80×10^{-3}	0.0%
$2^4\mathrm{K}$	94.4%	95.0%	0.6%	1.80×10^{-3}	0.1%
$2^5\mathrm{K}$	97.2%	97.5%	0.3%	1.81×10^{-3}	0.4%
$2^6\mathrm{K}$	98.6%	98.8%	0.2%	1.81×10^{-3}	0.4%

random feature weights $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d_f})$, and computing the expected diffusion activation matrix

$$\mathbf{U} = \frac{1}{m} \sum_{j=1}^{m} \mathbb{E}\left[\sigma\left(\frac{\mathbf{W}\boldsymbol{x}_t^j}{\sqrt{d}} \right) \sigma\left(\frac{\mathbf{W}\boldsymbol{x}_t^j}{\sqrt{d}} \right)^{\intercal} \right].$$

Here, the nonlinearity σ is taken to be the hyperbolic tangent function, and

$$\boldsymbol{x}_t^j(\xi) = e^{-t}\boldsymbol{x}^j + \sqrt{\Delta_t}\,\xi, \qquad \xi \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d),$$

denotes the Ornstein–Uhlenbeck evolved sample at diffusion time $t = 0.01$. The expectation in $\mathbf{U}$ is approximated by Monte Carlo using 1024 samples.

We fit an algebraic spectral curve $P(z, m) = 0$ to the empirical Stieltjes transform of the $n_0 = 2^2\mathrm{K}$ submatrix, using degrees $(\deg_m, \deg_z) = (7, 5)$ and moment constraints up to degree four. The fitted curve is then evolved by free decompression from $n_0 = 2^2\mathrm{K}$ to $n = 2^6\mathrm{K}$. For the edge diagnostics, we do not infer edges from the reconstructed density alone; instead, we evolve branch-point candidates directly from the spectral curve using Proposition G.1, and identify the corresponding physical spectral edges as described in Appendix G.1. The evolution detects one cusp, corresponding to the splitting of the leftmost component into a spike-like bulk and a neighboring small-eigenvalue bulk, consistent with the cusp mechanism in Proposition G.2.

The spectrum of $\mathbf{U}$ has a pronounced multi-bulk structure over several orders of magnitude. In particular, the leftmost component is an extremely narrow, spike-like bulk carrying most of the spectral mass: it contains about 80% of the mass at $n = 4000$, increasing to about 98% at $n = 64{,}000$, while occupying only a very small interval near the origin. Importantly, in our computation this component is not modeled as an atom; it is treated as an absolutely-continuous bulk and evolved by the same free-decompression procedure as the rest of the spectrum. Nevertheless, its mass follows the atom-law prediction of Proposition G.5 with high accuracy. Table I.3 reports this comparison, together with the center of mass of the leftmost bulk.

We also compare the locations of the bulk edges detected from empirical histograms with the edges evolved by free decompression; see Figure 6. Since the spectrum spans several orders of magnitude, edge errors are measured on the logarithmic λ-axis. At the initial size $n = 2^2\mathrm{K}$, the two leftmost components have not yet separated, while at larger sizes they split into a spike-like leftmost bulk and a neighboring small-eigenvalue bulk. The only edge excluded from the comparison is the left edge of the spike-like bulk. This edge is governed by finite-size extreme-eigenvalue effects rather than by the macroscopic free-decompression flow: initially it follows the Tracy–Widom-type behavior analyzed in Appendix H and illustrated in Figure H.1. The remaining edges, which correspond to the macroscopic bulk boundaries, are reported in Table I.4.

Table I.4: Errors of the evolved bulk edges for the diffusion-model spectrum. The entries compare empirically detected histogram edges with the corresponding edges evolved by free decompression, measured on the logarithmic λ-axis. At $n = 2^2\mathrm{K}$, the two leftmost components have not yet separated, so the entries under Bulk 2 are omitted; after the split, Bulk 1 denotes the spike-like leftmost component and Bulk 2 denotes the neighboring small-eigenvalue bulk. The left edge a_1 of the spike-like bulk is excluded because it is governed by finite-size extreme-eigenvalue effects rather than by the macroscopic free-decompression flow; see Appendix H and Figure H.1.

Size	Bulk 1		Bulk 2		Bulk 3		Bulk 4	
n	a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4
$2^2\mathrm{K}$	0.0%	0.0%	—	—	0.0%	0.2%	0.0%	0.1%
$2^3\mathrm{K}$	—	0.3%	0.1%	0.3%	0.2%	0.5%	0.2%	0.1%
$2^4\mathrm{K}$	—	0.1%	0.2%	0.2%	0.1%	0.2%	0.2%	0.1%
$2^5\mathrm{K}$	—	0.8%	0.7%	0.3%	0.1%	0.6%	0.1%	0.1%
$2^6\mathrm{K}$	—	1.2%	0.4%	0.2%	0.0%	0.8%	0.1%	0.0%

Appendix J Implementation and Reproducibility Guide

The algorithms in this paper are implemented through the `freealg.AlgebraicForm` interface of the open-source Python package `freealg`.[5] The package also includes several closed-form free-probability distributions, such as compound free Poisson and free Lévy laws, which are useful for synthetic experiments and validation. Listing 1 shows a minimal setup for a compound free Poisson example: it constructs the distribution, visualizes its spectral-curve branches, generates a large matrix realization, and then subsamples a smaller principal submatrix. We found that error induced by sampling different submatrices was negligible, being orders of magnitude smaller than the deterministic errors reported in our tables.

The synthetic distribution in Listing 1 is used only to provide a controlled demonstration. In applications, the user may instead start directly from a different input matrix, as in the empirical Hessian and diffusion-model examples in Appendices I.4 and I.5. Given only the smaller matrix A_{small}, Listing 2 fits an algebraic relation $P(z, m) = 0$, estimates the corresponding density, and applies free decompression to infer the density and atoms at the larger matrix size $n = 8000$. The same interface can also evolve densities over an array of matrix sizes and track lower-dimensional spectral features such as edges, cusps, and atom masses; the atom and edge evolution for this example is shown in Figure 1.

Notebook files that reproduce the figures appearing in this work can be found in the codebase. Further details on function arguments and class parameters are available in the package documentation. All experiments were performed on a consumer grade machine, with an AMD Ryzen 7 5800X processor, 128GB DDR4 RAM, running Fedora Linux.

[5] `freealg` is available for installation from PyPI (`https://pypi.org/project/freealg`), the documentation can be found at `https://ameli.github.io/freealg`, and the source code is available at `https://github.com/ameli/freealg`.

Listing 1: Synthetic compound free Poisson setup using `freealg`. The distribution object is used only to generate a controlled matrix example and visualize the associated spectral-curve branches.

```
# Install freealg with "pip install freealg"
import freealg as fa
import numpy as np

# Create an object for the Compound Free Poisson distribution (see Appendix I.1)
dist = fa.distributions.CompoundFreePoisson(t=[2, 5.5], w=[0.75, 0.25], lam=0.1)

# Visualize all branches and their gluing as in Figure I.1
x = np.linspace(-1, 9, 200)
y = np.linspace(-3, 3, 200)
dist.plot_branches(x, y)

# Generate a matrix realization of size n = 8000 for this distribution
A_large = dist.matrix(size=8000)

# Sample the above matrix to a smaller submatrix of size n_0 = 1000
A_small = fa.submatrix(A_large, 1000)
```

Listing 2: Free-decompression workflow using `freealg.AlgebraicForm`. Starting from the smaller matrix A_{small}, the code fits an algebraic spectral curve, estimates the density, decompresses to a larger matrix size, and tracks edges and cusps.

```
# Create an algebraic-form object for the small matrix
af = fa.AlgebraicForm(A_small)

# Fit the polynomial P with deg_m(P) = 3 and deg_z(P) = 1 (see Algorithm D.1)
af.fit(deg_m=3, deg_z=1)

# Estimate the density of the small matrix from P using Algorithm E.1
x = np.linspace(0, 10, 200)
rho_small = af.density(x, plot=True)

# Free decompression of density and atoms corresponding to a larger matrix of
# size n = 8000 using Algorithm F.1  (see Figure 3 ).
rho_large, atoms_large = af.decompress(
    size=8000, x=x, kind='free', return_atoms=True, plot=True)

# Evolve spectral edges from the small to the large matrix size using Algorithm G.1
# Also find cusp points where the bulks split (see Figure 1)
sizes = np.linspace(1000, 8000, 100)
t = np.log(sizes / 1000)
edges, cusps = af.edge(t, kind='free')
```

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We provide proofs of theoretical results and numerical examples where appropriate

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: This is discussed in an explicit limitations section in the conclusion

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: While some of the theorems in the main body are stated for ease of exposition and readership, complete statements and proofs are given in the appendices.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Full experimental details as well as a reproducibility guide are provided in the appendices. The software is publicly available on GitHub (see Appendix J).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: A reproducibility guide is provided in the appendix, and the software we developed is linked in a public GitHub repository.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: This is provided in the results section and appendices

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Our algorithms are deterministic. While subsamples are used for our initial data, we found that the induced errors were negligible compared to the error's deterministic component, so did not report them explicitly. This is stated in the reproducibility guide in the appendix.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All computation was done on a desktop-class machine, whose configuration is stated in the reproducibility guide in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: No ethics violations were observed

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The computational tools we develop are for computing eigenvalues of large matrices. We believe there is very little scope for this to have negative societal impact, at least for applications that we are aware of.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: We are not aware of any potential misuse of our tools.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Existing code bases that we used are open source, and we cited the datasets where appropriate.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: Code is provided via a public GitHub repository

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were not used for core method development

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.