Spectral Estimation with Deformed Decompression

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

Sample covariance matrices are fundamental objects in machine learning and statistics, with valuable information encoded into their eigenvalues. When treating modern problems, the scale of these matrices can become prohibitive for tractable computation. This creates a critical need for "spectral decompression": a method for inferring the eigenspectrum of a very large-scale model using only information from its smaller realizations. Existing methods fall into two main categories: those relying on spectral inversion; and those that rely on Free Decompression techniques. On the one hand, spectral inversion is a fundamentally ill-posed problem that is notoriously unstable, often failing to yield reliable results in finite-size settings. On the other hand, while Free Decompression was recently proposed to address this scaling, its reliance on the Nica-Speicher framework restricts its use to unitarily invariant ensembles, a symmetry condition that subsampled covariance matrices fail to satisfy. In this work, we introduce Deformed Decompression (DD), a novel framework for the stable, forward-extrapolation of spectral densities. By leveraging properties of the companion matrix, DD recasts the spectral decompression problem as a directed evolution of the Stieltjes transform of the companion matrix. This approach bypasses the instabilities of traditional inversion and provides a robust methodology for scaling covariance matrices and even their generalized variants (including Hessian matrices) that lack unitary invariance. We demonstrate the utility of our framework through experiments on a variety of random matrix examples, showing that DD successfully captures the complex spectral signatures of large-scale systems.

1 Introduction

Covariance matrices and their generalizations are fundamental objects in data analysis, arising whenever dependence, similarity, variability, or second-order structure needs to be represented. Common statistical practice treats the sample covariance as a reasonable reflection of the underlying population covariance. While true when the covariates are fixed and few in quantity relative to the number of datapoints, in the setting where the covariate dimension scales with data, the behavior of the sample covariance becomes fundamentally random, and tools from random matrix theory (RMT) are required for its treatment [14, 17]. This scenario is increasingly relevant in modern machine learning, where resource constraints can simultaneously restrict access to data and compute, and these components need to be delicately scaled accordingly [31, 32]. In this regime, a commonly encountered problem is the estimation of a population matrix; this is a notoriously difficult, ill-posed problem, but methods have been developed to provide reasonable estimates under strong regularization [15, 33, 34].

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

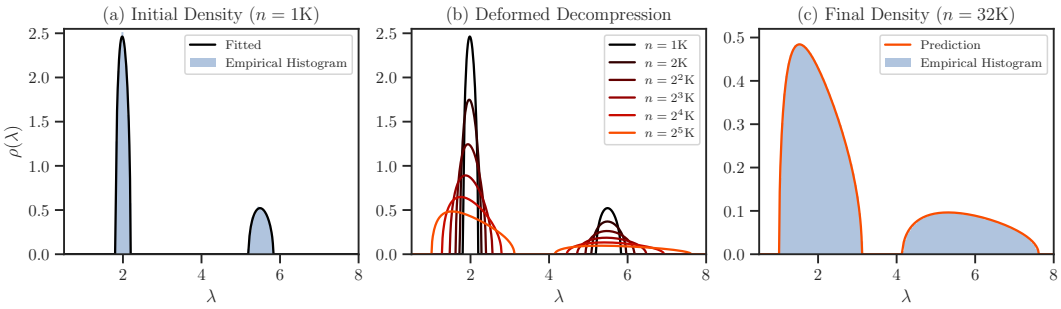


Figure 1: Example of deformed decompression for recovering the ESD of an $n \times n$ matrix $\mathbf{A}$ with $n = 32{,}000$, generated from a deformed Marchenko–Pastur law with atoms $(t_1, t_2) = (2, \frac{11}{2})$, weights $(w_1, w_2) = (\frac{3}{4}, \frac{1}{4})$, and target aspect ratio $\frac{1}{10}$. (a) ESD of the sampled $n_0 \times n_0$ matrix $\mathbf{A}_0$ with $n_0 = 1000$ (histogram) together with the fitted density (solid curve). (b) DD flow from n_0 to n. (c) Predicted ESD at size $n = 32{,}000$, compared with the benchmark ESD of the original matrix.

Our objective is different. We consider the case where the number of covariates is prohibitively large, rendering it difficult or impossible to form or extract fine-grained spectral information. For raw data covariances, this may be encountered, for example, in the analysis of genetic datasets [5, 52, 62]. Fisher information and Hessian matrices of large statistical and ML models have similar structure; for linear models these are precisely (generalized) covariance matrices [35], providing an important starting point to model their non-linear counterparts in ML [26, 47, 48, 49, 58]. The spectra of such matrices are of primary interest for downstream tasks, determining the concentration of variance across directions, presence of near-null modes, numerical stability of inverse problems, and the behavior of determinant-based objectives used in likelihoods and Bayesian inference [27, 28, 29, 56]. Mirroring the ubiquity of these matrices is their difficulty: they are often dense, highly ill-conditioned and non-isotropic, and sensitive to sampling, regularization, and model misspecification. In high dimensions, their empirical spectra may contain sharp edges, spikes, heavy tails, and clusters near zero [20, 21]. These features are often the most important for downstream analysis, e.g., log-determinants and inverse traces, and yet they remain the hardest to estimate accurately [3]. Classical approaches rely on low-rank approximations, including row/column sampling techniques and Nystrom-type approximations [18, 19, 25] or matrix-vector products [50, 55, 61]. While effective when the leading eigenspace dominates [1, 23], these classical approaches depend on assumptions that can radically fail under the heavy-tailed spectral properties often encountered in modern ML models [4, 30, 37, 38, 39, 40]. The bias in low-rank methods toward large eigenvalues severely underrepresents near-singular directions that are critical for computing determinants, inverses, and for uncertainty quantification [4, 27, 60].

Recent work on free decompression (FD) [2, 4] proposes a different strategy. By leveraging free probability and random matrix theory, FD evolves (or "decompresses") the empirical spectral density (ESD) of a sampled submatrix to estimate the ESD of a larger, inaccessible matrix. While this technique opens the door for spectral information to be obtained from very large "impalpable" matrices, there are classes of matrices to which FD does not apply (contrary to the original claim of Ameli et al. [2]). Crucially, among them is the important class of (generalized) covariance matrices (GCMs), which do not possess the unitary invariance properties required by FD. This raises the question of whether frameworks similar to FD can be developed to handle other classes of matrices.

Contributions. By directly appealing to the structure of covariance matrices and to the behavior of their *companion* Stieltjes and R-transforms, we derive *Deformed Decompression* (DD), a dual formulation of FD governed by a different but related partial differential equation (PDE). We show that DD unlocks the decompressibility of another wide class of matrices, including GCMs, and that it enables analogues of tools developed in [3]; this is illustrated in Figure 1. In doing so, we

- derive *Deformed Decompression* (DD), a new method for spectral expansion of GCMs;
- develop tools to track the dynamics of spectral edges, merging and splitting of bulks, as well as the evolution of spike components, without access to the full distribution;
- provide stability and error bounds for spectral expansion and approximation under our approach;
- give empirical examples that demonstrate the utility and accuracy of our methods;

- publish our implementation in the open source software project `freealg`.[1]

The remainder of this work is structured as follows. Section 2 briefly discusses some related work and its limitations. Section 3 introduces GCMs and their associated invariance property, develops DD flow, and gives stability and error bounds. Section 4 describes how this expansion can be achieved under a polynomial assumption, and provides related formulae for the behavior of spectral edges and atoms. Section 5 applies this method on a range of numerical examples. We conclude and discuss limitations and future work in Section 6. Additional information is provided in a set of appendices.

2 Background and Related Work

As the scale of scientific computing and machine learning problems continues to grow, it is becoming increasingly important to analyze matrices so large that they cannot be formed, in order to extract information for training, evaluation and diagnostics [23, 26, 28, 48, 49]. For example, Hessian matrices for modern language models contain on the order of 10^{24} elements [16]. Observing that heavy-tailed structures pose significant challenges for matrix-vector product strategies, Ameli et al. [4] proposed a statistical extrapolation approach to estimate log-determinants of large-scale kernel matrices using linear regression on small principal submatrices. While surprisingly accurate for neural tangent kernels of neural network models, its bespoke construction relies heavily on the mathematical structure of log-determinants of kernel matrices. Its rigidity necessitates new parallel methods for each desired quantity and each matrix class.

For the more general case, methods have been recently proposed to obtain approximations of the *entire spectral distribution* for arbitrary random matrices [2]. For example, using approximate Stieltjes transforms of submatrices as initial data, the FD method solves a partial differential equation to estimate the spectral distribution of the full matrix. While successful for simple and regularized problems, the fitting procedure and PDE evolution initially posed numerical challenges. These issues were largely overcome in followup work [3]. Drawing from the polynomial method [51] and knowledge of ML random matrix models [48, 49], the authors impose a polynomial ansatz on the Stieltjes transforms. This assumption accommodates a wider class of matrices, including multi-scale distributions with multiple bulks and spikes, and allows for tracking of moments, spectral edges, and the locations and weights of atoms. Such tools are particularly relevant for GCMs, where near-null directions and multi-scale eigenvalue structures are intrinsic features rather than numerical artifacts. However, this work also identified an imprecise assumption in [2]: the submatrix sampling used to seed the process can be too structured to guarantee recovery of the original matrix. Consequently, while the tools in [2, 3] apply broadly, certain matrix classes, notably GCMs, remain beyond their reach.

While not explicitly intended for these purposes, several analytic tools for GCMs exist, and could, in principle, be composed to obtain a similar procedure to the one presented here. Notably, Dobriban [17] leverages the same Silverstein equation we rely on to obtain a procedure for estimating the spectrum of large GCMs. Unlike our method, however, this procedure requires a-priori knowledge of the limiting covariance matrix, severely restricting its direct applicability. To estimate this population covariance matrix, Ledoit & Wolf [33, 34] provide mechanisms based on shrinkage estimators. However, the underlying problem is itself notoriously ill-posed (this is seen in Appendix C.2) and sensitive to minor noise fluctuations. Their methods consequently rely on strong regularization which, when used for preprocessing, introduces significant smoothing error, eliminating many of the structural features that make these matrices interesting.

Finally, we mention the foundational work of Marčenko & Pastur [41], who derived a PDE for the Stieltjes transform of deformed Marchenko–Pastur distributions, parameterizing the underlying aspect ratio as a "time" parameter. This is closely related to our own, more malleable equations.

Connection with Free Decompression. FD relies on a key invariance given by the Nica-Speicher theorem [44], which requires the matrix of interest to be embedded in an asymptotically free sequence of matrices. Ameli et al. [2] assert that uniform random subsampling of rows and columns ensures this property, but this claim seems to stem from an erroneous reading of Nica [43]. In reality, this invariance is only guaranteed if the projection is Haar-uniform [57]. This misconception creates a

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

gap: the techniques in Ameli et al. [3] work well when projected permutations exhibit asymptotic freeness, yet numerical tools to verify this condition are currently lacking. Similarly, while traffic probability [11, 36] offers analytic criteria for subsampling invariance, a clean analogue of FD within that framework remains unclear. Instead, we propose a technique targeting covariance matrices by exploiting a different invariance under the *companion R-transform*, providing an analogous toolkit to algebraic free decompression (AFD). It turns out that this invariance and Nica-Speicher can both hold only for Marchenko-Pastur laws, and so submatrices of nontrivial covariance matrices cannot satisfy the Nica-Speicher property.

3 Deformed Decompression

To develop our method, we consider the class of sample covariance matrices of the form

$$\mathbf{S} \coloneqq \frac{1}{n}\mathbf{X}\mathbf{X}^\intercal, \qquad \mathbf{X} \coloneqq \mathbf{\Sigma}_p^{\frac{1}{2}}\mathbf{Z},$$

where $\mathbf{Z} \in \mathbb{R}^{p\times n}$ has i.i.d. centered entries with unit variance, and $\mathbf{\Sigma}_p \in \mathbb{R}^{p\times p}$ is symmetric positive semidefinite. As our goal is to decompress the matrix $\mathbf{S}$ in a stable way, we consider it as embedded in a family of matrices $\mathbf{S}_p$ (see, e.g. Ameli et al. [2]) with the following common property.

Assumption 1 (Convergence of Population Covariance Structure)**.** The matrix $\mathbf{\Sigma}_p$ is a deterministic or independent population matrix whose empirical spectral distribution (ESD) converges weakly as $n \to \infty$ to a measure $\nu_{\mathbf{\Sigma}}$ with support on $[0, \infty)$.

To obtain a smaller realization of this covariance family, consider a larger matrix $\mathbf{S}_q$ with $q > p$, holding the number of samples n fixed, and let $\mathbf{P}_I \in \mathbb{R}^{p\times q}$ select p coordinates. The sampled matrix and its restricted population covariance are

$$\mathbf{S}_I = \mathbf{P}_I\mathbf{S}_q\mathbf{P}_I^\intercal, \qquad \mathbf{\Sigma}_I = \mathbf{P}_I\mathbf{\Sigma}_q\mathbf{P}_I^\intercal.$$

When $\mathbf{Z}$ has standard Gaussian entries, or when $\mathbf{\Sigma}_q$ is diagonal, $\mathbf{S}_I$ is again a sample covariance matrix of the same form, with population covariance $\mathbf{\Sigma}_I$. For the sampled matrix to belong to the same limiting covariance family, we additionally require that the ESD of $\mathbf{\Sigma}_I$ converges weakly to $\nu_{\mathbf{\Sigma}}$. This holds, for example, under uniform random coordinate sampling of diagonal population covariances with a convergent ESD and a nonvanishing sampling fraction, since the sampled population eigenvalues preserve their limiting proportions. Thus, under this condition, principal-submatrix sampling preserves the limiting population covariance law while changing the dimension of the covariance matrix.

The scalar parameter of interest is the limiting aspect ratio $c \coloneqq \lim_{n\to\infty} \frac{p}{n} \in (0, \infty)$, controlling the geometry of the underlying data matrix $\mathbf{X}$. In this limit, the ESD of $\mathbf{S}$ converges to a deterministic limit ν_c described by a *deformed Marchenko-Pastur law* [41], given in terms of the free multiplicative convolution $\nu_c = \nu_{\mathbf{\Sigma}} \boxtimes \mathrm{MP}_c$. Here, MP_c denotes the Marchenko–Pastur law with aspect ratio $c > 0$, and we remark that the mass at the origin is $\max\{\nu_{\mathbf{\Sigma}}(\{0\}), 1 - 1/c\}$. If $\nu_{\mathbf{\Sigma}}$ consists of a finite number of atoms, ν_c may exhibit multiple deformed Marchenko–Pastur bulk components.

For a fixed initial aspect ratio $c > 0$, consider a target aspect ratio τc with $\tau > 1$ (the *decompression ratio*). The measure $\nu_{\tau c}$ corresponds to the limiting spectrum for a matrix in the same form as $\mathbf{S}$, but with τ-times the number of rows and columns. Our objective is to construct a map, which we refer to as ***deformed decompression***[2] (DD), sending $\nu_c \mapsto \nu_{\tau c}$ without requiring knowledge of the population matrix $\mathbf{\Sigma}_p$ or its limiting spectral distribution $\nu_{\mathbf{\Sigma}}$. Provided this can be achieved, we can extrapolate from the spectrum of subsampled covariance matrices to the spectrum of the original.

To construct this map, let m denote the *Stieltjes transform* of ν_c and consider its *companion Stieltjes transform* $\underline{m}(z) = cm + \frac{c-1}{z}$. The R-transform corresponding to m is formally defined as the function R satisfying $R(-m(z)) = z + m(z)^{-1}$. Let $\underline{R}$ be the corresponding R-transform for the companion Stieltjes transform $\underline{m}$. Likewise, we let $\underline{R}_\tau$ denote the corresponding object for $\nu_{\tau c}$. In terms of this companion R-transform, we are able to obtain the following characterization of deformed decompression, proved in Appendix C.

[2]The name is derived from the involvement of the multiplicative convolution, commonly considered a form of *deformation*, as in "deformed Marchenko-Pastur law".

Theorem 1. *The deformed decompression map is prescribed by* $\underline{R}_\tau(z) = \tau \underline{R}(z)$.

This is to be compared to the Nica–Speicher Theorem for free decompression [2, 44], which is defined by a similar operation $R_\tau(z) = R(\tau z)$ in the R-transform. The origins of Theorem 1 can be understood in two ways. The first considers the *S-transform*, which is defined in Appendix B and uniquely determines a probability measure. It has the convenient property of factorizing through free multiplicative convolution: if S_c is the S-transform of ν_c, $S_{\mathbf{\Sigma}}$ is the S-transform of $\mathbf{\Sigma}$, and S_{MP_c} is the S-transform of MP_c, then $S_c(w) = S_{\mathbf{\Sigma}}(w) S_{\mathrm{MP}_c}(w)$. Since $S_{\mathrm{MP}_c}(w) = (1+cw)^{-1}$ is known explicitly, this relation effectively solves for the S-transform of the population matrix. However, performing deconvolution in this way is infamously ill-posed. Instead, taking advantage of the invariance of $S_{\mathbf{\Sigma}}$ to the aspect ratio, we may relate the S-transforms at two different aspect ratios c and $c\tau$ by $S_{\tau c}(w)(1+c\tau w) = S_c(w)(1+cw)$. This turns out to be better behaved, as seen in Appendix C.2. Through a series of algebraic manipulations (see Proposition C.2), the S-transform may be related to the companion R-transform $\underline{R}$, revealing Theorem 1.

The second, perhaps more convenient, interpretation of Theorem 1 arises by considering the companion matrix itself. The companion matrix for $\mathbf{S}$ is given by $\underline{\mathbf{S}} = \frac{1}{n}\mathbf{X}^\intercal\mathbf{X} = \frac{1}{n}\mathbf{Z}^\intercal\mathbf{\Sigma}_p\mathbf{Z}$, and its limiting spectral distribution is prescribed in terms of its Stieltjes transform $\underline{m}$. If $\mathbf{Z}$ is Gaussian, we may consider $\mathbf{\Sigma}_p = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$ with random ordering, due to the unitary invariance of $\mathbf{Z}$. Since $\underline{\mathbf{S}} = \frac{1}{n}\sum_{i=1}^p \sigma_i \boldsymbol{z}_i \boldsymbol{z}_i^\intercal$, where $\boldsymbol{z}_i$ is the ith row of $\mathbf{Z}$, increasing the aspect ratio c by increasing p is tantamount to adding rank-1 updates. In the large-dimensional limit, the companion spectral law is compound free Poisson with rate c and jump distribution $\nu_{\mathbf{\Sigma}}$, whose R-transform scales linearly with c, revealing Theorem 1.

The Deformed Decompression Flow. With a clearer picture of how the limiting spectral laws of covariance matrices vary with aspect ratio, we can now derive the flow equations for DD. Fix an initial aspect ratio $c > 0$, and let m_1 and $\underline{m}_1$ be the regular and companion Stieltjes transforms, respectively. Letting m_τ and $\underline{m}_\tau$ denote the corresponding transforms for the deformed decompression $\nu_{\tau c}$ under a ratio τ, we obtain inviscid Burgers' equations governing their dynamics, analogous to those for FD [2]. The proof of Theorem 2 is presented in Appendix C.1.

Theorem 2. *Let* $m_\tau(z)$ *denote the Stieltjes transform under the deformed decompression flow for initial aspect ratio* c*, and let* $\underline{m}_\tau(z)$ *denote its companion. Then*

$$\tau \underline{m}_\tau \frac{\partial}{\partial \tau}\underline{m}_\tau + (z\underline{m}_\tau + 1)\frac{\partial}{\partial z}\underline{m}_\tau = 0. \tag{1}$$

In terms of the Stieltjes transform, this becomes

$$(cz\tau m_\tau + c\tau - 1)\frac{\partial m_\tau}{\partial \tau} + c(m_\tau z + 1)\left(z\frac{\partial m_\tau}{\partial z} + m_\tau\right) = 0.$$

While we avoid the PDEs for computing m_τ in our implementation, they prove useful as analytic tools. In practice, the initial Stieltjes transform is approximated from a sampled matrix, raising the question of how small approximation errors propagate under DD. By linearizing the flow, we show that, under suitable conditions, these perturbations grow at most polynomially in τ over finite decompression intervals. The following proposition makes this precise, with proof given in Appendix D.

Proposition 3. *Let* $\underline{m}_\tau$ *be a companion Stieltjes transform of a density* ρ *such that* $\|\rho\|_{L^\infty} < +\infty$*. Let* η_1 *denote an initial perturbation to* $\underline{m}_1$ *at* $\tau = 1$*, and let* η_τ *denote its evolution under the linearization of the deformed decompression flow* (1)*. Under the conditions of Proposition D.1 and the uniform support bound of Remark D.1, for any fixed* $y > 0$ *and finite* $T > 1$*, there exists a constant* $\nu > 0$ *such that for any* $1 < \tau \le T$*, it holds* $\|\eta_\tau\|_{L^2(\Gamma_y)} \le \tau^\nu \|\eta_1\|_{L^2(\Gamma_y)}$.

Deformed Decompression is not Free Decompression. Both deformed decompression and free decompression obey related first-order complex transport flows. While obtained by distinct derivations, it is natural to consider deeper connections between them. It turns out that the two processes are virtually complementary. The following proposition formalizes this, stating that the only nontrivial distributions for which FD and DD are valid simultaneously belong to the Marchenko-Pastur family, reflecting the self-dual nature of these limiting distributions under the operation $\mathbf{X}\mathbf{X}^\intercal \mapsto \mathbf{X}^\intercal\mathbf{X}$.

Proposition 4. *Let $\{m(\tau, z)\}_{\tau \geq 1}$ be a family of Stieltjes transforms of probability measures $\{\nu_\tau\}_{\tau \geq 1}$ satisfying the PDEs for both the free decompression and deformed decompression flows for initial ratio $c > 0$. Then for any $\tau \geq 1$, ν_τ is a scaled Marchenko-Pastur distribution.*

Proposition 4 contrasts with the intended universality of free decompression implied in Ameli et al. [2]; for performing decompression of an arbitrary matrix $\mathbf{A}$, either free decompression *or* deformed decompression may succeed. Both tools are necessary for handling general impalpable matrices.

Generalized Covariance Matrices. Assume that $\mathbf{D}_n \in \mathbb{R}^{n \times n}$ is a diagonal matrix comprised of non-negative elements, and consider the matrix $\mathbf{H}$ and its companion given by

$$\mathbf{H} = \frac{1}{n} \mathbf{\Sigma}_p^{1/2} \mathbf{Z} \mathbf{D}_n \mathbf{Z}^\intercal \mathbf{\Sigma}_p^{1/2}, \qquad \underline{\mathbf{H}} = \frac{1}{n} \mathbf{D}_n^{1/2} \mathbf{Z}^\intercal \mathbf{\Sigma}_p \mathbf{Z} \mathbf{D}_n^{1/2}. \tag{2}$$

Note that, unlike regular covariance matrices, the companion matrix of $\mathbf{H}$ is of the same general form. Matrices in this generalized class commonly appear as Hessians under non-quadratic loss functions [35]. GCMs are conceptually distinct from the covariance matrix model treated so far where $\mathbf{D} = \mathbf{I}$. Unfortunately, there exists no single map that can send $\nu_c \mapsto \nu_{\tau c}$ for any arbitrary $\mathbf{D}$ (see discussion in Appendix F). Fortunately, in many cases, including Hessian matrices for models with cross-entropy loss [35, p. 9], the eigenvalues of $\mathbf{D}$ are bounded within a small interval or exhibit limited dispersion. The following proposition asserts that deformed decompression may still perform very well in this regime, as the limiting Stieltjes transform $m_{\mathbf{\Sigma}, \mathbf{D}}$ of $\mathbf{H}$, can be well-approximated by the corresponding transform of a covariance matrix model $m_{\tilde{\mathbf{\Sigma}}, \mathbf{I}}$.

Proposition 5. *Let $\Omega \subset \mathbb{C}^+$ be a bounded region on the upper half of the complex plane with positive distance from the real axis. There exists another limiting family of covariance matrices $\tilde{\mathbf{\Sigma}}$ such that*

$$\sup_{z \in \Omega} |m_{\mathbf{\Sigma}, \mathbf{D}}(z) - m_{\tilde{\mathbf{\Sigma}}, \mathbf{I}}(z)| = \mathcal{O}(\sigma_{\mathbf{D}}^2),$$

after taking $n \to \infty$, where $\sigma_{\mathbf{D}}^2$ is the limiting variance in the spectral distribution of $\mathbf{D}$.

4 Deformed Decompression on Spectral Curves

The strong analogy between DD and FD turns out to be highly useful for implementing DD. While the original FD implementation in Ameli et al. [2] demonstrated its utility across an impressive number of tasks, it still relied on an implicit assumption that the Stieltjes transform is a root of a quadratic polynomial. Ameli et al. [3] extended their earlier implementation by allowing for Stieltjes transforms that satisfy any polynomial relation. Their strategy casts the decompression process along a *spectral curve*, circumventing the need to solve a PDE or perform analytic continuation. Importantly, this enabled practical FD on a far wider array of matrices relevant to machine learning tasks, and provided tools to approximate a number of important spectral quantities. Fortunately, under this same assumption a parallel toolkit can be developed for DD. In order to do so, as in Ameli et al. [3], we impose the following ansatz.

Assumption 2. 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,

$$P(z, m(z)) = 0. \tag{3}$$

Let d_z denote the degree of P in z, and s its degree in m. While the Stieltjes transform encodes the underlying spectral distribution, the invariance that we exploit to perform DD is in terms of its companion Stieltjes transform. The co-ordinate change $(\zeta, \underline{m}_1(\zeta)) \mapsto (z, \underline{m}_\tau(z))$ given by the characteristic curves of (1) provides a more natural spectral curve to follow the effect of DD on the underlying polynomial (3). This is captured in our core computational Theorem 6, whose proof can be found as Theorem C.8 in Appendix C.

Theorem 6. *Let m be a Stieltjes transform satisfying Assumption 2, $\tau \geq 1$ be a decompression ratio, and m_τ be the deformed decompression of m by the factor τ. Then every point $(z, m_\tau(z)) \in \mathbb{C}^2$ satisfies*

$$P\left(\zeta, c^{-1}\left(\underline{m}_\tau - \frac{c-1}{\zeta}\right)\right) = 0, \quad \textit{with} \quad \zeta = \tau^{-1}\left(z + \underline{m}_\tau^{-1}\right) - \underline{m}_\tau^{-1}. \tag{4}$$

4.1 Computing and Evolving the Companion Stieltjes Transform

The downside to working with (3) is that at each $z \in \mathbb{C}$, $P(z, \cdot)$ has s possible solutions for m up to multiplicity. The *physical branch* corresponding to the Stieltjes transform of interest satisfies the Herglotz property $\Im m(z) > 0$ for $z \in \mathbb{C}^+$ and the normalization $m(z) = -z^{-1} + O(z^{-2})$ as $|z| \to \infty$. Fitting P from a real matrix $\mathbf{A}$ and identifying this correct branch was a key challenge overcome in Ameli et al. [3, Appendices D and E], and we deploy their methods here.

Evolving the Physical Branch. DD shares significant similarities and differences with FD. On the one hand, casting DD as evolution on a spectral curve allows us to adapt many AFD methods. However, this evolution is arguably simpler, as Biane's Subordination Theorem [9] guarantees that transport on the DD-induced curve maps the upper-half plane to itself. Consequently, while the spectral curve for FD provides automatic analytic continuation, DD computations need only consider the physical branch of the companion Stieltjes transform.

The evolved algebraic relation in (4) gives a characterization of candidate roots $m_\tau(z)$; however, it does not distinguish between the physical and non-physical branches. To obtain the physical root, we follow an analogous procedure to AFD [3] by tracking the known physical solution at $\tau = 1$ along the spectral curves. For a fixed query point z, we work in the spectral-curve coordinates (ζ, w) where $w = \underline{m}_\tau$, $\underline{P}(\zeta, w) = P(\zeta, c^{-1}(w - \zeta^{-1}(c-1))) = 0$, and view the characteristic relation $\zeta - \tau^{-1}(z + w^{-1}) + w^{-1} = 0$ as a constraint whose derivative in τ for fixed z is $\dot{\zeta} + \tau^{-2}(z + w^{-1}) + (\tau^{-1} - 1)w^{-2}\dot{w} = 0$. This recasts (ζ, w) as the coupled system

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

Note that at $\tau = 1$, we have $\zeta = z$ and $w = \underline{m}_1(z)$, where $\underline{m}_1(z)$ has previously been obtained by a physical-branch selection procedure. Fixing z and taking the derivative in τ yields the tangent equations

$$\begin{bmatrix} \partial_\zeta \underline{P}(\zeta, w) & \partial_w \underline{P}(\zeta, w) \\ 1 & (\tau^{-1} - 1)w^{-2} \end{bmatrix} \begin{bmatrix} \dot{\zeta} \\ \dot{w} \end{bmatrix} = \begin{bmatrix} 0 \\ -\tau^{-2}(z + w^{-1}) \end{bmatrix}.$$

We can generate a prediction of the next decompression-step by approximating this system of geometric ordinary differential equations. A Newton projection is then used to correct the predictor back onto the constrained set $(F_1, F_2) = (0, 0)$. When the local geometry becomes stiff, the steps are appropriately subdivided. Following this corrector step, the decompressed Stieltjes transform is recovered from its companion via $m_\tau(z) = \frac{1}{\tau c}\left(w - \frac{\tau c - 1}{z}\right)$. A full description of this geometric predictor–corrector scheme is provided in Appendix C.7.

4.2 Spectral Quantities

DD also allows for the tracking of the evolution of a number of spectral quantities of interest. In particular, the atomic weights, distributional edge locations, as well as moments can all be computed independently from the flow of the underlying distribution.

Atoms. An atom located at $x_\circ$ is characterized as 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 DD, nonzero atom locations are pushed away from zero, and we can track their weights explicitly as functions of the decompression ratio τ. The precise dynamics are encoded in the following proposition, whose proof is given in Appendix E.3.

Proposition 7 (Dynamics of Atoms under Deformed Decompression)**.** *Let $x_\circ \in \mathbb{R}$ be an atom of the spectral measure ν with initial mass $w_\circ \in (0, 1)$. Under deformed decompression with parameter τ,*

the following hold. If $x_\circ \neq 0$, then the atom location is pushed forward to

$$x_\circ(\tau) = \tau x_\circ, \quad \textit{and mass evolves as} \quad w(\tau) = w_\circ - \frac{\tau - 1}{c\tau}.$$

If $x_\circ = 0$, then $x_\circ(\tau) = 0$, and its mass evolves as $w(\tau) = \max\{w_\circ, 1 - \frac{1}{c\tau}\}$. For $x_\circ \neq 0$, the formula describes the atom while $w(\tau) > 0$. When the predicted mass reaches zero, the pole disappears from the physical branch.

Spectral Edges. To track the support of the spectral density, whose bulk regions may split or merge as the spectral density evolves, we will assume that the spectral distribution can be decomposed into a finite set of atoms and absolutely-continuous bulk components. The bulk support can then be written as $I(\tau) = \bigcup_{j=1}^{k} I_j(\tau)$, where each $I_j(\tau) = [a_j(\tau), b_j(\tau)]$, with spectral edges being the boundary points $\{a_j(\tau), b_j(\tau)\}_{j=1}^{k}$. Assumption 2 characterizes these points in terms of the *branch points* of the corresponding spectral curve, where different branches meet, satisfying the constraint equations

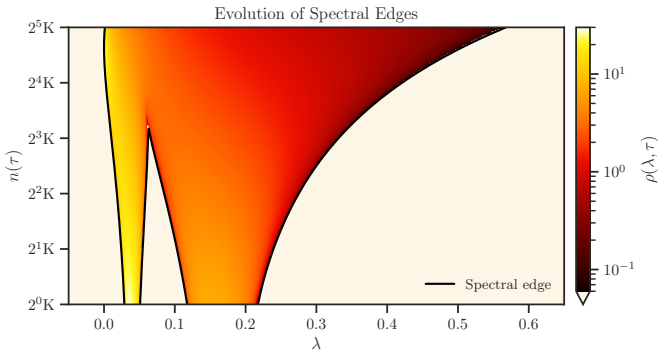


Figure 2: Evolution of spectral edges (solid curve) corresponding to the density in Figure 3. Color in log scale represents the evolved density, with beige background having zero density (spectral gaps).

$$P(\zeta_*, y_*) = 0,$$
$$\tau\left(c\zeta_* y_* + c - 1\right)^2 \partial_y P(\zeta_*, y_*) + (\tau - 1)\left(c\zeta_*^2\, \partial_\zeta P(\zeta_*, y_*) + (c-1)\partial_y P(\zeta_*, y_*)\right) = 0,$$

where y belongs to the physical branch. This is made precise as Proposition E.1 in Appendix E.1. Our procedure for finding the spectral edges is described in Algorithm E.1.

Merging and Splitting Bulks. Under deformed decompression, neighboring bulk regions may merge. Viewed in reverse, this corresponds to a bulk splitting into multiple bulks. These transitions correspond to second-order critical points of the spectral projection. These points are identified in Proposition E.2 in Appendix E.2 as satisfying

$$P(\zeta_*, y_*) = 0,$$
$$\tau_*(c\zeta_* y_* + c - 1)^2 \partial_y P + (\tau_* - 1)\left(c\zeta_*^2 \partial_\zeta P + (c-1)\partial_y P\right) = 0,$$
$$(c\zeta_* y_* + c - 1)\left[2\zeta_*(\partial_\zeta P)(\partial_y P)^2 + \zeta_*^2\left(\partial_{\zeta\zeta} P(\partial_y P)^2 - 2\partial_{\zeta y} P\, \partial_\zeta P\, \partial_y P + \partial_{yy} P(\partial_\zeta P)^2\right)\right]$$
$$- 2\partial_y P\left((c-1)\partial_y P + c\zeta_*^2 \partial_\zeta P\right)\left(y_* \partial_y P - \zeta_* \partial_\zeta P\right) = 0.$$

Moments. Spectral moments can also be evolved without reconstructing the density. As shown in Appendix E.4, the first moment remains constant, $\mu_1(\tau) = \mu_1(1)$, while the variance grows linearly with the decompression ratio: $\mathrm{Var}(\nu_{\tau c}) = \mathrm{Var}(\nu_c) + c(\tau - 1)\mu_1(1)^2$. More generally, for $k \geq 1$, the k-th moment is a polynomial in τ of degree at most $k - 1$ (see Proposition E.5).

5 Numerical Experiments

To demonstrate DD, we will consider three examples. The first mirrors the canonical example we used to derive our method in Section 3, an approximation of a deformed Marchenko–Pastur law. Then, we consider a generalized covariance matrix of the form given in (2), the Hessian of a logistic regression model. Finally, we treat the covariance matrix of the STL-10 image dataset [12].

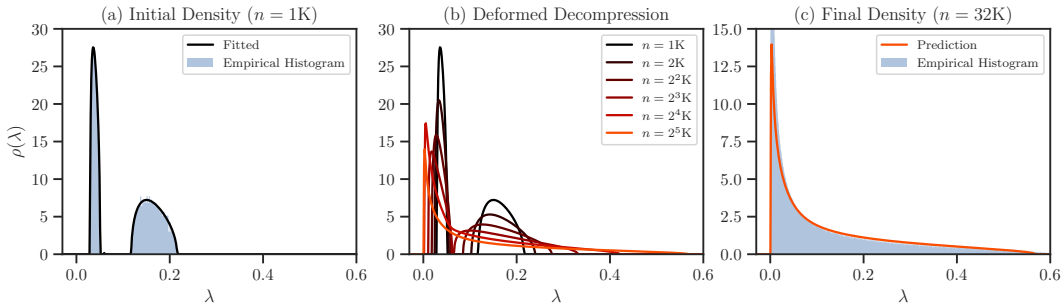


Figure 3: Example of deformed decompression for recovering the ESD of an $n \times n$ Hessian matrix $\mathbf{A}$ following [35], here with $n = 32{,}000$. (a) ESD of the sampled $n_0 \times n_0$ matrix $\mathbf{A}_0$ with $n_0 = 1000$ (histogram) and the fitted density (solid curve). (b) Deformed decompression flow from n_0 to n. (c) Final predicted ESD at $n = 32{,}000$, compared with the benchmark histogram of the original matrix.

Deformed Marchenko–Pastur. The deformed Marchenko–Pastur law is the canonical model for deformed decompression explicitly used in its construction as the limiting spectral distribution of covariance matrices. Detailed background and description on this distribution are given in Appendix G.1. To generate the matrix of interest, we generated an $n \times n$ matrix $\mathbf{A}$ with $n = 32{,}000$ from a deformed Marchenko–Pastur law with atoms $(t_1, t_2) = (2, \frac{11}{2})$, weights $(w_1, w_2) = (\frac{3}{4}, \frac{1}{4})$, and target aspect ratio $\frac{1}{10}$. A $n_0 \times n_0$ matrix $\mathbf{A}_0$ with $n_0 = 1000$ was subsampled from the original matrix, and its empirical Stieltjes transform generated from its spectral decomposition, fit with $d_z = 2$ and $s = 3$. DD was performed, with visualization of the flow given in Figure 1.

Hessian Matrices. We now consider the Hessian of a logistic regression loss function, where the features have non-uniform covariance. This model is taken from [35] and is a bona-fide GCM with $\mathbf{D}$ a diagonal matrix comprised of the second derivatives of the loss. The procedure used for training and Hessian formation is given in Appendix G.2. To perform DD, once the $n \times n$ matrix $\mathbf{H}$ with $n = 32{,}000$ was obtained, a $n_0 \times n_0$ submatrix $\mathbf{H}_0$ with $n_0 = 1000$ was sampled from the original matrix. The initial Stieltjes transform was fit from its empirical analogue with $d_z = 2$ and $s = 4$. DD was performed, and its flow is visualized in Figure 3. The evolution of its edges, demonstrating the merging of two bulk regions, is demonstrated in Figure 2. Table G.2 provides distributional distances in terms of normalized Wasserstein-1 and Gaussian-kernel maximum mean discrepancy, as well as relative error of estimates on the first two moments of distributions obtained during the flow. Although not a covariance matrix, DD still appears to perform well. This is consistent with the small variance in the entries of $\mathbf{D}$, as seen in Figure G.5. A further example of a Hessian matrix with $n = 16{,}000$ is given in Appendix G.2.

STL-10 Dataset Covariance Matrix. Finally, we consider a Gram covariance matrix obtained from a real dataset, the STL-10 image dataset. Details on this dataset and the procedure used to form the matrix, as well as additional figures and information on its edge behavior, are given in Appendix G.3. To perform DD, we formed the Gram covariance matrix $\mathbf{C} = p^{-1}\mathbf{X}\mathbf{X}^\intercal$ from $n = 8000$ STL-10 images with feature dimension $p = 27{,}648$, and then sampled an $n_0 \times n_0$ principal submatrix $\mathbf{C}_0$ with $n_0 = 1000$. DD was performed by fitting the initial Stieltjes transform with $d_z = 3$ and $s = 12$, and the distribution was evolved up to $\tau = 8$. This is visualized in Figure 4, which depicts ridge-plots of the estimated densities during the flow on log-scale, compared to the ESDs of subsamples.

Table 1 summarizes comparisons of runtime of eigenvalue computation both under DD and direct calculation of subsamples. Distributional errors for log-eigenvalues were also computed in terms of normalized Wasserstein-1 distance W_1/L, along with Gaussian-kernel maximum mean discrepancy [24, 46], and first-moment errors. Despite the substantial decrease in cost, the incurred errors are relatively small.

6 Conclusions

We introduce deformed decompression, a novel method for approximating spectral distributions of large generalized covariance matrices from smaller sampled submatrices. Due to their inherent

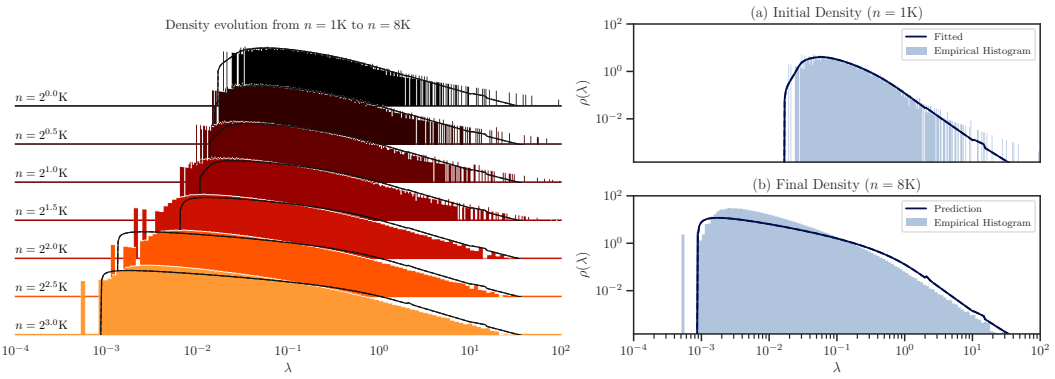


Figure 4: DD recovering the ESD of an 8000×8000 covariance matrix $\mathbf{A}$ of the STL-10 dataset. *Left:* ridge plot of the DD density evolution on log scale (solid lines), compared with ESD of sampled submatrices. *Right:* (a) ESD of the sampled 1000×1000 submatrix, with density fit; and (b) final predicted ESD, compared with the ESD of the original matrix.

structure, these matrices were previously not treatable by existing decompression methods. We also develop tools to track spectral edges and bulk mergers, atomic locations and weights, and arbitrary moments. DD opens the door for the extraction of spectral information in settings where the matrices themselves cannot explicitly be formed. These include covariance matrices of datasets with extremely large numbers of covariates, and covariance-like structures, such as the Fisher information or Hessian, for large models.

Limitations. While our method broadens the class of matrices that can be treated using decompression techniques, it only provides approximations for GCMs in terms of a nearby covariance matrix. In particular, our theory applies to positive-semidefinite covariance and curvature matrices, but does not extend to indefinite Hessians in general. As such, there still exist broad classes of matrices that cannot be treated using either FD or DD. Moreover, without prior analytical knowledge of the subsampled matrix, it remains unclear how to determine whether FD or DD may apply. While the reverse operation of deformed *compression*, where $0 < \tau < 1$, may be of independent interest for its potential recovery of the population covariance matrix, we found it to be too unstable in preliminary experiments.

Table 1: Process times for obtaining eigenvalues, as well as corresponding distributional distances and relative error on the first moment for the STL-10 covariance matrix considered in Section 5.

Size	Process Time (sec)		Accuracy Metrics		
n	Direct	DD (ours)	W_1/L	MMD	$\Delta\mu_1/\mu_1$
2^0K	4.2	$4.2 + 0.0$	0.00%	0.23%	0.51%
2^1K	14.0	$4.2 + 2.1$	0.58%	2.84%	0.25%
2^2K	69.5	$4.2 + 2.7$	1.42%	1.24%	0.05%
2^3K	579.6	$4.2 + 3.7$	3.70%	2.34%	0.03%

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] Adams, R. P., Pennington, J., Johnson, M. J., Smith, J., Ovadia, Y., Patton, B., & Saunderson, J. (2018). Estimating the spectral density of large implicit matrices. *arXiv preprint arXiv:1802.03451*.

[2] 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).

[3] Ameli, S., van der Heide, C., Hodgkinson, L., & Mahoney, M. W. (2026). Free decompression with algebraic spectral curves. In *The Fortieth Annual Conference on Neural Information Processing Systems*.

[4] 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).

[5] Aparicio, L., Bordyuh, M., Blumberg, A. J., & Rabadan, R. (2020). A random matrix theory approach to denoise single-cell data. *Patterns*, 1(3), 100035.

[6] Arizmendi, O., Hasebe, T., & Kitagawa, Y. (2026). Free multiplicative convolution with an arbitrary measure on the real line. *Probability Theory and Related Fields*.

[7] Bai, Z. & Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics. Springer New York, NY, second edition.

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

[9] Biane, P. (1998). Processes with free increments. *Mathematische Zeitschrift*, 227, 143–174.

[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] Cébron, G., Dahlqvist, A., & Male, C. (2024). Traffic distributions and independence II: universal constructions for traffic spaces. *Documenta Mathematica*, 29, 39–114.

[12] Coates, A., Ng, A. Y., & Lee, H. (2011). An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research* (pp. 215–223).

[13] Couillet, R. & Hachem, W. (2014). Analysis of the limiting spectral measure of large random matrices of the separable covariance type. *Random Matrices: Theory and Applications*, 3(04), 1450016.

[14] Couillet, R. & Liao, Z. (2022). *Random matrix methods for machine learning*. Cambridge University Press.

[15] Couillet, R. & McKay, M. (2014). Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. *Journal of Multivariate Analysis*, 131, 99–120.

[16] DeepSeek-AI (2026). *DeepSeek-V4: Towards Highly Efficient Million-Token Context Intelligence*. Technical report, DeepSeek. Accessed: 2026-05-07.

[17] Dobriban, E. (2015). Efficient computation of limit spectra of sample covariance matrices. *Random Matrices: Theory and Applications*, 04(04), 1550019.

[18] Drineas, P., Kannan, R., & Mahoney, M. W. (2006). Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition. *SIAM Journal on Computing*, 36(1), 184–206.

[19] Drineas, P. & Mahoney, M. W. (2016). RandNLA: Randomized numerical linear algebra. *Communications of the ACM*, 59(6), 80–90.

[20] El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics*, 36(6), 2757–2790.

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

[22] Erdős, L., Krüger, T., & Schröder, D. (2020). Cusp universality for random matrices I: Local law and the complex Hermitian case. *Communications in Mathematical Physics*, 378(2), 1203–1278.

[23] Gardner, J., Pleiss, G., Weinberger, K., Bindel, D., & Wilson, A. G. (2018). GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. In *Advances in Neural Information Processing Systems*, volume 31: Curran Associates, Inc.

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

[25] Gittens, A. & Mahoney, M. W. (2016). Revisiting the Nyström method for improved large-scale machine learning. *Journal of Machine Learning Research*, 17(117), 1–65.

[26] Hayase, T. & Karakida, R. (2021). The spectrum of Fisher information of deep networks achieving dynamical isometry. In A. Banerjee & K. Fukumizu (Eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research* (pp. 334–342).

[27] 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).

[28] 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*.

[29] 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*.

[30] Hodgkinson, L., Wang, Z., & Mahoney, M. W. (2025). Models of heavy-tailed mechanistic universality. In A. Singh, M. Fazel, D. Hsu, S. Lacoste-Julien, F. Berkenkamp, T. Maharaj, K. Wagstaff, & J. Zhu (Eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research* (pp. 23290–23329).

[31] Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., de Las Casas, D., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., van den Driessche, G., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., Vinyals, O., Rae, J., & Sifre, L. (2022). Training compute-optimal large language models. In *Advances in Neural Information Processing Systems*, volume 35 (pp. 30016–30030).

[32] Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

[33] Ledoit, O. & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2), 1024–1060.

[34] Ledoit, O. & Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, 48(5), 3043–3065.

[35] Liao, Z. & Mahoney, M. (2021). Hessian eigenspectra of more realistic nonlinear models. In *Advances in Neural Information Processing Systems*, volume 34 (pp. 20104–20117).

[36] Male, C. (2020). Traffic distributions and independence: permutation invariant random matrices and the three notions of independence. *Memoirs of the American Mathematical Society*, 267(1300).

[37] Martin, C. H. & Mahoney, M. W. (2019). Traditional and heavy-tailed self regularization in neural network models. In *Proceedings of the 36th International Conference on Machine Learning* (pp. 4284–4293).

[38] Martin, C. H. & Mahoney, M. W. (2020). Heavy-tailed Universality predicts trends in test accuracies for very large pre-trained deep neural networks. In *Proceedings of the 20th SIAM International Conference on Data Mining* (pp. 505–513).

[39] 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.

[40] Martin, C. H., Peng, T. S., & Mahoney, M. W. (2021). Predicting trends in the quality of state-of-the-art neural networks without access to training or testing data. *Nature Communications*, 12, 4122.

[41] 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.

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

[43] Nica, A. (1993). Asymptotically free families of random unitaries in symmetric groups. *Pacific Journal of Mathematics*, 157(2), 295–310.

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

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

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

[47] Pandit, P., Wang, Z., & Zhu, Y. (2025). Universality of kernel random matrices and kernel regression in the quadratic regime. *Journal of Machine Learning Research*, 26(224), 1–73.

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

[49] Pennington, J. & Worah, P. (2017). Nonlinear random matrix theory for deep learning. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 30: Curran Associates, Inc.

[50] Potapczynski, A., Finzi, M., Pleiss, G., & Wilson, A. G. (2023). CoLA: Exploiting compositional structure for automatic and efficient numerical linear algebra. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Advances in Neural Information Processing Systems*, volume 36 (pp. 43894–43917).

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

[52] Schäfer, J. & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1).

[53] Silverstein, J. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55(2), 331–339.

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

[55] Ubaru, S., Chen, J., & Saad, Y. (2017). Fast estimation of $\mathrm{tr}(f(A))$ via stochastic Lanczos quadrature. *SIAM Journal on Matrix Analysis and Applications*, 38(4), 1075–1099.

[56] Ubaru, S. & Saad, Y. (2018). Applications of trace estimation techniques. In *High Performance Computing in Science and Engineering* (pp. 19–33). Springer International Publishing.

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

[58] Wang, Z. & Zhu, Y. (2024). Deformed semicircle law and concentration of nonlinear random matrices for ultra-wide neural networks. *The Annals of Applied Probability*, 34(2), 1896–1947.

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

[60] Wilson, J., van der Heide, C., Hodgkinson, L., & Roosta, F. (2025). Uncertainty quantification with the empirical neural tangent kernel. In *Advances in Neural Information Processing Systems*, volume 38 (pp. 148970–149008).

[61] Yao, Z., Gholami, A., Keutzer, K., & Mahoney, M. W. (2020). PyHessian: Neural networks through the lens of the Hessian. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 581–590).

[62] Zhou, Y.-H., Marron, J. S., & Wright, F. A. (2018). Eigenvalue significance testing for genetic association. *Biometrics*, 74(2), 439–447.

Appendices

Contents

A Notation and Conventions **15**

B Transforms in Random Matrix Theory **17**

C Deformed Decompression on Algebraic Spectral Curves **19**

C.1 The Key Theorem 19

C.2 The Free Binomial Distribution 21

C.3 Deriving the Partial Differential Equations 23

C.4 Deformed Decompression and Free Decompression 24

C.5 Formal Deformed Decompression along Spectral Curves 25

C.6 Polynomial Evolution under Deformed Decompression 26

C.7 Deformed Decompression on Spectral Curves 27

D Stability Analysis **28**

E Evolution of Spectral Quantities **31**

E.1 Evolving Spectral Edges 31

E.2 Spectral Cusps and The Mechanism of Merging Bulks 32

E.3 Dynamics of Atoms 35

E.4 Evolution of Moments 36

F Generalized Covariance Models **38**

G Random Matrix Models **41**

G.1 Deformed Marchenko–Pastur Law 41

G.2 Hessian of Logistic Regression Model 46

G.3 STL-10 Image Dataset 47

H Implementation and Reproducibility Guide **48**

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.

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 [59, Section 2.5].

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
	$\mathrm{trace}(\cdot)$	Standard trace of a matrix
	$\mathrm{tr}_n(\cdot)$	Normalized trace, $n^{-1}\,\mathrm{trace}(\cdot)$
	$\mathbf{Q}_n$	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)
	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, $I = \mathrm{supp}(\rho) \subsetneq \mathbb{R}$
	m	Stieltjes transform of measure, $m(z) = \int_{\mathbb{R}} \frac{\nu(\mathrm{d}y)}{y-z}$, $\mathbb{C} \setminus \mathrm{supp}(\nu) \to \mathbb{C}$
	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 $x \mapsto cx$
	$K(z)$	K-transform, $K(z) = R(z) + z^{-1}$
	$M(z)$	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 $\mathcal{C} = \{(z, m) \in \mathbb{C}^2 \mid P(z, m) = 0\}$
Deformed 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 $t = \log(\tau)$
	c	Initial aspect ratio; the target aspect ratio at scale τ is τc
	$m_1, \underline{m}_1$	Initial ordinary and companion Stieltjes transforms at $\tau = 1$
	$m_\tau, \underline{m}_\tau$	Evolved ordinary and companion Stieltjes transforms at scale τ
Notation	$\mathbb{C}^{\pm}$	Upper (lower) half complex plane
	$\Re, \Im$	Real and imaginary part of complex variable
	$\mathrm{p.v.}$	Cauchy principal value

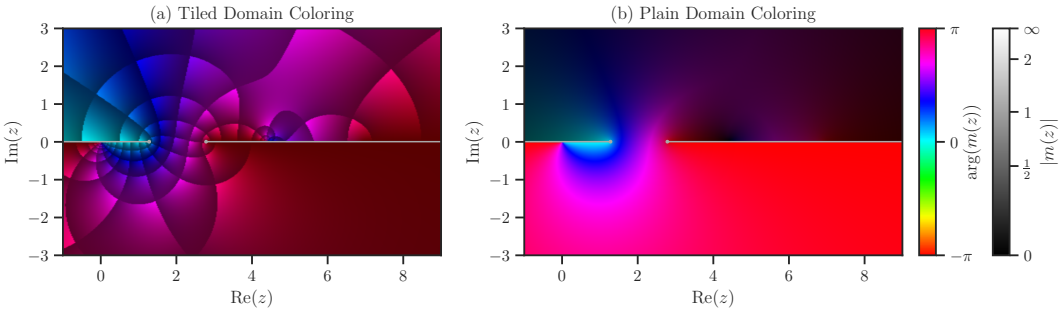


Figure A.1: Domain coloring of a Stieltjes-transform branch for the deformed Marchenko–Pastur example in Figure G.1. Right: plain domain coloring, with hue encoding $\arg m(z)$ and brightness encoding $|m(z)|$. Left: tiled domain coloring, where phase and logarithmic magnitude are periodically reset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

Appendix B Transforms in Random Matrix Theory

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 define $\rho(x) := \nu_{\mathrm{ac}}(\mathrm{d}x)/\mathrm{d}x$ as its density, given by the Radon–Nikodym derivative. In this section, we outline the various transforms of ν that each uniquely identify it. Our discussion will proceed formally without attention to domains of convergence. In the following, ν will be a fixed, but arbitrary, probability measure.

Definition B.1 (Stieltjes Transform)**.** 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, that is, $\Im m(z) > 0$ for $z \in \mathbb{C}^+$ where $\mathbb{C}^+ := \{z \in \mathbb{C} \mid \Im z > 0\}$, 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 $|z| \to \infty$.

Recovering the Density. Whenever the limits exist, let $m^{\pm}(x) := \lim_{y \to 0^{\pm}} m(x+iy)$. If $\nu(\mathrm{d}x) = \rho(x)\,\mathrm{d}x$ in a neighborhood of x, the Sokhotski–Plemelj formula [42, 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 by

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

which is the inversion of the Stieltjes transform. In the remainder, the support of the absolutely continuous part of ν is denoted by $I := \operatorname{supp}(\rho)$, and is assumed to be comprised of finitely many disjoint compact intervals,

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

where $k \geq 1$, and $a_j < b_j < a_{j+1}$. Here, k is the number of spectrum bulks. Atomic components of ν produce isolated poles of m on the real axis and must be treated separately.

The Stieltjes transform m is one example of a transform that uniquely prescribes the underlying probability measure ν. There are several other important transforms that are all related to the Stieltjes transform by algebraic relations. All transforms are to be discussed with reference to the same measure ν, hence the dependence on the measure is dropped from the notation. In other sections, it will be convenient to state transforms with respect to particular measures using subscripts.

The first transform is the *R-transform* which is required for the statement of Theorem 1. Our presentation follows Couillet & Liao [14, Definition 5].

Definition B.2 (R-transform)**.** The *R-transform* of ν is defined as the solution $R(z)$ to

$$m(R(z) + z^{-1}) = -z,$$

for all z where such a solution exists. Since $m(z) \sim -1/z$ as $|z| \to \infty$, $R(z)$ is defined in a neighbourhood of $z = 0$ where it has a power series expansion in terms of the *free cumulants* r_k:

$$R(z) = \sum_{k=1}^{\infty} r_k z^{k-1}.$$

The primary significance of the R-transform in random matrix theory is its behavior under free additive convolution. For two measures ν_1 and ν_2, the free additive convolution $\nu_1 \boxplus \nu_2$ (representing the distribution of the eigenvalues of the sum of freely independent matrices with spectral distributions ν_1 and ν_2) has R-transform satisfying [14, Lemma 2.19]

$$R_{\nu_1 \boxplus \nu_2}(z) = R_{\nu_1}(z) + R_{\nu_2}(z).$$

On some occasions, the algebraic manipulations render directly working with the R-transform notationally inconvenient. This is often rectified by working with the K-transform instead.

Definition B.3 (K-transform)**.** The *K-transform* is the functional inverse of $-m$, that is, $-m(K(w)) = w$. It is defined in terms of the R-transform by

$$K(w) = \frac{1}{w} + R(w).$$

The Stieltjes transform has a convenient Laurent series expansion for large $|z|$ in terms of the moments $\mu_n = \int x^n \nu(\mathrm{d}x)$ of ν:

$$m(z) = -\frac{1}{z} \sum_{n=0}^{\infty} \frac{\mu_n}{z^n}, \tag{B.3}$$

Therefore, a closely related transform to the Stieltjes transform is the (truncated) moment generating function, which is a direct power series in the moments.

Definition B.4 (Moment Generating Function)**.** For a fixed probability measure ν, we let $\mu_n = \int x^n \nu(\mathrm{d}x)$ denote the n-th moment of ν for integer n. The *(truncated) moment generating function* of ν is given by

$$M(z) = \sum_{n=1}^{\infty} \mu_n z^n = \int_{\mathbb{R}} \frac{tz}{1 - tz} \, \nu(\mathrm{d}t),$$

Here, the integral representation follows from the geometric series. From (B.3), it is clear that

$$m(z) = -\frac{1}{z}(1 + M(z^{-1})), \qquad M(z) = -1 - z^{-1} m(z^{-1}). \tag{B.4}$$

The moment generating function is the first transform on the path to the S-transform. The intermediate stage is referred to as the chi transform.

Definition B.5 (Chi Transform)**.** The chi transform $\chi(z)$ is the functional inverse of the moment generating function $M(z)$ so that $M(\chi(z)) = \chi(M(z)) = z$. As with the R-transform, this is defined formally, and is to be understood as defined over the domain where M is invertible.

Definition B.6 (S-transform)**.** The S-transform $S(z)$ is given in terms of the chi transform by

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

The S-transform is related to the moment generating function $M(z)$ by

$$z = \frac{M(z) S(M(z))}{1 + M(z)}. \tag{B.5}$$

Analogous to the R-transform, the significance of the S-transform is its behavior for multiplicative free convolution. For two measures ν_1 and ν_2, one of which has support entirely on $[0, \infty)$, the free multiplicative convolution $\nu_1 \boxtimes \nu_2$ (representing the distribution of the eigenvalues of the symmetric product $\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2}$ of freely independent matrices $\mathbf{A}$, $\mathbf{B}$ with spectral distributions ν_1 and ν_2, respectively) has S-transform satisfying [14, Lemma 2.19]

$$S_{\nu_1 \boxtimes \nu_2}(z) = S_{\nu_1}(z) S_{\nu_2}(z).$$

The R- and S-transforms are related by the following [3, Lemma F.1]:

$$y = zS(z) \iff z = yR(y). \tag{B.6}$$

Companion Matrix. A critical object in our exposition is the *companion matrix*: for a matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^\intercal$, the companion matrix of $\mathbf{S}$ is $\underline{\mathbf{S}} = \mathbf{X}^\intercal\mathbf{X}$. For rectangular $\mathbf{X}$, the shape of $\underline{\mathbf{S}}$ will differ from $\mathbf{S}$, but by the Weinstein–Aronszajn identity, their eigenspectra only differ in the number of zeros. In terms of the Stieltjes transform, this corresponds to the inclusion/removal of an atom at zero. The weight of this atom is determined by the aspect ratio of the sides of $\mathbf{X}$, denoted by $c > 0$.

Definition B.7 (Companion Stieltjes Transform)**.** For a fixed aspect ratio $c > 0$, the Stieltjes transform of the *companion* $\underline{m}$ is given in terms of the Stieltjes transform m by

$$\underline{m}(z) = cm(z) + \frac{c-1}{z}.$$

Here and throughout, the underline notation will be used to represent the corresponding object for the companion.

There is no straightforward relationship between the R-transform or K-transform and that of the companion. However, for the other transforms we have the following lemma.

Lemma B.1. *For a given aspect ratio $c > 0$, the companion moment generating function $\underline{M}$, chi transform $\underline{\chi}$, and S-transform $\underline{S}$ satisfy*

$$\underline{M}(z) = cM(z), \qquad \underline{\chi}(z) = \chi(z/c), \qquad \underline{S}(z) = \frac{z+1}{z+c} S\left(\frac{z}{c}\right).$$

Proof. Since $m(z) = -z^{-1}(1 + M(z^{-1}))$ and likewise for $\underline{m}(z)$,

$$-\frac{1}{z}(1 + \underline{M}(z^{-1})) = -\frac{c}{z}(1 + M(z^{-1})) + \frac{c-1}{z},$$

and by simplifying terms, we find that $\underline{M}(z^{-1}) = cM(z^{-1})$. For the chi transform, since $z = \underline{M}(\underline{\chi}(z)) = cM(\underline{\chi}(z))$, it follows that $\chi(z/c) = \underline{\chi}(z)$. Finally,

$$\underline{S}(z) = \frac{z+1}{z}\underline{\chi}(z) = \frac{z+1}{z}\chi\left(\frac{z}{c}\right) = \frac{z+1}{z} \cdot \frac{z/c}{z/c+1} S\left(\frac{z}{c}\right) = \frac{z+1}{z+c} S\left(\frac{z}{c}\right).$$

□

Appendix C Deformed Decompression on Algebraic Spectral Curves

In this section, we cover the mathematical fundamentals for deformed decompression. Our exposition begins in Appendix C.1 by proving our key Theorem 1, both in terms of free probability, and more rigorously using the Silverstein equation. Appendix C.2 proves that, for $0 < c \leq 1$, every probability measure with a well-defined S-transform admits forward deformed decompression for any $\tau \geq 1$. The PDEs in Theorem 2 are derived in Appendix C.3. Using these PDEs, Appendix C.4 shows that the Marchenko–Pastur distribution is the only probability measure that correctly evolves under both free decompression and deformed decompression, implying that the two operations are fundamentally distinct.

Our more practical exposition begins in Appendix C.5 with a derivation of the spectral curves along which the evolved companion Stieltjes transform $\underline{m}_\tau$ lie. Appendix C.6 develops a polynomial relation for $\underline{m}_\tau$ similar to Ameli et al. [3], although we find this to be primarily a theoretical curiosity with limited practical utility. Instead, solving for $\underline{m}_\tau$ is achieved using a predictor-corrector procedure described in Appendix C.7.

C.1 The Key Theorem

Our exposition begins with a complete proof of the key S-transform relationship which underpins the deformed decompression procedure. Recall that our generalized covariance matrix $\mathbf{S}$ and its companion $\underline{\mathbf{S}}$ satisfy

$$\mathbf{S} = \frac{1}{n}\boldsymbol{\Sigma}_p^{1/2}\mathbf{Z}\mathbf{Z}^\intercal\boldsymbol{\Sigma}_p^{1/2}, \qquad \underline{\mathbf{S}} = \frac{1}{n}\mathbf{Z}^\intercal\boldsymbol{\Sigma}_p\mathbf{Z}.$$

It is a classical result in free probability theory that if $\mathbf{Z}$ is comprised of independent standard normal random variables, under Assumption 1, the spectral distribution of $\mathbf{S}$ converges to the free multiplicative convolution of $\nu_{\mathbf{\Sigma}}$ and a Marchenko-Pastur distribution, that is,

$$\nu_{\mathbf{S}} = \nu_{\mathbf{\Sigma}} \boxtimes \nu_{\mathrm{MP}_c}.$$

In terms of the S-transform which factorizes free multiplicative convolution,

$$S_{\mathbf{S}}(w) = S_{\mathbf{\Sigma}}(w) S_{\mathrm{MP}_c}(w) = \frac{1}{1+cw} S_{\mathbf{\Sigma}}(w).$$

The free probability proofs of this result rely strongly on the unitary invariance of $\mathbf{Z}$ to establish that $\{\{\mathbf{Z}_n\}_{n=1}^{\infty}, \{\mathbf{Z}_n^{\intercal}\}_{n=1}^{\infty}\}$ and $\{\mathbf{\Sigma}_{p(n)}\}_{n=1}^{\infty}$ are asymptotically free [57, Theorem 2.3]. However, it turns out that the asymptotic distribution does not rely on the precise nature of the random variables which comprise $\mathbf{Z}$. If the elements of $\mathbf{Z}$ are instead only independent and identically distributed with unit variance, universality takes hold and the resulting limiting spectral distribution of $\mathbf{S}$ turns out the same. The proof of this particular finding is due to Silverstein [53] and follows a much different approach involving careful bounds on deterministic equivalents. A modern discussion of the proof is presented in Couillet & Liao [14]. We present the complete result below in the form most convenient for our purposes.

Theorem C.1 (Silverstein)**.** *Let $\mathbf{Z}_n$ be a $p \times n$ matrix for $p \equiv p(n)$ comprised of independent and identically distributed real-valued entries with unit variance. Let $\mathbf{\Sigma}_p$ be a $p \times p$ Hermitian non-negative definite deterministic matrix. Define $\mathbf{X}_n = \frac{1}{\sqrt{n}} \mathbf{\Sigma}_p^{1/2} \mathbf{Z}_n$, and let the sample covariance matrix be $\mathbf{S}_n = \mathbf{X}_n \mathbf{X}_n^{\intercal}$. Under the limit $n \to \infty$ with $p(n)/n \to c \in (0, \infty)$ and Assumption 1, almost surely, the ESD of $\mathbf{S}_n$ converges weakly to a deterministic probability measure ν with S-transform satisfying*

$$S(w) = \frac{1}{1+cw} S_{\mathbf{\Sigma}}(w).$$

Proof. This theorem is a restatement of the original equation of Silverstein [53]. Indeed, by Silverstein [53, Theorem 1.1], under the stated hypotheses, the ESD of $\mathbf{S}_n$ converges weakly to a deterministic probability measure ν whose Stieltjes transform satisfies the fixed point equation

$$m(z) = \int \frac{1}{t(1-c-czm(z)) - z} \nu_{\mathbf{\Sigma}}(\mathrm{d}t). \tag{C.1}$$

Now, since $m(z) = -z^{-1}(1 + M(z^{-1}))$ where M is the moment generating function, $1 - c - czm(z) = 1 + cM(z^{-1})$. Substituting this back into (C.1) gives

$$m(z) = \int \frac{1}{t(1+cM(z^{-1})) - z} \nu_{\mathbf{\Sigma}}(\mathrm{d}t).$$

Performing the change of variables $z \mapsto z^{-1}$ and using the identity $M(z) = -1 - z^{-1} m(z^{-1})$ reveals

$$\begin{aligned} M(z) &= -1 + \int \frac{1}{1 - tz(1+cM(z))} \nu_{\mathbf{\Sigma}}(\mathrm{d}t) \\ &= \int \frac{tz(1+cM(z))}{1 - tz(1+cM(z))} \nu_{\mathbf{\Sigma}}(\mathrm{d}t). \end{aligned}$$

But recalling the integral representation for $M_{\mathbf{\Sigma}}$, this implies

$$M(z) = M_{\mathbf{\Sigma}}(z(1+cM(z))).$$

Now, let $w = M(z)$ so that $z = \chi(w)$. Then $w = M_{\mathbf{\Sigma}}(\chi(w)(1+cw))$ which implies $\chi_{\mathbf{\Sigma}}(w) = \chi(w)(1+cw)$. Immediately, we find that $S_{\mathbf{\Sigma}}(w) = S(w)(1+cw)$, which implies the result. □

Since $S_{\mathbf{\Sigma}}$ is independent of the aspect ratio, the S-transform for the measure ν_c (the spectral distribution of $\mathbf{S}$ with aspect ratio c), which we shall denote by S, is related to the S-transform for the measure $\nu_{\tau c}$ for some decompression ratio $\tau > 1$ by

$$S(w)(1+cw) = S_{\tau}(w)(1+c\tau w).$$

Consequently, solving for S_τ gives ***the S-transform relation for deformed decompression***:

$$S_\tau(w) = \frac{1+cw}{1+c\tau w} S(w). \tag{C.2}$$

Finally, to get to Theorem 1 from (C.2), we make use of the following proposition.

Proposition C.2. *The following three relationships are equivalent.*

(i) $S_\tau(w) = \dfrac{1+cw}{1+c\tau w} S(w)$

(ii) $\underline{S}_\tau(w) = \tau^{-1} \underline{S}(w/\tau)$

(iii) $\underline{R}_\tau(z) = \tau \underline{R}(z)$

Proof. First, we show that (i) $\implies$ (ii). From Lemma B.1,

$$\underline{S}_\tau(w) = \frac{w+1}{w+c\tau} S_\tau\left(\frac{w}{c\tau}\right) = \frac{w+1}{w+c\tau} \cdot \frac{1+w/\tau}{1+w} S\left(\frac{w}{c\tau}\right) = \frac{1}{\tau} \cdot \frac{w/\tau+1}{w/\tau+c} S\left(\frac{w}{c\tau}\right),$$

but since $\underline{S}_1(w) = \frac{w+1}{w+c} S(w/c)$, we obtain (ii). Now, (ii) implies that $z := w\underline{S}_\tau(w) = w/\tau \underline{S}(w/\tau)$. But recall from (B.6) that this implies $z\underline{R}_\tau(z) = w$ and $z\underline{R}(z) = w/\tau$. Equating in w, $z\underline{R}(z) = \frac{1}{\tau} z\underline{R}_\tau(z)$, from whence (iii) is immediate. Since all operations presented are formally reversible, this also implies that (iii) $\implies$ (ii) $\implies$ (i). □

C.2 The Free Binomial Distribution

In this section, we show that for $0 < c \leq 1$, *every* probability measure with a well-defined S-transform admits forward deformed decompression for any $\tau \geq 1$. This is in stark contrast to free decompression outlined in Ameli et al. [2, 3]. Indeed, a probability measure is only infinitely freely decompressible if it is freely infinitely divisible [3, Proposition F.3], which is a fairly restrictive class [8]. The main tool in this proof is the construction of the *free binomial distribution* which has precisely the S-transform required to ensure that deformed decompression is simply free multiplicative convolution with the free binomial distribution.

Definition C.1. Let $c \in (0, 1]$ and $\tau \geq 1$. The *free binomial distribution* $\beta_{c,\tau}$ is the probability measure with S-transform

$$S_{c,\tau}(w) = \frac{1+cw}{1+c\tau w}.$$

In Proposition C.3 we show that the free binomial distribution exists and describe its properties.

Proposition C.3. *The free binomial distribution is well-defined and is given by*

$$\mathrm{d}\beta_{c,\tau} = \rho_{c,\tau}\mathrm{d}x + \max\{0, 1-(c\tau)^{-1}\}\delta_0 + \max\{0, 1-c^{-1}+(c\tau)^{-1}\}\delta_\tau,$$

where δ_x is an atomic measure at x and ρ is the density

$$\rho(x) = \frac{\sqrt{(x_+ - x)(x - x_-)}}{2\pi c x(\tau - x)}, \quad x \in [x_-, x_+],$$

where $x_\pm = 1+c\tau-2c\pm 2\sqrt{c(1-c)(\tau-1)}$ prescribe the edges of the support. It also has Stieltjes transform $m_{c,\tau}$ satisfying

$$cz(z-\tau)\, m_{c,\tau}(z)^2 + ((2c-1)z + (1-c\tau))\, m_{c,\tau}(z) - (1-c) = 0. \tag{C.3}$$

Proof. To establish existence, let P and Q be freely independent projections with traces c and τ^{-1}. Let $\beta_{c,\tau}$ be the spectral law of τPQP on the range of P, equipped with the normalized trace. Since the S-transform of a projection with trace a is $(1+w)/(a+w)$, multiplicativity and the compression relation give

$$S_{\beta_{c,\tau}}(w) = \frac{c(1+w)}{\tau(1+cw)} S_P(cw) S_Q(cw) = \frac{1+cw}{1+c\tau w}.$$

Thus, $\beta_{c,\tau}$ is a probability measure with the required S-transform.

We next derive its density and atoms. Here, we drop the c, τ subscript for brevity. Using (B.5), we can derive an algebraic relation for the moment generating function. Indeed,

$$M(z)(1 + cM(z)) = z(1 + M(z))(1 + c\tau M(z)),$$

and so

$$M(z^{-1})(1 + cM(z^{-1})) = z^{-1}(1 + M(z^{-1}))(1 + c\tau M(z^{-1})).$$

From (B.4), it follows that

$$(zm(z) + 1)(1 - czm(z) - c) = m(z)(1 - c\tau zm(z) - c\tau).$$

Expanding reveals that

$$m(z) + 1 - cz^2 m(z)^2 - 2czm(z) - c = m(z) - c\tau zm(z)^2 - c\tau m(z),$$

and rearranging terms arrives at (C.3). Solving using the quadratic formula reveals two possible branches for the Stieltjes transform:

$$m_{c,\tau}(z) = \frac{-((2c-1)z + (1-c\tau)) \pm \sqrt{\Delta(z)}}{2cz(z-\tau)},$$

where $\Delta(z) = ((2c-1)z + (1-c\tau))^2 + 4c(1-c)z(z-\tau)$ is the discriminant. Simplifying,

$$\Delta(z) = z^2 + 2(2c - 1 - c\tau)z + (1 - c\tau)^2.$$

For z on the real line, $\Delta(z) < 0$ if $z \in [x_-, x_+]$ where

$$\begin{aligned} x_\pm &= 1 + c\tau - 2c \pm \sqrt{(2c - 1 - c\tau)^2 - (1 - c\tau)^2} \\ &= 1 + c\tau - 2c \pm 2\sqrt{c(1-c)(\tau - 1)} \end{aligned}$$

This prescribes the support of the density. Possible atoms occur where the highest-order coefficient in m vanishes, namely at $z = 0$ and $z = \tau$. Their masses, when positive, are given by $((2c-1)z + (1-c\tau))/(2cz - c\tau)$ evaluated at the corresponding locations. By the Stieltjes inversion formula, the density ρ in the interval $[x_-, x_+]$ is given by

$$\rho(x) = \frac{\sqrt{-\Delta(x)}}{2\pi cx(\tau - x)},$$

which completes the proof. □

Using the free binomial distribution, it is straightforward to show that all measures are infinitely deformed decompressible.

Proposition C.4 (All Measures are Decompressible)**.** *Let ν be an arbitrary probability measure with S-transform S. For any $c \in (0, 1]$ and $\tau \geq 1$, there exists a probability measure ν_τ with S-transform S_τ such that $(1 + c\tau w)S_\tau(w) = (1 + cw)S(w)$.*

Proof. From Proposition C.3, for any $c \in (0, 1]$ and $\tau \geq 1$, there exists a probability measure $\beta_{c,\tau}$ such that the S-transform of $\beta_{c,\tau}$ is given by $S_{c,\tau}(w) = \frac{1+cw}{1+c\tau w}$. The construction $\nu_\tau = \beta_{c,\tau} \boxtimes \nu$ is always well-defined [6] and has S-transform of the desired form. □

The construction above guarantees forward evolution for $0 < c \leq 1$ and $\tau \geq 1$; outside this range, $S_{c,\tau}$ need not correspond to a probability measure. This does not imply that $S_\tau(w)$ is necessarily invalid (for example, if ν is a Marchenko–Pastur distribution, then $S_\tau(w)$ is a valid S-transform), but suggests that additional conditions on the initial law are required for the reverse operation. This is consistent with the relative ill-posedness of recovering the underlying population spectral distribution by taking $\tau \to 0$.

C.3 Deriving the Partial Differential Equations

Let ν be a probability measure and let $\{\nu_\tau\}_{\tau \geq 1}$ be a family of probability measures obtained by deformed decompression such that for each τ, denoting by $\underline{R}_\tau$ the companion R-transform of ν_τ, $\underline{R}_\tau(z) = \tau \underline{R}(z)$, where $\underline{R}$ is the companion R-transform of ν. For each $\tau \geq 1$, let m_τ denote the Stieltjes transform corresponding to ν_τ, and $\underline{m}_\tau$ its companion Stieltjes transform. In the following, we look to obtain PDEs to describe the evolution of the deformed decompression flow. Our derivations begin with the PDE for the companion Stieltjes transform. Similar to Ameli et al. [2] in the case of free decompression, it is most convenient to express this first equation under the change of variables $\tau = e^t \in [1, \infty)$.

Proposition C.5. *The companion Stieltjes transforms $\underline{m}(t, z) = \underline{m}_{e^t}(z)$ satisfy the PDE*

$$\frac{\partial \underline{m}}{\partial t} + \frac{\partial \underline{m}}{\partial z}\left(z + \frac{1}{\underline{m}}\right) = 0.$$

Proof. For brevity, we let $\underline{R}(t, z) = \underline{R}_{e^t}(z)$ and consider the K-transform $K(t, z) = \underline{R}(t, z) + z^{-1}$. Since $\underline{R}(t, z) = e^t \underline{R}(0, z)$, it follows that

$$\frac{\partial K}{\partial t}(t, z) = e^t \underline{R}(0, z) = K(t, z) - \frac{1}{z}.$$

Since $-\underline{m}(t, K(t, w)) = w$,

$$\frac{\partial \underline{m}}{\partial t}(t, K(t, w)) + \frac{\partial \underline{m}}{\partial z}(t, K(t, w))\left(K(t, w) - \frac{1}{w}\right) = 0.$$

Letting $w = -\underline{m}(t, z)$ recovers the result. □

Next, we consider the Stieltjes transform itself. This particular equation is best expressed in the original τ parameterization. A change of variables shows that

$$\tau \underline{m} \frac{\partial \underline{m}}{\partial \tau} + \frac{\partial \underline{m}}{\partial z}(\underline{m} z + 1) = 0 \tag{C.4}$$

Proposition C.6. *The Stieltjes transforms $m_\tau(z)$ satisfy the PDE*

$$(c\tau z m + c\tau - 1)\frac{\partial m}{\partial \tau} + c(zm + 1)\left(z\frac{\partial m}{\partial z} + m\right) = 0. \tag{C.5}$$

Proof. Recalling that $\underline{m} = c\tau m + (c\tau - 1)/z$,

$$\begin{aligned}
\frac{\partial \underline{m}}{\partial \tau} &= cm + c\tau \frac{\partial m}{\partial \tau} + \frac{c}{z} \\
\frac{\partial \underline{m}}{\partial z} &= c\tau \frac{\partial m}{\partial z} - \frac{c\tau - 1}{z^2} \\
z\underline{m} + 1 &= c\tau(zm + 1).
\end{aligned}$$

For the terms in the PDE (C.4), we have

$$\begin{aligned}
\tau \underline{m} \frac{\partial \underline{m}}{\partial \tau} &= c\tau \left(c\tau m + \frac{c\tau - 1}{z}\right)\left(m + \tau \frac{\partial m}{\partial \tau} + \frac{1}{z}\right) \\
\frac{\partial \underline{m}}{\partial z}(z\underline{m} + 1) &= \left(c\tau \frac{\partial m}{\partial z} - \frac{c\tau - 1}{z^2}\right) c\tau(zm + 1).
\end{aligned}$$

Substituting into the PDE from (C.4) and dividing by $c\tau$ gives

$$\left(c\tau m + \frac{c\tau - 1}{z}\right)\left(m + \tau \frac{\partial m}{\partial \tau} + \frac{1}{z}\right) + \left(c\tau \frac{\partial m}{\partial z} - \frac{c\tau - 1}{z^2}\right)(zm + 1) = 0.$$

Multiplying by z to clear the fractions, we obtain

$$(c\tau z m + c\tau - 1)\left(m + \tau \frac{\partial m}{\partial \tau} + \frac{1}{z}\right) + \left(c\tau z \frac{\partial m}{\partial z} - \frac{c\tau - 1}{z}\right)(zm + 1) = 0.$$

Expanding the products yields

$$(c\tau zm + c\tau - 1)m + (c\tau zm + c\tau - 1)\tau\frac{\partial m}{\partial \tau} + c\tau m + \frac{c\tau - 1}{z}$$
$$+ c\tau z(zm + 1)\frac{\partial m}{\partial z} - (c\tau - 1)m - \frac{c\tau - 1}{z} = 0.$$

Observe that the terms $(c\tau - 1)m$ and $\frac{c\tau - 1}{z}$ cancel and the remaining non-derivative terms simplify to $c\tau zm^2 + c\tau m = c\tau m(zm + 1)$. After these reductions, we find

$$(c\tau zm + c\tau - 1)\tau\frac{\partial m}{\partial \tau} + c\tau z(zm + 1)\frac{\partial m}{\partial z} + c\tau m(zm + 1) = 0.$$

Finally, dividing by τ and factoring $c(zm + 1)$ from the last two terms, we arrive at the result. □

C.4 Deformed Decompression and Free Decompression

Given that deformed decompression appears to display similarities in form to free decompression given in Ameli et al. [2], in this section, we discuss when the two operations coincide. It turns out that deformed decompression and free decompression coincide only in the special case where the initial distribution ν is a scaled Marchenko–Pastur distribution.

Proposition C.7. *Suppose that $m_\tau(z)$ satisfies both the deformed decompression flow (C.5) and the free decompression flow given by the PDE*

$$\tau m\frac{\partial m}{\partial \tau} = -m^2 + \frac{\partial m}{\partial z}. \tag{C.6}$$

Then there exists some $\sigma^2 \geq 0$ and some c such that for $\lambda = c\tau$,

$$\lambda\sigma^2 zm_\tau^2 + (z - \sigma^2(1 - \lambda))m_\tau + 1 = 0, \tag{C.7}$$

which is the defining relation for the Stieltjes transform of the Marchenko-Pastur distribution with aspect ratio λ and variance parameter σ^2.

Proof. From (C.6), we find that $\frac{\partial m}{\partial \tau} = \frac{1}{\tau m}(\frac{\partial m}{\partial z} - m^2)$. Substituting this into (C.5) yields the ordinary differential equation

$$(c\tau zm + c\tau - 1)(m' - m^2) + \tau mc(mz + 1)(zm' + m) = 0.$$

Expanding and simplifying terms reveals

$$\frac{\mathrm{d}m}{\mathrm{d}z} = \frac{m^2}{1 - c\tau(mz + 1)^2},$$

which is separable. However, observe that it will be far more convenient for the numerator and denominator in this expression to be reversed. To allow for this, consider instead the inverse function $z(m)$ which satisfies

$$\frac{\mathrm{d}z}{\mathrm{d}m} = \frac{1 - c\tau(mz + 1)^2}{m^2}.$$

Now let $v(m) = mz(m) + 1$ so that

$$\frac{\mathrm{d}v}{\mathrm{d}m} = z + m\frac{\mathrm{d}z}{\mathrm{d}m} = z + \frac{1 - c\tau v^2}{m} = \frac{v}{m}(1 - c\tau v).$$

Since the $m \equiv 0$ case is ruled out, consider the $v \equiv 0$ case, which implies that $m(z) = -1/z$, satisfying (C.7) with $\sigma^2 = 0$. If $c\tau v = 1$, then $m(z) = -\frac{1-(c\tau)^{-1}}{z}$, which fails to be a Stieltjes transform of a probability measure. With these cases ruled out, we may solve for v. Since

$$\frac{1}{v(1 - c\tau v)} = \frac{1}{v} + \frac{c\tau}{1 - c\tau v},$$

it follows that, for some constant C,

$$\log v - \log(1 - c\tau v) = \log m + C.$$

Letting $C = \log\sigma^2$ for some $\sigma^2 > 0$, we find that

$$\frac{v}{1 - \lambda v} = \sigma^2 m.$$

Substituting $v = mz + 1$ and rearranging yields (C.7). □

C.5 Formal Deformed 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

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

From Theorem 1, the operation of deformed decompression coincides with the map $\nu \mapsto \nu_\tau$ where the companion R-transforms $\underline{R}$ and $\underline{R}_\tau$ of ν and ν_τ, respectively, satisfy

$$\underline{R}_\tau(z) = \tau \underline{R}(z), \qquad \tau > 1.$$

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 deformed decompression. In doing so, we can track solutions to deformed decompression while avoiding any potential issues with m and/or R being multivalued. This is captured in Theorem C.8.

Theorem C.8. *For a polynomial $P(z, m)$, define the following spectral curves:*

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

$$\mathcal{C}_{\underline{m}} = \left\{(z, w) \in \mathbb{C}^2 \mid (z, c^{-1}(w - (c-1)z^{-1})) \in \mathcal{C}_m\right\}, \qquad \textit{(graph of companion } \underline{m}\textit{)}$$

$$\mathcal{C}_{\underline{R}} = \left\{(w, r) \in \mathbb{C}^2 \mid (r - 1/w, w) \in \mathcal{C}_{\underline{m}}\right\}, \qquad \textit{(R-transform of } \underline{m}\textit{)}$$

$$\mathcal{C}_{\underline{R}_\tau} = \left\{(w, r) \in \mathbb{C}^2 \mid (w, r/\tau) \in \mathcal{C}_{\underline{R}}\right\}, \qquad \textit{(R-transform of } \underline{m}_\tau\textit{)}$$

$$\mathcal{C}_{\underline{m}_\tau} = \left\{(z, \omega) \in \mathbb{C}^2 \mid \omega \neq 0 \textit{ and } (\omega, z + 1/\omega) \in \mathcal{C}_{\underline{R}_\tau}\right\}. \qquad \textit{(graph of } \underline{m}_\tau\textit{)}$$

Then

$$\mathcal{C}_{\underline{m}_\tau} \subseteq \mathcal{C}_{P_\tau} := \left\{(z, \omega) \in \mathbb{C}^2 \;\middle|\; P\left(\zeta, \frac{1}{c}\left(\omega - \frac{c-1}{\zeta}\right)\right) = 0 \quad \textit{where } \zeta = \frac{1}{\tau}\left(z + \frac{1}{\omega}\right) - \frac{1}{\omega}\right\}.$$

Proof. We begin by unpacking the conditions for a point (z, ω) to belong to $\mathcal{C}_{\underline{m}_\tau}$. By definition (excluding the $\omega = 0$ case), $(z, \omega) \in \mathcal{C}_{\underline{m}_\tau}$ if and only if

$$(\omega, z + 1/\omega) \in \mathcal{C}_{\underline{R}_\tau}.$$

According to the definition of $\mathcal{C}_{\underline{R}_\tau}$, this implies that

$$\left(\omega, \frac{1}{\tau}\left(z + \frac{1}{\omega}\right)\right) \in \mathcal{C}_{\underline{R}}.$$

By the definition of the companion R-transform curve $\mathcal{C}_{\underline{R}}$, setting $w = \omega$ and $r = \frac{1}{\tau}(z + \frac{1}{\omega})$, this implies that

$$\left(\frac{1}{\tau}\left(z + \frac{1}{\omega}\right) - \frac{1}{\omega}, \omega\right) \in \mathcal{C}_{\underline{m}}.$$

Denote the first coordinate as ζ, so $\zeta = \frac{1}{\tau}(z + \frac{1}{\omega}) - \frac{1}{\omega}$. We now have $(\zeta, \omega) \in \mathcal{C}_{\underline{m}}$. By the definition of the companion Stieltjes curve $\mathcal{C}_{\underline{m}}$, this means that there exists an underlying point in the original curve $\mathcal{C}_m$ satisfying

$$\left(\zeta, \frac{1}{c}\left(\omega - \frac{c-1}{\zeta}\right)\right) \in \mathcal{C}_m.$$

Finally, since belonging to $\mathcal{C}_m$ implies being a root of the annihilating polynomial P, we conclude that

$$P\left(\zeta, \frac{1}{c}\left(\omega - \frac{c-1}{\zeta}\right)\right) = 0.$$

Consequently, any $(z, \omega) \in \mathcal{C}_{\underline{m}_\tau}$ must satisfy this polynomial relation, meaning $(z, \omega) \in \mathcal{C}_{P_\tau}$. □

Theorem C.8 immediately implies Theorem 6, as we identify m_τ to be the deformed decompression of m by the factor τ.

C.6 Polynomial Evolution under Deformed Decompression

In this section, we use Theorem C.8 to derive a bivariate polynomial Q_τ such that the companion Stieltjes transform $\underline{m}_\tau$ under deformed decompression satisfies $Q_\tau(z, \underline{m}_\tau(z)) = 0$ for almost all z. This process is performed analogously to Ameli et al. [3, Corollary F.5] as a theoretical curiosity; we will find the resulting polynomial to exhibit high numerical instability in its roots.

Corollary C.9. *Suppose that the standard Stieltjes transform $m : \mathbb{C}^+ \to \mathbb{C}^+$ satisfies an algebraic relation of the form $P(z, m(z)) = 0$ 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 $\mathcal{C}_{\underline{m}_\tau}$ defined in Theorem C.8,

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

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

$$Q_\tau(z, \omega) := \sum_{p=0}^{d_z+s} \sum_{q=0}^{d_z+2s} c_{pq}^{(\tau)} z^p \omega^q,$$

and the coefficients $c_{pq}^{(\tau)}$ are defined as

$$c_{pq}^{(\tau)} := \sum_{j=0}^{d_z} c_{j,k} \tau^{-j} c^{-k} \sum_{u=0}^{p} \binom{s-k+j}{u} \binom{k}{p-u} (1-\tau)^{s-k+j-u} (1-\tau c)^{k-p+u},$$

where $k = q - p + j - d_z$, and we adopt the convention that $c_{jk} = 0$ if $k \notin \{0, \dots, s\}$ and $\binom{n}{r} = 0$ if $r < 0$ or $r > n$.

Proof. From Theorem C.8, we have the inclusion

$$\mathcal{C}_{\underline{m}_\tau} \subseteq \left\{ (z, \omega) \in \mathbb{C}^2 \,\middle|\, P\left(\zeta, \frac{1}{c}\left(\omega - \frac{c-1}{\zeta}\right)\right) = 0 \quad \text{where} \quad \zeta = \frac{1}{\tau}\left(z + \frac{1}{\omega}\right) - \frac{1}{\omega} \right\}.$$

For $(z, \omega) \in \mathcal{C}_{\underline{m}_\tau}$, we have $\omega \neq 0$. We first simplify the arguments evaluated by the original polynomial P. The first argument ζ can be rewritten with a common denominator as

$$\zeta = \frac{z}{\tau} + \frac{1-\tau}{\tau\omega} = \frac{z\omega + 1 - \tau}{\tau\omega}.$$

Let M denote the second argument of the polynomial P, which effectively maps back to the standard Stieltjes transform domain. Substituting our expression for ζ into M, we obtain

$$\begin{aligned} M &= \frac{1}{c}\left(\omega - \frac{c-1}{\frac{z\omega+1-\tau}{\tau\omega}}\right) = \frac{1}{c}\left(\omega - \frac{\tau\omega(c-1)}{z\omega+1-\tau}\right) \\ &= \frac{\omega(z\omega+1-\tau) - \tau\omega(c-1)}{c(z\omega+1-\tau)} \\ &= \frac{\omega(z\omega+1-\tau c)}{c(z\omega+1-\tau)}. \end{aligned}$$

Since (z, ω) originates from the curve ensuring these mapped coordinates are a root of P, substituting these rational functions into the polynomial expansion yields

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

To clear the denominators, we multiply the entire equation by $\omega^{d_z}(z\omega + 1 - \tau)^s$. Because the summation indices satisfy $j \leq d_z$ and $k \leq s$, this guarantees all terms become proper bivariate polynomials in z and ω:

$$\omega^{d_z}(z\omega + 1 - \tau)^s P(\zeta, M) = \sum_{j=0}^{d_z} \sum_{k=0}^{s} c_{jk} \tau^{-j} c^{-k} \omega^{d_z - j + k} (z\omega + 1 - \tau)^{s-k+j} (z\omega + 1 - \tau c)^k.$$

We define the left-hand side of this expression as the unnormalized polynomial $Q_\tau(z, \omega)$. To extract its explicit coefficients, we expand the binomial products in terms of $z\omega$. Using the binomial theorem on both factors,

$$\begin{aligned}(z\omega + 1 - \tau)^{s-k+j}(z\omega + 1 - \tau c)^k = &\left(\sum_{u=0}^{s-k+j} \binom{s-k+j}{u} (z\omega)^u (1-\tau)^{s-k+j-u} \right) \\ &\times \left(\sum_{v=0}^{k} \binom{k}{v} (z\omega)^v (1-\tau c)^{k-v} \right).\end{aligned}$$

Taking the Cauchy product of these two sums, the coefficient of $(z\omega)^p$ where $p = u + v$ is given by

$$C_{j,k,p} = \sum_{u=0}^{p} \binom{s-k+j}{u} \binom{k}{p-u} (1-\tau)^{s-k+j-u} (1-\tau c)^{k-p+u}.$$

Substituting this combinatorial term back into our expression for $Q_\tau(z, \omega)$ gives:

$$Q_\tau(z, \omega) = \sum_{j=0}^{d_z} \sum_{k=0}^{s} \sum_{p=0}^{s+j} c_{jk} \tau^{-j} c^{-k} C_{j,k,p} z^p \omega^{d_z - j + k + p}.$$

Finally, we change summation variables to group the terms strictly by powers of z and ω. Let p remain the power of z, and let $q = d_z - j + k + p$ be the power of ω. Consequently, we can isolate k as $k = q - p + j - d_z$. Reindexing the sums in terms of p and q yields the final polynomial form:

$$Q_\tau(z, \omega) = \sum_{p=0}^{d_z+s} \sum_{q=0}^{d_z+2s} c_{pq}^{(\tau)} z^p \omega^q,$$

as required. □

Unlike free decompression [3], we are unable to find conditions which guarantee that the degree of Q_τ, once simplified, does not exceed the degree of P. Indeed, the substantial increase in the degree of Q_τ over the initial polynomial P often renders Q_τ unusable in practice, as small numerical errors in the coefficients result in vast discrepancies in the roots.

C.7 Deformed Decompression on Spectral Curves

Analogous to Ameli et al. [3], identifying the desired solution $\underline{m}_\tau$ to the deformed decompression flow as a root of the polynomial Q_τ is not sufficient for reconstructing $\underline{m}_\tau$, as it remains unclear which root to select. Unlike free decompression, however, deformed decompression benefits from a key geometric property. Biane's Subordination Theorem [9] guarantees that the transport on the curve induced by deformed decompression maps the upper half plane into itself. Consequently, if we start on the physical branch, the continuation remains robustly physical. To identify this correct root, we track the known physical sheet at $\tau = 1$ along the deformed decompression flow, analogous to the geometric procedure used in Ameli et al. [3] for free decompression.

Let the initial fitted spectral curve be given by

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

where $y = m_1(\zeta)$ denotes a branch of the initial ordinary Stieltjes transform. Let c denote the initial aspect ratio, so the target aspect ratio at scale τ is τc. For a fixed target query point z, let $w = \underline{m}_\tau(z)$ be the decompressed companion Stieltjes transform value. Along the spectral curves of

the deformation, this value is conserved, so $w = \underline{m}_1(\zeta)$. The ordinary and companion transforms on the initial curve are related by $w = cy + \zeta^{-1}(c-1)$. From Theorem 6, the change of variables is given by

$$\zeta = \tau^{-1}(z + w^{-1}) - w^{-1}.$$

Substituting the relation for w into this inverse map yields the evolved algebraic relation characterizing the candidate solutions. However, to track the physical solution directly, we cast (ζ, w) as a coupled system:

$$F_1(\zeta, w) \coloneqq P\left(\zeta, c^{-1}\left(w - \frac{c-1}{\zeta}\right)\right) = 0 \tag{C.8a}$$

$$F_2(\zeta, w; \tau, z) \coloneqq \zeta - \tau^{-1}z - (\tau^{-1} - 1)w^{-1} = 0. \tag{C.8b}$$

At $\tau = 1$, the system (C.8) reduces to $\zeta = z$ and $w = \underline{m}_1(z)$, where the initial physical point is already known from the branch selection procedure adopted from Ameli et al. [3]. The task is to continue this solution in τ while keeping z fixed, using a predictor-corrector approach along the constrained curve.

Differentiating (C.8) with respect to τ provides the predictor direction. For $F_1 = 0$, we will define for brevity,

$$\underline{P}(\zeta, w) \coloneqq P\left(\zeta, c^{-1}\left(w - \frac{c-1}{\zeta}\right)\right).$$

For $F_2 = 0$, differentiation with respect to τ yields

$$\dot{\zeta} + \tau^{-2}(z + w^{-1}) + (\tau^{-1} - 1)w^{-2}\dot{w} = 0.$$

This reveals the 2×2 tangent system

$$\begin{bmatrix} \partial_\zeta \underline{P}(\zeta, w) & \partial_w \underline{P}(\zeta, w) \\ 1 & (\tau^{-1} - 1)w^{-2} \end{bmatrix} \begin{bmatrix} \dot{\zeta} \\ \dot{w} \end{bmatrix} = \begin{bmatrix} 0 \\ -\tau^{-2}(z + w^{-1}) \end{bmatrix}. \tag{C.9}$$

The system (C.9) provides the predictor direction for advancing the physical point from one decompression ratio to the next. The predicted point is then corrected by a Newton projection onto the nonlinear system $(F_1, F_2) = (0, 0)$. As in Ameli et al. [3], steps in τ are adaptively subdivided when the local geometry becomes stiff to prevent jumping between nearby sheets. After the correction step converges, the value of the decompressed ordinary Stieltjes transform is recovered from the companion transform via the relation

$$m_\tau(z) = (\tau c)^{-1}\left(w - \frac{\tau c - 1}{z}\right).$$

The complete predictor-corrector procedure is summarized in Algorithm C.1, and allows us to coherently evolve the density while remaining strictly on the physical branch.

Appendix D Stability Analysis

To analyze the stability of $\underline{m}_\tau$ it is convenient to reparameterize in terms of $t = \log(\tau)$. Noting that $\frac{\mathrm{d}\tau}{\mathrm{d}t} = \tau$, our PDE (1) becomes

$$\underline{m}_t \frac{\partial \underline{m}_t}{\partial t} + (z\underline{m}_t + 1)\frac{\partial \underline{m}_t}{\partial z} = 0. \tag{D.1}$$

We then consider the PDE operator's effect on the perturbed field $\underline{m}_t + \delta\eta_t$, and derive a PDE governing the evolution of the perturbation direction η_t.

We are then able to quantify the impact of this perturbation under the horizontal slice-wise L^2-norm in the upper half complex plane, defined by

$$\|\eta_t(\cdot + iy)\|^2_{L^2(\Gamma_y)} \coloneqq \int_{\mathbb{R}} |\eta_t(x + iy)|^2 \, \mathrm{d}x,$$

where $\Gamma_y \coloneqq \{z = x + iy \mid x \in \mathbb{R}\}$ is slice at height $y > 0$. This is a natural choice of norm for our analysis, since recovery of the spectral distribution requires evaluation over slices slightly above the real axis.

We establish a Grönwall-type inequality controlling the growth of the perturbation norm from its initial value. This is made precise in the following proposition.

Algorithm C.1: Deformed decompression by geometric continuation on a spectral curve

Input : Polynomial $P \in \mathbb{R}[z, m]$ such that $P(z, m_1(z)) = 0$ (Ameli et al. [3, Alg. D.1]);
Initial aspect ratio $c > 0$;
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 $m_1(z)$ on $P(z, m) = 0$ (Ameli et al. [3, Alg. 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 $j = 1, \dots, N_z$ **do**
  $m_j^{(0)} \leftarrow m_1(z_j)$ *// Select physical root using Ameli et al. [3, Appendix E]*
  $w_j^{(0)} \leftarrow c m_j^{(0)} + (c-1) z_j^{-1}$ *// Compute initial companion transform $\underline{m}_1$*
  $\zeta_j^{(0)} \leftarrow z_j$ *// At $\tau = 1$, the map is identity in z*
  $m_{\tau_0}(z_j) \leftarrow m_j^{(0)}$

// Continue the physical sheet in decompression ratio τ

for $\ell = 1, \dots, T$ **do**
  for $j = 1, \dots, N_z$ **do**
    Set $(\zeta, w) \leftarrow (\zeta_j^{(\ell-1)}, w_j^{(\ell-1)})$.
    Define $\underline{P}(\zeta, w) \coloneqq P\left(\zeta, c^{-1}\left(w - \frac{c-1}{\zeta}\right)\right)$.
    // Predict along the tangent of the constrained system
    Solve

$$\begin{bmatrix} \partial_\zeta \underline{P}(\zeta, w) & \partial_w \underline{P}(\zeta, w) \\ 1 & (\tau_{\ell-1}^{-1} - 1) w^{-2} \end{bmatrix} \begin{bmatrix} \dot{\zeta} \\ \dot{w} \end{bmatrix} = \begin{bmatrix} 0 \\ -\tau_{\ell-1}^{-2}(z_j + w^{-1}) \end{bmatrix}.$$

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

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

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

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

Proposition D.1. *Suppose $\rho(t, \cdot) \in L^2(\mathbb{R})$, and let $m_t(z)$ be its companion Stieltjes transform defined on $\mathbb{C}^+$. Let η_t be a Gâteaux perturbation direction of $\underline{m}_t$, with perturbation $\delta\eta_t$ for $0 < \delta \ll 1$, and suppose η_t satisfies the Bernstein inequality*

$$\|\partial_x \eta_t(z)\|_{L^2(\Gamma_y)} \le M \|\eta_t(z)\|_{L^2(\Gamma_y)}, \quad z \in \Gamma_y, \forall y > 0 \tag{D.2}$$

meaning that $\eta_t(\cdot + iy)$ lies in a M-band-limited Paley–Wiener space on each slice, ensuring appropriate perturbation decay. Then, for each $y > 0$, there holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\eta_t\|^2_{L^2(\Gamma_y)} \le (1 + C_y(t))\ \|\eta_t\|^2_{L^2(\Gamma_y)}, \tag{D.3}$$

where

$$C_y(t) \coloneqq \sup_{z\in\Gamma_y} \Re\, \frac{\partial_z \underline{m}_t(z)}{\underline{m}_t(z)^2} + 2M \sup_{z\in\Gamma_y} \left|\Im\left(z + \frac{1}{\underline{m}_t(z)}\right)\right|. \tag{D.4}$$

Proof. Write $\underline{m}'_t = \underline{m}_t + \delta\eta_t$, where $\eta_t \in H^2(\mathbb{C}^+)$ is the usual Hardy space, for a small perturbation with the boundary condition $\eta_t(z) = 0$ at $|z| \to \infty$. Plugging $\underline{m}'_t$ into the PDE (D.1) and discarding the $\mathcal{O}(\delta^2)$ terms as $\delta \to 0$ gives the linear perturbation equation

$$\frac{\partial \eta_t}{\partial t} = -\left(z + \frac{1}{\underline{m}_t}\right)\frac{\partial \eta_t}{\partial z} + \frac{1}{\underline{m}_t^2}\frac{\partial \underline{m}_t}{\partial z}\eta_t.$$

This is a linear transport PDE, with transport velocity $\left(z + \frac{1}{\underline{m}_t}\right)$ and growth coefficient $\frac{1}{\underline{m}_t^2}\frac{\partial \underline{m}_t}{\partial z}$.

Since η_t is holomorphic, $\frac{\partial \eta_t}{\partial z} = \frac{\partial \eta_t}{\partial x}$ along Γ_y, so on these slices we have

$$\frac{\partial \eta_t}{\partial t} = -\left(z + \frac{1}{\underline{m}_t}\right)\frac{\partial \eta_t}{\partial x} + \frac{1}{\underline{m}_t^2}\frac{\partial \underline{m}_t}{\partial z}\eta_t.$$

To obtain an energy estimate, we multiply by the complex conjugate $\overline{\eta_t}$, integrate over Γ_y and take the real part:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\eta_t\|^2_{L^2(\Gamma_y)} = \Re\int_{\Gamma_y} -\left(z + \frac{1}{\underline{m}_t}\right)\frac{\partial \eta_t}{\partial x}\overline{\eta_t}\mathrm{d}x + \int_{\Gamma_y} \Re\left(\frac{1}{\underline{m}_t^2}\frac{\partial \underline{m}_t}{\partial z}\right)|\eta_t|^2\mathrm{d}x.$$

Decomposing the velocity into its real and imaginary components as $z + \frac{1}{\underline{m}_t} = A + iB$, we observe

$$\Re\left(-\left(z + \frac{1}{\underline{m}_t}\right)\frac{\partial \eta_t}{\partial x}\overline{\eta_t}\right) = -\frac{1}{2}A\frac{\partial|\eta_t|^2}{\partial x} + B\Im\left(\frac{\partial \eta_t}{\partial x}\overline{\eta_t}\right)$$

Integrating this by parts, keeping in mind that the boundary terms vanish, we find

$$\Re\int_{\Gamma_y} -\left(z + \frac{1}{\underline{m}_t}\right)\frac{\partial \eta_t}{\partial x}\overline{\eta_t}\mathrm{d}x = \frac{1}{2}\int_{\Gamma_y}\frac{\partial A}{\partial x}|\eta_t|^2\mathrm{d}x + \int_{\Gamma_y} B\Im\left(\frac{\partial \eta_t}{\partial x}\overline{\eta_t}\right)\mathrm{d}x.$$

The Bernstein inequality (D.2) estimates the second term as

$$\left|\int_{\Gamma_y} B\Im\left(\frac{\partial \eta_t}{\partial x}\overline{\eta_t}\right)\mathrm{d}x\right| \le M\sup_{\Gamma_y}|B|\|\eta_t\|^2_{L^2(\Gamma_y)}.$$

Noting that $\frac{\partial A}{\partial x} = 1 - \Re\left(\frac{1}{\underline{m}_t^2}\frac{\partial \underline{m}_t}{\partial z}\right)$, we find that

$$\frac{1}{2}\frac{\partial A}{\partial x} + \Re\left(\frac{1}{\underline{m}_t^2}\frac{\partial \underline{m}_t}{\partial z}\right) = \frac{1}{2} + \frac{1}{2}\Re\left(\frac{1}{\underline{m}_t^2}\frac{\partial \underline{m}_t}{\partial z}\right).$$

Combining these terms we obtain the required result

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\eta_t\|^2_{L^2(\Gamma_y)} \le \left(1 + \sup_{z\in\Gamma_y}\Re\,\frac{\partial_z \underline{m}_t(z)}{\underline{m}_t(z)^2} + 2M\sup_{z\in\Gamma_y}\left|\Im\left(z + \frac{1}{\underline{m}_t(z)}\right)\right|\right)\ \|\eta_t\|^2_{L^2(\Gamma_y)}.$$

Using (D.4) and applying Grönwall's inequality we can further estimate

$$\|\eta_t\|^2_{L^2(\Gamma_y)} \le \|\eta_0\|^2_{L^2(\Gamma_y)}\exp\left(\int_0^t (1 + C_y(s))\,\mathrm{d}s\right).$$

□

Remark D.1. For compactly supported companion spectral measures, $C_y(t)$ is finite for every fixed $y > 0$. Indeed, $\Im\underline{m}_t(z) > 0$ on $\mathbb{C}^+$, and the Stieltjes expansion $\underline{m}_t(z) = -z^{-1} + O(z^{-2})$ implies $\partial_z\underline{m}_t(z)/\underline{m}_t(z)^2 \to 1$ and $z + 1/\underline{m}_t(z) = O(1)$ as $|z| \to \infty$. Thus, both suprema in (D.4) are finite on Γ_y. If the companion supports remain uniformly bounded over a finite decompression interval, so does $C_y(t)$. The bound need not remain finite as $y \to 0^+$. △

Appendix E Evolution of Spectral Quantities

Deformed decompression acts on spectral densities by evolving their Stieltjes transforms in a controlled way. While this gives access to the entire spectral distribution of a larger matrix, it is often the case that we only care about certain spectral quantities, such as condition numbers, spectral moments, atomic locations and weights, or cusp points where the bulk densities either merge or split. The spectral-curve viewpoint and its corresponding change-of-variable formulae allow for convenient calculation of these quantities.

E.1 Evolving Spectral Edges

Spectral edges are important quantities when modeling the spectrum of large matrices, with the most obvious applications being the estimation of the condition number or detecting when a family of matrices is becoming singular. More recently, the location of internal edges in multi-bulk distributions have been linked to training dynamics of diffusion models [10]. Writing the support of a distribution with k separate bulk components as

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

then the edges are simply the endpoints of the support subintervals $\{a_j(\tau), b_j(\tau)\}_{j=1}^{k}$.

Under Assumption 2 these edges correspond to branch points of the spectral curve

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

corresponding to points where multiple sheets (given by roots $y = y_i(\zeta)$) meet. These points have the convenient characterization as satisfying

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

describing how sheets glue across branch cuts. The branch points of interest to us are those that connect the physical sheet to non-physical sheets, since branch points connecting non-physical sheets do not correspond to the distribution of interest's support. This means that care must be taken to select the correct branch points.

With this in mind, we are able to track branch points under deformed decompression. This is formalized in the following proposition.

Proposition E.1 (Evolving Branch Points under Deformed Decompression)**.** *Let m be a Stieltjes transform satisfying Assumption 2, let $\tau \geq 1$, and let m_τ be the deformed decompression of m by the factor τ. Let*

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

Then, away from the singular chart loci, the branch-point candidates of the deformed decompressed curve are obtained from solutions $(\zeta_*, y_*) \in \mathcal{C}$ satisfying

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

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

The corresponding point in the deformed decompressed z-plane is

$$z_* = \tau\zeta_* + \frac{(\tau - 1)\zeta_*}{c\zeta_* y_* + c - 1}.$$

In particular, real solutions $z_* \in \mathbb{R}$ of (E.1) include the evolving spectral edges of the real density under deformed decompression.

Proof. Set $u \coloneqq \underline{m}_\tau$, and $y \coloneqq c^{-1}\big(u - \frac{c-1}{\zeta}\big)$. Then $u = cy + \frac{c-1}{\zeta} = \frac{c\zeta y + c - 1}{\zeta}$. The deformed decompression relation $P\big(\zeta, c^{-1}\big(\underline{m}_\tau - \frac{c-1}{\zeta}\big)\big) = 0$ therefore pulls back to the original curve $P(\zeta, y) = 0$. Moreover, from $\zeta = \tau^{-1}\big(z + u^{-1}\big) - u^{-1}$, we obtain

$$z = \tau\zeta + (\tau - 1)u^{-1} = \tau\zeta + \frac{(\tau - 1)\zeta}{c\zeta y + c - 1}.$$

Hence the branch points of the deformed decompressed curve are the critical points of the projection

$$(\zeta, y) \mapsto z(\zeta, y) = \tau\zeta + \frac{(\tau-1)\zeta}{c\zeta y + c - 1}$$

restricted to $\mathcal{C}$. Along $\mathcal{C}$, implicit differentiation gives

$$\partial_\zeta P(\zeta, y) + \partial_y P(\zeta, y)\frac{\mathrm{d}y}{\mathrm{d}\zeta} = 0, \qquad \text{so, whenever } \partial_y P \neq 0, \qquad \frac{\mathrm{d}y}{\mathrm{d}\zeta} = -\frac{\partial_\zeta P}{\partial_y P}. \tag{E.2}$$

Writing $D \coloneqq c\zeta y + c - 1$, we have $z(\zeta, y) = \tau\zeta + (\tau-1)\frac{\zeta}{D}$. Differentiating along $\mathcal{C}$,

$$\frac{\mathrm{d}z}{\mathrm{d}\zeta} = \tau + (\tau-1)\frac{D - \zeta D'}{D^2}.$$

Since $D' = cy + c\zeta\frac{\mathrm{d}y}{\mathrm{d}\zeta}$, we obtain

$$D - \zeta D' = c\zeta y + c - 1 - \zeta\left(cy + c\zeta\frac{\mathrm{d}y}{\mathrm{d}\zeta}\right) = c - 1 - c\zeta^2\frac{\mathrm{d}y}{\mathrm{d}\zeta}.$$

Therefore

$$\frac{\mathrm{d}z}{\mathrm{d}\zeta} = \tau + (\tau-1)\frac{c - 1 - c\zeta^2\frac{\mathrm{d}y}{\mathrm{d}\zeta}}{D^2}.$$

Substituting (E.2) gives

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

Thus $\mathrm{d}z/\mathrm{d}\zeta = 0$ is equivalent to

$$\tau D^2 \partial_y P + (\tau-1)\left(c\zeta^2\partial_\zeta P + (c-1)\partial_y P\right) = 0.$$

Since $D = c\zeta y + c - 1$, the critical-point equations are precisely (E.1). The pushforward formula for z_* follows from the expression for $z(\zeta, y)$.

Finally, we note that the algebraic system gives branch-point candidates. The spectral edges are the real-valued candidates whose cuts connect the physical sheet to a non-physical sheet. □

While Proposition E.1 holds for all branch points, we are concerned only with the selection of a particular subset, the spectral edges $\{a_j(\tau), b_j(\tau)\}_{j=1}^k$. For fixed $\tau \geq 1$, the system (E.1) admits solutions

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

where typically $p \geq k$. The real-valued subset $\mathcal{B}(\tau) \cap \mathbb{R}$ provides edge candidates, which can be selected as the ones connecting the physical to a non-physical sheet. This is realized as a nontrivial jump across the real axis at that point in the Stieltjes inversion formula. Algorithm E.1 summarizes the resulting procedure, and is a straight-forward adaptation of [3, Algorithm G.1] to the DD coordinate change.

E.2 Spectral Cusps and The Mechanism of Merging Bulks

Write $I_j(\tau) \coloneqq [a_j(\tau), b_j(\tau)]$ for the j-th bulk region and assume that at $\tau = 1$ the bulks are disjoint,

$$a_1(1) < b_1(1) < a_2(1) < b_2(1) < \cdots < a_k(1) < b_k(1),$$

so that $I_j(1) \cap I_{j+1}(1) = \emptyset$. After decompression to τ_*, merging or splitting of bulks can be characterized as the density developing a *cusp* (corresponding to a cubic singularity at a sharp local minimum), associated with the Pearcey universality class [22]. These colliding (or emerging) branch points correspond to a higher-order critical point of the projection of the spectral curve. Similar to the tracking of edges, under Assumption 2 there is a set of algebraic conditions that characterize these cusp points. The following proposition makes this precise.

Algorithm E.1: Evolution of spectral edges under algebraic deformed decompression

Input : Polynomial $P \in \mathbb{R}[z, m]$ with $\deg_m(P) = s$ such that $P(z, m_1(z)) = 0$ (see Ameli et al. [3, Algorithm E.1]);
Initial aspect ratio $c > 0$;
Decompression ratio grid $1 = \tau_0 < \tau_1 < \cdots < \tau_T$;
Initial support $I(1) = \bigcup_{j=1}^{k_0} [a_j(1), b_j(1)]$ (optional, for initialization);
Oracle for the physical branch $m_1(z)$ on $\mathbb{C} \setminus I$ (see [3, Algorithm E.1]); *// Only needed if $I(1)$ is provided.*
Tolerance $\epsilon > 0$.

Output : Edge trajectories $\{a_j(\tau_\ell), b_j(\tau_\ell)\}$ stored as $T + 1$ by $2k_0$ with `NaN`s after merging;
Active bulk count $k(\tau_\ell)$ (number of connected components).

// Define the edge system from Proposition E.1

Define the functions

$$G_1(\zeta, y; \tau) \coloneqq P(\zeta, y)$$

$$G_2(\zeta, y; \tau) \coloneqq \tau \left(c\zeta y + c - 1\right)^2 \partial_y P(\zeta, y) + (\tau - 1) \left(c\zeta^2 \, \partial_\zeta P(\zeta, y) + (c - 1)\partial_y P(\zeta, y) \right).$$

// Initialization at $\tau_0 = 1$

if *initial edges $I(1)$ 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 near each endpoint
    ▷ Choose Herglotz-consistent root $y_0^{(j)} \leftarrow m_1(\zeta_0^{(j)})$.
else
  Obtain initial guesses $\{(\zeta_0^{(j)}, y_0^{(j)})\}_{j=1}^{2k_0}$ by solving $(G_1, G_2) = (0, 0)$ at τ_0 using multiple seeds.

// Refine the initial guesses

Solve $(G_1, G_2) = (0, 0)$ at τ_0, store the initial edge locations $x_0^{(j)} = \zeta_0^{(j)}$. Set $k(\tau_0) \leftarrow k_0$.

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

for $\ell = 1, \ldots, T$ **do**
  $\tau \leftarrow \tau_\ell$
  foreach $j = 1, \cdots, 2k_0$ *not yet discarded* **do**
    Use previous solution $(\zeta_{\ell-1}^{(j)}, y_{\ell-1}^{(j)})$ as initial guess.
    Apply damped Newton to solve $(G_1(\zeta, y; \tau_\ell), G_2(\zeta, y; \tau_\ell)) = (0, 0)$ for $(\zeta_\ell^{(j)}, y_\ell^{(j)})$.
    $z_\ell^{(j)} \leftarrow \tau \zeta_\ell^{(j)} + \frac{(\tau-1)\zeta_\ell^{(j)}}{c\zeta_\ell^{(j)} y_\ell^{(j)} + c - 1}$ *// Pushforward to the z-plane*
    if $|\Im z_\ell^{(j)}| < \epsilon$ **then** Store $x_\ell^{(j)} \leftarrow \Re z_\ell^{(j)}$ **else** mark failure.
  // Handle merged edges of disjoint bulks (inner edges annihilate)
  $\mathcal{P} \leftarrow \{a_1, b_1, \ldots, a_{k_0}, b_{k_0}\}$
  foreach *adjacent pair* (b_j, a_{j+1}) *in* $\mathcal{P}$ **do**
    if $b_j(\tau_\ell) > a_{j+1}(\tau_\ell)$ **then** $(b_j(\tau_\ell), a_{j+1}(\tau_\ell)) \leftarrow (\texttt{NaN}, \texttt{NaN})$
  $k(\tau_\ell) \leftarrow$ number of remaining connected components.

return *Spectral edges* $\left\{ \{a_j(\tau_\ell), b_j(\tau_\ell)\}_{j=1}^{k_0} \right\}_{\ell=0}^{T}$ *and active bulk count* $\{k(\tau_\ell)\}_{\ell=0}^{T}$;

Proposition E.2 (Cusp Point at Merging Bulks under Deformed Decompression)**.** *Suppose two consecutive real edges collide at $\tau = \tau_*$, so that $x_* = b_j(\tau_*) = a_{j+1}(\tau_*)$. Let $(\zeta_*, y_*) \in \mathcal{C}$, where $\mathcal{C} = \{(\zeta, y) \in \mathbb{C}^2 : P(\zeta, y) = 0\}$, be the pullback of x_* under the deformed-decompression*

projection

$$z(\zeta, y) = \tau_*\zeta + \frac{(\tau_* - 1)\zeta}{c\zeta y + c - 1}.$$

Then (ζ_*, y_*) is a higher-order critical point of this projection restricted to $\mathcal{C}$. Hence, away from the singular chart loci, it satisfies

$$P(\zeta_*, y_*) = 0, \tag{E.3a}$$

$$\tau_*(c\zeta_* y_* + c - 1)^2 \partial_y P + (\tau_* - 1)\left(c\zeta_*^2 \partial_\zeta P + (c-1)\partial_y P\right) = 0, \tag{E.3b}$$

$$\begin{aligned}(c\zeta_* y_* + c - 1)\left[2\zeta_*(\partial_\zeta P)(\partial_y P)^2 + \zeta_*^2\left(\partial_{\zeta\zeta} P(\partial_y P)^2 - 2\partial_{\zeta y} P\,\partial_\zeta P\,\partial_y P + \partial_{yy} P(\partial_\zeta P)^2\right)\right] \\ - 2\partial_y P\left((c-1)\partial_y P + c\zeta_*^2 \partial_\zeta P\right)\left(y_*\partial_y P - \zeta_* \partial_\zeta P\right) = 0.\end{aligned} \tag{E.3c}$$

Here all derivatives are evaluated at (ζ_*, y_*). Moreover,

$$x_* = \tau_*\zeta_* + \frac{(\tau_* - 1)\zeta_*}{c\zeta_* y_* + c - 1}.$$

Proof. Let $D(\zeta, y) = c\zeta y + c - 1$. For deformed decompression, the projection to the z-plane is $z(\zeta, y) = \tau\zeta + \frac{(\tau-1)\zeta}{D(\zeta,y)}$. A cusp is a critical point of multiplicity two of this map restricted to $\mathcal{C}$. Thus, along a local branch $y = y(\zeta)$, we impose

$$z'(\zeta_*) = 0, \qquad z''(\zeta_*) = 0, \qquad P(\zeta_*, y_*) = 0.$$

Implicit differentiation of $P(\zeta, y(\zeta)) = 0$ gives

$$y' = -\frac{\partial_\zeta P}{\partial_y P}, \qquad \text{and} \qquad y'' = -\frac{\partial_{\zeta\zeta} P(\partial_y P)^2 - 2\partial_{\zeta y} P\,\partial_\zeta P\,\partial_y P + \partial_{yy} P(\partial_\zeta P)^2}{(\partial_y P)^3}.$$

Now $z' = \tau + (\tau - 1)\frac{D - \zeta D'}{D^2}$. Since $D - \zeta D' = c - 1 - c\zeta^2 y'$, the condition $z' = 0$, after substituting the formula for y', becomes

$$\tau D^2 \partial_y P + (\tau - 1)\left(c\zeta^2 \partial_\zeta P + (c-1)\partial_y P\right) = 0.$$

Next set $N = c - 1 - c\zeta^2 y'$. Then $z' = \tau + (\tau - 1)\frac{N}{D^2}$, so $z'' = (\tau - 1)\frac{N'D - 2ND'}{D^3}$. Thus $z'' = 0$ is equivalent to $N'D - 2ND' = 0$. Substituting

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

and clearing denominators yields the third equation in (E.3). Finally, evaluating the projection gives

$$x_* = \tau_*\zeta_* + \frac{(\tau_* - 1)\zeta_*}{c\zeta_* y_* + c - 1}.$$

□

Equation (E.3) allows us to solve for the triple (ζ_*, y_*, τ_*), recovering both the merging (or splitting) ratio τ_* and the location x_*. This is more accurate than using a discrete time grid to monitor edge crossing. After the identification of the cusp time τ_*, the decompressed distribution can be treated as having one fewer (or more) bulk regions.

E.3 Dynamics of Atoms

There is also a convenient characterization of atomic components of a distribution, as simple poles of the Stieltjes transform. Around such a pole at $z = x_\circ$, the weights can be computed in terms of the residue $w = -\operatorname{Res}_{z=x_\circ} m(z) = \lim_{z\to x_\circ}(x_\circ - z)m(z)$.

Writing the polynomial defining the spectral curve $P(z, m) = 0$ as $P(z, m) \coloneqq \sum_{j=0}^{s} a_j(z)m^j$, a pole at a finite real point $x_\circ$ necessarily satisfies $x_\circ$ is a root of the leading coefficient, i.e., $a_s(x_\circ) = 0$. By expanding the Stieltjes transform around such a point, we are able to track changes in both the location and weights under deformed decompression in the following way.

Proposition E.3 (Dynamics of Atoms under Deformed Decompression)**.** *Let $x_\circ \in \mathbb{R}$ be an atom of the spectral measure ν with initial mass $w_\circ \in (0, 1)$. Under deformed decompression with parameter τ, the following hold. If $x_\circ \neq 0$, then the atom location is pushed forward to*

$$x_\circ(\tau) = \tau x_\circ,$$

and mass evolves as

$$w(\tau) = w_\circ - \frac{\tau - 1}{c\tau}.$$

If $x_\circ = 0$, then $x_\circ(\tau) = 0$, and its mass evolves as $w(\tau) = \max\{w_\circ, 1 - \frac{1}{c\tau}\}$. For $x_\circ \neq 0$, the formula describes the atom while $w(\tau) > 0$. When the predicted mass reaches zero, the pole disappears from the physical branch.

Remark E.1. For an atom at zero, the weight remains at $w_\circ$ until the rank-deficiency threshold $1 - (c\tau)^{-1}$ exceeds $w_\circ$, after which it follows this threshold. For $x_\circ$ away from zero, the atom location is pushed away from the origin, while its weight decreases until extinction, if it occurs. $\triangle$

Proof. We use the deformed-decompression change of variables $u \coloneqq \underline{m}_1(\zeta) = \underline{m}_\tau(z)$, and $y = c^{-1}\big(u - \frac{c-1}{\zeta}\big)$, so that $u = cy + \frac{c-1}{\zeta}$, and $z = \tau\zeta + (\tau - 1)u^{-1}$. First suppose $x_\circ \neq 0$. Near the initial atom, $y(\zeta) = -\frac{w_\circ}{\zeta - x_\circ} + \mathcal{O}(1)$. Since $x_\circ \neq 0$, the term $(c - 1)/\zeta$ is analytic at $x_\circ$. Hence

$$u(\zeta) = cy(\zeta) + \frac{c-1}{\zeta} = -\frac{cw_\circ}{\zeta - x_\circ} + \mathcal{O}(1),$$

and also

$$u(\zeta)^{-1} = -\frac{\zeta - x_\circ}{cw_\circ} + \mathcal{O}\Big((\zeta - x_\circ)^2\Big).$$

Substituting into the pushforward map gives

$$z = \tau\zeta + (\tau - 1)u^{-1} = \tau x_\circ + \left(\tau - \frac{\tau - 1}{cw_\circ}\right)(\zeta - x_\circ) + \mathcal{O}\Big((\zeta - x_\circ)^2\Big).$$

Thus the atom is pushed forward to $x_\circ(\tau) = \tau x_\circ$. Moreover, if $\alpha \coloneqq \tau - \frac{\tau-1}{cw_\circ}$, then

$$z - \tau x_\circ = \alpha(\zeta - x_\circ) + \mathcal{O}\Big((\zeta - x_\circ)^2\Big).$$

Consequently, $u(z) = -\frac{cw_\circ\alpha}{z-\tau x_\circ} + O(1)$. On the decompressed side,

$$u(z) = \underline{m}_\tau(z) = \tau c\, m_\tau(z) + \frac{\tau c - 1}{z}.$$

Since $\tau x_\circ \neq 0$, the term $(\tau c - 1)/z$ is analytic at $z = \tau x_\circ$. Hence, if the pushed-forward atom has mass $w(\tau)$,

$$u(z) = -\frac{\tau c\, w(\tau)}{z - \tau x_\circ} + O(1).$$

Comparing residues gives $\tau c\, w(\tau) = cw_\circ\alpha$, and so

$$w(\tau) = \frac{w_\circ}{\tau}\left(\tau - \frac{\tau - 1}{cw_\circ}\right) = w_\circ - \frac{\tau - 1}{c\tau}.$$

Now suppose $x_\circ = 0$. Then $y(\zeta) = -\frac{w_\circ}{\zeta} + O(1)$, so the initial companion transform satisfies

$$u(\zeta) = cy(\zeta) + \frac{c-1}{\zeta} = -\frac{\kappa_\circ}{\zeta} + O(1), \qquad \kappa_\circ \coloneqq 1 - c + cw_\circ.$$

Since $\underline{R}_\tau = \tau \underline{R}_1$, the companion law evolves by a free additive convolution power. Its atom at zero therefore has weight

$$\kappa(\tau) = \max\{\tau\kappa_\circ - (\tau - 1), 0\}.$$

On the decompressed side, $\underline{m}_\tau(z) = \tau c\, m_\tau(z) + \frac{\tau c - 1}{z}$, so comparing residues at zero gives $\kappa(\tau) = 1 - \tau c + \tau c\, w(\tau)$. Consequently,

$$w(\tau) = \frac{\kappa(\tau) + \tau c - 1}{\tau c} = \max\left\{w_\circ, 1 - \frac{1}{c\tau}\right\}.$$

The atom remains located at the origin. □

E.4 Evolution of Moments

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

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

Much like the free decompression case in Ameli et al. [3], these moments can be evolved without reconstructing ν_τ. The deformed decompression PDE in Proposition C.6 induces a closed triangular system of ODEs for $\mu_n(\tau)$, allowing each moment to be computed from lower-order moments.

Lemma E.4 (ODEs for Moments)**.** *The evolution of the moment $\mu_n(\tau)$ under deformed decompression is governed by the system of triangularly coupled ODEs*

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}\tau} = c\sum_{j=1}^{n-1} \mu_{n-j}(\tau)\left(j\mu_j(\tau) - \tau\frac{\mathrm{d}\mu_j}{\mathrm{d}\tau}\right), \qquad n \geq 1, \tag{E.4}$$

and, in particular, $\mu_0(\tau) = 1$ and $\mu_1(\tau) = \mu_1(1)$ for all $\tau \geq 1$.

Proof. We begin with the Stieltjes transform PDE for deformed decompression (C.5), so

$$(cz\tau m_\tau + c\tau - 1)\frac{\partial m_\tau}{\partial \tau} + c(m_\tau z + 1)\left(z\frac{\partial m_\tau}{\partial z} + m_\tau\right) = 0.$$

For large $|z|$, $m_\tau(z)$ has an expansion as a convergent Laurent series $m_\tau(z) = -\sum_{n=0}^{\infty} \mu_n(\tau) z^{-n-1}$. Since $\mu_0(\tau) = 1$,

$$1 + zm_\tau = -\sum_{n=1}^{\infty} \frac{\mu_n(\tau)}{z^n}, \quad \text{and} \quad z\frac{\partial m_\tau}{\partial z} + m_\tau = \sum_{n=0}^{\infty} \frac{n\mu_n(\tau)}{z^{n+1}}.$$

Therefore, by the Cauchy product, the second term of the PDE (C.5) becomes

$$\begin{aligned} c(m_\tau z + 1)\left(z\frac{\partial m_\tau}{\partial z} + m_\tau\right) &= -c\left(\sum_{k=1}^{\infty} \frac{\mu_k(\tau)}{z^k}\right)\left(\sum_{j=0}^{\infty} \frac{j\mu_j(\tau)}{z^{j+1}}\right) \\ &= -c\sum_{n=1}^{\infty}\left(\sum_{j=0}^{n-1} j\mu_{n-j}(\tau)\mu_j(\tau)\right)\frac{1}{z^{n+1}}. \end{aligned} \tag{E.5}$$

For the first term of the PDE, there is $cz\tau m_\tau + c\tau - 1 = -c\tau\sum_{n=1}^{\infty}\mu_n(\tau)z^{-n} - 1$ and $\frac{\partial m_\tau}{\partial \tau} = -\sum_{n=0}^{\infty}\mu_n'(\tau)z^{-n-1}$. Their Cauchy product yields

$$(cz\tau m_\tau + c\tau - 1)\frac{\partial m_\tau}{\partial \tau} = \sum_{n=0}^{\infty}\frac{\mu_n'(\tau)}{z^{n+1}} + c\tau\sum_{n=1}^{\infty}\left(\sum_{j=0}^{n-1}\mu_{n-j}(\tau)\mu_j'(\tau)\right)\frac{1}{z^{n+1}}. \tag{E.6}$$

Combining (E.6) and (E.5) with (C.5) and equating the coefficient of z^{-n-1} with zero gives

$$\mu_n'(\tau) + c\tau \sum_{j=0}^{n-1} \mu_{n-j}(\tau)\mu_j'(\tau) - c\sum_{j=0}^{n-1} j\mu_{n-j}(\tau)\mu_j(\tau) = 0. \tag{E.7}$$

For $n = 0$, this gives $\mu_0'(\tau) = 0$, as expected. Consequently, this implies that the lower limit in the sums of (E.7) can start from $j = 1$. For $n = 1$, this becomes an empty sum, and so $\mu_1'(\tau) = 0$, implying $\mu_1(\tau)$ is constant. □

The triangular structure of Lemma E.4 implies that the moments can be solved for recursively. Specifically, $\mu_n(\tau)$ takes the form of a polynomial in τ:

$$\mu_n(\tau) = \sum_{k=0}^{n-1} \kappa_{n,k}\tau^k,$$

where the coefficients $\kappa_{n,k}$ are determined recursively from the initial moments $\mu_k(\tau = 1)$. This recurrence is established in the following proposition.

Proposition E.5 (Moment Recurrence)**.** *For any $\tau \geq 1$ and $n \geq 1$, the n-th moment $\mu_n(\tau)$ is a polynomial in τ of degree at most $n-1$ given by $\mu_n(\tau) = \sum_{k=0}^{n-1} \kappa_{n,k}\tau^k$ where $\kappa_{1,0} = \mu_1(1)$ and for $n \geq 2$,*

$$\kappa_{n,k} = \begin{cases} \dfrac{c}{k}\displaystyle\sum_{i=1}^{n-1}\sum_{j=0}^{k-1}(i-j)\kappa_{i,j}\kappa_{n-i,k-1-j} & \text{if } 1 \leq k \leq n-1, \\ \mu_n(1) - \displaystyle\sum_{j=1}^{n-1}\kappa_{n,j} & \text{if } k = 0, \end{cases}$$

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

Proof. First, we substitute the polynomial ansatz $\mu_n(\tau) = \sum_{k=0}^{n-1} \kappa_{n,k}\tau^k$ into the system of ODEs obtained in Lemma E.4. The left-hand side gives

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}\tau} = \sum_{k=1}^{n-1} k\kappa_{n,k}\tau^{k-1} = \sum_{p=0}^{n-2}(p+1)\kappa_{n,p+1}\tau^p. \tag{E.8}$$

For the right-hand side, starting with the bracketed term,

$$i\mu_i(\tau) - \tau\frac{\mathrm{d}\mu_i}{\mathrm{d}\tau} = \sum_{j=0}^{i-1} i\kappa_{i,j}\tau^j - \sum_{j=1}^{i-1} j\kappa_{i,j}\tau^j = \sum_{j=0}^{i-1}(i-j)\kappa_{i,j}\tau^j.$$

Multiplying this by $\mu_{n-i}(\tau) = \sum_{m=0}^{n-i-1} \kappa_{n-i,m}\tau^m$ gives

$$\begin{aligned}\mu_{n-i}(\tau)\left(i\mu_i(\tau) - \tau\frac{\mathrm{d}\mu_i}{\mathrm{d}\tau}\right) &= \left(\sum_{m=0}^{n-i-1}\kappa_{n-i,m}\tau^m\right)\left(\sum_{j=0}^{i-1}(i-j)\kappa_{i,j}\tau^j\right) \\ &= \sum_{p=0}^{n-2}\left(\sum_{j=0}^{p}(i-j)\kappa_{i,j}\kappa_{n-i,p-j}\right)\tau^p, \end{aligned} \tag{E.9}$$

where we define $p = m + j$ and adopt the convention that $\kappa_{a,b} = 0$ when $b \geq a$. Substituting (E.8) and (E.9) into the ODE (E.4) gives

$$\sum_{p=0}^{n-2}(p+1)\kappa_{n,p+1}\tau^p = c\sum_{i=1}^{n-1}\sum_{p=0}^{n-2}\left(\sum_{j=0}^{p}(i-j)\kappa_{i,j}\kappa_{n-i,p-j}\right)\tau^p.$$

Equating the coefficients of τ^p for $0 \leq p \leq n-2$ on both sides gives

$$(p+1)\kappa_{n,p+1} = c\sum_{i=1}^{n-1}\sum_{j=0}^{p}(i-j)\kappa_{i,j}\kappa_{n-i,p-j}.$$

Setting $k = p + 1$ yields the recursive formula for $\kappa_{n,k}$ for $1 \leq k \leq n-1$. Finally, evaluating the moment representation at the initial state $\tau = 1$ requires $\mu_n(1) = \sum_{k=0}^{n-1}\kappa_{n,k}$, which rearranges to provide the expression for $\kappa_{n,0}$. □

Appendix F Generalized Covariance Models

In this section, we consider the generalized covariance matrix model and study the difference between its limiting Stieltjes transform and that of a corresponding covariance matrix model which can be treated using deformed decompression. Our exposition and tools closely follow Couillet & Hachem [13]; for further details on the generalized covariance model, see Couillet & Liao [14]. Since covariance matrices are a subset of generalized covariance matrices, for the sake of brevity, we shall temporarily adopt a more general notation, and later connect our findings to the central model presented in Section 3.

A *generalized covariance matrix (GCM)* is of the general form

$$\mathbf{S}_n \coloneqq \frac{1}{n}\mathbf{A}_n^{1/2}\mathbf{Z}_n\mathbf{B}_n\mathbf{Z}_n^\intercal\mathbf{A}_n^{1/2},$$

where $\mathbf{Z}_n \in \mathbb{R}^{p\times n}$ is comprised of independent and identically distributed elements with zero mean and unit variance, and $\mathbf{A}_n \in \mathbb{R}^{p\times p}$, $\mathbf{B}_n \in \mathbb{R}^{n\times n}$ are deterministic positive-semidefinite matrices. Consider the limit $n \to \infty$ with $p \equiv p(n)$ where $p(n)/n \to c > 0$, where c is the aspect ratio. In this regime, we assume the following.

Assumption F.1. The empirical spectral measures of $\mathbf{A}_n$ and $\mathbf{B}_n$ converge weakly to the probability measures ν and η, respectively, with support on $[0, \infty)$.

Under Assumption F.1, from Couillet & Hachem [13, Proposition 1.1], the empirical spectral distribution of $\mathbf{S}_n$ converges weakly to a probability measure; we shall denote the Stieltjes transform of this measure by m_ν^η. We will generally keep ν fixed, and consider the effect of changing η. For a fixed probability measure η, we shall compare m_ν^Λ between the following choices for Λ:

- An arbitrary probability measure η with support on $[0, \infty)$. This is the complete generalized covariance model.
- A Dirac measure $\delta_{\bar\eta}$ assigning total probability mass to the point $\bar\eta = \int x\eta(\mathrm{d}x)$. This corresponds to the covariance model approximation, where $\mathbf{B}_n$ is replaced with $\frac{1}{n}\mathsf{trace}(\mathbf{B}_n)\mathbf{I}$.

For each $z \in \mathbb{C}^+$, define $\delta_1(z), \tilde\delta_1(z) \in \mathbb{C}^+$ as the unique solution to the master equations

$$\delta_1(z) = c\int \frac{t}{-z(1+\tilde\delta_1(z)t)}\nu(\mathrm{d}t), \qquad \tilde\delta_1(z) = \int \frac{t}{-z(1+\delta_1(z)t)}\eta(\mathrm{d}t).$$

By Couillet & Hachem [13, eqn. (5)], the Stieltjes transform m_ν^η is given by

$$m_\nu^\eta(z) = -z^{-1} - c^{-1}\delta_1(z)\tilde\delta_1(z).$$

Next, we define the same objects for the Dirac case: for each $z \in \mathbb{C}^+$, let $\delta_2(z), \tilde\delta_2(z) \in \mathbb{C}^+$ satisfy

$$\delta_2(z) = c\int \frac{t}{-z(1+\tilde\delta_2(z)t)}\nu(dt), \qquad \tilde\delta_2(z) = \frac{\bar\eta}{-z(1+\delta_2(z)\bar\eta)},$$

with corresponding Stieltjes transform $m_\nu^{\delta_{\bar\eta}}$ given by

$$m_\nu^{\delta_{\bar\eta}}(z) = -z^{-1} - c^{-1}\delta_2(z)\tilde\delta_2(z).$$

Our main theorem for this section bounds the discrepancy between m_ν^η and $m_\nu^{\delta_{\bar\eta}}$.

Theorem F.1. *Let η have finite second moment, and let $\Omega \subset \mathbb{C}^+$ be a bounded region on the complex plane with positive distance from the real axis. There exists $K_\Omega > 0$ such that for σ_η^2 the variance of η,*

$$\sup_{z\in\Omega}|m_\nu^\eta(z) - m_\nu^{\delta_{\bar\eta}}(z)| \le K_\Omega \cdot \sigma_\eta^2.$$

Proof. Let $z \in \mathbb{C}^+$ be fixed and define the function

$$f_z(t) = \frac{t}{-z(1+\delta_2(z)t)}, \qquad t \in [0,\infty),$$

noting that $\tilde{\delta}_2(z) = f_z(\bar{\eta})$. The primary source of the error between m_ν^η and $m_\nu^{\delta_{\bar{\eta}}}$ is captured in the term $B(z)$, defined as the difference

$$B(z) = \int f_z(t)\eta(\mathrm{d}t) - f_z(\bar{\eta}).$$

To bound $B(z)$, a second-order Taylor expansion of $f_z(t)$ around $\bar{\eta}$ will suffice:

$$f_z(t) = f_z(\bar{\eta}) + f_z'(\bar{\eta})(t-\bar{\eta}) + \frac{1}{2}f_z''(\xi_t)(t-\bar{\eta})^2, \qquad \xi_t \in (\bar{\eta}, t).$$

Integrating with respect to η, the linear term cancels, and

$$|B(z)| \le \frac{\sigma_\eta^2}{2} \sup_{t\ge 0} |f_z''(t)|. \tag{F.1}$$

The second derivative of f_z is given by $f_z''(t) = \frac{2\delta_2(z)}{z(1+\delta_2(z)t)^3}$. From the proof of Couillet & Hachem [13, Proposition 2.1], for any $t \ge 0$ and $z \in \mathbb{C}^+$, $|z(1+\delta_2(z)t)| \ge \mathrm{Im}(z) > 0$. Consequently, $|1+\delta_2(z)t|^{-3} \le \frac{|z|^3}{\mathrm{Im}(z)^3}$. Furthermore, Couillet & Hachem [13, Lemma 3.1] guarantees that there exists $K_\Omega^{(1)} > 0$ such that

$$\sup_{z\in\Omega} \max\{|\delta_1(z)|, |\delta_2(z)|, |\tilde{\delta}_1(z)|, |\tilde{\delta}_2(z)|\} \le K_\Omega^{(1)}.$$

Letting $K_\Omega^{(2)} = \sup_{z\in\Omega} \frac{|z|^2}{\mathrm{Im}(z)^3}$,

$$\sup_{t\ge 0} |f_z''(t)| \le 2K_\Omega^{(1)} K_\Omega^{(2)}. \tag{F.2}$$

Now let $\Delta\delta = \delta_1 - \delta_2$ and $\Delta\tilde{\delta} = \tilde{\delta}_1 - \tilde{\delta}_2$, and observe that

$$\begin{aligned}\Delta\tilde{\delta} &= \int \frac{t}{-z(1+\delta_1(z)t)}\eta(\mathrm{d}t) - \frac{\bar{\eta}}{-z(1+\delta_2(z)\bar{\eta})} \\ &= \int \left(\frac{t}{-z(1+\delta_1(z)t)} - \frac{t}{-z(1+\delta_2(z)t)}\right)\eta(\mathrm{d}t) + B(z).\end{aligned}$$

Combining the fractions yields

$$\begin{aligned}\frac{t}{-z(1+\delta_1(z)t)} - \frac{t}{-z(1+\delta_2(z)t)} &= \frac{t(1+\delta_2(z)t) - t(1+\delta_1(z)t)}{-z(1+\delta_1(z)t)(1+\delta_2(z)t)} \\ &= \frac{t^2}{z}\frac{\delta_1(z)-\delta_2(z)}{(1+\delta_1(z)t)(1+\delta_2(z)t)}.\end{aligned}$$

Consequently,

$$\Delta\tilde{\delta} = z\Delta\delta \int \frac{t^2}{z^2(1+\delta_1(z)t)(1+\delta_2(z)t)}\eta(\mathrm{d}t) + B(z) =: z\Gamma_\eta(z)\Delta\delta + B(z). \tag{F.3}$$

Applying the same fraction combination to $\Delta\delta$ now under the shared measure ν,

$$\begin{aligned}\Delta\delta &= c\int \left[\frac{t}{-z(1+\tilde{\delta}_1(z)t)} - \frac{t}{-z(1+\tilde{\delta}_2(z)t)}\right]\nu(\mathrm{d}t) \\ &= z\Delta\tilde{\delta}\cdot c\int \frac{t^2}{z^2(1+\tilde{\delta}_1(z)t)(1+\tilde{\delta}_2(z)t)}\nu(\mathrm{d}t) \\ &=: z\Gamma_\nu(z)\Delta\tilde{\delta}.\end{aligned}$$

Substituting $\Delta\delta$ into (F.3) reveals

$$\Delta\tilde{\delta}(1 - z^2\Gamma_\eta(z)\Gamma_\nu(z)) = B(z), \tag{F.4}$$

where

$$\Gamma_\eta(z) = \int \frac{t^2}{z^2(1+\delta_1(z)t)(1+\delta_2(z)t)}\eta(\mathrm{d}t)$$
$$\Gamma_\nu(z) = \int \frac{ct^2}{z^2(1+\tilde{\delta}_1(z)t)(1+\tilde{\delta}_2(z)t)}\nu(\mathrm{d}t).$$

Since the difference in the Stieltjes transforms $\Delta m = m_\nu^\eta - m_\nu^{\delta_{\bar{\eta}}}$ satisfies

$$c\,|\Delta m(z)| = \left|\delta_1\tilde{\delta}_1 - \delta_2\tilde{\delta}_2\right| = \left|\delta_1\Delta\tilde{\delta} + \tilde{\delta}_2\Delta\delta\right| \leq \left|\Delta\tilde{\delta}\right|\left|\delta_1 + \tilde{\delta}_2 z\Gamma_\nu(z)\right| = \frac{|B(z)|\left|\delta_1 + \tilde{\delta}_2 z\Gamma_\nu(z)\right|}{|1 - z^2\Gamma_\eta(z)\Gamma_\nu(z)|},$$

this implies from (F.1), (F.2), and (F.4), that

$$\begin{aligned}\sup_{z\in\Omega}|\Delta m(z)| &\leq \frac{1}{c}\frac{\sigma_\eta^2 K_\Omega^{(1)} K_\Omega^{(2)}}{|1 - z^2\Gamma_\eta(z)\Gamma_\nu(z)|}\left|\delta_1 + \tilde{\delta}_2 z\Gamma_\nu(z)\right| \\ &\leq \frac{1}{c}\sigma_\eta^2 (K_\Omega^{(1)})^2 K_\Omega^{(2)} \cdot \frac{1 + \sup_{z\in\Omega}|z\Gamma_\nu(z)|}{\inf_{z\in\Omega}|1 - z^2\Gamma_\eta(z)\Gamma_\nu(z)|}\end{aligned}$$

It only remains to deal with the last fraction. First, we establish that $\inf_{z\in\Omega}|1 - z^2\Gamma_\eta(z)\Gamma_\nu(z)| > 0$. To do so, let

$$\gamma_1(z_1,z_2) = \int \frac{ct^2}{z_1 z_2(1+\tilde{\delta}_1(z_1)t)(1+\tilde{\delta}_1(z_2)t)}\nu(\mathrm{d}t)$$
$$\gamma_2(z_1,z_2) = \int \frac{ct^2}{z_1 z_2(1+\tilde{\delta}_2(z_1)t)(1+\tilde{\delta}_2(z_2)t)}\nu(\mathrm{d}t)$$
$$\tilde{\gamma}_1(z_1,z_2) = \int \frac{t^2}{z_1 z_2(1+\delta_1(z_1)t)(1+\delta_1(z_2)t)}\eta(\mathrm{d}t)$$
$$\tilde{\gamma}_2(z_1,z_2) = \int \frac{t^2}{z_1 z_2(1+\delta_2(z_1)t)(1+\delta_2(z_2)t)}\eta(\mathrm{d}t)$$

By the Cauchy-Schwarz inequality,

$$|\Gamma_\eta(z)| \leq \tilde{\gamma}_1(z,\bar{z})^{1/2}\tilde{\gamma}_2(z,\bar{z})^{1/2}$$
$$|\Gamma_\nu(z)| \leq \gamma_1(z,\bar{z})^{1/2}\gamma_2(z,\bar{z})^{1/2}.$$

Together this reveals that

$$|z^2\Gamma_\eta(z)\Gamma_\nu(z)| \leq \sqrt{|z|^2\gamma_1(z,\bar{z})\tilde{\gamma}_1(z,\bar{z})}\cdot\sqrt{|z|^2\gamma_2(z,\bar{z})\tilde{\gamma}_2(z,\bar{z})}.$$

Let $h_i(z) = |z|^2\gamma_i(z,\bar{z})\tilde{\gamma}_i(z,\bar{z})$ for $i = 1, 2$. From Couillet & Hachem [13, eqn. (8)], $h_i(z) < 1$ for any $z \in \mathbb{C}^+$. Since h_i is also continuous on $\mathbb{C}^+$, by the Extreme Value Theorem, h_i must achieve its maximum on $\bar{\Omega}$, and $\max_{i=1,2}\sup_{z\in\Omega} h_i(z) \leq 1 - \epsilon$ for some $\epsilon > 0$. Consequently,

$$\inf_{z\in\Omega}|1 - z^2\Gamma_\eta(z)\Gamma_\nu(z)| \geq 1 - \sup_{z\in\Omega}\sqrt{h_1(z)h_2(z)} \geq \epsilon.$$

Finally, note that

$$|z\Gamma_\nu(z)| \leq |z|\gamma_1(z,\bar{z})^{1/2}\gamma_2(z,\bar{z})^{1/2}.$$

Since $|z| \leq M$ for $z \in \Omega$ for some $M > 0$ and Couillet & Hachem [13, Lemma 3.2] implies that $\gamma_i(z,\bar{z})$ is uniformly bounded on Ω (by M_γ, say), this completes the proof with

$$K_\Omega = (c\epsilon)^{-1}(K_\Omega^{(1)})^2 K_\Omega^{(2)}(1 + MM_\gamma).$$

□

Theorem F.1 immediately implies Proposition 5 with $\mathbf{A}_n = \mathbf{\Sigma}_p$, $\mathbf{B}_n = \mathbf{D}$, and $\tilde{\mathbf{\Sigma}}_p = \frac{1}{n}\mathsf{trace}(\mathbf{D})\mathbf{\Sigma}_p$.

For a fixed $y > 0$, Theorem F.1 also provides a bound on the Poisson-smoothed spectral densities

$$\rho^{\Lambda}_{\nu,y}(x) \coloneqq \pi^{-1}\Im m^{\Lambda}_{\nu}(x+iy) \quad \text{for} \quad \Lambda \in \{\eta, \delta_{\bar{\eta}}\}.$$

Indeed, for any compact interval $I \subset \mathbb{R}$ and any Ω as in Theorem F.1 containing $\{x+iy : x \in I\}$,

$$\sup_{x\in I}\left|\rho^{\eta}_{\nu,y}(x) - \rho^{\delta_{\bar{\eta}}}_{\nu,y}(x)\right| \le \frac{K_\Omega}{\pi}\sigma_\eta^2.$$

This follows directly by taking imaginary parts of the Stieltjes-transform bound. Since the bound applies at a fixed height $y > 0$, it does not by itself provide control of the unsmoothed density, exact spectral edges, or log-determinants.

Given the fine level of granularity available for examining GCMs, one might consider whether it is feasible to develop a generalization of deformed decompression that is directly capable of extrapolating from the limiting spectral measure ν_c of $\mathbf{S}_n$ to that of a larger aspect ratio $\nu_{\tau c}$. This turns out to be impossible, at least, using the available techniques. To see why, assume that $\mathbf{Z}_n$ is comprised of i.i.d. standard normal random variables. The S-transform $S(z)$ of the limiting spectral distribution of $\mathbf{S}_n$ should factorize into three parts:

$$S(z) = S_{\mathbf{A}}(z)S_{\mathrm{MP}_c}(z)S_{\mathbf{B}}(cz),$$

where the presence of c in the S-transform for $\mathbf{B}$ addresses the change in scale due to the different sizes of $\mathbf{A}$ and $\mathbf{B}$. In practice, we have knowledge of S and S_{MP_c}, but not of $S_{\mathbf{A}}$ or $S_{\mathbf{B}}$. Due to the presence of c in the scale factor for $S_{\mathbf{B}}$, to extrapolate from smaller to larger matrix sizes in the GCM setting requires solving for *both* $S_{\mathbf{A}}$ and $S_{\mathbf{B}}$. With only one equation and two unknowns, this is unachievable from the information available. While it may be possible to develop some extrapolation procedure using this framework, the resulting method cannot behave as a single-parameter flow.

Appendix G Random Matrix Models

This section describes the matrix models and Gram covariance matrices used in our numerical experiments. We first use a closed-form deformed Marchenko–Pastur model as a controlled benchmark, where the limiting spectral law is known algebraically and the effect of changing the aspect ratio can be tested directly (Appendix G.1). We then consider Hessian matrices arising from logistic regression (Appendix G.2) and an empirical image Gram covariance matrix formed from the STL-10 dataset (Appendix G.3).

G.1 Deformed Marchenko–Pastur Law

The deformed Marchenko–Pastur law is the canonical model for deformed decompression. It arises as the limiting eigenvalue distribution of a sample covariance matrix with a nontrivial population covariance [7, Section 9.7]. Let $\mathbf{Z} \in \mathbb{R}^{p\times n}$ have independent centered entries with variance one, let $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ be a deterministic positive semidefinite matrix, and define

$$\mathbf{X} \coloneqq \mathbf{\Sigma}^{1/2}\mathbf{Z}, \qquad \mathbf{S} \coloneqq \frac{1}{n}\mathbf{X}\mathbf{X}^{\intercal}.$$

Assume that $p/n \to c \in (0,\infty)$, and that the empirical spectral distribution of $\mathbf{\Sigma}$ converges weakly to a compactly supported probability measure $\nu_{\mathbf{\Sigma}}$ on $[0,\infty)$. Then, under standard assumptions, the empirical spectral distribution of $\mathbf{S}$ converges weakly to a deterministic probability measure ν_c, given by the free multiplicative convolution

$$\nu_c = \nu_{\mathbf{\Sigma}} \boxtimes \mathrm{MP}_c, \tag{G.1}$$

where MP_c denotes the Marchenko–Pastur law with aspect ratio $c > 0$. The mass at the origin is $\max\{\nu_{\mathbf{\Sigma}}(\{0\}), 1 - 1/c\}$.

For the explicit algebraic examples used in this paper, we take the population law $\nu_{\mathbf{\Sigma}}$ to have finitely many atoms,

$$\nu_{\mathbf{\Sigma}}(\mathrm{d}t) \coloneqq \sum_{i=1}^{r} w_i \delta_{t_i}(\mathrm{d}t), \qquad t_i > 0, \qquad w_i > 0, \qquad \sum_{i=1}^{r} w_i = 1. \tag{G.2}$$

A convenient characterization of ν_c is given in terms of the *companion* Stieltjes transform $\underline{m}$

$$\underline{m}(z) \coloneqq cm(z) + \frac{c-1}{z}. \tag{G.3}$$

This companion transform reflects the rectangular geometry of the data matrix: it is the Stieltjes transform of the limiting empirical law of $n^{-1}\mathbf{X}^\intercal\mathbf{X}$, and it satisfies the Silverstein self-consistent equation [53]

$$-\frac{1}{\underline{m}(z)} = z - c\int_{\mathbb{R}} \frac{t}{1+t\underline{m}(z)}\,\nu_{\mathbf{\Sigma}}(\mathrm{d}t). \tag{G.4}$$

For the finite-atom population law (G.2), this reduces to

$$-\frac{1}{\underline{m}(z)} = z - c\sum_{i=1}^{r} \frac{w_i t_i}{1+t_i\underline{m}(z)}. \tag{G.5}$$

Multiplying (G.5) by the common denominator $\underline{m}\prod_{i=1}^{r}(1+t_i\underline{m})$ yields the algebraic constraint

$$P(z,\underline{m}) \coloneqq (1+z\underline{m})\prod_{i=1}^{r}(1+t_i\underline{m}) - c\underline{m}\sum_{i=1}^{r} w_i t_i \prod_{j\neq i}(1+t_j\underline{m}) = 0. \tag{G.6}$$

Thus, when $\nu_{\mathbf{\Sigma}}$ has r atoms, the companion Stieltjes transform is an algebraic function of degree $r+1$. The physical branch is selected by the Herglotz property $\Im\underline{m}(z) > 0$ for $z \in \mathbb{C}^+$, equivalently $\Im m(z) > 0$, together with the normalization at infinity.

Example G.1 (Two-atom deformed Marchenko–Pastur law)**.** For the two-atom population law $\nu_{\mathbf{\Sigma}} = w_1\delta_{t_1} + w_2\delta_{t_2}$, with $w_1 + w_2 = 1$, the companion polynomial (G.6) is cubic:

$$P(z,\underline{m}) = \underline{a}_3(z)\underline{m}^3 + \underline{a}_2(z)\underline{m}^2 + \underline{a}_1(z)\underline{m} + \underline{a}_0(z) = 0, \tag{G.7}$$

with coefficients

$$\begin{aligned}
\underline{a}_3(z) &= zt_1t_2,\\
\underline{a}_2(z) &= z(t_1+t_2) + t_1t_2(1-c),\\
\underline{a}_1(z) &= z + (t_1+t_2) - c(w_1t_1 + w_2t_2),\\
\underline{a}_0(z) &= 1.
\end{aligned} \tag{G.8}$$

Equivalently, using (G.3), the ordinary Stieltjes transform satisfies

$$P(z,m) = a_3(z)m^3 + a_2(z)m^2 + a_1(z)m + a_0(z) = 0, \tag{G.9}$$

where

$$\begin{aligned}
a_3(z) &= c^2t_1t_2z^2,\\
a_2(z) &= c(t_1+t_2)z^2 - 2ct_1t_2(1-c)z,\\
a_1(z) &= z^2 + (c(w_2t_1+w_1t_2) + (c-1)(t_1+t_2))\,z + (c-1)^2t_1t_2,\\
a_0(z) &= z - (1-c)(w_2t_1+w_1t_2).
\end{aligned} \tag{G.10}$$

In particular, the two-atom deformed Marchenko–Pastur law has $s = 3$ algebraic sheets and $d_z = 2$. △

The two-atom case is also a useful example for visualizing the spectral curve associated with the algebraic Stieltjes transform. Since (G.9) is cubic, each point z away from branch points has three algebraic roots, m_1, m_2, m_3, displayed on the z-plane in the first row of Figure G.1. Only one of these roots is the physical Stieltjes branch (here, m_1), selected by the Herglotz property and the normalization at infinity. The other two roots (m_2 and m_3) are nonphysical branches, but they are not extraneous: they encode the analytic continuation of the physical branch across the real branch cuts. For the physical branch, these branch cuts coincide with the support intervals of the density on the real axis. The lower panels of Figure G.1 show the half-plane gluings that describe how the physical sheet continues through the cuts.

The same information can be viewed as a three-sheeted algebraic surface, shown schematically in Figure G.2. In this representation, the branch cuts are not endpoints of the Stieltjes transform, but

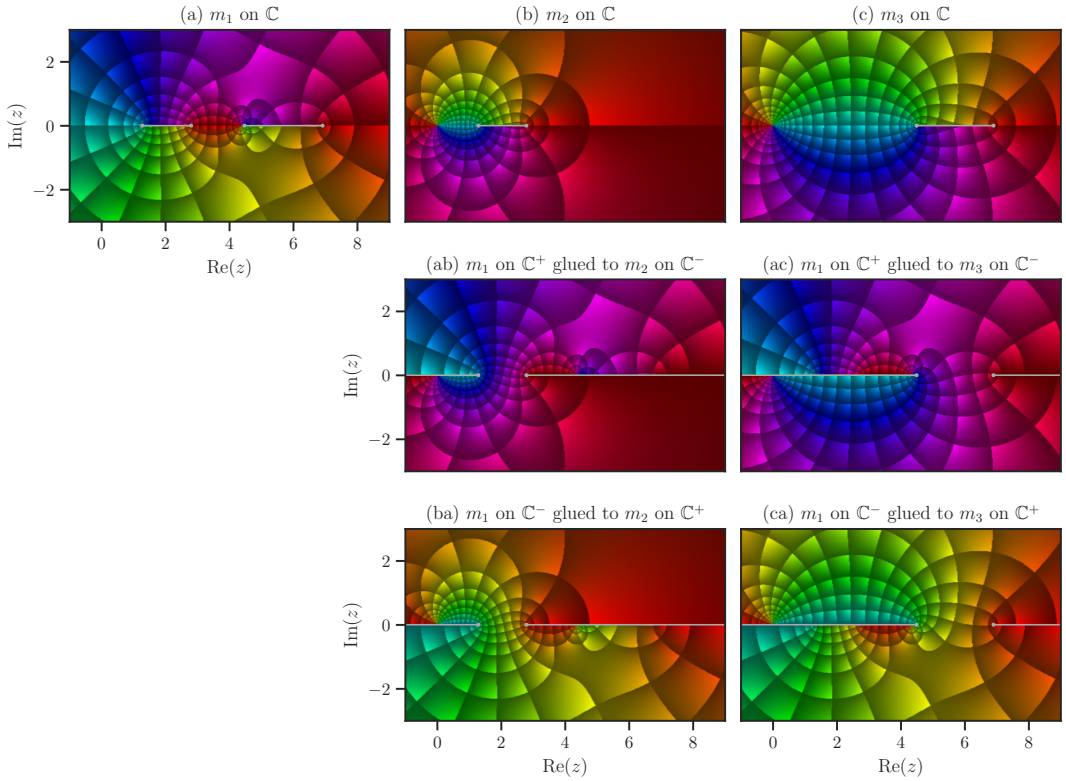


Figure G.1: Branches and analytic continuations of the cubic deformed Marchenko–Pastur Stieltjes transform in (G.9). The first row shows the three algebraic branches m_1, m_2, m_3 on the z-plane, with branch cuts marked on the real axis. The physical branch is m_1. The lower panels show the possible 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.

seams where the physical sheet is attached to nonphysical sheets. Thus, crossing a branch cut does not terminate the analytic continuation; it moves the continuation from one sheet of the spectral curve to another. In this larger geometric picture, the physical Stieltjes branch is only one visible part of a connected algebraic object, and the continuation can be followed smoothly across sheets by passing through the appropriate cuts. This viewpoint is central to the polynomial method, since the algebraic relation keeps track of all branches simultaneously, even though only one branch represents the physical Stieltjes transform on $\mathbb{C}^+$.

Relation to compound free Poisson. The deformed Marchenko–Pastur law is closely related to the compound free Poisson law used in free decompression; see [45, p. 218] and [3, Section I.1]. In the finite-atom case, the companion law associated with $\nu_c = \nu_{\mathbf{\Sigma}} \boxtimes \mathrm{MP}_c$ is precisely the compound free Poisson law with jump distribution $\nu_{\mathbf{\Sigma}}$ and rate c. This relation is most transparent at the matrix level. The nonzero eigenvalues of $n^{-1}\mathbf{X}\mathbf{X}^\intercal$ and $n^{-1}\mathbf{X}^\intercal\mathbf{X}$ agree, but the rectangular dimensions determine how many zero eigenvalues are added to one companion relative to the other. In the regime $c < 1$, the companion law has an additional atom at zero [54, Section 22.5.4]. More precisely, if $\underline{\nu}_c$ denotes the companion law associated with ν_c, then

$$\underline{\nu}_c = c\nu_c + (1-c)\delta_0.$$

Hence, for $c < 1$, the absolutely continuous part of $\underline{\nu}_c$ has the same nonzero support as ν_c, but its density is scaled by the factor c, and the remaining mass $1 - c$ is placed at the origin.

This relationship is illustrated in Figure G.3 for the same two-atom parameters used above. The deformed Marchenko–Pastur law has two absolutely continuous bulks with total mass one and no

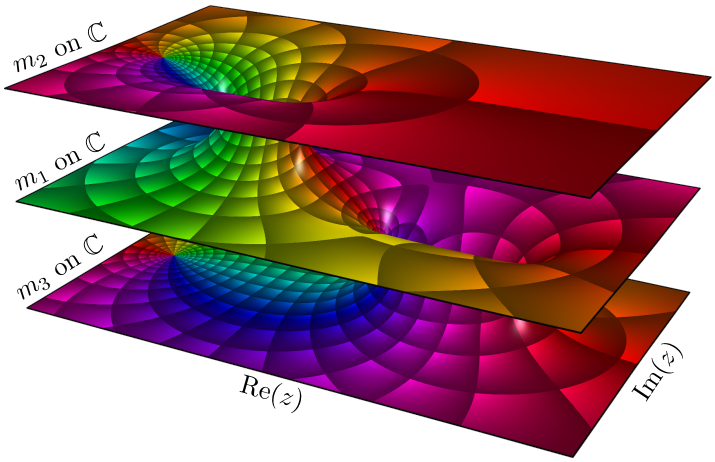


Figure G.2: Three-sheet visualization of the cubic deformed Marchenko–Pastur spectral curve from Figure G.1. The middle sheet represents the physical branch m_1, while the upper and lower sheets represent the nonphysical branches m_2 and m_3. The sheets are vertically separated for visualization and 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.

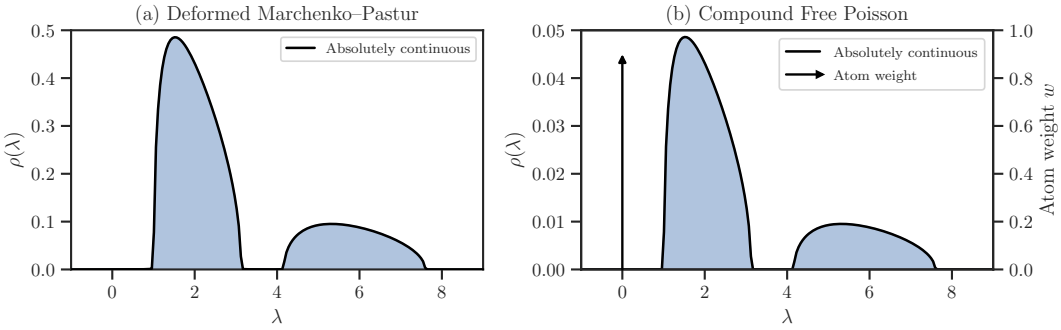


Figure G.3: Comparison between the deformed Marchenko–Pastur law and its compound-free-Poisson companion for $(t_1, t_2) = (2, 11/2)$, $(w_1, w_2) = (3/4, 1/4)$, and $c = \lambda = 1/10$. Panel (a) shows the deformed Marchenko–Pastur density ν_c, whose absolutely continuous part has total mass one and no atom at the origin. Panel (b) shows the corresponding compound free Poisson law. Its absolutely continuous part has the same nonzero support and the same profile scaled by c, while the remaining mass $1 - c = 0.9$ appears as an atom at the origin. In panel (b), the left ordinate corresponds to the continuous density, and the right ordinate corresponds to the atom weight indicated by the vertical arrow at the origin.

atom at the origin, since $c = 0.1 < 1$. Its compound-free-Poisson companion has the same two nonzero bulks after a vertical rescaling, but those bulks carry only mass $c = 0.1$; the remaining mass $1 - c = 0.9$ appears as an atom at zero. Thus, visually, the two nonzero density profiles look the same up to the scale of the vertical axis, even though the measures are different.

Despite this close algebraic and visual relationship, the two models belong to different decompression geometries. The compound free Poisson law is governed by the free-decompression R-transform invariance, while the deformed Marchenko–Pastur law is governed by the companion Silverstein structure used in deformed decompression. This distinction is not only formal: it changes the entire evolution path from a sampled matrix to the target matrix. Even when the final large-matrix laws are companion-related, the sampled submatrices used to initialize FD and DD generally have different spectra. Consequently, their initial supports and edge locations can be different, and the spectral edges follow different trajectories before reaching the final companion-related nonzero support.

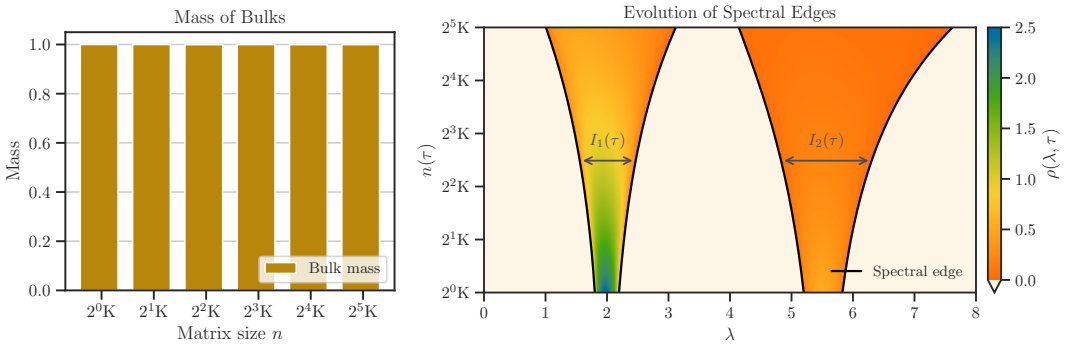


Figure G.4: Mass and edge evolution for the two-atom deformed Marchenko–Pastur example under deformed decompression from $n = 2^0\text{K}$ to $n = 2^5\text{K}$. Panel (a) shows that the full mass remains in the absolutely continuous bulks throughout the DD flow; unlike the compound-free-Poisson companion under free decompression, no mass is transferred to an atom at the origin. Panel (b) shows the evolution of the four spectral edges over the same range of matrix sizes. The color field shows the DD-evolved density $\rho(\lambda, n)$, while the light background marks spectral gaps. The edge curves are computed independently from the algebraic branch-point evolution in Proposition E.1, not extracted from the density plot. Their agreement with the visible boundary of the density support confirms the consistency between the evolved density and the DD edge dynamics.

Table G.1: Numerical validation of deformed decompression against exact deformed Marchenko–Pastur laws at increasing decompression ratios. All errors are percentages. For $\tau = 2^{10}$, the laws have an atom at zero, so the reported errors compare their normalized absolutely continuous components.

τ	W_1/L (%)	MMD (%)	Mean error (%)	Std error (%)
1	4.1×10^{-7}	7.2×10^{-7}	5.7×10^{-7}	7.3×10^{-7}
2^2	3.7×10^{-7}	7.2×10^{-7}	5.8×10^{-7}	7.1×10^{-7}
2^4	3.3×10^{-7}	7.8×10^{-7}	6.5×10^{-7}	6.7×10^{-7}
2^6	3.9×10^{-7}	1.1×10^{-6}	1.1×10^{-6}	5.9×10^{-7}
2^8	1.4×10^{-5}	4.5×10^{-5}	6.5×10^{-5}	2.4×10^{-6}
2^{10}	4.9×10^{-8}	1.3×10^{-7}	1.3×10^{-7}	9.6×10^{-8}

Figure G.4 shows this behavior for the deformed Marchenko–Pastur example. Panel (a) tracks the mass throughout the DD flow from $n = 2^0\text{K}$ to $n = 2^5\text{K}$. In contrast to the compound-free-Poisson FD example in [3, Figure 1], where mass is transferred between the absolutely continuous bulks and an atom at the origin, the deformed MP law here has no atom for $c < 1$, and the full mass remains in the absolutely continuous component throughout the evolution. Panel (b) tracks the four spectral edges of the two-bulk density over the same range of matrix sizes. The color field shows the DD-evolved density $\rho(\lambda, \tau)$, while the light background indicates spectral gaps where the density vanishes. The edge curves are not extracted from the plotted density; they are evolved separately using the algebraic branch-point evolution equation in Proposition E.1. Their agreement with the visible boundary of the density support provides a consistency check between the evolved density and the independently evolved DD edges.

Large-Ratio Numerical Validation. To assess the numerical accuracy of DD independently of finite-matrix sampling, we initialize the method from an analytically specified two-atom deformed Marchenko–Pastur model and compare the evolved distributions against their exact limiting laws for decompression ratios $\tau = 2^{2j}$, $j = 0, \ldots, 5$. Table G.1 reports the normalized Wasserstein-1 distance, maximum mean discrepancy (MMD), and errors in the first moment and standard deviation.

The discrepancies remain extremely small across all tested ratios, with the largest reported error below $7 \times 10^{-5}\%$. The nonzero errors at $\tau = 1$ already reflect numerical fitting and Stieltjes inversion, rather than error from the DD evolution. These results demonstrate the numerical consistency of the

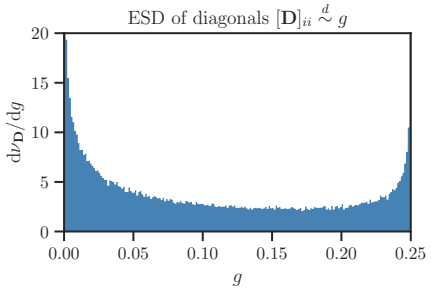


Figure G.5: Empirical distribution of the diagonal curvature weights D_{ii} for the logistic-regression Hessian model. These weights are samples of the random curvature $g = \ell''(y, h)$ in the generalized Hessian model of Liao & Mahoney [35]. The Hessian is $\mathbf{H} = N^{-1}\mathbf{X}\mathbf{D}\mathbf{X}^\intercal$, with $N = 120{,}000$ samples and feature dimension $p = 16{,}000$.

algebraic DD implementation with its exact limiting laws over large decompression ratios. They do not measure the finite-sample errors incurred when initializing DD from an empirical submatrix.

G.2 Hessian of Logistic Regression Model

We next consider Hessian matrices from trained logistic regression models, following the generalized sample-covariance setting studied by Liao & Mahoney [35]. In this model, the Hessian has the form

$$\mathbf{H} = \frac{1}{N}\mathbf{X}\mathbf{D}\mathbf{X}^\intercal,$$

where $\mathbf{X} \in \mathbb{R}^{p\times N}$ is the feature matrix, N is the number of samples, p is the feature dimension, and $\mathbf{D} = \mathrm{diag}(D_{11}, \ldots, D_{NN})$ is a diagonal matrix of curvature weights. This structure is a natural test case for deformed decompression, since the spectrum depends not only on the feature covariance but also on the empirical law of the diagonal weights in $\mathbf{D}$.

We generated two Hessian ensembles with different ambient sizes. The first has $N = 100{,}000$ samples and $p = 32{,}000$ features, with results given in Section 5. The second, reported here, has $N = 120{,}000$ samples and $p = 16{,}000$ features. For both experiments, the columns $\boldsymbol{x}_i$ of $\mathbf{X}$ were sampled with zero-mean Gaussian entries, with half of the features having variance 1 and the other half variance 4. Ground-truth weights $\boldsymbol{w}^\star \in \mathbb{R}^p$ were sampled from a standard normal distribution and used to define the separating hyperplane $\{\boldsymbol{x} \mid \boldsymbol{x}^\intercal\boldsymbol{w}^\star = 0\}$. Labels were first generated as $y_i = \mathbf{1}_{\{\boldsymbol{x}_i^\intercal\boldsymbol{w}^\star>0\}}$, and then 10% of the labels were randomly flipped.

In both experiments, the trainable weights $\boldsymbol{w}_0$ were initialized as $\mathcal{N}(\mathbf{0}, 10^{-2}\mathbf{I})$. We trained the logistic regression model with cross-entropy loss using the Adam optimizer with learning rate 0.1, zero weight decay, and 2000 iterations. After training, the predictive probabilities were computed from the fitted logits, and the diagonal matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$ was formed with entries $D_{ii} = \sigma_i(1-\sigma_i)$. The Hessian matrix was then explicitly formed as $\mathbf{H} = N^{-1}\mathbf{X}\mathbf{D}\mathbf{X}^\intercal$, giving a $p\times p$ matrix. Although the curvature weights in $\mathbf{D}$ depend on the same data used to form $\mathbf{X}$, Liao & Mahoney [35] show that, under their Gaussian model assumptions with fixed model parameters, this dependence does not alter the limiting spectral distribution of the corresponding independent-weight model. Here, the model parameters are learned from the same data, so we assess the accuracy of DD empirically.

Figure G.5 shows the empirical distribution of the diagonal curvature weights D_{ii}. These weights are the samples of the random curvature term $g = \ell''(y, h)$ in the generalized Hessian model of Liao & Mahoney [35]. The distribution is nonconstant, so the Hessian is not merely a rescaled sample covariance matrix; rather, its spectrum reflects the interaction between the feature covariance and the curvature profile induced by the trained logistic model.

For the DD reconstruction, the evolving matrix size is the Hessian dimension. Thus, in Figures G.6 and G.7, $n = p = 16{,}000$, while $n_0 = 1000$ denotes the sampled principal-submatrix size. Figure G.6 shows the deformed-decompression reconstruction for the $n = 16{,}000$ Hessian. Starting from the spectrum of a sampled $n_0 = 1000$ principal submatrix, we fit an algebraic spectral curve and evolve it to the target size $n = 16{,}000$. The initial fitted density, the intermediate DD flow, and

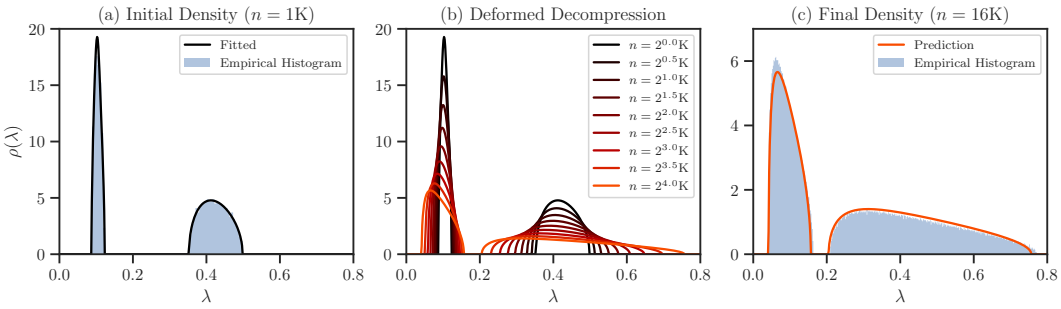


Figure G.6: Deformed-decompression reconstruction of the Hessian spectrum for the logistic-regression model with target matrix size $n = 16{,}000$. (a) ESD of the sampled $n_0 \times n_0$ principal submatrix with $n_0 = 1000$ (histogram), together with the fitted algebraic density (solid curve). (b) DD evolution from n_0 to n. (c) Final DD-predicted density at size $n = 16{,}000$, compared with the empirical spectrum obtained by direct eigendecomposition of the full Hessian.

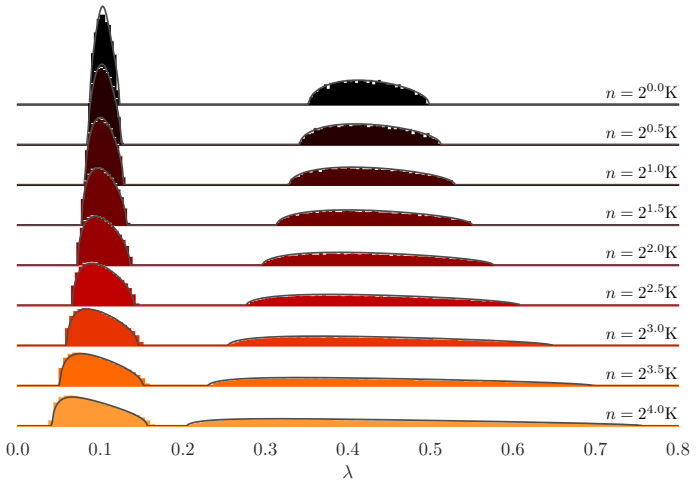


Figure G.7: Ridge-plot view of the deformed-decompression evolution in Figure G.6. Each row corresponds to an intermediate matrix size; solid curves show the DD-predicted spectral densities, while filled histograms show the empirical spectra of sampled submatrices at the same sizes. The stacked view highlights how the full density evolves from the $n_0 = 1000$ submatrix to $n = 16{,}000$

the final comparison with the directly computed empirical spectrum show that the predicted density tracks the full-size Hessian spectrum accurately.

Figure G.7 presents the same evolution as a ridge plot, making the gradual deformation of the spectral density across matrix sizes visible. This view emphasizes that DD does not simply shift the density, but evolves the full spectral shape as the effective aspect ratio changes.

Finally, Table G.2 reports runtime and accuracy metrics for the $N = 100{,}000$, $n = p = 32{,}000$ Hessian example described in Section 5. Direct eigendecomposition becomes rapidly more expensive as the matrix size grows, while DD only requires the eigendecomposition of the initial 2^0K-sized submatrix followed by a comparatively inexpensive spectral evolution. The distributional distances, moment errors, and edge errors remain small across the tested range, indicating that DD captures both the bulk density and the lower-dimensional spectral quantities of the Hessian.

G.3 STL-10 Image Dataset

We next consider an empirical Gram covariance matrix formed from natural images in the STL-10 dataset [12]. STL-10 contains 96×96 RGB images and includes a large unlabeled split, making it a convenient source of real high-dimensional observations for testing deformed decompression. We

Table G.2: Comparison of direct eigendecomposition with deformed decompression (DD) for the $n = 32{,}000$ logistic-regression Hessian matrix described in Section 5. The Direct column reports the process time for computing the eigenvalues at each matrix size. In the DD column, the first term is the time to compute eigenvalues of the initial 2^0K-sized submatrix, and the second term is the DD evolution time. Distributional distances compare the empirical spectrum with the DD reconstruction, with W_1/L denoting the Wasserstein-1 distance normalized by the spectral range L, and MMD denoting the Gaussian-kernel maximum mean discrepancy on eigenvalues. Moment errors are relative errors of the first moment and standard deviation. Edge errors compare empirically detected spectral edges with DD-evolved edges. All reported errors are shown as percentages.

Size	Process Time (sec)		Distributional Distance		Moments Rel. Error		Edges Error	
n	Direct	DD (ours)	W_1/L	MMD	$\Delta\mu_1/\mu_1$	$\Delta\sigma/\sigma$	Mean	Max
2^0K	3.6	$3.6 + 0.0$	0.00%	0.04%	0.00%	0.03%	0.1%	0.2%
2^1K	13.1	$3.6 + 1.0$	0.02%	1.01%	2.15%	0.07%	0.1%	0.3%
2^2K	75.3	$3.6 + 2.3$	0.03%	1.19%	2.86%	0.05%	0.2%	0.5%
2^3K	576.7	$3.6 + 4.7$	0.05%	1.42%	4.36%	0.45%	0.4%	1.1%
2^4K	4221.8	$3.6 + 9.2$	0.05%	1.19%	4.14%	0.37%	1.3%	2.5%
2^5K	27897.1	$3.6 + 18.6$	1.31%	5.63%	4.24%	0.47%	2.8%	5.4%

vectorize each image into a row vector of length $p = 96 \times 96 \times 3 = 27{,}648$, preprocess the rows by removing per-image brightness and contrast, and then standardize the resulting features across the sampled images.

Let $\mathbf{X} \in \mathbb{R}^{n\times p}$ denote the resulting data matrix. We form the sample Gram covariance matrix $\mathbf{A} := \frac{1}{p}\mathbf{X}\mathbf{X}^\intercal$. This example differs from the closed-form deformed Marchenko–Pastur model in Appendix G.1: the underlying population law is not known analytically and the fitted algebraic curve is obtained directly from the empirical spectrum of a smaller principal submatrix. Starting from $n_0 = 1000$, we apply deformed decompression to predict the spectral density at larger matrix sizes up to $n = 8000$.

Table 1 summarizes the corresponding runtime and distributional errors. The direct baseline computes the empirical spectrum separately at each matrix size, while DD uses only the spectrum of the initial $n_0 = 2^0$K submatrix and then evolves the fitted algebraic curve to larger aspect ratios. Because the DD reconstruction is a continuous density and the empirical eigenvalues form a discrete measure, we compare them using probability metrics rather than pointwise density errors. The normalized Wasserstein distance W_1/L measures discrepancy on the log-spectral axis without requiring kernel density estimation, while MMD provides a complementary kernel-based comparison of the same log-eigenvalue distributions [24, 46]. Across the tested sizes, DD substantially reduces the cost relative to direct eigendecomposition while maintaining small W_1/L, MMD, and first-moment errors.

The behavior of the left edge across matrix sizes n is plotted in Figure G.8. It is particularly interesting that the same Tracy-Widom-then-exponential scaling regime described in Ameli et al. [3, Appendix H] is present here also. We remark that the justification provided there, in the form of Ameli et al. [3, Proposition H.2], is not dependent on the free decompression operation, and describes a universal left-edge behavior for positive-definite matrices where $\lambda_{\min}$ is approaching zero.

Appendix H Implementation and Reproducibility Guide

All numerical procedures in this paper are provided through the `freealg.AlgebraicForm` interface in the open-source Python package `freealg`.[3] The package includes tools for fitting algebraic spectral curves from empirical spectra, evaluating the associated Stieltjes transform, evolving densities under deformed decompression, and tracking lower-dimensional spectral quantities such as

[3] `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.

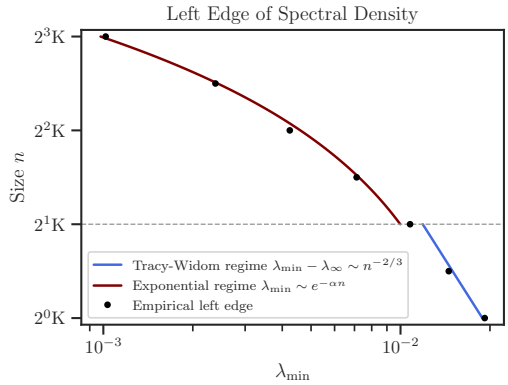


Figure G.8: Empirical (black) smallest eigenvalue $\lambda_{\min}$ of the STL-10 image dataset example, with Tracy–Widom (blue) predicted decay at rate $\asymp n^{-2/3}$ and exponential (red) predicted decay at rate $\asymp e^{-\alpha n}$ for fitted α.

edges and cusps. It also provides closed-form free-probability distributions, including deformed Marchenko–Pastur laws, which are useful for controlled validation experiments.

Listing 1 gives a minimal setup for the synthetic deformed Marchenko–Pastur experiment used in this paper. The example constructs the distribution, visualizes its spectral-curve branches, generates a large matrix realization, and extracts a smaller principal submatrix. Listing 2 then applies the `freealg.AlgebraicForm` workflow to the smaller matrix: it fits an algebraic curve, reconstructs the initial density, evolves the density to a larger matrix size using deformed decompression by setting `kind='deformed'`, and tracks the corresponding edges and cusp points. Further details on the package interface, supported input formats, and Stieltjes-transform fitting routines are given in Ameli et al. [2] and Ameli et al. [3].

The scripts and notebooks used to generate the figures and tables in this paper are included in the code repository. Additional details on function arguments, class options, and reproducibility settings are provided in the package documentation. The experiments were run on a consumer-grade workstation with an AMD Ryzen 7 5800X processor, 128GB DDR4 RAM, and Fedora Linux.

Listing 1: Synthetic deformed Marchenko-Pastur 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 deformed Marchenko--Pastur distribution (see Appendix G.1)
dist = fa.distributions.DeformedMarchenkoPastur(t=[2.0, 5.5], w=[0.75, 0.25], c=0.1)

# Visualize all branches and their gluing as in Figure G.1
x = np.linspace(-1, 9, 300)
y = np.linspace(-3, 3, 300)
dist.plot_branches(x, y)

# Generate a matrix realization of size n = 8000 for this distribution
A_large = dist.matrix(8000, seed=0)

# Sample the above matrix to a smaller submatrix of size n_0 = 1000
A_small = fa.submatrix(A_large, 1000, seed=0)
```

Listing 2: Deformed 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
ratio = dist.ratio * (A_small.shape[0] / A_large.shape[0])
af = fa.AlgebraicForm(A_small, ratio=ratio, support=None)

# Fit the polynomial P with deg_m(P) = 3 and deg_z(P) = 2
af.fit(deg_m=3, deg_z=2)

# Estimate the density of the small matrix from P
x = np.linspace(0, 10, 200)
rho_small = af.density(x, plot=True)

# Deformed decompression of density and atoms corresponding to a larger matrix of
# size n = 8000 using Algorithm C.1  (similar to Figure 1 ).
rho_large = af.decompress(size=8000, x=x, kind='deformed', plot=True)

# Evolve spectral edges similar to Figure G.4 using Algorithm E.1
# This also finds cusp points, if any occur.
sizes = np.linspace(1000, 8000, 100)
t = np.log(sizes / 1000)
edges, cusps = af.edge(t, kind='deformed')
```

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 H).

   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.