Proceedings of Machine Learning Research vol 134:1–43, 2021 34th Annual Conference on Learning Theory

Rank-one matrix estimation: analytic time evolution of gradient descent dynamics

Antoine Bodin ANTOINE.BODIN@EPFL.CH
Communication Theory Laboratory,
School of Computer and Communication Sciences,
Ecole Polytechnique Fédérale de Lausanne

Nicolas Macris NICOLAS.MACRIS@EPFL.CH
Communication Theory Laboratory,
School of Computer and Communication Sciences,
Ecole Polytechnique Fédérale de Lausanne

Editors: Mikhail Belkin and Samory Kpotufe

Abstract

We consider a rank-one symmetric matrix corrupted by additive noise. The rank-one matrix is formed by an n-component unknown vector on the sphere of radius $\sqrt{n}$, and we consider the problem of estimating this vector from the corrupted matrix in the high dimensional limit of n large, by gradient descent for a quadratic cost function on the sphere. Explicit formulas for the whole time evolution of the overlap between the estimator and unknown vector, as well as the cost, are rigorously derived. In the long time limit we recover the well known spectral phase transition, as a function of the signal-to-noise ratio. The explicit formulas also allow to point out interesting transient features of the time evolution. Our analysis technique is based on recent progress in random matrix theory and uses *local versions* of the semi-circle law.

Keywords: Gradient Descent, Rank-one Matrix Estimation, Phase Transitions, Local Semi-circle Law

1. Introduction

Gradient descent dynamic is at the root of machine learning methods, and in particular, its stochastic version augmented by various ad-hoc methods, has been very successful at finding ”good” minima of cost functions Lecun et al. (1998). However, rigorous detailed results on the full time evolution of the dynamics are scarce even for simple models and usual gradient descent. In this contribution, we show how to completely solve for the whole time evolution for a simple paradigm of non-linear estimation; the problem of estimating a rank-one spike embedded in noise.

Let $\theta^* \in \mathbb{S}^{n-1}(\sqrt{n})$ a *hidden* vector on the $n-1$ dimensional sphere of radius $\sqrt{n}$, i.e., $\theta^* = (\theta_1^*, \ldots, \theta_n^*)^T$ and $\|\theta^*\|_2^2 = n$. We consider the *data* matrix Y with elements $Y = \theta^*\theta^{*T} + \sqrt{\frac{n}{\lambda}}\xi$ where $\lambda > 0$ is the signal-to-noise parameter and $\xi = (\xi_{i,j})_{1\leq i,j\leq n}$ a symmetric random noise matrix with i.i.d $\xi_{i,j}$ for $i \leq j$. The goal is to recover θ^* given that Y and λ are known. This model is usually considered for a gaussian noise symmetric matrix $\xi_{ij} \sim \mathcal{N}(0,1)$, $i \leq j$, and is variously called the noisy rank-one matrix estimation problem or the spiked Wigner model. In this paper all the results hold under the general assumption that $\mathbb{E}\xi_{ij} = 0$, $\mathbb{E}\xi_{ij}^2 = 1 + O(\delta_{ij})$ and for all integers p we have $\mathbb{E}|\xi_{ij}|^p$ finite.[1]

1. The notation $O(\delta_{ij})$ means that the second moment of off-diagonal elements is 1 but the variance of diagonal elements can be different. For example $\xi_{ij} \sim \mathcal{N}(0,1)$, $i < j$, and $\xi_{ii} \sim \mathcal{N}(0,2)$, corresponds to Wigner's Gaussian Orthogonal Ensemble. We refer to the general case as the generalized Wigner ensemble.

We consider the cost function ($\|\cdot\|_F$ the Frobenius norm)

$$\mathcal{H}(\theta) = \frac{1}{2n^2}\left\|Y - \theta\theta^T\right\|_F^2 - \frac{1}{2n^2}\left\|Y - \theta^*\theta^{*T}\right\|_F^2 \tag{1}$$

(normalized so that, $\mathcal{H}(\theta^*) = 0$, and at the same time, the limit $n \to +\infty$ is well defined) and want to characterize the time evolution of the estimator for θ^* provided by gradient descent dynamics on the sphere. In gradient descent, an initial (deterministic) vector $\theta_0 \in \mathbb{S}^{n-1}(\sqrt{n})$ is updated through the autonomous ordinary differential equation

$$\frac{d\theta_t}{dt} = -\eta\big(\nabla_\theta \mathcal{H}(\theta_t) - \frac{\theta_t}{n}\langle\theta_t, \nabla_\theta\mathcal{H}(\theta_t)\rangle\big) \tag{2}$$

where $\eta \in \mathbb{R}_+^*$ is a learning rate. The second term on the right hand side enforces the constraint $\theta_t \in \mathbb{S}^{n-1}(\sqrt{n})$ at all times (see Appendix E). The main quantities of interest to be computed are the time evolutions of the cost $\mathcal{H}(\theta_t)$ and overlap $q(t) = n^{-1}\langle\theta^*, \theta_t\rangle$ in the high-dimensional limit $n \to +\infty$. We note that the overlap is equivalent to the mean-square-error $n^{-1}\|\theta_t - \theta^*\|_2^2 = 2\big(1 - \frac{\langle\theta^*,\theta_t\rangle}{n}\big)$.

Contribution: We compute the full time evolution of the cost and overlap in the scaling limit $\lim_{n\to+\infty} \mathcal{H}(\theta_{t=\tau n/\eta})$ and $\lim_{n\to+\infty} \frac{\langle\theta^*,\theta_{t=\tau n/\eta}\rangle}{n}$ for all $\tau > 0$. Explicit formulas are expressed solely in terms of a modified Bessel function of first order in theorems 1 and 2 (section 2). The formulas allow to explore the asymptotic behavior as $\tau \to +\infty$, as well as transient behavior by computing one and two dimensional integrals numerically (section 2). In the long time limit we recover (analytically) as expected the phase transition at $\lambda = 1$ with a limiting value of the overlap equal to $sign(\langle\theta^*, \theta_0\rangle)\sqrt{1 - 1/\lambda}\mathbb{1}(\lambda > 1)$. This is the well known BBP-like phase transition found in the spectral method Péché (2004); Féral and Péché (2006); Baik et al. (2005). The transient behavior also exhibits interesting features. For example, depending on the magnitude of the initial overlap $n^{-1}\langle\theta^*, \theta_0\rangle$ and $\lambda > 1$ for intermediate times we find that the overlap may display a maximum and then decrease to its limiting value. Such results may therefore give guidelines for applying early stopping during gradient descent to get a better estimate of the signal. We note that in the asymptotic limit of large n we require an initial overlap which is bounded away from zero uniformly in n. There are interesting situations where the signal θ^* has some structure and this is not an unnatural situation. These points are further discussed in section 2.2.

On the technical side the analysis is based on a set of integro-differential equations (derived in section 3) satisfied by matrix elements of the resolvent of the noise matrix $\langle\theta^*, (\frac{1}{\sqrt{n}}\xi - z)^{-1}\theta_t\rangle$ and $\langle\theta_t, (\frac{1}{\sqrt{n}}\xi - z)^{-1}\theta_t\rangle$, $z \in \mathbb{C} \setminus \mathbb{R}$. These quantities concentrate with respect to the probability law of the noise matrix as $n \to +\infty$ (for deterministic θ^* and θ_0). The main steps to prove concentration are explained in section 4. They combine concentration properties of the matrix elements of the resolvents with an adaptation of Gronwall type arguments to the integro-differential equations. Concentration of matrix elements of resolvents of random matrices amount to study the spectrum on a *local* scales. Such results are only a decade old in random matrix theory and go under the name of *local* semi-circle laws Erdős et al. (2008); Bloemendal et al. (2014); Benaych-Georges and Knowles (2016). They have found many applications and here we provide one more. In section 5 we present an exact analysis of the integro-differential equations and deduce the formulas for the time evolution of the overlap and cost.

Related Work: The statistical limits of the symmetric as well as non-symmetric spiked Wigner model have been elucidated in great detail in the Bayesian framework in a series of works Korada and Macris (2009); Barbier et al. (2016); Lelarge and Miolane (2018); Miolane (2017) where expressions for mutual information (in the form of low-dimensional variational problems) and minimum-mean-square-error are rigorously computed. This analysis has also been carried on for estimation of low-rank tensors corrupted by additive gaussian noise Lesieur et al. (2017); Barbier et al. (2017); Perry et al. (2020). The dynamical behaviour under Approximate Message Passing (AMP) has also been investigated in detail and, depending on the exact model and prior, large computational-to-statistical gaps are found Barbier et al. (2016); Lesieur et al. (2017). We note that these settings are different from the one of the present paper in that θ^* as well as θ_0 are random. When the prior of the spike is unbiased with zero mean (for example uniform on the sphere or binary) an initial strictly positive overlap (uniformly in n), is necessary to start the AMP algorithm, much like gradient descent, and hence the initial condition cannot be chosen at random. In this connection, the behaviour of AMP under spectral initialization has been derived in the work Montanari and Venkataramanan (2021). We note that spectral initialization is not an option for us because it yields a stationary point of gradient flow (see appendix F for a justification).

Starting with the early work of Burer and Monteiro (2005, 2003) the efficiency of gradient descent techniques has been uncovered in recent years for a host of low-rank matrix recovery modern problems, e.g., in PCA, low-rank matrix factorization, matrix completion, phase retrieval, phase synchronization, Ge et al. (2017b); Bhojanapalli et al. (2016); Ge et al. (2017a); De Sa et al. (2015); Park et al. (2017); Ling et al. (2019); Bandeira et al. (2016). We also refer to Chi et al. (2019) for a general review and references. Underpinning the efficiency of gradient descent in such non-convex problems, is a high-level result Lee et al. (2016), stating that when the landscape satisfies a *strict saddle property* (i.e., critical points are strict saddles or minima) gradient descent with sufficiently small discrete step size and random initialization will converge almost surely to a minimum Lee et al. (2016). The spiked Wigner models falls in this category at least for n finite: critical points of the cost function on the sphere $\mathcal{S}^{n-1}(\sqrt{n})$ are the eigenvectors of Y and it is easy to show that almost surely (with respect to the noise matrix ξ) the largest eigenvector is a minimum while all the other ones are strict saddles. Therefore gradient descent will converge for small enough step size to the largest eigenvector and the spectral properties of Y imply that for $\lambda > 1$ with high probability this largest eigenvector has an overlap with θ^* close to $\pm\sqrt{1-1/\lambda}$ (these known facts are briefly reviewed in Appendix F).

While these approaches are able to provide guarantees and convergence rates of gradient descent and variants thereof, they do not provide the full time-evolution and do not say much about intermediate or transient times. This is what we achieve in this paper for the admittedly simple Wigner spiked models. We believe that the techniques used here can be extended to other problems of interest in regression and learning. Recently, pure gradient descent was studied for the much harder optimization of the cost of a mixed matrix-tensor inference problem Sarao Mannelli et al. (2019); Mannelli et al. (2019) (see also Sarao Mannelli et al. (2020) for Langevin dynamics) and it was shown how the structure of saddles and minima determines the phase transition thresholds. This is based on a set of very sophisticated integro-differential CSHCK equations Crisanti et al. (1993); Cugliandolo and Kurchan (1993) with a long history in the framework of Langevin dynamics on spin-glass landscapes in statistical physics. While these derivation of the CSHCK equations for the inference problem are non-rigorous and their solution entirely numerical, they contain in principle the whole time evolution of the system (in the context of spin-glasses the formalism has been made rigorous Ben Arous et al. (2004)). The integro-differential equations and methods of the present paper are entirely *different* (and involve different objects) even when specializing to the matrix case.

We note that for the mixed matrix-tensor case the CSHCK formalism is quite intractable, but nevertheless in the pure matrix case it should be possible to retrieve our final analytical solution as (partly) done in Cugliandolo and Dean (1995) for the spherical spin-glass. We briefly comment on possible extensions of our formalism in the conclusion.

Organization of the paper: The main theorems and illustrations of analytical formulas for the whole time-evolution of the overlap and cost are presented in section 2. The heart of the method presented here is contained in sections 3 (derivation of integro-differential equations), 4 (local semi-circle laws and concentration of solutions), 5 (analytical solution of integro-differential equations). Appendices contain proofs, of intermediate results and technical material.

In the rest of the paper it is understood that the noise matrix ξ satisfies: (i) $\mathbb{E}\xi_{ij} = 0$, (ii) $\mathbb{E}\xi_{ij}^2 = 1 + O(\delta_{ij})$, (iii) $\mathbb{E}|\xi_{ij}|^p$ finite for all $p \in \mathbb{N}$. We use the notations $H = n^{-1/2}\xi$, $\mathbb{P}$ for its probability law, and $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$ for convergence in probability, i.e., $\lim_{n\to+\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$ for any $\epsilon > 0$.

2. Analytical solutions and illustrations

We solve gradient descent dynamics (2) in the scaling limit $t = \tau n/\eta$, with fixed $\tau > 0$ and $n \to +\infty$. The main quantities that we determine in the scaling limit are the overlap $q(\tau) = \frac{1}{n}\langle\theta^*, \theta_{n\tau/\eta}\rangle$ and the cost $\mathcal{H}(\theta_{n\tau/\eta})$. We remark that the overlap is directly linked to the mean-square error $n^{-1}\|\theta^* - \theta_{n\tau/\eta}\|^2 = 2(1 - q(\tau))$.

The initial condition θ_0 is fixed such that $q(0) = \alpha$ where $\alpha \in [-1, 1]$ is independent of n. It will become clear that: (i) If θ_t is a solution with initial condition $q(0) = \alpha$ then $-\theta_t$ is a solution with $q(0) = -\alpha$; (ii) For $\alpha = 0$ the solution remains trivial $q(\tau) = 0$. Therefore the reader can keep in mind that $\alpha > 0$ (all the analysis is valid for any α though).

2.1. Main results

The solution of the gradient descent dynamics can be entirely expressed thanks to a scaled moment generating function of Wigner's semi-circle law $\mu_{\rm sc}(s) = \frac{1}{2\pi}\sqrt{4 - s^2}\,\chi_{[-2,2]}(s)$,

$$M_\lambda(\tau) = \int_{-\infty}^{\infty} ds\, \mu_{\rm sc}(s) e^{s\frac{\tau}{\sqrt{\lambda}}} \tag{3}$$

Setting $s = 2\cos\theta$ we have $M_\lambda(\tau) = 2\int_0^\pi \frac{d\theta}{\pi}(\sin\theta)^2 e^{\frac{2\tau}{\sqrt{\lambda}}\cos\theta}$. Integration by parts then shows that $M_\lambda(\tau) = \frac{\sqrt{\lambda}}{\tau} I_1(\frac{2\tau}{\sqrt{\lambda}})$ where $I_1(x) = \int_0^\pi \frac{d\theta}{\pi}(\cos\theta)e^{x\cos\theta}$ is a modified Bessel function of the first kind.

Theorem 1 (Time evolution of the overlap) *Let $\theta_0 \in \mathbb{S}^{n-1}(\sqrt{n})$ an initial condition such that $q(0) = \frac{1}{n}\langle\theta^*, \theta_0\rangle = \alpha$ for a fixed $\alpha \in [-1, +1]$. The overlap converges in probability to a deterministic limit:*

$$q(\tau) \xrightarrow[n\to\infty]{\mathbb{P}} \bar{q}(\tau) = \frac{\hat{q}(\tau)}{\sqrt{\hat{p}(\tau)}} \tag{4}$$

where

$$\hat{q}(\tau) = \alpha e^{(1+\frac{1}{\lambda})\tau}\big[1 - \frac{1}{\lambda}\int_0^\tau ds\, e^{-(1+\frac{1}{\lambda})s} M_\lambda(s)\big] \tag{5}$$

and

$$\hat{p}(\tau) = M_\lambda(2\tau) + 2\alpha \int_0^\tau ds\, \hat{q}(s) M_\lambda(2\tau - s) + \int_0^\tau \int_0^\tau du dv\, \hat{q}(u)\hat{q}(v) M_\lambda(2\tau - u - v). \quad (6)$$

Theorem 2 (Time evolution of the cost) *Under the same conditions as in theorem 1 the cost converges to a deterministic limit* $\mathcal{H}(\theta_{\tau n/\eta}) \xrightarrow[n\to\infty]{\mathbb{P}} 1 - \frac{1}{2}\frac{d}{d\tau}\{\ln \hat{p}(\tau)\}$.

Using asymptotic properties of the Bessel function and the Laplace method it is possible to calculate the asymptotics of the integrals in (5) and (6) for $\tau \to +\infty$. We find for the overlap $\lim_{\tau\to\infty} \bar{q}(\tau) = \mathrm{sign}(\alpha)\sqrt{1-\lambda^{-1}}\mathbb{1}(\lambda \geq 1)$. The overlap displays the well known phase transition at $\lambda = 1$ also predicted by the spectral method. The asymptotic values can also be derived independently from theorem 1 by directly looking at the stationary equation $\nabla_\theta \mathcal{H}(\theta_\infty) - \frac{\theta_\infty}{n}\langle\theta_\infty, \nabla_\theta \mathcal{H}(\theta_\infty) = 0$. This is discussed in Appendix G for completeness. It is also possible to go one step further in the asymptotics to argue that at the transition $\lambda = 1$ the power law behavior holds $\bar{q}(\tau) \sim (\frac{2}{\pi\tau})^{1/4}$ (see Appendix I).

Besides the transition at $\lambda = 1$, for finite λ, a detailed analysis of the equations of theorem 1 which are described also in Appendix I allows to derive the first order asymptotic behavior of $\bar{q}$ for large τ. These tedious calculations are carried out analytically in detail and checked numerically. Specifically, in the regime $1 < \lambda < +\infty$ we find

$$\bar{q}(\tau) - \mathrm{sign}(\alpha)\sqrt{1-\frac{1}{\lambda}} \sim \frac{\mathrm{sign}(\alpha)}{2\sqrt{\pi}\lambda^{\frac{1}{4}}\sqrt{1-\frac{1}{\lambda}}\left(1-\frac{1}{\sqrt{\lambda}}\right)^2}\tau^{-\frac{3}{2}}e^{-(1-\frac{1}{\sqrt{\lambda}})^2\tau} \quad (7)$$

As for $0 < \lambda < 1$, we retrieve a power law behavior:

$$\bar{q}(\tau) \sim \frac{\alpha\left(\frac{2}{\pi}\right)^{\frac{1}{4}}}{\lambda^{\frac{5}{8}}\left(1-\frac{1}{\sqrt{\lambda}}\right)^2\sqrt{1-\alpha^2+\frac{\alpha^2}{\lambda(\frac{1}{\sqrt{\lambda}}-1)^2}}}\tau^{-\frac{3}{4}} \quad (8)$$

The noise-less regime $\lambda = +\infty$ is an elementary case for which the overlap can be obtained very simply. Taking the inner product of (2) with θ^* we find the differential equation (for $t = \tau n/\eta$) $\frac{dq(\tau)}{d\tau} = q(\tau) - q(\tau)^3$, $q(0) = \alpha$, which has the solution $q(\tau) = \alpha(\alpha^2 + (1-\alpha^2)e^{-2\tau})^{-1/2}$. As we will see, in the noisy case there is no closed form first order ODE for $q(\tau)$ and we must solve integro-differential equations for suitable generating functions (or an infinite hierarchy of coupled differential equations for generalized overlaps). As a sanity check, we can verify that theorem 1 leads to the same expression when $\lambda \to +\infty$. Explicitly, we find $\lim_{\lambda\to+\infty}\hat{q}(\tau) = \alpha e^\tau$ and $\lim_{\lambda\to+\infty}\hat{p}(\tau) = 1 - \alpha^2 + \alpha^2 e^{2\tau}$ which implies the noiseless expression for the overlap.

2.2. Discussion and numerical experiments

Theorems 1 and 2 provide theoretical predictions for the full time evolution of the overlap and risk in the high dimensional limit $n \to +\infty$. In this section (and Appendix J) we briefly illustrate and discuss this time evolution. Moreover in Appendix J we also compare the theoretical predictions with simulations of discrete step size gradient descent for runs over multiple samples of ξ.

2.2.1. CHOICE OF THE INITIAL CONDITION

Given $\theta^* \in \mathbb{S}^{n-1}(\sqrt{n})$ if we choose the initial condition θ_0 uniformly at random we expect α a random variable of zero mean and standard deviation $n^{-1/2}$. To analyze this case one should deal with finite n corrections to the dynamics which is beyond the scope of this paper. Numerical plots of our formulas (fig 1a, 5a) show when $\alpha \to 0$ gradient flow kicks-off at larger and larger times; this suggests that if $\alpha \sim n^{-1/2}$ gradient flow kicks-off once a large enough time-scale has elapsed. Our analysis is presumably valid beyond this time-scale (we do not have a proof of this claim), but estimating this time-scale is open. As mentioned in the introduction AMP suffers from similar issues. However, there are many interesting situations where the signal has some *structure* which is partially known, and it is then very natural to have $\alpha > 0$ (uniformly in n). For example signals which may have a non-zero empirical expectation $\rho > 0$, for instance with components distributed as $\mathrm{Ber}(\frac{\rho+1}{2})$ in $\{-1, 1\}$. Then we can take the initial all-one vector $\theta_0 = \mathbb{1}_n$, and thus $\alpha = \rho > 0$ which naturally kicks-off the gradient flow, and our analysis applies.

2.2.2. TIME EVOLUTION OF THE OVERLAP

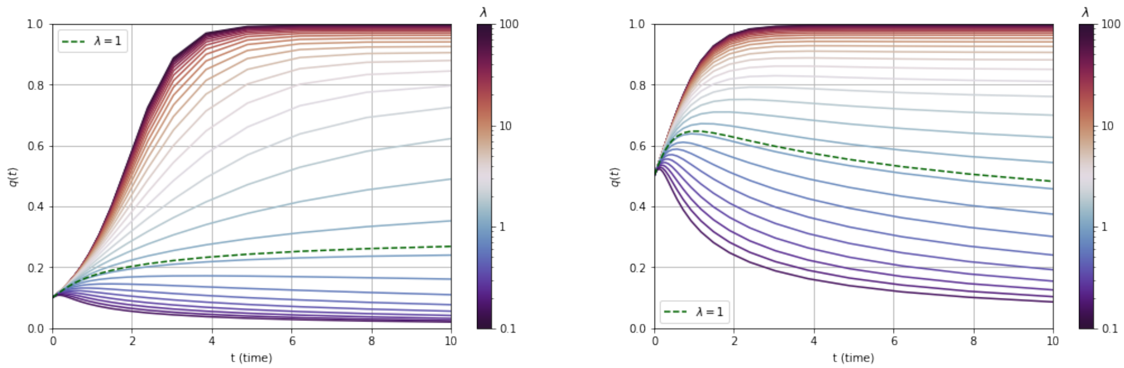
Figure 1 shows the theoretical overlap at all times $\tau \in \mathbb{R}^+$ for two initial conditions $\alpha = 0.1$ and $\alpha = 0.5$ and any signal-to-noise ratio λ. Let us say a few words about the *transient behaviors* that are observed.

On the one hand, the closer α gets to 0, the longer it takes for the gradient descent to "kick-in": the overlap stays longer close to 0 before reaching its asymptotic behavior. An additional example for $\alpha = 0.01$ illustrates this fact in Appendix J. On the other hand, we clearly see that when the initial overlap α is not too close to 0, the time evolution is *not* monotone even for $\lambda > 1$, and a specific bump is reached at early times where the overlap reaches a maximum before dropping down to its limit. In fact this is clearly suggested by (7) for $\alpha < 1$. This can be seen in particular in the case $\alpha = 0.5$ in figure 1 (b). This suggests that in practice, in such situations, it is worth using early-stopping techniques to optimize the estimation of the signal. The increase of the overlap above the spectral estimate for finite times is a consequence of the side information $\alpha > 0$ that standard PCA does not have. As a side note we mention that in the Bayesian setting with known prior the information theoretic overlap is at least as good or better than PCA (for a $\mathcal{N}(0, 1)$ prior they are equal).

In the case $\lambda = 1$ one can show that $\hat{q}(\tau) = \alpha\,(I_0(2\tau) + I_1(2\tau))$ (with modified Bessel functions of the first kind) and it is numerically much easier to evaluate the asymptotic behavior of $q(\tau)$. The calculation yields $q(\tau) \sim \left(\frac{2}{\pi\tau}\right)^{\frac{1}{4}}$ (see Appendix I). Furthermore plotting a family of curves with $\lambda = 1$ and $\alpha \in (0, 1)$ in figure 2, it appears that this asymptote also seems to act as an *upper-bound*.

2.2.3. TIME EVOLUTION OF THE COST

We also have predictions for the evolution of cost at any time for any values of (α, λ). This is illustated in figure 3. As seen in the analysis of section A, Equ. (40) the cost has two additive contributions basically interpreted as $q(\tau)^2$ and $p_1(\tau) = n^{-1}\langle\theta_\tau, H\theta_\tau\rangle$. The second contribution equals $n^{-1}\mathrm{Tr}H\theta_\tau\theta_\tau^T$ can be interpreted as a similarity measure of the reconstructed matrix $\theta_\tau\theta\tau^T$ and the noise matrix H, and is thus a "proxy" for assessing over-fitting in this particular setting. Interestingly, in the depicted example where $\lambda = 2, \alpha = 0.1$, $p_1(\tau)$ is shown to decrease the risk at early stages at a fast rate, until it slightly "heals" for $\tau \geq 3$. Conversely, when $\alpha = 0.5$, we see



(a) $\alpha = 0.1$ (b) $\alpha = 0.5$

Figure 1: Overlap as a function of time according to theorem 1 for two initial conditions and different signal-to-noise ratios. Thick dotted line corresponds to $\lambda = 1$ and tends to zero slowly as $(2/\pi\tau)^{1/4}$. For $\lambda < 1$ the curves tend to zero and for $\lambda > 1$ they tend to $\sqrt{1 - 1/\lambda}$.

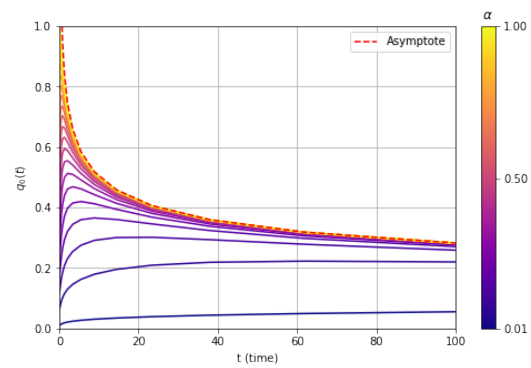


Figure 2: Overlap comparison for $\lambda = 1$ with a range of values for α

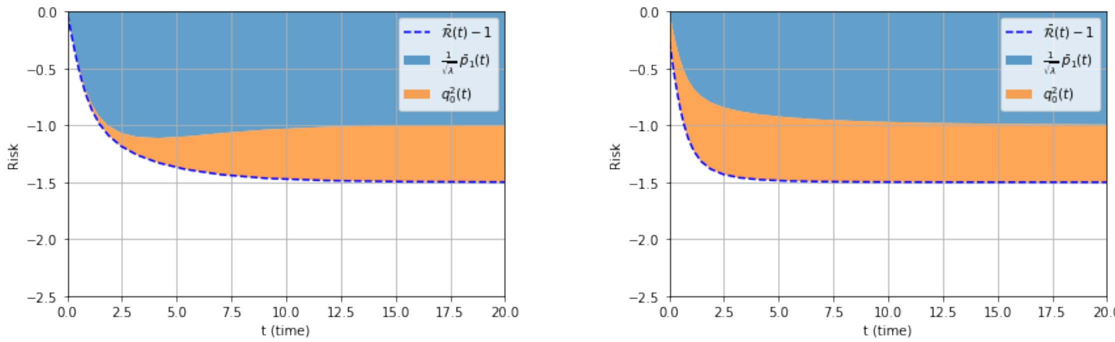
$p_1(\tau)$ does not decrease as much in early stages, and the healing phenomenon does not occur. At the same time, as observed on 1 (b) $q(\tau)$ is not monotonous: it increases at early stages and decreases down to its limiting value later.

3. Integro-differential equations

We study gradient descent in a regime where $t = \tau n/\eta$, $n \to +\infty$, with τ fixed. Abusing slightly notation we set $\theta_{\tau n/\eta} \to \theta_\tau$ so that equation (2) reads

$$\frac{d\theta_\tau}{d\tau} = -n\nabla_\theta \mathcal{H}(\theta_\tau) + \theta_\tau \langle \theta_\tau, \nabla_\theta \mathcal{H}(\theta_\tau)\rangle = \frac{1}{n^2} Y\theta_\tau - \frac{1}{n^2}\langle \theta_\tau, Y\theta_\tau\rangle \theta_\tau \tag{9}$$

We define $H = n^{-1/2}\xi$ the suitably normalized noise matrix. Besides the basic overlap $q(\tau) = \frac{1}{n}\langle \theta^*, \theta_\tau\rangle$, another one also plays an important role, namely $p_1(\tau) = \frac{1}{n}\langle \theta_\tau, H\theta_\tau\rangle$.



(a) $\alpha = 0.1$ (b) $\alpha = 0.5$

Figure 3: Cost evolution for $\lambda = 5$

Using $Y = \theta^*\theta^{*T} + \frac{n}{\sqrt{\lambda}}H$ we find

$$\frac{d\theta_\tau}{d\tau} = q(\tau)\theta^* + \frac{1}{\sqrt{\lambda}}H\theta_\tau - \left(q(\tau)^2 + \frac{p_1(\tau)}{\sqrt{\lambda}}\right)\theta_\tau \tag{10}$$

It is not possible to write down a closed set of equations that involve only $q(\tau)$ and $p_1(\tau)$, but only for a hierarchy of such objects, or for their generating functions. We now introduce these generating functions and then give the closed set of equations which they satisfy.

The $n \times n$ matrix $H = n^{-1/2}\xi$ is drawn with the probability law $\mathbb{P}$. Fix any small $\delta > 0$ and let $\mathcal{S}_\delta^n$ the set of realizations of H such that all eigenvalues fall in an interval $I_\delta = [-2-\delta, 2+\delta]$. Then $\mathbb{P}(\mathcal{S}_\delta^n) \to 1$ as $n \to +\infty$ (see for example Erdós (2011)). In the rest of this section it is understood that $H \in \mathcal{S}_\delta^n$. In particular the resolvent matrix[2] $\mathcal{R}(z) = (H - zI)^{-1}$ is well defined for $z \in \mathbb{C} \setminus I_\delta$ if $H \in \mathcal{S}_\delta^n$.

For any contour $\mathcal{C} = \{z \in \mathbb{C} \mid z = \rho e^{i\theta}, \theta \in [0, 2\pi]\}$ with $\rho > 2 + \delta$ we can define three generating functions

$$Q_\tau(z) = \frac{1}{n}\langle\theta_\tau, \mathcal{R}(z)\theta^*\rangle, \quad P_\tau(z) = \frac{1}{n}\langle\theta_\tau, \mathcal{R}(z)\theta_\tau\rangle, \quad R(z) = \frac{1}{n}\langle\theta^*, \mathcal{R}(z)\theta^*\rangle. \tag{11}$$

From standard holomorphic functional calculus for matrices (see for example Dunford and Schwartz (1988)) we have

$$q(\tau) = -\oint_{\mathcal{C}} \frac{dz}{2\pi i} Q_\tau(z), \quad p_1(\tau) = -\oint_{\mathcal{C}} \frac{dz}{2\pi i} z P_\tau(z). \tag{12}$$

Note that these two overlaps are part of a hierarchy of overlaps $q_k(\tau) \equiv \frac{\langle\theta^*, H^k\theta_\tau\rangle}{n} = -\oint_{\mathcal{C}} \frac{dz}{2\pi i} z^k Q_\tau(z)$ and $p_k(\tau) \equiv \frac{\langle\theta_\tau, H^k\theta_\tau\rangle}{n} = -\oint_{\mathcal{C}} \frac{dz}{2\pi i} z^k P_\tau(z)$, $k \geq 1$, which can all be calculated by the methods of this paper (note $q(\tau)$ corresponds to $k = 0$).

2. Here I is the identity $n \times n$ matrix and we will slightly abuse notation by omitting it and simply write $(H - z)^{-1}$.

Proposition 3 *For any realization $H \in \mathcal{S}_\delta$ and any $z \in \mathbb{C} \setminus I_\delta$ the generating functions (11) satisfy the integro-differential equation*

$$\begin{cases} \frac{d}{d\tau} Q_\tau(z) = q(\tau) R(z) + \frac{1}{\sqrt{\lambda}}(z Q_\tau(z) + q(\tau)) - \left(q^2(\tau) + \frac{1}{\sqrt{\lambda}} p_1(\tau)\right) Q_\tau(z) \\ \frac{1}{2}\frac{d}{d\tau} P_\tau(z) = q(\tau) Q_\tau(z) + \frac{1}{\sqrt{\lambda}}(z P_\tau(z) + 1) - \left(q^2(\tau) + \frac{1}{\sqrt{\lambda}} p_1(\tau)\right) P_\tau(z) \end{cases} \tag{13}$$

where $q(\tau) = -\oint_{\mathcal{C}} \frac{dz}{2\pi i} Q_\tau(z)$ and $p_1(\tau) = -\oint_{\mathcal{C}} \frac{dz}{2\pi i} z P_\tau(z)$.

Proof Let us derive the first equation. Using (10)

$$\begin{aligned} \frac{d}{d\tau} Q_\tau(z) = \frac{1}{n}\langle \theta^*, (H-z)^{-1} \frac{d\theta_\tau}{d\tau}\rangle &= \frac{q(\tau)}{n}\langle \theta^*, (H-z)^{-1}\theta^*\rangle + \frac{1}{n\sqrt{\lambda}}\langle \theta^*, (H-z)^{-1} H\theta_\tau\rangle \\ &- \left(q(\tau)^2 + \frac{p_1(\tau)}{\sqrt{\lambda}}\right)\frac{1}{n}\langle \theta^*, (H-z)^{-1}\theta_\tau\rangle \end{aligned} \tag{14}$$

Using $(H-z)^{-1}H = I + z(H-z)^{-1}$ in the second term in the right hand side, we immediately get the first equation of (13). Let us now derive the second equation. Again using (10) and since $(H - zI)$ is a symmetric matrix

$$\begin{aligned} \frac{d}{d\tau} P_\tau(z) = \frac{2}{n}\langle \theta_\tau, (H-z)^{-1} \frac{d\theta_\tau}{d\tau}\rangle &= 2\frac{q(\tau)}{n}\langle \theta_\tau, (H-z)^{-1}\theta^*\rangle + \frac{2}{n\sqrt{\lambda}}\langle \theta_\tau, (H-z)^{-1} H\theta_\tau\rangle \\ &- 2\left(q(\tau)^2 + \frac{p_1(\tau)}{\sqrt{\lambda}}\right)\frac{1}{n}\langle \theta_\tau, (H-z)^{-1}\theta_\tau\rangle \end{aligned} \tag{15}$$

Thus using again $(H-z)^{-1}H = I + z(H-z)^{-1}$ and $\langle \theta_\tau, \theta_\tau\rangle = 1$ we get the second equation of (13). ■

4. Concentration results

We introduce the Stieltjes transform of the semi-circle law $\mu_{\rm sc}(s) = \frac{1}{2\pi}\sqrt{4-s^2}\chi_{[-2,2]}(s)$,

$$G_{\rm sc}(z) = \int_{\mathbb{R}} ds \frac{\mu_{\rm sc}(s)}{s-z} = \frac{1}{2}(-z + \sqrt{z^2-4}), \quad z \in \mathbb{C} \setminus [-2,2]. \tag{16}$$

It is a classical result of random matrix theory Erdós (2011) that, for any $z \in \mathbb{C} \setminus [-2,2]$,

$$\frac{1}{n}\mathrm{Tr}\mathcal{R}(z) \xrightarrow[n\to\infty]{\mathbb{P}} G_{\rm sc}(z)$$

. However here we will need convergence in probability of *matrix elements* of the resolvent (for given z and also uniformly in z). This tool is provided by recent results in random matrix theory that go under the name of *local semi-circle laws* Bloemendal et al. (2014).

Recall that $\mathcal{S}_\delta^n$ is the set of realizations of $H = \frac{1}{\sqrt{n}}\xi$ with eigenvalues in $I_\delta = [-2-\delta, 2+\delta]$, $\delta > 0$, and that $\lim_{n\to+\infty}\mathbb{P}(\mathcal{S}_\delta^n) = 1$. It will be convenient to use the notation $\mathbb{P}_\delta$ for the conditional probability law of H *conditioned* on the event $H \in \mathcal{S}_\delta^n$.

4.1. Initial condition analysis

We first derive natural initial conditions for the integro-differential equations (13) when $\frac{1}{n}\langle\theta_0, \theta^*\rangle = q(0) = \alpha \in [-1, +1]$. We claim (corollary 5 below) that the initial conditions $Q_0(z), P_0(z)$ as well as $R(z)$ concentrate on explicit functions $\bar{Q}_0(z), \bar{P}_0(z), \bar{R}(z)$. The main tool is the following proposition which we prove in section B (based on a theorem in Bloemendal et al. (2014)):

Proposition 4 *Fix $\delta > 0$, $\epsilon > 0$. For any fixed $z \in \mathbb{C} \setminus I_\delta$ and any deterministic sequence of unit vectors $u^{(n)}, v^{(n)} \in \mathbb{S}^{n-1}(1)$ the n-sphere of unit radius, we have*

$$\lim_{n\to\infty} \mathbb{P}_\delta\big(|\langle u^{(n)}, \mathcal{R}(z)v^{(n)}\rangle - \langle u^{(n)}, v^{(n)}\rangle G_{sc}(z)| > \epsilon\big) = 0. \tag{17}$$

Applying this proposition to the three pairs of unit vectors $(u^{(n)}, v^{(n)}) = (\frac{\theta_0}{\sqrt{n}}, \frac{\theta^*}{\sqrt{n}})$, $(\frac{\theta_0}{\sqrt{n}}, \frac{\theta_0}{\sqrt{n}})$, and $(\frac{\theta^*}{\sqrt{n}}, \frac{\theta^*}{\sqrt{n}})$ we directly obtain

Corollary 5 *Fix $\alpha \in [-1, +1]$ and θ_0 such that $\frac{1}{n}\langle\theta_0, \theta^*\rangle = \alpha$. For $z \in \mathbb{C}\setminus\mathbb{I}_\delta$ we have convergence in probability of $Q_0(z), P_0(z), R(z)$ to the Stieljes transform of the semi-circle law:*

$$Q_0(z) \xrightarrow[n\to\infty]{\mathbb{P}_\delta} \bar{Q}_0(z) = \alpha G_{sc}(z), \quad P_0(z) \xrightarrow[n\to\infty]{\mathbb{P}_\delta} \bar{P}_0(z) = G_{sc}(z), \quad R(z) \xrightarrow[n\to\infty]{\mathbb{P}_\delta} \bar{R}(z) = G_{sc}(z). \tag{18}$$

4.2. Concentration of the overlap for finite times

We consider the integro-differential equations (13) for the limiting initial conditions $(\bar{Q}_0(z), \bar{P}_0(z)) = (\alpha G_{\rm sc}(z), G_{\rm sc}(z))$ and limiting $\bar{R}(r) = G_{\rm sc}(z)$. More explicitly we define $\bar{Q}_\tau(z)$, $\bar{P}_\tau(z)$ as the (holomorphic over $z \in \mathbb{C} \setminus I_\delta$) solutions of

$$\begin{cases} \frac{d}{d\tau}\bar{Q}_\tau(z) = \bar{q}(\tau)\bar{R}(z) + \frac{1}{\sqrt{\lambda}}(z\bar{Q}_\tau(z) + \bar{q}(\tau)) - \left(\bar{q}^2(\tau) + \frac{1}{\sqrt{\lambda}}\bar{p}_1(\tau)\right)\bar{Q}_\tau(z) \\ \frac{1}{2}\frac{d}{d\tau}\bar{P}_\tau(z) = \bar{q}(\tau)\bar{Q}_\tau(z) + \frac{1}{\sqrt{\lambda}}(z\bar{P}_\tau(z) + 1) - \left(\bar{q}^2(\tau) + \frac{1}{\sqrt{\lambda}}\bar{p}_1(\tau)\right)\bar{P}_\tau(z) \end{cases} \tag{19}$$

where *by definition* $\bar{q}(\tau) = -\oint_{\mathcal{C}} \frac{dz}{2\pi i}\bar{Q}_\tau(z)$ and $\bar{p}_1(\tau) = -\oint_{\mathcal{C}} \frac{dz}{2\pi i} z\bar{P}_\tau(z)$, and the initial conditions are $\bar{Q}_0(z) = \alpha G_{\rm sc}(z)$, $\bar{P}_0(z) = G_{\rm sc}(z)$. The explicit calculation of the solutions $\bar{Q}_\tau(z)$, $\bar{P}_\tau(z)$ in section 5 shows that they exist and they are holomorphic for $z \in \mathbb{C} \setminus I_\delta$.

One can show that the concentration result of corollary 5 extends to all finite times. This can be done by a Grönwall stability type argument. A difficulty with respect to the standard argument is that here we deal with an integro-differential equation instead of purely ordinary differential equation. For this reason we need a *uniform* (over z) concentration result which strengthens proposition 4. The following is proved in section B.

Proposition 6 *Fix $\delta > 0$, $\epsilon > 0$. Recall $\mathcal{C} = \{z \in \mathbb{C} \mid z = \rho e^{i\theta}, \theta \in [0, 2\pi]\}$ for $\rho \geq 2 + \delta$. For any deterministic sequence of unit vectors $u^{(n)}, v^{(n)} \in \mathbb{S}^{n-1}(1)$ the n-sphere of unit radius, we have*

$$\lim_{n\to\infty} \mathbb{P}_\delta\big(\sup_{z\in\mathcal{C}} |\langle u^{(n)}, \mathcal{R}(z)v^{(n)}\rangle - \langle u^{(n)}, v^{(n)}\rangle G_{sc}(z)| > \epsilon\big) = 0. \tag{20}$$

Applying this proposition to appropriate pairs of unit vectors as previously we get directly:

Corollary 7 *Fix $\alpha \in [-1,+1]$ and θ_0 such that $\frac{1}{n}\langle\theta_0,\theta^*\rangle = \alpha$. Let $\mathcal{C} = \{z \in \mathbb{C} \mid z = \rho e^{i\theta}, \theta \in [0,2\pi]\}$ for some $\rho \geq 2+\delta$. Recall $\bar{Q}_0(z) = \alpha G_{sc}(z)$, $\bar{P}_0(z) = G_{sc}(z)$, $\bar{R}(z) = G_{sc}(z)$. Then $\sup_{z\in\mathcal{C}} |Q_0(z) - \bar{Q}_0(z)|$, $\sup_{z\in\mathcal{C}} |P_0(z) - \bar{P}_0(z)|$, $\sup_{z\in\mathcal{C}} |R(z) - \bar{R}(z)|$ all converge in $\mathbb{P}_\delta$-probability to zero.*

In section C this corollary is used to prove:

Proposition 8 *Fix $\alpha \in [-1,+1]$ and θ_0 such that $\frac{1}{n}\langle\theta_0,\theta^*\rangle = \alpha$. Fix any $T > 0$. We have convergences at any $\tau \in [0,T]$ of the following overlaps to the deterministic limits $q(\tau) \xrightarrow[n\to\infty]{\mathbb{P}} \bar{q}(\tau)$, $p_1(\tau) \xrightarrow[n\to\infty]{\mathbb{P}} \bar{p}_1(\tau)$ where here convergence is with respect to the probability law $\mathbb{P}$ of the generalized Wigner ensemble.*

Remark 9 *With a bit more work the proof of this corollary can be strengthened to also show that for any $z \in \mathbb{C}\setminus I_\delta$ and $\tau \in [0,T]$ we have convergence in probability of $Q_\tau(z), P_\tau(z)$ to the deterministic solutions of the integro-differential equations (19), i.e., $Q_\tau(z) \xrightarrow[n\to\infty]{\mathbb{P}_\delta} \bar{Q}_\tau(z)$, $P_\tau(z) \xrightarrow[n\to\infty]{\mathbb{P}_\delta} \bar{P}_\tau(z)$, as well as convergence of all overlaps $q_k(\tau), p_k(\tau) \xrightarrow[n\to\infty]{\mathbb{P}} \bar{q}_k(\tau), \bar{p}_k(\tau)$ ($k \geq 1$). Since we will not need these results we omit their proof.*

5. Solution of integro-differential equations and overlap

In this section we analyze (19) for $z \in \mathbb{C}\setminus I_\delta$ with the initial conditions $(\bar{Q}_0(z), \bar{P}_0(z), \bar{R}(z)) = (\alpha G_{\rm sc}(z), G_{\rm sc}(z), G_{\rm sc}(z))$. In the process we obtain $\bar{q}(\tau) \equiv -\int_{\mathbb{C}} \frac{dz}{2\pi i}\bar{Q}_\tau(z)$.

Proof [Proof of formulas (5) and (6) in theorem 1] We use a change of variable $\hat{Q}_\tau(z) = e^{F(\tau)}\bar{Q}_\tau(z)$ and $\hat{P}_\tau(z) = e^{2F(\tau)}\bar{P}_\tau(z)$ with $F(\tau) = \int_0^\tau ds\Big(\bar{q}^2(s) + \frac{1}{\sqrt{\lambda}}\bar{p}_1(s)\Big)$. Similarly, we define also $\hat{q}(\tau) = e^{F(\tau)}\bar{q}(\tau), \hat{p}(\tau) = e^{2F(\tau)}$. We have $\bar{q}(\tau) = \hat{q}(\tau)/\sqrt{\hat{p}(\tau)}$, and therefore in order to determine the overlap it suffices to determine $\hat{q}(\tau)$ and $\hat{p}(\tau)$. With the change of variables equations (13) become

$$\begin{cases} \frac{d}{d\tau}\hat{Q}_\tau(z) = \hat{q}(\tau)\left(\bar{R}(z) + \frac{1}{\sqrt{\lambda}}\right) + \frac{z}{\sqrt{\lambda}}\hat{Q}_\tau(z) \\ \frac{1}{2}\frac{d}{d\tau}\hat{P}_\tau(z) = \hat{q}(\tau)\hat{Q}_\tau(z) + \frac{1}{\sqrt{\lambda}}\hat{p}(\tau) + \frac{z}{\sqrt{\lambda}}\hat{P}_\tau(z) \end{cases} \tag{21}$$

We analyze these equations in the Laplace domain. Recall the Laplace transformation $\mathcal{L}f(p) = \int_0^{+\infty} d\tau e^{-p\tau} f(\tau)$, $\mathrm{Re}\, p > a \in \mathbb{R}_+$, which is well defined as long as $|f(\tau)| \leq e^{a\tau}$. All functions involved below in Laplace transforms satisfy this requirement for some $a \in \mathbb{R}_+$ large enough independent of n. It will often be convenient to use the notations $\mathcal{L}(f(t))(p) = \int_0^{+\infty} d\tau e^{-p\tau} f(\tau)$, $\mathcal{L}Q_p(z) = \int_0^{+\infty} d\tau e^{-p\tau} Q_\tau(z)$, $\mathcal{L}P_p(z) = \int_0^{+\infty} d\tau e^{-p\tau} P_\tau(z)$.

A) Derivation of (5) *for $\hat{q}(\tau)$.* Taking the Laplace transform of the first equation in (21)

$$p\mathcal{L}\hat{Q}_p(z) - \hat{Q}_0(z) = \mathcal{L}\hat{q}(p)\big(\bar{R}(z) + \frac{1}{\sqrt{\lambda}}\big) + \frac{z}{\sqrt{\lambda}}\mathcal{L}\hat{Q}_p(z) \tag{22}$$

Notice that $\hat{Q}_0(z) = e^{F(0)}\bar{Q}_0(z) = \alpha G_{\rm sc}(z)$ and $\bar{R}(z) = G_{\rm sc}(z)$, and hence we can re-arrange the terms,

$$\mathcal{L}\hat{Q}_p(z) = \alpha\frac{\sqrt{\lambda}G_{\rm sc}(z)}{p\sqrt{\lambda} - z} + \mathcal{L}\hat{q}(p)\frac{\sqrt{\lambda}G_{\rm sc}(z) + 1}{p\sqrt{\lambda} - z}. \tag{23}$$

Now, assuming $\mathrm{Re}\, p > \frac{2+\delta}{\sqrt{\lambda}}$ (recall $\mathrm{Re}\, p > 0$) leaves the point $p\sqrt{\lambda}$ outside the contour $\mathcal{C}$. Using Fubini first and the definition of $\hat{q}$ secondly

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i} \int_0^{+\infty} d\tau e^{-p\tau} \hat{Q}_\tau(z) = \int_0^{+\infty} d\tau e^{-p\tau} \oint_{\mathcal{C}} \frac{dz}{2\pi i} \hat{Q}_\tau(z) = -\int_0^{+\infty} d\tau e^{-p\tau} \hat{q}(\tau) \tag{24}$$

Thus we have $\oint_{\mathcal{C}} \frac{dz}{2\pi i} \mathcal{L}\hat{Q}_p(z) = -\mathcal{L}\hat{q}(p)$ on the left side of (23) while a straightforward calculation using Fubini on compact sets shows

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i} \frac{G_{\rm sc}(z)}{p\sqrt{\lambda} - z} = \frac{1}{2\pi i} \oint_{\mathcal{C}} \int_{-2}^{2} \frac{\mu_{\rm sc}(l) \mathrm{d}l \mathrm{d}z}{(l - z)(p\sqrt{\lambda} - z)} = \int_{-2}^{2} \frac{\mu_{\rm sc}(l) \mathrm{d}l}{l - p\sqrt{\lambda}} = G_{\rm sc}(p\sqrt{\lambda}) \tag{25}$$

So taking Cauchy integration formula on both sides of (23) we get

$$-\mathcal{L}\hat{q}(p) = \alpha\sqrt{\lambda} G_{\rm sc}(p\sqrt{\lambda}) + \mathcal{L}\hat{q}(p)\sqrt{\lambda} G_{\rm sc}(p\sqrt{\lambda}) \tag{26}$$

Thus we find:

$$\mathcal{L}\hat{q}(p) = -\frac{\alpha G_{\rm sc}(p\sqrt{\lambda})}{\frac{1}{\sqrt{\lambda}} + G_{\rm sc}(p\sqrt{\lambda})} = \alpha \frac{1 + \frac{1}{\sqrt{\lambda}} G_{\rm sc}(p\sqrt{\lambda})}{p - (1 + \frac{1}{\lambda})} \tag{27}$$

where the last equality can be checked from the explicit expression (16) of $G_{\rm sc}(z)$. It remains to invert this equation in the time domain. To do so we first notice that

$$G_{\rm sc}(p\sqrt{\lambda}) = -\frac{1}{\sqrt{\lambda}} \int_{-2}^{2} ds \mu_{\rm sc}(s) \int_0^{+\infty} d\tau\, e^{(\frac{s}{\sqrt{\lambda}} - p)\tau} = -\frac{1}{\sqrt{\lambda}} \int_0^{+\infty} d\tau\, e^{-p\tau} M_\lambda(\tau) \tag{28}$$

where we recall that $M_\lambda(\tau)$ is the scaled moment generating function of the semi-circle law (3). The interchange of integrals in the third equality is justified by Fubini. Using $\mathcal{L}(e^{(1+\frac{1}{\lambda})\tau})(p) = (p - (1 + \frac{1}{\lambda}))^{-1}$, equation (27) becomes $\mathcal{L}\hat{q}(p) = \alpha \mathcal{L}(e^{(1+\frac{1}{\lambda})t})(p) - \frac{\alpha}{\lambda} \mathcal{L}(e^{(1+\frac{1}{\lambda})t}) \mathcal{L} M_\lambda(p)$. This is easily transformed back in the time-domain using standard properties of the Laplace transform to get (5).

B) A useful identity. For the derivation of $\hat{p}(\tau)$ we will need the following identity derived in Appendix H

$$-\oint_{\mathcal{C}} \frac{dz}{2\pi i} \hat{Q}_u(z) e^{\frac{z(\tau - u)}{\sqrt{\lambda}}} = \alpha M_\lambda(2\tau - u) + \int_0^u ds \hat{q}(s) M_\lambda(2\tau - u - s) \tag{29}$$

where we recall $\mathcal{C}$ is the circle with center the origin and radius $\rho > 2 + \delta$.

C) Derivation of $\hat{p}(\tau)$. Taking the Laplace transform of the second equation in (21) we find

$$\frac{1}{2}(p\mathcal{L}\hat{P}_p(z) - \hat{P}_0(z)) = \mathcal{L}(\hat{q}(\tau)\hat{Q}_\tau(z))(p) + \frac{1}{\sqrt{\lambda}} \mathcal{L}\hat{p}(p) + \frac{z}{\sqrt{\lambda}} \mathcal{L}\hat{P}_p(z) \tag{30}$$

and using $\hat{P}_0(z) = e^{F(0)} \bar{P}_0(z) = G_{\rm sc}(z)$ we can rearrange the terms to get

$$\mathcal{L}\hat{P}_p(z) = \frac{1}{p - \frac{z\sqrt{\lambda}}{2}} \left(G_{\rm sc}(z) + 2\sqrt{\lambda} \mathcal{L}(\hat{q}(\tau)\hat{Q}_\tau(z))(p) + \frac{2}{\sqrt{\lambda}} \mathcal{L}\hat{p}(p) \right). \tag{31}$$

Then using $(p - \frac{2z}{\sqrt{\lambda}})^{-1} = \mathcal{L}(e^{\frac{2zt}{\sqrt{\lambda}}})(p)$ and

$$2\mathcal{L}(e^{\frac{2zt}{\sqrt{\lambda}}})\mathcal{L}(\hat{q}(t)\hat{Q}_t(z)) = \mathcal{L}\big(2\int_0^t \hat{q}(u)\hat{Q}_u(z)e^{\frac{2z(t-u)}{\sqrt{\lambda}}}du\big), \tag{32}$$

and replacing in (31) we get

$$\mathcal{L}\hat{P}_p(z) = \mathcal{L}\big(e^{\frac{2z\tau}{\sqrt{\lambda}}}\big)(p)G_{\rm sc}(z) + 2\mathcal{L}\bigg(\int_0^\tau \hat{q}(u)\hat{Q}_u(z)e^{\frac{2z(\tau-u)}{\sqrt{\lambda}}}du\bigg)(p) + \frac{2}{\sqrt{\lambda}}\mathcal{L}(e^{\frac{2z\tau}{\sqrt{\lambda}}})(p)\mathcal{L}\hat{p}(p). \tag{33}$$

Now we take $\mathrm{Re}\, p > 4/\sqrt{\lambda}$ and choose the contour $\mathcal{C}$, encircling the interval I_δ, but such that it does not encircle the point $z = \frac{1}{2}p\sqrt{\lambda}$, and integrate each term of (33) along this contour. First note that the contribution of the last term vanishes since $\mathcal{L}(e^{\frac{2z\tau}{\sqrt{\lambda}}})(p) = (p - \frac{2z}{\sqrt{\lambda}})^{-1}$ and the pole $z = \frac{1}{2}p\sqrt{\lambda}$ lies in the exterior of $\mathcal{C}$. Then there remains

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i}\mathcal{L}\hat{P}_p(z) = \oint_{\mathcal{C}} \frac{dz}{2\pi i}\mathcal{L}(e^{\frac{2z\tau}{\sqrt{\lambda}}})(p)G_{\rm sc}(z) + 2\oint_{\mathcal{C}} \frac{dz}{2\pi i}\mathcal{L}\bigg(\int_0^\tau \hat{q}(u)\hat{Q}_u(z)e^{\frac{2z(\tau-u)}{\sqrt{\lambda}}}du\bigg)(p). \tag{34}$$

For the left hand side we have

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i}\int_0^{+\infty} d\tau e^{-p\tau}\hat{P}_\tau(z) = \int_0^{+\infty} d\tau e^{-p\tau}\oint_{\mathcal{C}} \frac{dz}{2\pi i}\hat{P}_\tau(z) = -\int_0^{+\infty} d\tau e^{-p\tau}\hat{p}(\tau) \tag{35}$$

where the first equality follows from Fubini and the second by functional calculus Dunford and Schwartz (1988). For the first term on the right hand side of (34) we find (see Appendix H for details)

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i}G_{\rm sc}(z)\int_0^{+\infty} d\tau e^{-p\tau}e^{\frac{z\tau}{\sqrt{\lambda}}} = -\int_0^{+\infty} d\tau e^{-p\tau}M_\lambda(\tau). \tag{36}$$

Finally it remains to treat the last contour integral in (34). Using again Fubini and (29) we find

$$\begin{aligned}
&\oint_{\mathcal{C}} \frac{dz}{2\pi i}\int_0^{+\infty} d\tau e^{-p\tau}\int_0^\tau \hat{q}(u)\hat{Q}_u(z)e^{\frac{2z(\tau-u)}{\sqrt{\lambda}}}du = \int_0^{+\infty} d\tau e^{-p\tau}\int_0^\tau \hat{q}(u)\oint_{\mathcal{C}} \frac{dz}{2\pi i}\hat{Q}_u(z)e^{\frac{2z(\tau-u)}{\sqrt{\lambda}}}du\\
&= -\int_0^{+\infty} d\tau e^{-p\tau}\int_0^\tau du\hat{q}(u)\bigg[\alpha M_\lambda(2\tau-u) + \int_0^u ds\hat{q}(s)M_\lambda(2\tau-u-s)\bigg]\\
&= -\int_0^{+\infty} d\tau e^{-p\tau}\bigg[\alpha\int_0^\tau du\hat{q}(u)M_\lambda(2\tau-u) + \frac{1}{2}\int_0^\tau\int_0^\tau duds\hat{q}(s)M_\lambda(2\tau-u-s)\bigg]
\end{aligned} \tag{37}$$

Putting together (34), (35), (36), (37) we obtain (6) in the Laplace domain. Going back to the time domain we obtain (6). ■

6. Conclusion and future work

Tracking gradient descent dynamics and their variants for different scores and loss functions can be used to provide meaningful insights on a learning algorithm and for example, help monitor its progress and avoid over-fitting. As computational capabilities increase with distributed systems allowing for bigger datasets and larger systems to be treated, a good understanding of the dynamics can help account for computational cost.

We have seen in this work that for the rank-one matrix recovery problem in the regime of large dimensions, probabilistic concentrations naturally occur that can be captured by the local semi-circle laws in random matrix theory obtained in the last decade. In particular, suitable generating functions constructed out of the resolvent of the noise matrix concentrate around the solutions of a set of deterministic integro-differential equations. We have been able to completely solve these equations thereby tracking the dynamics for all times. It is also observed that the analytical solution provides a good approximation for the expected behavior of the learning algorithm, even for dimensions as low as $n < 100$.

The method and integro-differential equations derived here can be generalized to different models. For instance, we will show in forthcoming work how it is possible to apply it to certain neural-network architectures, and in particular the random feature models. This allows us to better understand the dynamical emergence of interesting behaviors such as the double descent phenomenon. The generalisation is possible, in essence, when the dynamics can be captured by spectral properties of some ”resolvent matrix”. Depending on the system though, performing random matrix averages can be arbitrarily complicated. For problems where the dynamics is not captured by some resolvent matrix, such as a genuine tensor problem (with tensor of order greater equal than three) it is not so clear how to proceed since there are no obvious spectral notions for tensors. One option would be to approach the problem by looking at the dynamics of the alternating least square method.

Acknowledgment

The work of A. B has been supported by Swiss National Science Foundation grant no 200020 182517.

References

Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 09 2005. doi: 10.1214/009117905000000233. URL https://doi.org/10.1214/009117905000000233.

Afonso S. Bandeira, Nicolas Boumal, and Vladislav Voroninski. On the low-rank approach for semidefinite programs arising in synchronization and community detection. volume 49 of *Proceedings of Machine Learning Research*, pages 361–382, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR. URL http://proceedings.mlr.press/v49/bandeira16.html.

J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems (NIPS) 29*, pages 424–432. 2016.

J. Barbier, N. Macris, and L. Miolane. The Layered Structure of Tensor Estimation and its Mutual Information. In *55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, September 2017.

Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo-kurchan equations for dynamics of spin-glasses. *Probability Theory and Related Fields*, 136, 10 2004. doi: 10.1007/s00440-005-0491-y.

Florent Benaych-Georges and Antti Knowles. Lectures on the local semicircle law for Wigner matrices. working paper or preprint, 2016. URL https://hal.archives-ouvertes.fr/hal-01258444.

Srinadh Bhojanapalli, Behnam Neyshabur, and Nathan Srebro. Global optimality of local search for low rank matrix recovery. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 3880–3888, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

Alex Bloemendal, László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Isotropic local laws for sample covariance and generalized wigner matrices. *Electron. J. Probab.*, 19:53 pp., 2014. doi: 10.1214/EJP.v19-3054. URL https://doi.org/10.1214/EJP.v19-3054.

Samuel Burer and Renato Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming, Series B*, 95:329–357, 02 2003. doi: 10.1007/s10107-002-0352-8.

Samuel Burer and Renato Monteiro. Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming*, 103:427–444, 07 2005. doi: 10.1007/s10107-004-0564-1.

Yuejie Chi, Yue M. Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, Oct 2019.

A. Crisanti, H. Horner, and H. Sommers. The sphericalp-spin interaction spin-glass model. *Zeitschrift für Physik B Condensed Matter*, 92:257–271, 1993.

L F Cugliandolo and D S Dean. Full dynamical solution for a spherical spin-glass model. *Journal of Physics A: Mathematical and General*, 28(15):4213–4234, Aug 1995. ISSN 1361-6447. doi: 10.1088/0305-4470/28/15/003. URL http://dx.doi.org/10.1088/0305-4470/28/15/003.

L. F. Cugliandolo and J. Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Phys. Rev. Lett.*, 71:173–176, Jul 1993. doi: 10.1103/PhysRevLett.71.173. URL https://link.aps.org/doi/10.1103/PhysRevLett.71.173.

Christopher De Sa, Kunle Olukotun, and Christopher Ré. Global convergence of stochastic gradient descent for some non-convex matrix problems. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 2332–2341. JMLR.org, 2015.

N Dunford and J T Schwartz. *Linear Operators*. Wiley Classics Library, 1988.

Laszló Erdós. Universality of wigner random matrices: a survey of recent results. *Russian Mathematical Surveys*, 66(3):507–626, Jun 2011. ISSN 1468-4829. doi: 10.1070/rm2011v066n03abeh004749. URL http://dx.doi.org/10.1070/RM2011v066n03ABEH004749.

László Erdós, Benjamin Schlein, and Horng-Tzer Yau. Local semicircle law and complete delocalization for wigner random matrices. *Communications in Mathematical Physics*, 287(2): 641–655, Sep 2008. ISSN 1432-0916. doi: 10.1007/s00220-008-0636-9. URL http://dx.doi.org/10.1007/s00220-008-0636-9.

Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large wigner matrices. *Communications in Mathematical Physics*, 272, 06 2006. doi: 10.1007/s00220-007-0209-3.

Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. In *Proceedings of the 34th International Conference on Machine Learning*, ICML'17, pages 1233–1242. JMLR.org, 2017a.

Rong Ge, Chi Jin, and Yi Zheng. No spurious local minima in nonconvex low rank problems: A unified geometric analysis. volume 70 of *Proceedings of Machine Learning Research*, pages 1233–1242, International Convention Centre, Sydney, Australia, 06–11 Aug 2017b. PMLR. URL http://proceedings.mlr.press/v70/ge17a.html.

Satish Babu Korada and Nicolas Macris. Exact solution of the gauge symmetric p-spin glass model on a complete graph. *Journal of Statistical Physics*, 136(2):205–230, 2009.

Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pages 2278–2324, 1998.

Jason D. Lee, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. Gradient descent only converges to minimizers. volume 49 of *Proceedings of Machine Learning Research*, pages 1246–1257, Columbia University, New York, New York, USA, 23-pmlr-v49-lee1626 Jun 2016. PMLR. URL http://proceedings.mlr.press/v49/lee16.html.

Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3-4):859–929, 2018.

Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory, ISIT 2017, Aachen, Germany, June 25-30, 2017*, pages 511–515. IEEE, 2017. doi: 10.1109/ISIT.2017.8006580. URL https://doi.org/10.1109/ISIT.2017.8006580.

Shuyang Ling, Ruitu Xu, and Afonso S. Bandeira. On the landscape of synchronization networks: A perspective from nonconvex optimization. *arXiv:1809.11083*, 2019.

Stefano Sarao Mannelli, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborova. Passed and spurious: Descent algorithms and local minima in spiked matrix-tensor models. volume 97 of *Proceedings of Machine Learning Research*, pages 4333–4342, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL http://proceedings.mlr.press/v97/mannelli19a.html.

Leo Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. 2017.

Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *The Annals of Statistics*, 49(1):321 – 345, 2021. doi: 10.1214/20-AOS1958. URL https://doi.org/10.1214/20-AOS1958.

Dohyung Park, Anastasios Kyrillidis, Constantine Carmanis, and Sujay Sanghavi. Non-square matrix sensing without spurious local minima via the Burer-Monteiro approach. volume 54 of *Proceedings of Machine Learning Research*, pages 65–74, Fort Lauderdale, FL, USA, 20–22 Apr 2017. PMLR. URL http://proceedings.mlr.press/v54/park17a.html.

Sandrine Péché. The largest eigenvalue of small rank perturbations of hermitian random matrices. *Probability Theory and Related Fields*, 134:127–173, 2004.

Amelia Perry, Alexander S. Wein, and Afonso S. Bandeira. Statistical limits of spiked tensor models. *Annales de l'Institut Henri Poincaré, Probabilitées et Statistiques*, 56(1):230 – 264, 2020. doi: 10.1214/19-AIHP960. URL https://doi.org/10.1214/19-AIHP960.

Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, and Lenka Zdeborová. Who is afraid of big bad minima? analysis of gradient-flow in spiked matrix-tensor models. In H. Wallach, H. Larochelle, A. Beygelzimer, F. Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8679–8689. Curran Associates, Inc., 2019.

Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborovà. Marvels and pitfalls of the Langevin algorithm in noisy high-dimensional inference. *Physical Review X*, 10(1), Mar 2020. ISSN 2160-3308. doi: 10.1103/physrevx.10.011057. URL http://dx.doi.org/10.1103/PhysRevX.10.011057.

Appendix A. Analysis of the cost

Proof [Proof of theorem 2] Expanding the Frobenius norm in the cost and using $\|\theta_\tau\|^2 = \|\theta^*\|^2 = n$ we find

$$\begin{aligned}\mathcal{H}(\theta_\tau) &= \frac{1}{2n^2}\{-2\mathrm{Tr}Y\theta_\tau\theta_\tau^T + \mathrm{Tr}(\theta_\tau\theta_\tau^T\theta_\tau\theta_\tau^T)\} - \frac{1}{2n^2}\{-2\mathrm{Tr}Y\theta^*\theta^{*T} + \mathrm{Tr}(\theta^*\theta^{*T}\theta^*\theta^{*T})\} \\ &= \frac{1}{n^2}\langle\theta^*, Y\theta^*\rangle - \frac{1}{n^2}\langle\theta_\tau, Y\theta_\tau\rangle. \end{aligned} \tag{38}$$

Using that $Y = \theta^*\theta^{*T} + \frac{n}{\sqrt{\lambda}}H$ (recall $H = \frac{1}{\sqrt{n}}\xi$) we get

$$\begin{aligned}\mathcal{H}(\theta_\tau) &= \big(1 + \frac{1}{\sqrt{\lambda}}\frac{\langle\theta^*, H\theta^*\rangle}{n}\big) - \big(\frac{\langle\theta_\tau, \theta^*\rangle^2}{n^2} + \frac{1}{\sqrt{\lambda}}\frac{\langle\theta_\tau, H\theta_t\rangle}{n}\big) \\ &= \big(1 + \frac{1}{\sqrt{\lambda}}\frac{\langle\theta^*, H\theta^*\rangle}{n}\big) - \big(q(\tau)^2 + \frac{p_1(\tau)}{\sqrt{\lambda}}\big). \end{aligned} \tag{39}$$

By the law of large numbers $\frac{\langle\theta^*, H\theta^*\rangle}{n} \xrightarrow[n\to\infty]{\mathbb{P}} 0$ and since $q(\tau) \xrightarrow[n\to\infty]{\mathbb{P}} \bar{q}(\tau)$ and $p_1(\tau) \xrightarrow[n\to\infty]{\mathbb{P}} \bar{p}_1(\tau)$ we have

$$\mathcal{H}(\theta_\tau) \xrightarrow[n\to\infty]{\mathbb{P}} 1 - \left(\bar{q}(\tau)^2 + \frac{\bar{p}_1(\tau)}{\sqrt{\lambda}}\right). \tag{40}$$

Now it remains to recall the definition of $F(\tau)$ and $\hat{p}_0(\tau) = e^{2F(\tau)}$, to see that

$$\bar{q}(\tau)^2 + \frac{\bar{p}_1(\tau)}{\sqrt{\lambda}} = \frac{dF(\tau)}{d\tau} = \frac{1}{2}\frac{d}{d\tau}\ln\hat{p}(\tau). \tag{41}$$

The result of the theorem follows from (40) and (41). ■

Appendix B. Proof of propositions 4 and 6

The proof is based the following *local* semi-circle law (theorem 2.12 in Bloemendal et al. (2014)):

Theorem 10 (isotropic local semi-circle law Bloemendal et al. (2014)) *For any $\omega \in (0,1)$ consider the following domain in the upper half-plane*

$$S(\omega, n) = \left\{z \in \mathbb{C} \mid |\mathrm{Re}(z)| \leq \frac{1}{\omega}, \frac{1}{n^{1-\omega}} \leq \mathrm{Im}(z) \leq \frac{1}{\omega}\right\}.$$

Then for all $\delta, D > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$, and any unit vectors $u, v \in \mathbb{S}_n(1)$:

$$\sup_{z\in S(\omega,n)} \mathbb{P}\left(|\langle u, \mathcal{R}(z)v\rangle - \langle u, v\rangle G_{sc}(z)| > n^\delta\left[\sqrt{\frac{\mathrm{Im}\,G_{sc}(z)}{n\,\mathrm{Im}\,z}} + \frac{1}{n\,\mathrm{Im}\,z}\right]\right) < \frac{1}{n^D} \tag{42}$$

where $\mathbb{P}$ is the probability law on the generalized Wigner matrix.

Proof [*Proof of proposition 4*] First we note that for $\operatorname{Im} z \neq 0$ since $\lim_{n\to+\infty} \mathbb{P}(\mathcal{S}_\delta^n) = 1$ we have

$$\lim_{n\to+\infty} \mathbb{P}\big(|\langle u, \mathcal{R}(z)v\rangle - \langle u, v\rangle G_{\text{sc}}(z)| > \epsilon\big) = \lim_{n\to+\infty} \mathbb{P}_\delta\big(|\langle u, \mathcal{R}(z)v\rangle - \langle u, v\rangle G_{\text{sc}}(z)| > \epsilon\big). \quad (43)$$

We consider fist the cases $\operatorname{Im} z$ strictly positive, negative, and then give the extra argument needed for $\operatorname{Im} z = 0$.

First we take $\operatorname{Im} z > 0$. We can find $n_1 \in \mathbb{N}, \omega \in (0,1)$ such that $z \in S(\omega, n_1)$ and henceforth, for all $n \geq n_1$, $z \in S(\omega, n)$. Taking $\delta = \frac{1}{4}$, $D = 1$ and applying theorem 10 yields the existence of n_0 such that for all $n \geq \max(n_0, n_1)$:

$$\mathbb{P}\left(|\langle u, \mathcal{R}(z)v\rangle - \langle u, v\rangle G_{\text{sc}}(z)| > n^{\frac{1}{4}}\left[\sqrt{\frac{\operatorname{Im} G_{\text{sc}}(z)}{n \operatorname{Im} z}} + \frac{1}{n \operatorname{Im} z}\right]\right) < \frac{1}{n}. \quad (44)$$

Set $l(n, z) = n^{\frac{1}{4}}\left[\sqrt{\frac{\operatorname{Im} G_{\text{sc}}(z)}{n \operatorname{Im} z}} + \frac{1}{n \operatorname{Im} z}\right]$. Since $\lim_{n\to\infty} l(n, z) = 0$, we can find n_2 such that for all $n \geq n_2$ we have $l(n, z) < \epsilon$. Thus for all $n \geq \max(n_0, n_1, n_2)$ we have the set inclusion in the generalized Wigner ensemble

$$\{H : |\langle u, \mathcal{R}(z)v\rangle - \langle u, v\rangle G_{\text{sc}}(z)| > \epsilon\} \subset \{H : |\langle u, \mathcal{R}(z)v\rangle - \langle u, v\rangle G_{\text{sc}}(z)|| > l(n, z)\} \quad (45)$$

and therefore

$$\mathbb{P}\big(|\langle u, \mathcal{R}(z)v\rangle - \langle u, v\rangle G_{\text{sc}}(z)| > \epsilon\big) < \frac{1}{n}. \quad (46)$$

Applying this inequality to a deterministic sequence $(u^{(n)}, v^{(n)})$ on the unit sphere and taking the limit $n \to \infty$ concludes the proof for $\operatorname{Im} z > 0$.

To deal with $\operatorname{Im} z < 0$ it suffices to remark that $|\langle u, \mathcal{R}(z)v\rangle - \langle u, v\rangle G_{\text{sc}}(z)| = |\langle u, \mathcal{R}(\bar{z})v\rangle - \langle u, v\rangle G_{\text{sc}}(\bar{z})|$. Alternatively one could use a version of theorem 10 for the lower half-plane.

Consider now $z = x$ with $x \in \mathbb{R} \setminus I_\delta$ and $H \in \mathcal{S}_\delta^n$. Take a complex number $x + iy$, $0 < y \leq \frac{\epsilon}{2}|x - (2+\delta)|^2$. From the mean value theorem we have

$$\begin{aligned} &|(\langle u, \mathcal{R}(x)v\rangle - \langle u, v\rangle G_{\text{sc}}(x)) - (\langle u, \mathcal{R}(x+iy)v\rangle - \langle u, v\rangle G_{\text{sc}}(x+iy_k))| \\ &\leq |y| \sup_{y>0} |\frac{d}{dy}\langle u, \mathcal{R}(x+iy)v\rangle|. \end{aligned} \quad (47)$$

Since for $H \in \mathcal{S}_\delta^n$

$$|\frac{d}{dy}\langle u, \mathcal{R}(x+iy)v\rangle| = |\langle u, (x+iy-H)^{-2}v\rangle| \leq \frac{1}{(x-(2+\delta))^2 + y^2} \quad (48)$$

we deduce from (47) and the triangle inequality

$$\begin{aligned} |\langle u, \mathcal{R}(x)v\rangle - \langle u, v\rangle G_{\text{sc}}(x)| &\leq |\langle u, \mathcal{R}(x+iy)v\rangle - \langle u, v\rangle G_{\text{sc}}(x+iy_k)| + \frac{y}{|x-(2+\delta)|^2} \\ &\leq |\langle u, \mathcal{R}(x+iy)v\rangle - \langle u, v\rangle G_{\text{sc}}(x+iy_k)| + \frac{\epsilon}{2}. \end{aligned} \quad (49)$$

Thus for realizations $H \in \mathcal{S}_\delta^n$, the event $|\langle u, \mathcal{R}(x)v\rangle - \langle u, v\rangle G_{\text{sc}}(x)| > \epsilon$ implies the event $|\langle u, \mathcal{R}(x+iy)v\rangle - \langle u, v\rangle G_{\text{sc}}(x+iy)| \geq \frac{\epsilon}{2}$ for any $0 < y \leq \frac{\epsilon}{2}|x-(2+\delta)|^2$. In other words

$$\mathbb{P}_\delta\big(|\langle u, \mathcal{R}(x)v\rangle - \langle u, v\rangle G_{\text{sc}}(x)| > \epsilon\big) \leq \mathbb{P}_\delta\big(|\langle u, \mathcal{R}(x+iy)v\rangle - \langle u, v\rangle G_{\text{sc}}(x+iy)| \geq \frac{\epsilon}{2}\big). \quad (50)$$

By the previous results for $\operatorname{Im} z > 0$ we conclude that these probabilities tend to zero as $n \to +\infty$. ■

Proof [*Proof of proposition 6*] The proof uses a discretization argument together with the union bound. Consider the discrete set of N points on the contour $\mathcal{C}$, $z_k = \rho e^{i\theta_k}$, $\theta_k = \frac{2\pi k}{N}$, $k = 0, \dots, N-1$. First, Observe that from the union bound

$$\mathbb{P}\big(\max_{k=0,\cdots,N} |\langle u^{(n)}, \mathcal{R}(z_k) v^{(n)}\rangle - \langle u^{(n)}, v^{(n)}\rangle G_{\mathrm{sc}}(z_k)| > \epsilon\big) \\ \leq \sum_{k=0}^{N} \mathbb{P}\big(|\langle u^{(n)}, \mathcal{R}(z_k) v^{(n)}\rangle - \langle u^{(n)}, v^{(n)}\rangle G_{\mathrm{sc}}(z_k)| > \epsilon\big) \tag{51}$$

thus from proposition (4)

$$\lim_{n\to+\infty} \mathbb{P}\big(\max_{k=0,\cdots,N} |\langle u^{(n)}, \mathcal{R}(z_k) v^{(n)}\rangle - \langle u^{(n)}, v^{(n)}\rangle G_{\mathrm{sc}}(z_k)| > \epsilon\big) = 0 \tag{52}$$

Second, for any $z = \rho e^{i\theta} \in \mathcal{C}$ there exist a θ_k such that $|\theta - \theta_k| \leq \frac{1}{N}$. Applying the triangle inequality $|b| \leq |a| + |b - a|$ for $a = \langle u^{(n)}, \mathcal{R}(z) v^{(n)}\rangle - \langle u^{(n)}, G_{\mathrm{sc}}(z) v^{(n)}\rangle$ and $b = \langle u^{(n)}, \mathcal{R}(z_k) v^{(n)}\rangle - \langle u^{(n)}, G_{\mathrm{sc}}(z_k) v^{(n)}\rangle$, and the mean value theorem, we get

$$|\langle u^{(n)}, \mathcal{R}(z_k) v^{(n)}\rangle - \langle u^{(n)}, G_{\mathrm{sc}}(z_k) v^{(n)}\rangle| \leq |\langle u^{(n)}, \mathcal{R}(z) v^{(n)}\rangle - \langle u^{(n)}, \mathcal{R}_H(z) v^{(n)}\rangle| \\ + \frac{1}{N} \sup_{\theta\in[0,2\pi]} |\frac{d}{d\theta}\langle u^{(n)}, \mathcal{R}(\rho e^{i\theta}) v^{(n)}\rangle| \tag{53}$$

We can take the supremum of the right hand side over $z \in \mathcal{C}$ and then the maximum of the right hand side over $k = 0, \dots, N-1$ to deduce

$$\max_{k=0,\dots,N} |\langle u^{(n)}, \mathcal{R}(z) v^{(n)}\rangle - \langle u^{(n)}, G_{\mathrm{sc}}(z) v^{(n)}\rangle| \leq \sup_{z\in\mathcal{C}} |\langle u^{(n)}, \mathcal{R}(z_k) v^{(n)}\rangle - \langle u^{(n)}, \mathcal{R}(z_k) v^{(n)}\rangle| \\ + \frac{1}{N} \sup_{\theta\in[0,2\pi]} |\frac{d}{d\theta}\langle u^{(n)}, \mathcal{R}(\rho e^{i\theta}) v^{(n)}\rangle| \tag{54}$$

Since

$$\frac{d}{d\theta}\langle u^{(n)}, \mathcal{R}(\rho e^{i\theta}) v^{(n)}\rangle = i\rho e^{i\theta}\langle u^{(n)}, (\rho e^{i\theta} - H)^{-2} v^{(n)}\rangle \tag{55}$$

we deduce from Cauchy-Schwarz, that with probability tending to one as $n \to +\infty$

$$\frac{1}{N} \sup_{\theta\in[0,2\pi]} |\frac{d}{d\theta}\langle u^{(n)}, \mathcal{R}(\rho e^{i\theta}) v^{(n)}\rangle| \leq \frac{\rho}{N(\rho-2)^2} \tag{56}$$

Therefore taking $N > \frac{2\rho}{\epsilon(\rho-2)^2}$ we find from (52), (56) and (54)

$$\lim_{n\to+\infty} \mathbb{P}\big(\sup_{z\in\mathcal{C}} |\langle u^{(n)}, \mathcal{R}(z) v^{(n)}\rangle - \langle u^{(n)}, v^{(n)}\rangle G_{\mathrm{sc}}(z)| \geq \frac{\epsilon}{2}\big) = 0 \tag{57}$$

for any $\epsilon > 0$. This concludes the proof. ■

Appendix C. Proof of proposition 8

We assume the condition $H \in \mathcal{S}_\delta^n$ so that $\|\mathcal{R}(z)\|_{\rm op} \leq (\rho-2)^{-1}$ for all $z \in \mathcal{C} = \{z \in \mathbb{C} \mid z = \rho e^{i\theta}, \theta \in [0, 2\pi]\}$ and $\rho > 2+\delta$. The condition is relaxed at the very end.

The proof of proposition 8 is based on a Gronwall type argument. As explained in section 4 the difficulty here is that we have an integro-differential equation instead of a plain ordinary differential equation and the usual Lipshitz condition is not a priori satisfied. For this reason, given that $H \in \mathcal{S}_\delta^n$, we need preliminary bounds on $\sup_{z\in\mathcal{C}} |Q_\tau(z)|$, $\sup_{z\in\mathcal{C}} |P_\tau(z)|$, $\sup_{z\in\mathcal{C}} |R(z)|$, $\sup_{z\in\mathcal{C}} |\bar{R}(z)|$ and on $\sup_{z\in\mathcal{C}} |\bar{Q}_\tau(z)|$, $\sup_{z\in\mathcal{C}} |\bar{P}_\tau(z)|$, for $\tau \in [0, T]$. Here we do not seek the best possible bounds but rather we just need that all quantities are bounded (with high probability for the first three).

For the first *four* quantities the bound easily follows from their definition (11). By Cauchy-Schwartz we obtain that $\sup_{z\in\mathcal{C}} |Q_\tau(z)|$, $\sup_{z\in\mathcal{C}} |P_\tau(z)|$ and $\sup_{z\in\mathcal{C}} |R(z)|$ are upper bounded by $(\rho-2)^{-1}$. For $\sup_{z\in\mathcal{C}} |\bar{R}(z)|$ we can use the integral representation to get the same (loose) bound.

The remaining *two* quantities are here defined through the solution of the integro-differential equation (19) which we take as a starting point to prove a bound. In section 5 we compute exactly the combination $\bar{q}(\tau)^2 + \frac{1}{\lambda}\bar{p}_1(\tau) \equiv \frac{1}{2}\ln \hat{p}(\tau)$ and find $\hat{p}(\tau)$ given by formula (6). It can be checked that this is a continuous function for any compact time interval, so $\sup_{\tau\in[0,T]} |\bar{q}(\tau)^2 + \frac{1}{\sqrt{\lambda}}\bar{p}_1(\tau)| \leq L_*(T) < +\infty$ for any $T > 0$ (in fact one can even take L_* independent of T but we will not need this information). Then, integrating the first equation in (19) over $[0, \tau]$, using the triangle inequality, and then taking suprema, we deduce

$$\sup_{z\in\mathcal{C}} |\bar{Q}_\tau(z)| \leq \sup_{z\in\mathcal{C}} |\bar{Q}_0(z)| + \big(\frac{\rho}{\rho-2} + \frac{2\rho}{\sqrt{\lambda}} + \rho^2 L_*(T)\big) \int_0^\tau ds \sup_{z\in\mathcal{C}} |\bar{Q}_s(z)| \tag{58}$$

Iterating this inequality a standard calculation yields any $\tau \in [0, T]$

$$\sup_{z\in\mathcal{C}} |\bar{Q}_\tau(z)| \leq \sup_{z\in\mathcal{C}} |\bar{Q}_0(z)|\, e^{T\left(\frac{\rho}{\rho-2} + \frac{2\rho}{\sqrt{\lambda}} + \rho^2 L_*(T)\right)} \leq \frac{\alpha}{\rho-2} e^{T\left(\frac{\rho}{\rho-2} + \frac{2\rho}{\sqrt{\lambda}} + \rho^2 L_*(T)\right)} \tag{59}$$

where we used $\bar{Q}_0(z) = \alpha G_{\rm sc}(z)$, and for $|G_{\rm sc}(z)| \leq \frac{1}{\rho-2}$ for $z \in \mathcal{C}$. The definition of $\bar{q}(\tau)$ in terms of a contour integral implies immediately $\sup_{\tau\in[0,T]} |\bar{q}(\tau)| \leq L(T)$ where $L(T)$ is the right hand side of (59) multiplied by ρ. Now, integrating the second equation in (19) over $[0, \tau]$, using the triangle inequality, and then taking suprema again, we deduce

$$\begin{aligned}\frac{1}{2}\sup_{z\in\mathcal{C}} |\bar{P}_\tau(z)| \leq & \frac{1}{2}\sup_{z\in\mathcal{C}} |\bar{P}_0(z)| + \frac{\alpha^2\rho\tau}{(\rho-2)^2} e^{2T\left(\frac{\rho}{\rho-2} + \frac{2\rho}{\sqrt{\lambda}} + \rho^2 L_*(T)\right)} + \frac{\tau}{\sqrt{\lambda}} \\ & + \big(\frac{\rho}{\sqrt{\lambda}} + L_*(T)\big) \int_0^\tau \sup_{z\in\mathcal{C}} |\bar{P}_s(z)| \end{aligned}\tag{60}$$

Again a standard calculation yields (using the initial condition $\bar{P}_0(z) = G_{\rm sc}(z)$)

$$\sup_{z\in\mathcal{C}} |\bar{P}_\tau(z)| \leq \left(\frac{1}{\rho-2} + \frac{2\alpha^2\rho T}{(\rho-2)^2} e^{2T\left(\frac{\rho}{\rho-2} + \frac{2\rho}{\sqrt{\lambda}} + \rho^2 L_*(T)\right)} + \frac{2\tau}{\sqrt{\lambda}}\right) e^{T(\frac{\rho}{\sqrt{\lambda}} + L_*(T))} \tag{61}$$

Note that this implies the bound $\sup_{\tau\in[0,T]} |\bar{p}_1(\tau)| \leq L_1(T)$ where $L_1(T)$ is the right hand side of (61) multiplied by ρ^2.

We now have all the elements to adapt a Gronwall type argument.

Proof [*Proof of proposition 8*] We start by deriving preliminary bounds We set $Q_\tau(z) - \bar{Q}_\tau(z) = \Delta_\tau^Q(z)$, $P_\tau(z) - \bar{P}_\tau(z) = \Delta_\tau^P(z)$, $R(z) - \bar{R}(z) = \Delta^R(z)$, $q(\tau) - \bar{q}(\tau) = \delta^q(\tau)$, $p_1(\tau) - \bar{p}_1(\tau) = \delta^{p_1}(\tau)$. Note for later use that all the $\sup_{z\in\mathcal{C}} |\cdot|$ of these differences are bounded by some finite positive constant depending only on ρ, α, λ, T. Taking the difference of (19) and (13) we find after a bit of algebra

$$\begin{aligned}\frac{d}{d\tau}\Delta_\tau^Q(z) = &\,\delta^q(\tau)\Delta^R(z) + \delta^q(\tau)\bar{R}(z) + \bar{q}(\tau)\Delta^R(z) + \frac{1}{\sqrt{\lambda}}\big(z\Delta_\tau^Q(z) + \delta^q(\tau)\big)\\ &- (q(\tau)+\bar{q}(\tau))\delta^q(\tau)\Delta_\tau^Q(z) - (q(\tau)+\bar{q}(\tau))\delta^q(\tau)\bar{Q}_\tau(z) - \bar{q}(\tau)^2\Delta_\tau^Q(z)\\ &- \frac{1}{\sqrt{\lambda}}\big(\delta^{p_1}(\tau)\Delta_\tau^Q(z) - \delta^{p_1}(\tau)\bar{Q}_\tau(z) - \bar{p}_1(\tau)\Delta_\tau^Q(z)\big)\end{aligned} \tag{62}$$

and

$$\begin{aligned}\frac{d}{d\tau}\Delta_\tau^P(z) = &\,\delta^q(\tau)\Delta_\tau^Q(z) + \delta^q(\tau)\bar{Q}_\tau(z) + \bar{q}(\tau)\Delta_\tau^Q(z) + \frac{1}{\sqrt{\lambda}}z\Delta_\tau^P(z)\\ &- (q(\tau)+\bar{q}(\tau))\delta^q(\tau)\Delta_\tau^P(z) - (q(\tau)+\bar{q}(\tau))\delta^q(\tau)\bar{P}_\tau(z) - \bar{q}(\tau)^2\Delta_\tau^P(z)\\ &- \frac{1}{\sqrt{\lambda}}\big(\delta^{p_1}(\tau)\Delta_\tau^P(z) - \delta^{p_1}(\tau)\bar{P}_\tau(z) - \bar{p}_1(\tau)\Delta_\tau^P(z)\big)\end{aligned} \tag{63}$$

After integrating the above equations over the interval $[0,\tau]$, using the triangle inequality, and the inequalities $|\delta^q(\tau)| \le \rho\sup_{z\in\mathcal{C}}|\Delta_\tau^Q(z)|$, $|\delta^{p_1}(\tau)| \le \rho^2\sup_{z\in\mathcal{C}}|\Delta_\tau^P(z)|$, $|q(\tau)| \le 1$, $\sup_{\tau\in[0,T]}|\bar{q}(\tau)| < L(T)$, $\sup_{\tau\in[0,T]}|\bar{p}_1(\tau)| < L_1(T)$, we deduce (with $L = \max(L(T), L_1(T))$)

$$\begin{aligned}\sup_{z\in\mathcal{C}}|\Delta_\tau^Q(z)| \le &\sup_{z\in\mathcal{C}}|\Delta_0^Q(z)| + \rho\sup_{z\in\mathcal{C}}|\Delta^R(z)|\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)| + \rho\sup_{z\in\mathcal{C}}|\bar{R}(z)|\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)|\\ &+ L\tau\sup_{z\in\mathcal{C}}|\Delta^R(z)| + \frac{2\rho}{\sqrt{\lambda}}\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)| + (1+L)\rho\int_0^\tau ds\,(\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)|)^2\\ &+ (1+L)\rho\int_0^\tau ds\,(\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)|)^2\sup_{z\in\mathcal{C}}|\bar{Q}_s(z)| + L^2\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)|\\ &+ \frac{\rho^2}{\sqrt{\lambda}}\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^P(z)|\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)| + \frac{\rho^2}{\sqrt{\lambda}}\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^P(z)|\sup_{z\in\mathcal{C}}|Q_s(z)|\\ &+ \frac{L}{\sqrt{\lambda}}\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)|\end{aligned} \tag{64}$$

and

$$\begin{aligned}\sup_{z\in\mathcal{C}}|\Delta_\tau^P(z)| \le &\sup_{z\in\mathcal{C}}|\Delta_0^P(z)| + \rho\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)|^2 + \rho\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)|\sup_{z\in\mathcal{C}}|\bar{Q}_s(z)|\\ &+ L\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)| + \frac{\rho}{\sqrt{\lambda}}\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^P(z)| + (1+L)\rho\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)|\sup_{z\in\mathcal{C}}|\Delta_s^P(z)|\\ &+ (1+L)\rho\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^Q(z)|\sup_{z\in\mathcal{C}}|\bar{P}_s(z)| + L^2\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^P(z)|\\ &+ \frac{\rho^2}{\sqrt{\lambda}}\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^P(z)|^2 + \frac{\rho^2}{\sqrt{\lambda}}\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^P(z)|\sup_{z\in\mathcal{C}}|\bar{P}_s(z)| + \frac{L}{\sqrt{\lambda}}\int_0^\tau ds\,\sup_{z\in\mathcal{C}}|\Delta_s^P(z)|\end{aligned} \tag{65}$$

Now, using (59) and (61) we can "linearize" the right hand side to obtain two inequalities of the form (where $C(\rho, \alpha, \lambda, T)$ is a suitable constant)

$$\sup_{z\in\mathcal{C}} |\Delta_\tau^Q(z)| \leq \sup_{z\in\mathcal{C}} |\Delta_0^Q(z)| + L\tau \sup_{z\in\mathcal{C}} \Delta^R(z) + C(\rho,\alpha,\lambda,T) \int_0^\tau ds \left\{\sup_{z\in\mathcal{C}} |\Delta_s^Q(z)| + \sup_{z\in\mathcal{C}} |\Delta_s^P(z)|\right\} \tag{66}$$

and

$$\sup_{z\in\mathcal{C}} |\Delta_\tau^P(z)| \leq \sup_{z\in\mathcal{C}} |\Delta_0^P(z)| + C(\rho,\alpha,\lambda,T) \int_0^\tau ds \left\{\sup_{z\in\mathcal{C}} |\Delta_s^Q(z)| + \sup_{z\in\mathcal{C}} |\Delta_s^P(z)|\right\} \tag{67}$$

Summing (66) and (67) and iterating the resulting integral inequality we deduce

$$\sup_{z\in\mathcal{C}} |\Delta_\tau^Q(z)| + \sup_{z\in\mathcal{C}} |\Delta_\tau^P(z)| \leq \left\{\sup_{z\in\mathcal{C}} |\Delta_0^Q(z)| + \sup_{z\in\mathcal{C}} |\Delta_0^P(z)| + LT \sup_{z\in\mathcal{C}} \Delta^R(z)\right\} e^{2TC(\rho,\alpha,\lambda,T)} \tag{68}$$

By corollary 7 we conclude that for $\tau \in [0,T]$ $\sup_{z\in\mathcal{C}} |\Delta_\tau^Q(z)|$ and $\sup_{z\in\mathcal{C}} |\Delta_\tau^P(z)|$ converge in $\mathbb{P}_\delta$-probability to zero.

Finally, we can look at the overlaps. Observe that $|q(\tau) - \bar{q}(\tau)| = |\int_{\mathcal{C}} \frac{dz}{2\pi i} \Delta_\tau^Q(z)|$ so $|q(\tau) - \bar{q}(\tau)| \leq \rho \sup_{z\in\mathcal{C}} |\Delta_\tau^Q(z)|$ and $|p_1(\tau) - \bar{p}_1(\tau)| = |\int_{\mathcal{C}} \frac{dz}{2\pi i} z \Delta_\tau^P(z)| \leq \rho^2 \sup_{z\in\mathcal{C}} |\Delta_\tau^P(z)|$. Therefore $|q(\tau) - \bar{q}(\tau)|$ and $|p_1(\tau) - \bar{p}_1(\tau)|$ converge with $\mathbb{P}_\delta$-probability to 0. But since $\lim_{n\to+\infty} \mathbb{P}(H \in \mathcal{S}_\delta^n) = 1$ it is easy to see (by the law of total probability) that $|q(\tau) - \bar{q}(\tau)|$ and $|p_1(\tau) - \bar{p}_1(\tau)|$ also converge with $\mathbb{P}$-probability to 0.

■

Appendix D. Laplace Transform applicability

Laplace transform can be applied appropriately with the condition of deriving a bound of the form $e^{a\tau}$ with $a > 0$ for the terms $\hat{q}(\tau), \hat{p}(\tau)$ first, and $\hat{Q}_\tau(z), \hat{P}_\tau(z)$ secondly. For $\hat{q}(\tau)$ because $M_\lambda(s)$ is positive on $[0,\tau]$ and $\alpha\hat{q}(\tau)$ remains positive at all time, we derive the bound

$$0 \leq |\hat{q}(\tau)| \leq e^{(1+\frac{1}{\lambda})\tau} \tag{69}$$

Next we find a bound for $M_\lambda(\tau)$ with the definition (3)

$$|M_\lambda(\tau)| \leq 2\int_0^\pi \frac{\mathrm{d}\theta}{\pi} |\sin(\theta)^2| e^{\cos(\theta)\frac{2\tau}{\sqrt{\lambda}}} \leq 2e^{\frac{2\tau}{\sqrt{\lambda}}} \tag{70}$$

For $\hat{p}(\tau)$ using the previous bound and $|\alpha| \leq 1$

$$|\hat{p}(\tau)| \leq |M_\lambda(2\tau)| + 2\int_0^\tau |\hat{q}(s)||M_\lambda(2\tau - s)|\mathrm{d}s + \int_0^\tau \int_0^\tau |\hat{q}(u)\hat{q}(v)||M_\lambda(2\tau - u - v)|\mathrm{d}u\mathrm{d}v \tag{71}$$

$$\leq 2e^{\frac{4\tau}{\sqrt{\lambda}}} + 4\int_0^\tau e^{(1+\frac{1}{\lambda})s + \frac{2}{\sqrt{\lambda}}(2\tau - s)}\mathrm{d}s + 2\int_0^\tau \int_0^\tau e^{(1+\frac{1}{\lambda})(u+v) + \frac{2}{\sqrt{\lambda}}(2\tau - u - v)}\mathrm{d}u\mathrm{d}v \tag{72}$$

Hence

$$\frac{1}{2}|\hat{p}(\tau)|e^{-\frac{4\tau}{\sqrt{\lambda}}} \leq 1 + 2\int_0^\tau e^{(1-\frac{1}{\sqrt{\lambda}})^2 s}\mathrm{d}s + \int_0^\tau\int_0^\tau e^{(1-\frac{1}{\sqrt{\lambda}})^2 u}e^{(1-\frac{1}{\sqrt{\lambda}})^2 v}\mathrm{d}u\mathrm{d}v \tag{73}$$

$$\leq \left(1 + \int_0^\tau e^{(1-\frac{1}{\sqrt{\lambda}})^2 s}\mathrm{d}s\right)^2 \tag{74}$$

$$\leq \left(1 + \frac{e^{(1-\frac{1}{\sqrt{\lambda}})^2\tau} - 1}{(1-\frac{1}{\sqrt{\lambda}})^2}\right)^2 \tag{75}$$

$$\leq e^{2(1-\frac{1}{\sqrt{\lambda}})^2\tau}\left(e^{-(1-\frac{1}{\sqrt{\lambda}})^2\tau} + \frac{1 - e^{-(1-\frac{1}{\sqrt{\lambda}})^2\tau}}{(1-\frac{1}{\sqrt{\lambda}})^2}\right)^2 \tag{76}$$

$$\leq e^{2(1-\frac{1}{\sqrt{\lambda}})^2\tau}\left(1 + \frac{1}{(1-\frac{1}{\sqrt{\lambda}})^2}\right)^2 \tag{77}$$

Hence with $C_\lambda = 2\left(1 + \frac{1}{(1-\frac{1}{\sqrt{\lambda}})^2}\right)^2$ we have the exponential bound $|\hat{p}(\tau)| \leq C_\lambda e^{2(1+\frac{1}{\lambda})\tau}$. Now going back to equations (21) we have the system

$$\begin{cases} \frac{d}{d\tau}e^{-\frac{z\tau}{\sqrt{\lambda}}}\hat{Q}_\tau(z) = e^{-\frac{z\tau}{\sqrt{\lambda}}}\hat{q}(\tau)\left(\bar{R}(z) + \frac{1}{\sqrt{\lambda}}\right) \\ \frac{1}{2}\frac{d}{d\tau}e^{\frac{-2z\tau}{\sqrt{\lambda}}}\hat{P}_\tau(z) = e^{\frac{-2z\tau}{\sqrt{\lambda}}}\hat{q}(\tau)\hat{Q}_\tau(z) + \frac{1}{\sqrt{\lambda}}\hat{p}(\tau)e^{\frac{-2z\tau}{\sqrt{\lambda}}} \end{cases} \tag{78}$$

Hence integrating over $[0,\tau]$ provides

$$\begin{cases} \hat{Q}_\tau(z) = \bar{Q}_0(z)e^{\frac{z\tau}{\sqrt{\lambda}}} + \left(\bar{R}(z) + \frac{1}{\sqrt{\lambda}}\right)\int_0^\tau \mathrm{d}s e^{\frac{z(\tau-s)}{\sqrt{\lambda}}}\hat{q}(s) \\ \hat{P}_\tau(z) = e^{\frac{2z\tau}{\sqrt{\lambda}}}\bar{P}_0(z) + 2\int_0^\tau \mathrm{d}s e^{\frac{2z(\tau-s)}{\sqrt{\lambda}}}\hat{q}(s)\hat{Q}_s(z) + \frac{2}{\sqrt{\lambda}}\int_0^\tau \mathrm{d}s\hat{p}(s)e^{\frac{2z(\tau-s)}{\sqrt{\lambda}}} \end{cases} \tag{79}$$

Notice again that we have $|G_{\mathrm{sc}}(z)| \leq \frac{1}{\rho-2}$ for $z \in \mathcal{C} = \{\, z \in \mathbb{C}|\ z = \rho e^{i\theta}, \theta \in [0, 2\pi]\}$ where $\rho > 2$. For $\hat{Q}_\tau(z)$ we find

$$|\hat{Q}_\tau(z)| \leq \frac{|\alpha|}{\rho-2}e^{\frac{\mathrm{Re}(z)\tau}{\sqrt{\lambda}}} + \left(\frac{1}{\rho-2} + \frac{1}{\sqrt{\lambda}}\right)\int_0^\tau \mathrm{d}s e^{\frac{\mathrm{Re}(z)(\tau-s)}{\sqrt{\lambda}}}e^{(1+\frac{1}{\lambda})s} \tag{80}$$

$$\leq e^{\frac{\rho\tau}{\sqrt{\lambda}}}\left(\frac{1}{\rho-2} + \left(\frac{1}{\rho-2} + \frac{1}{\sqrt{\lambda}}\right)\int_0^\tau \mathrm{d}s e^{(\frac{1}{\lambda} - \frac{\rho}{\sqrt{\lambda}})s + s}\right) \tag{81}$$

$$\leq e^{\frac{\rho\tau}{\sqrt{\lambda}}}\left(\frac{1}{\rho-2} + \left(\frac{1}{\rho-2} + \frac{1}{\sqrt{\lambda}}\right)e^{\frac{\tau}{\lambda}}\int_0^\tau \mathrm{d}s e^{s}\right) \tag{82}$$

$$\leq e^{\frac{\rho\tau}{\sqrt{\lambda}}}\left(\frac{1}{\rho-2} + \left(\frac{1}{\rho-2} + \frac{1}{\sqrt{\lambda}}\right)e^{(\frac{1}{\lambda}+1)\tau}(1 - e^{-\tau})\right) \tag{83}$$

$$\leq e^{(1+\frac{\rho}{\sqrt{\lambda}}+\frac{1}{\lambda})\tau}\left(\frac{2}{\rho-2} + \frac{1}{\sqrt{\lambda}}\right) \tag{84}$$

With $C'_{\rho,\lambda} = \frac{2}{\rho-2} + \frac{1}{\sqrt{\lambda}}$ we thus have $|\hat{Q}_\tau(z)| \leq C'_{\rho,\lambda} e^{(1+\frac{\rho}{\sqrt{\lambda}}+\frac{1}{\lambda})\tau}$ for any $z \in \mathcal{C}$. Similarly for $\hat{P}_\tau(z)$:

$$|\hat{P}_\tau(z)| \leq e^{\frac{2\rho\tau}{\sqrt{\lambda}}} \left(\frac{1}{\rho-2} + 2C'_{\rho,\lambda} \int_0^\tau e^{\frac{-2\rho s}{\sqrt{\lambda}} + (1+\frac{1}{\lambda})s + (1+\frac{\rho}{\sqrt{\lambda}}+\frac{1}{\lambda})s} \mathrm{d}s + \frac{2C_\lambda}{\sqrt{\lambda}} \int_0^\tau e^{\frac{-2\rho s}{\sqrt{\lambda}} + 2(1-\frac{1}{\sqrt{\lambda}})^2 s} \mathrm{d}s \right) \tag{85}$$

$$\leq e^{\frac{2\rho\tau}{\sqrt{\lambda}}} \left(\frac{1}{\rho-2} + 2C'_{\rho,\lambda} e^{2(1+\frac{1}{\lambda})\tau} + \frac{2}{\sqrt{\lambda}} e^{2(1+\frac{1}{\lambda})\tau} \right) \tag{86}$$

$$\leq e^{2(1+\frac{\rho}{\sqrt{\lambda}}+\frac{1}{\lambda})\tau} \left(\frac{1}{\rho-2} + 2C'_{\rho,\lambda} + \frac{2}{\sqrt{\lambda}} C_\lambda \right) \tag{87}$$

Hence with $C''_{\lambda,\rho} = \frac{1}{\rho-2} + 2C'_{\rho,\lambda} + \frac{2}{\sqrt{\lambda}} C_\lambda$ we find $|\hat{P}_\tau(z)| \leq C''_{\rho,\lambda} e^{2(1+\frac{\rho}{\sqrt{\lambda}}+\frac{1}{\lambda})\tau}$ for any $z \in \mathcal{C}$.

Appendix E. Enforcing the spherical constraint in gradient dynamics

The second term in equation (2) enforces the spherical constraint $\theta_t \in \mathbb{S}^{n-1}(\sqrt{n})$ at all times. This is well known but we briefly recall how to derive it for completeness. Since the n dimensional sphere is embedded in $\mathbb{R}^n$ the covariant gradient D_θ can be obtained by projecting the usual gradient ∇_θ on a tangent plane. This projection is obtained by subtracting the component along a radius of the sphere, i.e., $\frac{\theta}{\sqrt{n}} \langle \frac{\theta}{\sqrt{n}}, \nabla_\theta \mathcal{H}(\theta) \rangle$. Therefore gradient descent reads

$$\frac{d\theta_t}{dt} = \eta D_\theta \mathcal{H}(\theta_t) = \eta \big(\nabla_\theta \mathcal{H}(\theta_t) - \frac{\theta_t}{n} \langle \theta_t, \nabla_\theta \mathcal{H}(\theta_t) \rangle \big). \tag{88}$$

It is easily checked that $\frac{d\|\theta_t\|_2^2}{dt} = 0$ and since $\theta_0 \in \mathbb{S}^{n-1}(\sqrt{n})$ we have $\theta_t \in \mathbb{S}^{n-1}(\sqrt{n})$ for all times. Indeed

$$\frac{d\|\theta_t\|_2^2}{dt} = 2\langle \theta_t, \frac{d\theta_t}{dt} \rangle = 2\eta \big(\langle \theta_t, \nabla_\theta \mathcal{H}(\theta_t) \rangle - \frac{\langle \theta_t, \theta_t \rangle}{n} \langle \theta_t, \nabla_\theta \mathcal{H}(\theta_t) \rangle \big) = 0. \tag{89}$$

Appendix F. Strict saddle property

We say that the strict saddle property is satisfied if the critical points of the cost are *strict* saddles or minima (a strict saddle has by definition at least one strictly negative eigenvalue of the Hessian). It is known from Lee et al. (2016) that for a cost satisfying the strict saddle property, gradient descent with small enough discrete time steps converges to a minimum, almost surely with respect to the initial condition. In the present context (as shown below) the critical points are given by the eigenvectors of $A \equiv \frac{\sqrt{\lambda}}{n} Y = \frac{\sqrt{\lambda}}{n} \theta^* \theta^{*T} + \frac{1}{\sqrt{n}} \xi$ - call them $v_i \in \mathcal{S}^{n-1}(\sqrt{n})$, $i = 1, \ldots, n$ - and the Hessian at v_i is proportional to $\alpha_i I - A$ where α_i is the corresponding eigenvalue. For a random $n \times n$ matrix and *fixed* λ the spectrum is almost surely non-degenerate,[3] i.e., $\alpha_1 < \alpha_2 < \ldots < \alpha_n$, so the strict saddle property is almost surely satisfied. Moreover the top eigenvector v_n has positive definite Hessian and is a minimum, while for the other ones are strict saddles with non-zero positive and negative eigenvalues. Now, for $\lambda > 1$ we know, that for n large enough with high probability, $\{\alpha_1 < \cdots < \alpha_{n-1}\} \subset [-2, 2]$, $\alpha_n \approx \sqrt{\lambda} + 1/\sqrt{\lambda} > 2$ and $n^{-1}|\langle \theta_*, v_n \rangle\rangle| \approx \sqrt{1 - 1/\lambda}$ (where

3. However for a fixed realization when λ varies we can have eigenvalue crossings.

$a \approx b$ means $|a-b| = o_n(1)$) Péché (2004); Féral and Péché (2006). This explains that for $\lambda > 1$ gradient descent with a small enough discrete time steps will converge to v_n and the overlap approach $\pm\sqrt{1-1/\lambda}$.

The critical points on the sphere $\mathcal{S}^{n-1}(\sqrt{n})$ satisfy $D_\theta \mathcal{H}(\theta) = 0$ where $D_\theta = (1 - \frac{1}{n}\theta\theta^T)\nabla_\theta$ is the covariant derivative. We have

$$D_\theta \mathcal{H}(\theta) \propto \frac{1}{n}\langle \theta, A\theta\rangle\theta - A\theta = 0 \tag{90}$$

and has n solutions $\theta = v_i$, $i = 1, \ldots, n$. The Hessian matrix on the sphere is (up to a positive prefactor)

$$D_\theta D_\theta^T \mathcal{H}(\theta) \propto (1 - \frac{1}{n}\theta\theta^T)\big(\frac{1}{n}\langle\theta, A\theta\rangle I - A\big) \tag{91}$$

and for each critical point $\theta = v_i$ we find $D_\theta D_\theta^T \mathcal{R}(v_i) \propto \frac{1}{n^2}(\alpha_i I - A)$. This has $n-1$ eigenvectors v_j, $j \neq i$ (perpendicular to v_i and tangent to the sphere) with eigenvalues $\alpha_i - \alpha_j$, $j \neq i$, and one eigenvector v_i with 0 eigenvalue. For fixed λ there is no degeneracy $\alpha_1 < \alpha_2 < \ldots < \alpha_n$, almost surely and v_n is a minimum while v_j, $j \neq n$ are strict saddles.

Appendix G. Analysis of the stationary equation

The stationary equations corresponding to (13) are given by setting the time derivatives on the left hand side to zero.

$$\begin{cases} \bar{q}^\infty \left(\bar{R}(z) + \frac{1}{\sqrt{\lambda}}\right) + \left(\frac{z}{\sqrt{\lambda}} - (\bar{q}^\infty)^2 - \frac{1}{\sqrt{\lambda}}\bar{p}_1^\infty\right)\bar{Q}_\infty(z) = 0 \\ \bar{q}^\infty \bar{Q}_\infty(z) + \frac{1}{\sqrt{\lambda}} + \left(\frac{z}{\sqrt{\lambda}} - (\bar{q}^\infty)^2 - \frac{1}{\sqrt{\lambda}}\bar{p}_1^\infty\right)\bar{P}_\infty(z) = 0 \end{cases} \tag{92}$$

where $\bar{q}^\infty \equiv -\int_{\mathcal{C}} \frac{dz}{2\pi i}\,\bar{Q}_\infty(z)$, $\bar{p}_1^\infty \equiv -\int_{\mathcal{C}} \frac{dz}{2\pi i}\, z\bar{P}_\infty(z)$, $\bar{R}(z) = G_{\rm sc}(z)$, and $\mathcal{C} = \{z \in \mathbb{C} \mid z = \rho e^{i\theta}, \theta \in [0, 2\pi]\}$, $\rho > 2$. Here we show how to derive all possible solutions of these equations. One expects that the set of solutions contains the limiting solution for $\tau \to +\infty$ and we check that this is indeed the case.

From (92) we get

$$\begin{cases} \bar{Q}_\infty(z) = \bar{q}^\infty \frac{\sqrt{\lambda}G_{\rm sc}(z)+1}{\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty - z} \\ \bar{P}_\infty(z) = (\bar{q}^\infty)^2\sqrt{\lambda}\frac{\sqrt{\lambda}G_{\rm sc}(z)+1}{\left(\sqrt{\lambda}(\bar{q}^\infty)^2+\bar{p}_1^\infty - z\right)^2} + \frac{1}{\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty - z} \end{cases} \tag{93}$$

Let us first assume that $|\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty| \leq 2$. We integrate the second equation over the contour $\mathcal{C}$. One can show that integral of the first term on the right hand side vanishes. Thus we find the condition by $\bar{p}_1^\infty = \sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty$ which implies $\bar{q}^\infty = 0$. This implies in turn that $Q_\infty(z) = 0$, $P_\infty(z) = (\bar{p}_1^\infty - z)^{-1}$ and $|p_1^\infty| \leq 2$.

Now assume that $|\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty| > 2$. Integrating the first equation of (93) over $\mathcal{C}$ we find

$$\bar{q}^\infty = \sqrt{\lambda}\bar{q}^\infty \int_{z\in\mathcal{C}} \frac{dz}{2\pi i}\frac{G_{\rm sc}(z)}{z - (\sqrt{\lambda}(q^\infty)^2 + \bar{p}_1^\infty)} \tag{94}$$

The solution $\bar{q}^\infty = 0$ is again a possibility $Q_\infty(z) = 0$, $P_\infty(z) = (\bar{p}_1^\infty - z)^{-1}$ and $|\bar{p}_1^\infty| > 2$.

Now assume that $\bar{q}^\infty \neq 0$ (and still $|\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty| > 2$). Computing the contour integral we find the equation $1 = -\sqrt{\lambda} G_{\rm sc}(\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty)$ which provides a solution and a condition

$$\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty = \frac{1}{\sqrt{\lambda}} + \sqrt{\lambda} \tag{95}$$

$$\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty \geq \frac{2}{\sqrt{\lambda}} \tag{96}$$

Notice that the initial condition $\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty > 2$ is satisfied for all $\lambda \neq 1$, while $\frac{1}{\sqrt{\lambda}} + \sqrt{\lambda} \geq \frac{2}{\sqrt{\lambda}}$ is equivalent to $\lambda \geq 1$. So a solution can only exist when $\lambda > 1$. Integrating the second equation in (93) over $\mathcal{C}$ we find

$$-\frac{1}{\lambda} = (\bar{q}^\infty)^2 \int_{z\in\mathcal{C}} \frac{dz}{2\pi i} \frac{G_{\rm sc}(z)}{\Big(z - (\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty)\Big)^2} = -(\bar{q}^\infty)^2 \frac{dG_{\rm sc}(z)}{dz}\Big|_{\sqrt{\lambda}(\bar{q}^\infty)^2 + \bar{p}_1^\infty} \tag{97}$$

Then using the explicit expression of $G_{\rm sc}(z)$ we find that $(\bar{q}^\infty)^2 = 1 - \frac{1}{\lambda}$, with $\lambda > 1$. Furthermore we have from (93) and (95)

$$\begin{cases} \bar{Q}_\infty(z) = (1 - \frac{1}{\lambda}) \frac{\sqrt{\lambda} G_{\rm sc}(z) + 1}{\sqrt{\lambda} + \frac{1}{\sqrt{\lambda}} - z} \\ \bar{P}_\infty(z) = (1 - \frac{1}{\lambda})\sqrt{\lambda} \frac{\sqrt{\lambda} G_{\rm sc}(z) + 1}{\left(\sqrt{\lambda} + \frac{1}{\sqrt{\lambda}} - z\right)^2} + \frac{1}{\sqrt{\lambda} + \frac{1}{\sqrt{\lambda}} - z} \end{cases} \tag{98}$$

Note that multiplying the second equation in (98) by z and integrating over $\mathcal{C}$ yields $\bar{p}_1^\infty = \frac{2}{\sqrt{\lambda}}$. This is consistent with (95).

We conclude by noting that the solutions that are attainable from the time evolution when $\lambda > 1$ are $\{\bar{q}^\infty = 0, |\bar{p}_1^\infty| \leq 2\}$ and $\{\bar{q}^\infty = \pm\sqrt{1 - \frac{1}{\lambda}}, \bar{p}_1^\infty = \frac{2}{\lambda}\}$. The first one is "attained" from an initial condition with $\alpha = \frac{1}{n}\langle \theta^*, \theta_0 \rangle = 0$. In this case gradient descent "does not start" and $\bar{q}^\infty = \bar{q}(0) = 0$, $\bar{p}_1^\infty = \bar{p}_1(0) = \frac{1}{n}\langle \theta_0, H\theta_0 \rangle$ and $\bar{p}_1(0) \leq 2$ with high probability. The other two solutions correspond to the initial conditions $\alpha = \frac{1}{n}\langle \theta^*, \theta_0 \rangle$ with $\alpha > 0$ and $\alpha < 0$. When $\lambda \leq 1$, there is only one possible solution $\{\bar{q}^\infty = 0, |\bar{p}_1^\infty| \leq 2\}$.

Appendix H. Intermediate identities

We derive a number of identities requiring interchange of integrals.

A) Derivation of (29). To prove (29) we start with (23) in the form

$$\mathcal{L}\hat{Q}_p(z) = \alpha G_{\rm sc}(z) \mathcal{L}(e^{\frac{zt}{\sqrt{\lambda}}})(p) + \mathcal{L}\hat{q}(p) \mathcal{L}(e^{\frac{zt}{\sqrt{\lambda}}})(p)\big(G_{\rm sc}(z) + \frac{1}{\sqrt{\lambda}}\big) \tag{99}$$

and invert it back to the time domain

$$\hat{Q}_\tau(z) = \alpha G_{\rm sc}(z) e^{\frac{z\tau}{\sqrt{\lambda}}} + G_{\rm sc}(z) \int_0^\tau ds \hat{q}(s) e^{\frac{z(\tau-s)}{\sqrt{\lambda}}} + \frac{1}{\sqrt{\lambda}} \int_0^\tau ds \hat{q}(s) e^{\frac{z(\tau-s)}{\sqrt{\lambda}}}. \tag{100}$$

So this generating function is entirely known. Now we multiply this equation by $e^{\frac{z(\tau-u)}{\sqrt{\lambda}}}$ and integrate along $\mathcal{C}$. It is easy to see that, by Fubini's theorem, for the last term on the right hand side, the contour integral and the s-integral can be exchanged. Therefore the contour integral of the last term on the right hand side vanishes because $e^{\frac{z(2\tau-s-u)}{\sqrt{\lambda}}}$ is holomorphic in the whole complex plane. For the other two terms on the right hand side we use the semi-circle law representation of $G_{\text{sc}}(z)$ to obtain (see below for details) to obtain

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i} G_{\text{sc}}(z) e^{\frac{z(2\tau-u)}{\sqrt{\lambda}}} = -M_\lambda(2\tau - u) \tag{101}$$

and

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i} G_{\text{sc}}(z) \int_0^\tau ds \hat{q}(s) e^{\frac{z(2\tau-s-u)}{\sqrt{\lambda}}} = -\int_0^\tau ds \hat{q}(s) M_\lambda(2\tau - s - u). \tag{102}$$

Putting together (101), (102) and (100) we obtain the claimed identity (29).

B) Derivation of (101). From the semi-circle law representation of G_{sc}

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i} G_{\text{sc}}(z) e^{\frac{z(2\tau-u)}{\sqrt{\lambda}}} = \oint_{\mathcal{C}} \frac{dz}{2\pi i} \int_{-2}^{2} ds \frac{\mu_{\text{sc}}(s)}{s-z} e^{\frac{z(2\tau-u)}{\sqrt{\lambda}}} \tag{103}$$

It is easy to see that Fubini's theorem can be applied to interchange the integrals. Indeed the contour integral over $\mathcal{C}$ can be parametrized so that we then have two integrals with bounded functions over bounded intervals. So

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i} G_{\text{sc}}(z) e^{\frac{z(2\tau-u)}{\sqrt{\lambda}}} = \int_{-2}^{2} ds \mu_{\text{sc}}(s) \oint_{\mathcal{C}} \frac{dz}{2\pi i} \frac{e^{\frac{z(2\tau-u)}{\sqrt{\lambda}}}}{s-z} = -\int_{-2}^{2} ds \mu_{\text{sc}}(s) e^{\frac{s(2\tau-u)}{\sqrt{\lambda}}} = -M_\lambda(2\tau-u) \tag{104}$$

C) Derivation of (102). We proceed similarly. First,

$$\oint_{\mathcal{C}} \frac{dz}{2\pi i} G_{\text{sc}}(z) \int_0^\tau ds \hat{q}(s) e^{\frac{z(2\tau-s-u)}{\sqrt{\lambda}}} = \oint_{\mathcal{C}} \frac{dz}{2\pi i} \int_{-2}^{2} dx \int_0^\tau ds \mu_{\text{sc}}(x) \hat{q}(s) \frac{e^{\frac{z(2\tau-s-u)}{\sqrt{\lambda}}}}{x-z} \tag{105}$$

Again, it is clear that the contour integral can be parametrized so that we all integrals are over bounded intervals and all functions are bounded, so that Fubini's theorem applies. Thus

$$\begin{aligned}\oint_{\mathcal{C}} \frac{dz}{2\pi i} G_{\text{sc}}(z) \int_0^\tau ds \hat{q}(s) e^{\frac{z(2\tau-s-u)}{\sqrt{\lambda}}} &= \int_0^\tau ds \hat{q}(s) \int_{-2}^{2} dx \mu_{\text{sc}}(x) \oint_{\mathcal{C}} \frac{dz}{2\pi i} \frac{e^{\frac{z(2\tau-s-u)}{\sqrt{\lambda}}}}{x-z} \\ &= -\int_0^\tau ds \hat{q}(s) M_\lambda(2\tau - s - u)\end{aligned} \tag{106}$$

D) Derivation of (36). Again, using Fubini and then Cauchy's theorem,

$$\begin{aligned}
\oint_{\mathcal{C}} \frac{dz}{2\pi i} G_{\rm sc}(z) \int_0^{+\infty} d\tau e^{-p\tau} e^{\frac{z\tau}{\sqrt{\lambda}}} &= \int_0^{+\infty} d\tau e^{-p\tau} \oint_{\Gamma'} \frac{dz}{2\pi i} G_{\rm sc}(z) e^{\frac{z\tau}{\sqrt{\lambda}}} \\
&= \int_0^{+\infty} d\tau e^{-p\tau} \oint_{\mathcal{C}} \frac{dz}{2\pi i} e^{\frac{z\tau}{\sqrt{\lambda}}} \int_{-2}^{2} ds \frac{\mu_{\rm sc}(s)}{s-x} \\
&= \int_0^{+\infty} d\tau e^{-p\tau} \int_{-2}^{2} ds \mu_{\rm sc}(s) \oint \frac{dz}{2\pi i} \frac{e^{\frac{z\tau}{\sqrt{\lambda}}}}{s-z} \\
&= -\int_0^{+\infty} d\tau e^{-p\tau} \int_{-2}^{2} ds \mu_{\rm sc}(s) e^{\frac{s\tau}{\sqrt{\lambda}}} \\
&= -\int_0^{+\infty} d\tau e^{-p\tau} M_\lambda(\tau) \qquad (107)
\end{aligned}$$

Appendix I. Asymptotic analysis of $\bar{q}$

I.1. limit when $\lambda > 1$

We deduce the limiting behavior for $\lambda > 1$. The next order correction is given in I.3. Rewriting the first term from theorem 1, we have for $\tau \in \mathbb{R}^+$

$$e^{-(1+\frac{1}{\lambda})\tau} \hat{q}(\tau) = \alpha \left[1 - \frac{1}{\lambda} \int_0^{\tau} e^{-(1+\frac{1}{\lambda})s} M_\lambda(s) \mathrm{d}s \right]. \qquad (108)$$

We notice that in the limit $\tau \to \infty$, the right hand side of the integral is the laplace transform

$$\int_0^{\infty} e^{-(1+\frac{1}{\lambda})s} M_\lambda(s) \mathrm{d}s = \mathcal{L} M_\lambda \left(1 + \frac{1}{\lambda} \right) \qquad (109)$$

and we have seen the connection with resolvent in (28)

$$\mathcal{L} M_\lambda \left(1 + \frac{1}{\lambda} \right) = -\sqrt{\lambda} G_{\rm sc} \left((1 + \frac{1}{\lambda}) \sqrt{\lambda} \right). \qquad (110)$$

But $X^2 + (1+\frac{1}{\lambda})\sqrt{\lambda} X + 1 = 0$ has two roots: $\{-\sqrt{\lambda}; \frac{-1}{\sqrt{\lambda}}\}$. To ensure $G_{\rm sc}(z) \in \mathbb{C}_+$ when $z \in \mathbb{C}_+$, we have $-\sqrt{\lambda}$ for $\lambda < 1$ and $\frac{-1}{\sqrt{\lambda}}$ for $\lambda > 1$. Thus we conclude

$$\lim_{\tau \to \infty} e^{-(1+\frac{1}{\lambda})\tau} \hat{q}(\tau) = \begin{cases} 0 & (\lambda < 1) \\ \alpha(1 - \frac{1}{\lambda}) & (\lambda > 1) \end{cases} \qquad (111)$$

Therefore, in the regime $\lambda > 1$, we find the asymptotic behavior for $\tau \to \infty$

$$\hat{q}(\tau) \sim \alpha e^{(1+\frac{1}{\lambda})\tau} (1 - \frac{1}{\lambda}). \qquad (112)$$

A careful analysis of the terms entering $\hat{p}(\tau)$ shows the main contribution stems from the last term, on the square $\mathcal{C} = [\sqrt{\tau}, \tau]^2$ (as the integral can be neglected on $[0, \tau]^2 \setminus \mathcal{C}$):

$$\hat{p}(\tau) \simeq \int_{\sqrt{\tau}}^{\tau} \int_{\sqrt{\tau}}^{\tau} \hat{q}(u) q(v) M_\lambda(2\tau - u - v) \mathrm{d}u \mathrm{d}v. \qquad (113)$$

Using the approximation of $\hat{q}(t)$ in (112) for large $t \in \mathcal{C}$, we can further approximate

$$\hat{p}(\tau) \simeq \alpha^2 \left(1 - \frac{1}{\lambda}\right)^2 \int_{\sqrt{\tau}}^{\tau} \int_{\sqrt{\tau}}^{\tau} e^{(1+\frac{1}{\lambda})(u+v)\tau} M_\lambda(2\tau - u - v)\mathrm{d}u\mathrm{d}v \tag{114}$$

and a change of variables $u = \tau - x, v = \tau - y$ provides

$$\hat{p}(\tau) \simeq \alpha^2 e^{2(1+\frac{1}{\lambda})\tau} \left(1 - \frac{1}{\lambda}\right)^2 \iint_{[0,\tau(1-\frac{1}{\sqrt{\tau}})]^2} e^{-(1+\frac{1}{\lambda})(x+y)} M_\lambda(x+y)\mathrm{d}y\mathrm{d}x. \tag{115}$$

Now, notice the integral converges towards a non-zero value K_λ when $\tau \to \infty$

$$K_\lambda = \iint_{[0,\infty]^2} e^{-(1+\frac{1}{\lambda})(x+y)} M_\lambda(x+y)\mathrm{d}y\mathrm{d}x. \tag{116}$$

Using a further change of variable $s = x + y$ we find

$$K_\lambda = \int_{x=0}^{\infty} \int_{s=x}^{\infty} e^{-(1+\frac{1}{\lambda})s} M_\lambda(s)\mathrm{d}s\mathrm{d}x = \int_{s=0}^{\infty} \int_{x=0}^{s} e^{-(1+\frac{1}{\lambda})s} M_\lambda(s)\mathrm{d}x\mathrm{d}s. \tag{117}$$

Hence again, we find a connection with a Laplace transform (with a derivative from the additional s term inside the integral)

$$K_\lambda = \int_{s=0}^{\infty} e^{-(1+\frac{1}{\lambda})s} s M_\lambda(s)\mathrm{d}x = -(\mathcal{L}M_\lambda)' \left(1 + \frac{1}{\lambda}\right) \tag{118}$$

As $\mathcal{L}M_\lambda(p) = -\sqrt{\lambda}G_{\mathrm{sc}}(p\sqrt{\lambda})$, and considering that $G'_{\mathrm{sc}}(z) = -\frac{G_{\mathrm{sc}}(z)}{2G_{\mathrm{sc}}(z)+z}$, and that $G_{\mathrm{sc}}((1 + \frac{1}{\lambda})\sqrt{\lambda}) = -\frac{1}{\sqrt{\lambda}}$ in the case when $\lambda > 1$, we conclude

$$K_\lambda = \lambda \frac{\frac{1}{\sqrt{\lambda}}}{-2\frac{1}{\sqrt{\lambda}} + (1 + \frac{1}{\lambda})\sqrt{\lambda}} = \frac{1}{1 - \frac{1}{\lambda}}. \tag{119}$$

Finally, with (119) and (115) we find

$$\hat{p}(\tau) \sim \alpha^2 \left(1 - \frac{1}{\lambda}\right) e^{2(1+\frac{1}{\lambda})\tau} \tag{120}$$

and for $\alpha > 0$, we can conclude $\lim_{\tau\to\infty} \bar{q}(\tau) = \sqrt{1 - \frac{1}{\lambda}}$.

I.2. Asymptotic analysis of $\lambda < 1$

The case $\lambda < 1$ is computationally more involved as $\hat{q}(\tau)$ converges to 0, and hence we need to find the rate of convergence towards 0 of this term and that of $\hat{p}(\tau)$ in order to deduce the one from $\bar{q}(\tau)$. Though it is not the main topic of the paper, we provide some calculus elements to achieve this. We start with a lemma to find a suitable expression for $\hat{q}(\tau)$. Most of the calculations has been checked with Mathematica (a notebook is provided in the supplementary material).

Lemma 11 *$\hat{q}(\tau)$ has the following equivalent form:*

$$\hat{q}(\tau) = \alpha\left(1-\frac{1}{\lambda}\right)e^{(1+\frac{1}{\lambda})\tau}\mathbb{I}_{(1,+\infty)}(\lambda) + \frac{2\alpha}{\pi\lambda}e^{\frac{2}{\sqrt{\lambda}}\tau}\int_0^\pi e^{\frac{2}{\sqrt{\lambda}}(\cos(\theta)-1)\tau}\frac{\sin(\theta)^2}{(1+\frac{1}{\lambda})-\frac{2}{\sqrt{\lambda}}\cos(\theta)}\mathrm{d}\theta \tag{121}$$

Proof Starting with $\hat{q}(\tau)$ from (1), one can use a similar expression of M_λ

$$\frac{e^{-(1+\frac{1}{\lambda})\tau}}{\alpha}\hat{q}(\tau) = 1 - \frac{2}{\pi\lambda}\int_0^\pi\int_0^\tau e^{\left(\frac{2}{\sqrt{\lambda}}\cos(\theta)-(1+\frac{1}{\lambda})\right)s}\sin(\theta)^2\mathrm{d}s\mathrm{d}\theta \tag{122}$$

The inward integral can further be integrated (notice the constant term in the exponent is non-zero)

$$\frac{e^{-(1+\frac{1}{\lambda})\tau}}{\alpha}\hat{q}(t) = 1 - \frac{2}{\pi\lambda}\int_0^\pi\left(e^{\left(\frac{2}{\sqrt{\lambda}}\cos(\theta)-(1+\frac{1}{\lambda})\right)\tau}-1\right)\frac{\sin(\theta)^2}{\frac{2}{\sqrt{\lambda}}\cos(\theta)-(1+\frac{1}{\lambda})}\mathrm{d}\theta \tag{123}$$

Using proposition 12 with the constant $a = \frac{1+\frac{1}{\lambda}}{\frac{2}{\sqrt{\lambda}}} > 1$, one can simplify

$$a - \sqrt{a^2-1} = \sqrt{\lambda}\frac{1+\frac{1}{\lambda}-|1-\frac{1}{\lambda}|}{2} = \begin{cases}\frac{1}{\sqrt{\lambda}} & (\lambda > 1)\\ \sqrt{\lambda} & (\lambda < 1)\end{cases} \tag{124}$$

and we finally find

$$\frac{2}{\pi\lambda}\int_0^\pi\frac{\sin(\theta)^2}{\frac{2}{\sqrt{\lambda}}\cos(\theta)-(1+\frac{1}{\lambda})}\mathrm{d}\theta = \frac{1}{\pi\sqrt{\lambda}}\int_0^\pi\frac{\sin(\theta)^2}{\cos(\theta)-a}\mathrm{d}\theta = \begin{cases}-\frac{1}{\lambda} & (\lambda > 1)\\ -1 & (\lambda < 1)\end{cases} \tag{125}$$

using the solution (125) in (123) concludes the proof. ■

Proposition 12 *For any $a > 1$, we have:*

$$\int_0^\pi\frac{\sin(\theta)^2}{\cos(\theta)-a}\mathrm{d}\theta = \pi(\sqrt{a^2-1}-a) \tag{126}$$

Proof Bioche's rules suggest a change of variable $u = \tan(\frac{\theta}{2})$, we find on the left-hand side

$$\int_0^\pi\frac{\sin(\theta)^2}{\cos(\theta)-a}\mathrm{d}\theta = \int_0^\infty\frac{\left(\frac{2u}{u^2+1}\right)^2}{\frac{1-u^2}{1+u^2}-a}\frac{2\mathrm{d}u}{1+u^2} = \int_0^\infty\frac{8u^2}{[(1-a)-(1+a)u^2](1+u^2)^2}\mathrm{d}u \tag{127}$$

Using the constant $K = \frac{a-1}{a+1}$ (or equivalently $a = \frac{1+K}{1-K}$) we can rewrite

$$\int_0^\pi\frac{\sin(\theta)^2}{\cos(\theta)-a}\mathrm{d}\theta = -4(1-K)\int_0^\infty\frac{u^2}{(K+u^2)(1+u^2)^2}\mathrm{d}u \tag{128}$$

and make a classical partial fraction decomposition of the inward term of the integral

$$\frac{u^2}{(K+u^2)(1+u^2)^2} = \frac{1}{(1-K)^2}\left(\frac{u^2}{K+u^2} - \frac{u^2}{1+u^2}\right) - \frac{1}{1-K}\frac{u^2}{(1+u^2)^2} \tag{129}$$

$$= \frac{1}{(1-K)^2}\left(\frac{1}{1+u^2} - \frac{K}{K+u^2}\right) - \frac{1}{1-K}\left[\frac{1}{1+u^2} - \frac{1}{(1+u^2)^2}\right] \tag{130}$$

$$= \frac{K}{(1-K)^2}\left(\frac{1}{1+u^2} - \frac{1}{K+u^2}\right) + \frac{1}{1-K}\frac{1}{(1+u^2)^2} \tag{131}$$

Then on the one hand, with change of variable $u = \tan(x)$ we have:

$$\int_0^\infty \frac{\mathrm{d}u}{(1+u^2)^2} = \int_0^{\frac{\pi}{2}} \frac{\mathrm{d}x}{1+\tan^2(x)} = \int_0^{\frac{\pi}{2}} \cos^2(x)\mathrm{d}x = \frac{\pi}{4} \tag{132}$$

On the other hand, with change of variable $u = \sqrt{K}\tan(x)$ we have:

$$\int_0^\infty \frac{\mathrm{d}u}{K+u^2} = \int_0^{\frac{\pi}{2}} \mathrm{d}x = \frac{\pi}{2}\frac{1}{\sqrt{K}} \tag{133}$$

Thus:

$$-4(1-K)\int_0^\infty \frac{u^2\mathrm{d}u}{(K+u^2)(1+u^2)^2} = -\pi\left[\frac{2K}{1-K}\left(1-\frac{1}{\sqrt{K}}\right)+1\right] \tag{134}$$

and:

$$\frac{2K}{1-K}\left(1-\frac{1}{\sqrt{K}}\right)+1 = a - \sqrt{a^2-1} \tag{135}$$

■

Going back to the case $\lambda < 1$, we can simplify the expression from equation (121)

$$\hat{q}(\tau) = \frac{2\alpha}{\pi\lambda}e^{\frac{2}{\sqrt{\lambda}}\tau}\int_0^\pi e^{\frac{2}{\sqrt{\lambda}}(\cos(\theta)-1)\tau}\frac{\sin(\theta)^2}{(1+\frac{1}{\lambda})-\frac{2}{\sqrt{\lambda}}\cos(\theta)}\mathrm{d}\theta \tag{136}$$

Further, with $u = \frac{2}{\sqrt{\lambda}}(1-\cos(\theta))$ we rewrite (136) to apply Watson's lemma

$$\hat{q}(\tau) = \frac{2\alpha}{\pi\lambda}e^{\frac{2}{\sqrt{\lambda}}\tau}\left(\frac{\sqrt{\lambda}}{2}\right)^{\frac{1}{2}}\int_0^{\sqrt{\lambda}} e^{-u\tau}\frac{(u(2-\frac{\sqrt{\lambda}}{2}u))^{\frac{1}{2}}}{(1+\frac{1}{\lambda})-\frac{2}{\sqrt{\lambda}}(1-\frac{\sqrt{\lambda}}{2}u)}\mathrm{d}u \tag{137}$$

Therefore, Watson's lemma provides the asymptotic equivalence

$$\hat{q}(\tau) \sim \frac{2\alpha}{\pi\lambda}e^{\frac{2}{\sqrt{\lambda}}\tau}\left(\frac{\sqrt{\lambda}}{2}\right)^{\frac{3}{2}}\frac{2^{\frac{1}{2}}\Gamma(\frac{3}{2})\tau^{-\frac{3}{2}}}{(1+\frac{1}{\lambda})-\frac{2}{\sqrt{\lambda}}} \tag{138}$$

With $\Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$ we have therefore

$$\hat{q}(\tau) \sim \frac{\alpha\tau^{-\frac{3}{2}}e^{\frac{2}{\sqrt{\lambda}}\tau}}{2\sqrt{\pi}\lambda^{\frac{1}{4}}\left(1-\frac{1}{\sqrt{\lambda}}\right)^2} \tag{139}$$

The remaining term $\hat{p}(\tau)$ can further be analyzed by splitting each integral from theorem 1 and analyzing the terms with the asymptotic form $e^{\frac{4\tau}{\sqrt{\lambda}}}\tau^{-\frac{3}{2}}$. For instance, we get easily the first term for which we have

$$M_\lambda(2\tau) = \frac{\sqrt{\lambda}}{2\tau} I_1\left(\frac{4\tau}{\sqrt{\lambda}}\right) \sim \frac{\sqrt{\lambda}e^{\frac{4\tau}{\sqrt{\lambda}}}}{2t\sqrt{2\pi\frac{4\tau}{\sqrt{\lambda}}}} \sim \frac{\lambda^{\frac{3}{4}}e^{\frac{4\tau}{\sqrt{\lambda}}}}{2^{\frac{5}{2}}\sqrt{\pi}\tau^{\frac{3}{2}}} \tag{140}$$

The other terms require more technical considerations. We will use both former approximations from the equivalence relations (139) and (140). However, these approximations are only valid for large τ while the integral for the second term is applied on the whole range $[0, \tau]$. Therefore, we split the integration intervals into two segments, say $[0, \sqrt{\tau}]$ and $[\sqrt{\tau}, \tau]$, and apply the approximations in the domains where it is valid.

Starting with the second term, as $2\tau - s > \tau$ for all $s \in [0, \tau]$, we can already apply the relation (140) and split further the integrals:

$$\int_0^\tau \hat{q}(s)M_\lambda(2\tau - s)\mathrm{d}s \simeq \frac{\lambda^{\frac{3}{4}}e^{\frac{4}{\sqrt{\lambda}}\tau}}{2\sqrt{\pi}}\left[\int_0^{\sqrt{\tau}} \hat{q}(s)\frac{e^{-\frac{2}{\sqrt{\lambda}}s}}{(2\tau - s)^{\frac{3}{2}}}\mathrm{d}s + \int_{\sqrt{\tau}}^\tau \hat{q}(s)\frac{e^{-\frac{2}{\sqrt{\lambda}}s}}{(2\tau - s)^{\frac{3}{2}}}\mathrm{d}s\right] \tag{141}$$

Then the integrand on the first segment of (141) is further approximated using $\frac{1}{(2\tau - s)^{\frac{3}{2}}} = \frac{1}{(2\tau)^{\frac{3}{2}}}$. Indeed, as $s \leq \sqrt{\tau}$ we have $s = o(\tau)$. In the end we retrieve the laplace transform of $\hat{q}$:

$$\int_0^{\sqrt{\tau}} \hat{q}(s)\frac{e^{-\frac{2}{\sqrt{\lambda}}s}}{(2\tau - s)^{\frac{3}{2}}}\mathrm{d}s \simeq \frac{1}{(2\tau)^{\frac{3}{2}}}\int_0^{\sqrt{\tau}} \hat{q}(s)e^{-\frac{2}{\sqrt{\lambda}}s}\mathrm{d}s \simeq \frac{1}{(2\tau)^{\frac{3}{2}}}\mathcal{L}\hat{q}\left(\frac{2}{\sqrt{\lambda}}\right) \tag{142}$$

From (27) which remains valid at $z = 2$ with $G_{\mathrm{sc}}(2) = -1$, we can even derive further the constant term

$$\mathcal{L}\hat{q}\left(\frac{2}{\sqrt{\lambda}}\right) = \frac{-\alpha G_{\mathrm{sc}}(2)}{\frac{1}{\sqrt{\lambda}} + G_{\mathrm{sc}}(2)} = \frac{\alpha}{\frac{1}{\sqrt{\lambda}} - 1} \tag{143}$$

In the second segment of the integral in (141), we use the approximation from (139) and use change of variable $r = \frac{s}{\tau}$

$$\int_{\sqrt{\tau}}^\tau \hat{q}(s)\frac{e^{-\frac{2}{\sqrt{\lambda}}s}}{(2\tau - s)^{\frac{3}{2}}}\mathrm{d}s \simeq \frac{\alpha}{2\sqrt{\pi}\lambda^{\frac{1}{4}}\left[(1 + \frac{1}{\lambda}) - \frac{2}{\sqrt{\lambda}}\right]}\int_{\frac{1}{\sqrt{\tau}}}^1 \frac{1}{r^{\frac{3}{2}}(2 - r)^{\frac{3}{2}}}\frac{\tau}{\tau^{\frac{3}{2}+\frac{3}{2}}}\mathrm{d}r \tag{144}$$

The integral from the right side can be solved:

$$\int_{\frac{1}{\sqrt{\tau}}}^1 \frac{\mathrm{d}r}{r^{\frac{3}{2}}(2 - r)^{\frac{3}{2}}} = \frac{1 - \frac{1}{\sqrt{\tau}}}{\sqrt{\frac{1}{\sqrt{\tau}}\left(2 - \frac{1}{\sqrt{\tau}}\right)}} \sim \frac{\tau^{\frac{1}{4}}}{\sqrt{2}} \tag{145}$$

Putting things together with (145) in (144) we get

$$\int_{\sqrt{\tau}}^\tau \hat{q}(s)\frac{e^{-\frac{2}{\sqrt{\lambda}}s}}{(2\tau - s)^{\frac{3}{2}}}\mathrm{d}s \simeq \frac{\alpha\tau^{-\frac{7}{4}}}{2\sqrt{2\pi}\lambda^{\frac{1}{4}}\left[(1 + \frac{1}{\lambda}) - \frac{2}{\sqrt{\lambda}}\right]} \tag{146}$$

So, the main contribution comes from the first integral of equation (141) with the coefficient $\tau^{-\frac{3}{2}}$

$$2\alpha \int_0^\tau \hat{q}(s) M_\lambda(2\tau - s)\mathrm{d}s \sim \frac{\alpha^2 \lambda^{\frac{3}{4}} e^{\frac{4}{\sqrt{\lambda}}\tau} \tau^{-\frac{3}{2}}}{2^{\frac{3}{2}}\sqrt{\pi}\left(\frac{1}{\sqrt{\lambda}} - 1\right)} \tag{147}$$

The third term with the double-integral requires extending the previous calculation idea on each rectangle: $I_1 = [0, \sqrt{\tau}]^2$, $I_2 = [0, \sqrt{\tau}] \times [\sqrt{\tau}, \tau]$, $I_2' = [\sqrt{\tau}, \tau] \times [0, \sqrt{\tau}]$ and $I_3 = [\sqrt{\tau}, \tau]^2$. As we will see, only the integral on I_1 brings a contribution of order $\tau^{-\frac{3}{2}}$ and the others can be neglected.

Interval $I_1 = [0, \sqrt{\tau}]^2$ On this interval, $2\tau - u - v \gg 1$ so we consider

$$\iint_{I_1} \hat{q}(u)\hat{q}(v) M_\lambda(2\tau - u - v)\mathrm{d}u\mathrm{d}v \simeq \frac{\lambda^{\frac{3}{4}} e^{\frac{4}{\sqrt{\lambda}}\tau}}{2\sqrt{\pi}} \iint_{I_1} \hat{q}(u)\hat{q}(v) \frac{e^{-\frac{2}{\sqrt{\lambda}}(u+v)}}{(2\tau - u - v)^{\frac{3}{2}}}\mathrm{d}u\mathrm{d}v \tag{148}$$

also, on I_1 we have $\frac{1}{(2\tau-u-v)^{\frac{3}{2}}} \simeq \frac{1}{(2\tau)^{\frac{3}{2}}}$, thus we are left to consider:

$$\iint_{I_1} \hat{q}(u)\hat{q}(v) e^{-\frac{2}{\sqrt{\lambda}}(u+v)}\mathrm{d}u\mathrm{d}v \simeq \left[\mathcal{L}\hat{q}\left(\frac{2}{\sqrt{\lambda}}\right)\right]^2 \tag{149}$$

hence with (123) we find

$$\iint_{I_1} \hat{q}(u)\hat{q}(v) M_\lambda(2\tau - u - v)\mathrm{d}u\mathrm{d}v \simeq \frac{\alpha^2 \lambda^{\frac{3}{4}} e^{\frac{4}{\sqrt{\lambda}}\tau} \tau^{-\frac{3}{2}}}{2^{\frac{5}{2}}\sqrt{\pi}\left(\frac{1}{\sqrt{\lambda}} - 1\right)^2} \tag{150}$$

Interval $I_2 = [0, \sqrt{\tau}] \times [\sqrt{\tau}, \tau]$ here we still have $2\tau - u - v \gg 1$ but also $v \geq \sqrt{\tau} \gg 1$ so with (139) we first get

$$\iint_{I_2} \hat{q}(u)\hat{q}(v) M_\lambda(2\tau - u - v)\mathrm{d}u\mathrm{d}v \simeq \frac{\alpha \iint_{I_2} \hat{q}(u) v^{-\frac{3}{2}} e^{\frac{2}{\sqrt{\lambda}}v} M_\lambda(2\tau - u - v)\mathrm{d}u\mathrm{d}v}{2\sqrt{\pi}\lambda^{\frac{1}{4}}\left[(1 + \frac{1}{\lambda}) - \frac{2}{\sqrt{\lambda}}\right]} \tag{151}$$

and then:

$$\iint_{I_2} \hat{q}(u) v^{-\frac{3}{2}} e^{\frac{2}{\sqrt{\lambda}}v} M_\lambda(2\tau - u - v)\mathrm{d}u\mathrm{d}v \simeq \frac{\lambda^{\frac{3}{4}} e^{\frac{4}{\sqrt{\lambda}}\tau}}{2\sqrt{\pi}} \iint_{I_2} \hat{q}(u) v^{-\frac{3}{2}} e^{\frac{2}{\sqrt{\lambda}}v} \frac{e^{-\frac{2}{\sqrt{\lambda}}(u+v)}}{(2\tau - u - v)^{\frac{3}{2}}}\mathrm{d}u\mathrm{d}v \tag{152}$$

so

$$\iint_{I_2} \hat{q}(u) v^{-\frac{3}{2}} e^{\frac{2}{\sqrt{\lambda}}v} M_\lambda(2\tau - u - v)\mathrm{d}u\mathrm{d}v \simeq \frac{\lambda^{\frac{3}{4}} e^{\frac{4}{\sqrt{\lambda}}\tau}}{2\sqrt{\pi}} \int_0^{\sqrt{\tau}} \hat{q}(u) e^{-\frac{2}{\sqrt{\lambda}}u} \int_{\sqrt{\tau}}^\tau \frac{1}{v^{\frac{3}{2}}(2\tau - u - v)^{\frac{3}{2}}}\mathrm{d}v\mathrm{d}u \tag{153}$$

At fixed $u \in [0, \sqrt{\tau}]$ With change of variable $s = \frac{v}{\tau}$ we find

$$\int_{\sqrt{\tau}}^\tau \frac{1}{v^{\frac{3}{2}}(2\tau - u - v)^{\frac{3}{2}}}\mathrm{d}v = \int_{\frac{1}{\sqrt{\tau}}}^1 \frac{1}{\tau^{\frac{3}{2}} s^{\frac{3}{2}}(\tau(2 - s) - u)^{\frac{3}{2}}}\tau\mathrm{d}s = \frac{1}{\tau^2} \int_{\frac{1}{\sqrt{\tau}}}^1 \frac{1}{s^{\frac{3}{2}}\left((2 - s) - \frac{u}{\tau}\right)^{\frac{3}{2}}}\mathrm{d}s \tag{154}$$

Because $u \leq \sqrt{\tau}$ we have $\frac{u}{\tau} = o(1)$. Notice we have

$$\int_{\frac{1}{\sqrt{\tau}}}^{1} \frac{\mathrm{d}s}{s^{\frac{3}{2}}\left((2-s)-\frac{u}{\tau}\right)^{\frac{3}{2}}} = \left[\frac{2(\frac{u}{\tau}+2s-2)}{(\frac{u}{\tau}-2)^2\sqrt{s(2-s-\frac{u}{\tau})}}\right]_{\frac{1}{\sqrt{\tau}}}^{1} \tag{155}$$

$$= \frac{2}{(\frac{u}{\tau}-2)^2}\left[\frac{u}{\tau\sqrt{1-\frac{u}{\tau}}} - \frac{\frac{u}{\tau}+\frac{2}{\sqrt{\tau}}-2}{\sqrt{\frac{1}{\sqrt{\tau}}(2-\frac{1}{\sqrt{\tau}}-\frac{u}{\tau})}}\right] \sim \frac{\tau^{\frac{1}{4}}}{\sqrt{2}} \tag{156}$$

Hence we have a term in $\tau^{-\frac{7}{4}}$ so the term on I_2 can be neglected compared to I_1:

$$\iint_{I_2} \hat{q}(u)v^{-\frac{3}{2}}e^{\frac{2}{\sqrt{\lambda}}v}M_\lambda(2\tau-u-v)\mathrm{d}u\mathrm{d}v \sim \frac{\lambda^{\frac{3}{4}}e^{\frac{4}{\sqrt{\lambda}}\tau}\tau^{-\frac{7}{4}}}{2\sqrt{2\pi}}\mathcal{L}\hat{q}\left(\frac{2}{\sqrt{\lambda}}\right) \tag{157}$$

Notice finally that the interval $I_2' = [\sqrt{\tau}, \tau] \times [0, \sqrt{\tau}]$ is similar as the integrand is symmetric in its arguments.

Interval $I_3 = [\sqrt{\tau}, \tau]^2$ we can approximate both $\hat{q}(u), \hat{q}(v)$

$$\iint_{I_3} \hat{q}(u)\hat{q}(v)M_\lambda(2\tau-u-v)\mathrm{d}u\mathrm{d}v \simeq \frac{\alpha^2 \iint_{I_3}(uv)^{-\frac{3}{2}}e^{\frac{2}{\sqrt{\lambda}}(u+v)}M_\lambda(2\tau-u-v)\mathrm{d}u\mathrm{d}v}{4\pi\lambda^{\frac{1}{2}}\left[(1+\frac{1}{\lambda})-\frac{2}{\sqrt{\lambda}}\right]^2} \tag{158}$$

Let's focus on the right hand side integral

$$f(\tau) = \iint_{I_3}(uv)^{-\frac{3}{2}}e^{\frac{2}{\sqrt{\lambda}}(u+v)}M_\lambda(2\tau-u-v)\mathrm{d}u\mathrm{d}v \tag{159}$$

Now, using change of variable $u = \tau - x, v = \tau - y$ we have

$$f(\tau) = e^{\frac{4}{\sqrt{\lambda}}\tau}\iint_{[0,\ \tau(1-\frac{1}{\sqrt{\tau}})]^2}(\tau-x)^{-\frac{3}{2}}(\tau-y)^{-\frac{3}{2}}e^{\frac{-2}{\sqrt{\lambda}}(x+y)}M_\lambda(x+y)\mathrm{d}x\mathrm{d}y \tag{160}$$

with $s = x + y$

$$e^{\frac{-4}{\sqrt{\lambda}}\tau}f(\tau) = \int_0^{\tau(1-\frac{1}{\sqrt{\tau}})}\int_x^{x+\tau(1-\frac{1}{\sqrt{\tau}})}(\tau-x)^{-\frac{3}{2}}(\tau-s+x)^{-\frac{3}{2}}e^{\frac{-2}{\sqrt{\lambda}}s}M_\lambda(s)\mathrm{d}s\mathrm{d}x \tag{161}$$

$$= \int_0^{2\tau(1-\frac{1}{\sqrt{\tau}})}\int_{\max\left(s-\tau(1-\frac{1}{\sqrt{\tau}}),0\right)}^{\min\left(\tau(1-\frac{1}{\sqrt{\tau}}),s\right)}(\tau-x)^{-\frac{3}{2}}(\tau-s+x)^{-\frac{3}{2}}\mathrm{d}x e^{\frac{-2}{\sqrt{\lambda}}s}M_\lambda(s)\mathrm{d}s \tag{162}$$

$$= \int_0^{\tau(1-\frac{1}{\sqrt{\tau}})}\int_0^{s}(\tau-x)^{-\frac{3}{2}}(\tau-s+x)^{-\frac{3}{2}}\mathrm{d}x e^{\frac{-2}{\sqrt{\lambda}}s}M_\lambda(s)\mathrm{d}s \tag{163}$$

$$+ \int_{\tau(1-\frac{1}{\sqrt{\tau}})}^{2\tau(1-\frac{1}{\sqrt{\tau}})}\int_{s-\tau(1-\frac{1}{\sqrt{\tau}})}^{\tau(1-\frac{1}{\sqrt{\tau}})}(\tau-x)^{-\frac{3}{2}}(\tau-s+x)^{-\frac{3}{2}}\mathrm{d}x e^{\frac{-2}{\sqrt{\lambda}}s}M_\lambda(s)\mathrm{d}s \tag{164}$$

On the first integral, we find

$$\int_0^s (\tau-x)^{-\frac{3}{2}}(\tau-s+x)^{-\frac{3}{2}}\mathrm{d}x = \left[\frac{2(2x-s)}{(2\tau-s)^2\sqrt{(\tau-x)(\tau+x-s)}}\right]_0^s = \frac{4s}{(2\tau-s)^2\sqrt{(\tau-s)\tau}} \tag{165}$$

However, $s \le \tau-\sqrt{\tau}$ so $\sqrt{\tau} \le \tau - s$ and $\tau+\sqrt{\tau} \le 2\tau - s$ so:

$$\frac{4s}{(2\tau-s)^2\sqrt{(\tau-s)\tau}} \le \frac{4\tau(1-\frac{1}{\sqrt{\tau}})}{\tau^2(1+\frac{1}{\sqrt{\tau}})^2\tau^{\frac{1}{2}}\tau^{\frac{1}{4}}} = 4\tau^{-\frac{7}{4}}(1+o_\tau(1)) \tag{166}$$

Therefore, we find:

$$\int_0^{\tau(1-\frac{1}{\sqrt{\tau}})}\int_0^s (\tau-x)^{-\frac{3}{2}}(\tau-s+x)^{-\frac{3}{2}}\mathrm{d}x e^{\frac{-2}{\sqrt{\lambda}}s}M_\lambda(s)\mathrm{d}s \le 4\tau^{-\frac{7}{4}}(1+o_\tau(1))\mathcal{L}M_\lambda\left(\frac{2}{\sqrt{\lambda}}\right) \tag{167}$$

Noticeably, $\mathcal{L}M_\lambda\left(\frac{2}{\sqrt{\lambda}}\right) = -\sqrt{\lambda}G_{\text{sc}}(2) = \sqrt{\lambda}$. In the asymptotic limit, this term can be neglected due to $\tau^{-\frac{7}{4}}$ compared to $\tau^{-\frac{3}{2}}$.
Similarly, we find

$$\int_{s-\tau(1-\frac{1}{\sqrt{\tau}})}^{\tau(1-\frac{1}{\sqrt{\tau}})}(\tau-x)^{-\frac{3}{2}}(\tau-s+x)^{-\frac{3}{2}}\mathrm{d}x = \frac{4(2\tau-s-2\sqrt{\tau})}{(2\tau-s)^2\tau^{\frac{1}{4}}\sqrt{2\tau-\sqrt{\tau}-s}} \tag{168}$$

then, in $[\tau-\sqrt{\tau}, 2(\tau-\sqrt{\tau})]$ we approximate M_λ with its asymptotic expression. So we are left to evaluate

$$K(\tau) = \int_{\tau(1-\frac{1}{\sqrt{\tau}})}^{2\tau(1-\frac{1}{\sqrt{\tau}})} \frac{4(2\tau-s-2\sqrt{\tau})}{s^{\frac{3}{2}}(2\tau-s)^2\tau^{\frac{1}{4}}\sqrt{2\tau-\sqrt{\tau}-s}}\mathrm{d}s \tag{169}$$

Notice that $2\tau-s-2\sqrt{\tau} \le \tau(1-\frac{1}{\sqrt{\tau}})$, and $2\tau^{\frac{1}{2}} \le 2\,\tau-s$ and $\tau^{\frac{1}{2}} \le 2\tau-\sqrt{\tau}-s$, hence

$$0 \le K(\tau) \le \frac{4\tau(1-\frac{1}{\sqrt{\tau}})}{(2\tau^{\frac{1}{2}})^2\tau^{\frac{1}{4}}\sqrt{\tau^{\frac{1}{2}}}}\int_{\tau(1-\frac{1}{\sqrt{\tau}})}^{2\tau(1-\frac{1}{\sqrt{\tau}})}\frac{\mathrm{d}s}{s^{\frac{3}{2}}} \tag{170}$$

So

$$0 \le K(\tau) \le \frac{(1-\frac{1}{\sqrt{\tau}})}{\tau^{\frac{1}{2}}}\left[-\frac{2}{s^{\frac{1}{2}}}\right]_{\tau(1-\frac{1}{\sqrt{\tau}})}^{2\tau(1-\frac{1}{\sqrt{\tau}})} \tag{171}$$

with a change of variable $u = s-(\tau-\sqrt{\tau})$ we find

$$K(\tau) = \frac{1}{\tau^{\frac{1}{4}}}\int_0^{\tau(1-\frac{1}{\sqrt{\tau}})}\frac{4(\tau(1-\frac{1}{\sqrt{\tau}})-u)}{(\tau(1-\frac{1}{\sqrt{\tau}})+u)^{\frac{3}{2}}(\tau(1+\frac{1}{\sqrt{\tau}})-u)^2\sqrt{\tau-u}}\mathrm{d}u \tag{172}$$

with another change of variable $u = \tau r$ we find:

$$K(\tau) = \frac{4}{\tau^{\frac{9}{4}}}\int_0^{1-\frac{1}{\sqrt{\tau}}}\frac{(1-\frac{1}{\sqrt{\tau}}-r)}{(1-\frac{1}{\sqrt{\tau}}+r)^{\frac{3}{2}}(1+\frac{1}{\sqrt{\tau}}-r)^2\sqrt{1-r}}\mathrm{d}r \tag{173}$$

Though this integral can be completely solved, we are only interested in bounding it. In particular, we find:

$$K(\tau) \leq \frac{4}{\tau^{\frac{9}{4}}} \int_0^{1-\frac{1}{\sqrt{\tau}}} \frac{(1-r)}{(1-\frac{1}{\sqrt{\tau}})^{\frac{3}{2}}(1-r)^2\sqrt{1-r}} \mathrm{d}r = \frac{4}{\tau^{\frac{9}{4}}(1-\frac{1}{\sqrt{\tau}})^{\frac{3}{2}}} \int_0^{1-\frac{1}{\sqrt{\tau}}} \frac{\mathrm{d}r}{(1-r)^{\frac{3}{2}}} \tag{174}$$

So

$$K(\tau) \leq \frac{4}{\tau^{\frac{9}{4}}(1-\frac{1}{\sqrt{\tau}})^{\frac{3}{2}}} \left[\frac{2}{(1-r)^{\frac{1}{2}}}\right]_0^{1-\frac{1}{\sqrt{\tau}}} = \frac{8(1-\frac{1}{\tau^{\frac{1}{4}}})}{\tau^{\frac{8}{4}}(1-\frac{1}{\sqrt{\tau}})^{\frac{3}{2}}} = 8\tau^{-\frac{8}{4}}(1+o(1)) \tag{175}$$

In the end, the integral on I_3 can also be neglected.

conclusion summing up all the main contributions from (140), (147) and (150) we find

$$\begin{aligned} \lim_{\tau\to\infty} \tau^{\frac{3}{2}} \mathrm{e}^{-\frac{4\tau}{\sqrt{\lambda}}} \hat{p}(\tau) &= \frac{\lambda^{\frac{3}{4}}}{2^{\frac{5}{2}}\sqrt{\pi}} + \frac{\alpha^2\lambda^{\frac{3}{4}}}{2^{\frac{3}{2}}\sqrt{\pi}\left(\frac{1}{\sqrt{\lambda}}-1\right)} + \frac{\alpha^2\lambda^{\frac{3}{4}}}{2^{\frac{5}{2}}\sqrt{\pi}\left(\frac{1}{\sqrt{\lambda}}-1\right)^2} && (176) \\ &= \frac{\lambda^{\frac{3}{4}}}{2^{\frac{5}{2}}\sqrt{\pi}} \left[1+\alpha^2\left(\frac{2}{\frac{1}{\sqrt{\lambda}}-1} + \frac{1}{\left(\frac{1}{\sqrt{\lambda}}-1\right)^2}\right)\right] && (177) \end{aligned}$$

and thus:

$$\frac{1}{\sqrt{\hat{p}(\tau)}} \sim \frac{2^{\frac{5}{4}}\pi^{\frac{1}{4}}}{\lambda^{\frac{3}{8}}} \left[1-\alpha^2+\frac{\alpha^2}{\lambda(\frac{1}{\sqrt{\lambda}}-1)^2}\right]^{-\frac{1}{2}} \tau^{\frac{3}{4}} e^{-\frac{2}{\sqrt{\lambda}}\tau} \tag{178}$$

Using back (139) we find

$$\bar{q}(\tau) \sim \frac{\alpha\left(\frac{2}{\pi}\right)^{\frac{1}{4}}}{\lambda^{\frac{5}{8}}\left(1-\frac{1}{\sqrt{\lambda}}\right)^2\sqrt{1-\alpha^2+\frac{\alpha^2}{\lambda(\frac{1}{\sqrt{\lambda}}-1)^2}}} \tau^{-\frac{3}{4}} \tag{179}$$

Numerical evaluations from the functions of theorem 1 match correctly this expression for different values of (α, λ), see figure 4 (a) for instance.

I.3. Asymptotic analysis of $\lambda > 1$

Using the previous analysis for $\hat{q}(\tau)$ ((121) and (139)), we have an additional term:

$$\hat{q}(\tau) = \alpha\left(1-\frac{1}{\lambda}\right) e^{(1+\frac{1}{\lambda})\tau} + \frac{\alpha\tau^{-\frac{3}{2}} e^{\frac{2}{\sqrt{\lambda}}\tau}}{2\sqrt{\pi}\lambda^{\frac{1}{4}}\left(1-\frac{1}{\sqrt{\lambda}}\right)^2} + o(\tau^{-\frac{3}{2}} e^{\frac{2}{\sqrt{\lambda}}\tau}) \tag{180}$$

Now, for $\hat{p}(\tau)$, we have already seen the leading asymptotics in equation (120). For the next correction, we postulate through computer analysis that there exists a non-null constant $C \in \mathbb{R}_+^*$ such that it takes the form:

$$\hat{p}(\tau) = \alpha^2\left(1-\frac{1}{\lambda}\right) e^{2(1+\frac{1}{\lambda})\tau} \left[1-2\tau^{-\frac{3}{2}} e^{-2(1-\frac{1}{\sqrt{\lambda}})^2\tau}(C+o(1))\right] \tag{181}$$

Hence the expression:

$$\frac{1}{\sqrt{\hat{p}(\tau)}} = \frac{e^{-(1+\frac{1}{\lambda})\tau}}{|\alpha|\sqrt{1-\frac{1}{\lambda}}}\left[1+\tau^{-\frac{3}{2}}e^{-2(1-\frac{1}{\sqrt{\lambda}})^2\tau}(C+o(1))\right] \tag{182}$$

Putting things together, we find:

$$\bar{q}(\tau) = \text{sign}(\alpha)\sqrt{1-\frac{1}{\lambda}}\left(1+\frac{\tau^{-\frac{3}{2}}e^{-(1-\frac{1}{\sqrt{\lambda}})^2\tau}(1+o(1))}{2(1-\frac{1}{\lambda})\sqrt{\pi}\lambda^{\frac{1}{4}}\left(1-\frac{1}{\sqrt{\lambda}}\right)^2}\right)\left(1+\tau^{-\frac{3}{2}}e^{-2(1-\frac{1}{\sqrt{\lambda}})^2\tau}(C+o(1))\right) \tag{183}$$

Hence the exponential term in the expression of $\hat{q}$ dominates the one in the expression of $\hat{p}$. Therefore, expanding the asymptotic expansion provides the result:

$$\bar{q}(\tau) - \text{sign}(\alpha)\sqrt{1-\frac{1}{\lambda}} \sim \frac{\text{sign}(\alpha)}{2\sqrt{\pi}\lambda^{\frac{1}{4}}\sqrt{1-\frac{1}{\lambda}}\left(1-\frac{1}{\sqrt{\lambda}}\right)^2}\tau^{-\frac{3}{2}}e^{-(1-\frac{1}{\sqrt{\lambda}})^2\tau} \tag{184}$$

More specifically, equation (183) shows that the second order term of $\hat{q}$ dominates the one of $\frac{1}{\sqrt{\hat{p}}}$ when we compute the final contribution in equation (184). Therefore, this fact can be emphasized with the equivalent limiting behavior:

$$\bar{q}(\tau) - \text{sign}(\alpha)\sqrt{1-\frac{1}{\lambda}} \sim \frac{1}{|\alpha|\sqrt{1-\frac{1}{\lambda}}}\left(\hat{q}(\tau)e^{-(1+\frac{1}{\lambda})\tau} - \alpha\left(1-\frac{1}{\lambda}\right)\right) \tag{185}$$

This form is actually more convenient because a numerical evaluation $\bar{q}(\tau)$ for large τ requires extra precision and computational resources due to the double-integral within the $\hat{p}(\tau)$ term. Therefore, it appears to be easier to observe the equivalent behavior in (185) rather than in (184). To illustrate this phenomenon, one can evaluate:

$$\psi(\tau) = |\alpha|\sqrt{1-\frac{1}{\lambda}}\left(\bar{q}(\tau) - \text{sign}(\alpha)\sqrt{1-\frac{1}{\lambda}}\right)e^{(1-\frac{1}{\sqrt{\lambda}})^2\tau} \tag{186}$$

$$\phi(\tau) = \left(\hat{q}(\tau)e^{-(1+\frac{1}{\lambda})\tau} - \alpha\left(1-\frac{1}{\lambda}\right)\right)e^{(1-\frac{1}{\sqrt{\lambda}})^2\tau} \tag{187}$$

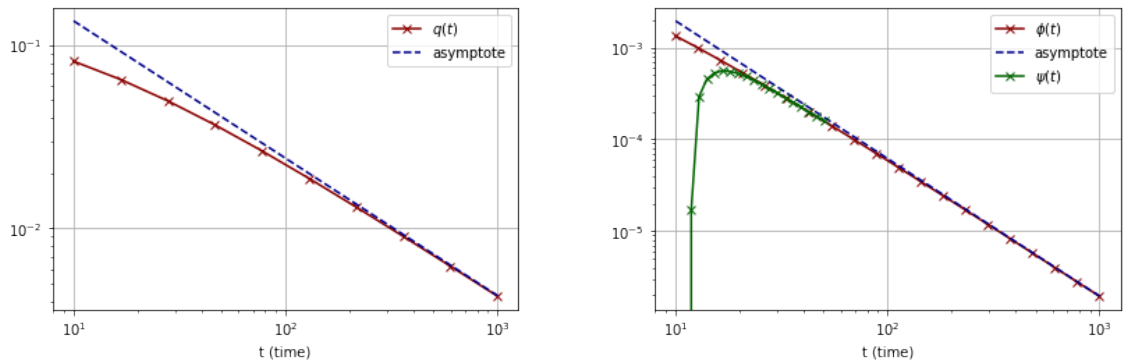
$$\mathcal{A}(\tau) = \frac{\alpha}{2\sqrt{\pi}\lambda^{\frac{1}{4}}\left(1-\frac{1}{\sqrt{\lambda}}\right)^2}\tau^{-\frac{3}{2}} \tag{188}$$

and expect to observe $\psi(\tau) \sim \phi(\tau) \sim \mathcal{A}(\tau)$ when $\tau \to \infty$ for any $\lambda > 1$ and $\alpha \neq 0$. See figure 4 (b) as an example where the computation of $\psi(\tau)$ had to be stopped earlier in time to cope with computational limits of the math library Scipy.

I.4. Asymptotic analysis for $\lambda = 1$

In the special case $\lambda = 1$ where the regime changes, one can write explicitly:

$$\hat{q}(\tau) = \alpha e^{2\tau}\left(1-1+e^{-2\tau}(I_0(2\tau)+I_1(2\tau))\right) = \alpha\left[I_0(2\tau)+I_1(2\tau)\right] \tag{189}$$



(*a*) $\lambda = 0.5$, $\alpha = 0.1$ (*b*) $\lambda = 5$, $\alpha = 0.1$

Figure 4: Example of a numerical evaluation of theorem 1 and comparisons with their respective asymptotes in log-scale for $\lambda < 1$ in (a) and $\lambda > 1$ in (b).

and we find the first term of the asymptotic expansion in $\tau \to \infty$:

$$\hat{q}(\tau) \sim \alpha \frac{e^{2\tau}}{\sqrt{\pi\tau}}. \tag{190}$$

Some further analysis lead us to a similar estimate for $\hat{p}(\tau)$

$$\sqrt{\hat{p}(\tau)} \sim |\alpha| \frac{e^{2t}}{(2\pi\tau)^{\frac{1}{4}}} \tag{191}$$

and thus to conclude using (190) (for $\alpha > 0$):

$$\bar{q}(\tau) \sim \left(\frac{2}{\pi\tau}\right)^{\frac{1}{4}} \tag{192}$$

Using similar arguments as the case $\lambda < 1$ (see section I.2), we can check that the main asymptotic contribution in $\tau^{-\frac{1}{2}}$ comes from the the third term of $\hat{p}$ on the interval I_3. Indeed, the first term in $M_1(2\tau)$ is obviously in $\tau^{-\frac{3}{2}}$.The second term can also be neglected, notice that we have:

$$\int_{\sqrt{\tau}}^{\tau} \hat{q}(s) \frac{e^{-2s}}{(2\tau - s)^{\frac{3}{2}}} \mathrm{d}s \sim \frac{\alpha}{\sqrt{\pi}} \int_{\sqrt{\tau}}^{\tau} \frac{\mathrm{d}s}{\sqrt{s}(2\tau - s)^{\frac{3}{2}}} \sim \frac{\alpha}{\sqrt{\pi}\tau} \tag{193}$$

Also we don't have a constant term with the laplace transform of $\hat{q}$. Instead for any $t > 0$

$$\int_0^t \hat{q}(s) e^{-2s} \mathrm{d}s = \frac{\alpha}{2}(e^{-2t}(1 + 4t)I_0(2t) + 4te^{-2t}I_1(2t) - 1) \tag{194}$$

In particular when $t = \sqrt{\tau}$ and $\tau \to \infty$:

$$\int_0^{\sqrt{\tau}} \hat{q}(s) e^{-2s} \mathrm{d}s \sim \frac{4\alpha\sqrt{\tau}}{\sqrt{4\pi\sqrt{\tau}}} \sim \frac{2\alpha\tau^{\frac{1}{4}}}{\sqrt{\pi}} \tag{195}$$

Hence with the additional term in $\tau^{-\frac{3}{2}}$ this gives a term in $\tau^{-\frac{5}{4}}$. We proceed similarly for the third term with the 4 segments I_1, I_2, I_2', I_3.

Interval $I_1 = [0, \sqrt{\tau}]^2$ Similar considerations using the result (195) lead to the asymptotics:

$$\iint_{I_1} \hat{q}(u)\hat{q}(v)e^{-2(u+v)}\mathrm{d}u\mathrm{d}v \sim \frac{4\alpha^2\tau^{\frac{1}{2}}}{\pi} \tag{196}$$

Hence with the additional term in $\tau^{-\frac{3}{2}}$ this gives a term in τ^{-1}.

Interval $I_2 = [0, \sqrt{\tau}] \times [\sqrt{\tau}, \tau]$ We get:

$$\iint_{I_2} \hat{q}(u)\hat{q}(v)M_1(2\tau - u - v)\mathrm{d}u\mathrm{d}v \simeq \frac{\alpha}{2\pi}\iint_{I_2} \hat{q}(u)\frac{e^{2v}}{\sqrt{v}}\frac{e^{2(2\tau-u-v)}}{(2\tau-u-v)^{\frac{3}{2}}}\mathrm{d}u\mathrm{d}v \tag{197}$$

We can compute further the integral considering $u = o(\tau)$:

$$\begin{aligned} \int_v \frac{1}{\sqrt{v}(2\tau-u-v)^{\frac{3}{2}}}\mathrm{d}v &= \frac{2}{2\tau-u}\left[\sqrt{\frac{v}{2\tau-u-v}}\right]_{\sqrt{\tau}}^{\tau} \\ &= \frac{2}{2\tau-u}\left[\sqrt{\frac{\tau}{2\tau-u}} - \sqrt{\frac{\sqrt{\tau}}{2\tau-u-\sqrt{\tau}}}\right] \\ &\sim \tau^{-1} \end{aligned} \tag{198}$$

Finally, using (195) gives:

$$\iint_{I_2} \hat{q}(u)\hat{q}(v)M_1(2\tau - u - v)\mathrm{d}u\mathrm{d}v \sim \frac{\alpha^2}{\pi^{\frac{3}{2}}}e^{4\tau}\tau^{-\frac{3}{4}} \tag{199}$$

Interval $I_3 = [\sqrt{\tau}, \tau]^2$ On this interval we have:

$$\iint_{I_3} \hat{q}(u)\hat{q}(v)M_1(2\tau - u - v)\mathrm{d}u\mathrm{d}v \simeq \frac{\alpha^2}{\pi}\iint_{I_3} e^{2(u+v)}\frac{I_1(2(2\tau-u-v))}{(2\tau-u-v)\sqrt{uv}}\mathrm{d}u\mathrm{d}v \tag{200}$$

Let's focus on the right hand side integral:

$$f(\tau) = \iint_{I_3} e^{2(u+v)}\frac{I_1(2(2\tau-u-v))}{(2\tau-u-v)\sqrt{uv}}\mathrm{d}u\mathrm{d}v \tag{201}$$

With $x = \tau - u$, $y = \tau - v$ we find:

$$e^{-4\tau}f(\tau) = \iint_{[0,\tau-\sqrt{\tau}]^2} e^{-2(x+y)}\frac{I_1(2(x+y))}{(x+y)\sqrt{(\tau-x)(\tau-y)}}\mathrm{d}x\mathrm{d}y \tag{202}$$

Now, consider further the change of variable: $x = (\tau - \sqrt{\tau})r$ and $y = (\tau - \sqrt{\tau})s$. we have:

$$\sqrt{\tau}e^{-4\tau}f(\tau) = \iint_{[0,1]^2} \sqrt{\tau}\frac{e^{-2(\tau-\sqrt{\tau})(r+s)}I_1(2(\tau-\sqrt{\tau})(r+s))}{(r+s)\sqrt{\left(\frac{1}{1-\frac{1}{\sqrt{\tau}}} - r\right)\left(\frac{1}{1-\frac{1}{\sqrt{\tau}}} - s\right)}}\mathrm{d}r\mathrm{d}s \tag{203}$$

Now, for all $r, s \in [0,1]^2 \setminus \{(0,0)\}$, we have:

$$\lim_{\tau\to\infty} \sqrt{\tau}\frac{e^{-2(\tau-\sqrt{\tau})(r+s)}I_1(2(\tau-\sqrt{\tau})(r+s))}{(r+s)\sqrt{\left(\frac{1}{1-\frac{1}{\sqrt{\tau}}} - r\right)\left(\frac{1}{1-\frac{1}{\sqrt{\tau}}} - s\right)}} = \frac{1}{\sqrt{4\pi}(r+s)^{\frac{3}{2}}\sqrt{(1-r)(1-s)}} \tag{204}$$

and it can be shown that this function is integrable:

$$\iint_{[0,1]^2} \frac{\mathrm{d}r\mathrm{d}s}{\sqrt{4\pi}(r+s)^{\frac{3}{2}}\sqrt{(1-r)(1-s)}} = \sqrt{\frac{\pi}{2}} \tag{205}$$

Further, for all $r, s \in [0,1]^2 \setminus \{(0,0)\}$ and for instance $\tau \geq 4$:

$$\sqrt{\tau} I_1(2(\tau - \sqrt{\tau})(r+s)) \leq \frac{1}{\sqrt{4\,\pi(1-\frac{1}{\sqrt{\tau}})(r+s)}} \leq \frac{\sqrt{2}}{\sqrt{4\,\pi(r+s)}} \tag{206}$$

and

$$\frac{1}{\sqrt{\left(\frac{1}{1-\frac{1}{\sqrt{\tau}}} - r\right)\left(\frac{1}{1-\frac{1}{\sqrt{\tau}}} - s\right)}} \leq \frac{1}{\sqrt{(1-r)(1-s)}} \tag{207}$$

Hence for all $\tau \geq 4$, the integrand is dominated by its limit times $\sqrt{2}$.

In conclusion, we have the main contribution term

$$\iint_{I_3} \hat{q}(u)\hat{q}(v) M_1(2\tau - u - v)\mathrm{d}u\mathrm{d}v \sim \frac{\alpha^2}{\sqrt{2\pi\tau}} \tag{208}$$

I.5. Asymptotic analysis conclusion

We have seen the case $\lambda < 1$ in (179) and $\lambda > 1$ in (184). So compared to the first case $\lambda < 1$, the convergence towards the limit is reached with an exponential term $\exp\{-(1-\frac{1}{\sqrt{\lambda}})^2\tau\}$ in the asymptotic limit for $\lambda > 1$. It confirms the result that the convergence happens faster as λ grows to infinity, and that the exponential term vanishes as λ gets close to 1 - with an additional singularity in the denominator.

Appendix J. Additional experiments

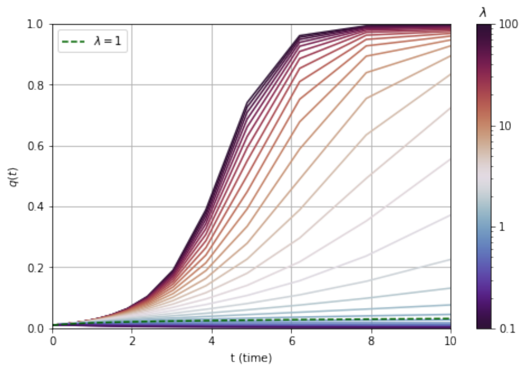
A python notebook is available in the supplementary material to reproduce all the examples.

J.1. Limiting gradient descent

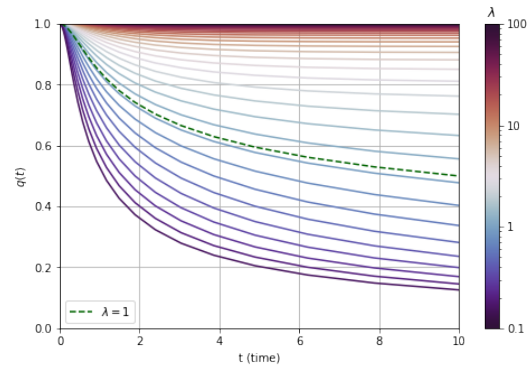
We illustrate the predicted time evolution for cases α very close to 0 and α very close to 1 in figure 5. Since $\alpha = 0$ leads to a null overlap evolution, a slight non-zero initial value of α is required to initiate the learning algorithm. The smaller the α the more is the asymptotic regime delayed. The opposite case $\alpha = 1$ brings another insight, namely when $\theta_0 = \pm\theta^*$ the effect of the noise inexorably disturbs the signal towards a lower limiting overlap (for $\lambda < \infty$).

J.2. Comparison with experimental gradient descent algorithm

The theoretical gradient descent prediction is compared with the experimental values when taking the data dimension n sufficiently large over multiple runs with new samples of the noise matrix. Discrete step size gradient descent is performed while keeping θ_t on $\mathcal{S}_d(\sqrt{n})$. We choose a $\delta_t > 0$ sufficiently small and consider discrete times $t_k = k\delta_t$ for $k \in \mathbb{N}$. We update θ_{t_k} in two steps: first

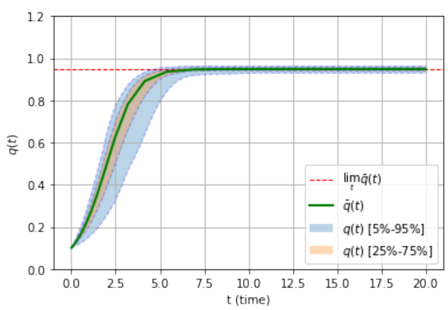


(*a*) $\alpha = 0.01$

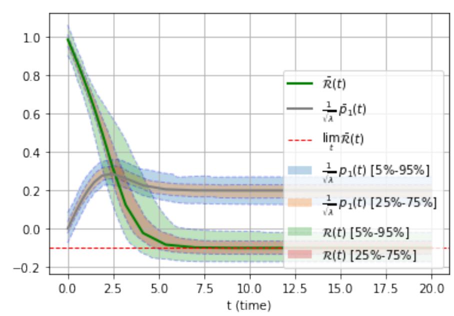


(*b*) $\alpha = 1.0$

Figure 5: Comparison of the overlap over time with different configurations of λ parameter, and between two different values of α.



(*a*) Overlap



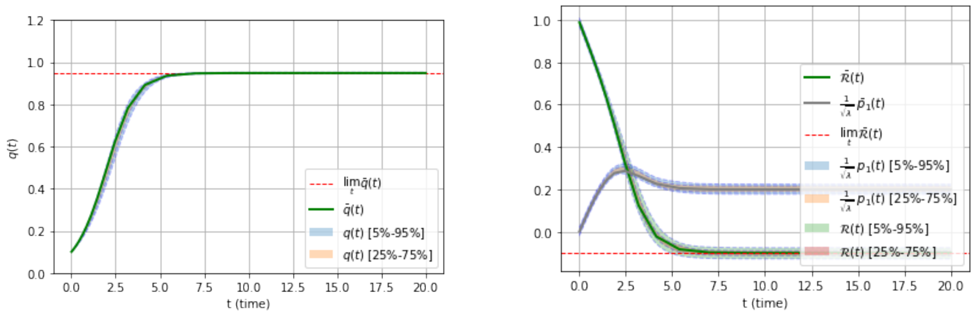
(*b*) Cost

Figure 6: $\lambda = 10, n = 70, \alpha = 0.1, \delta_t = 0.1$

with the gradient descent $\theta_{t_k+\frac{\delta t}{2}} = \theta_{t_k} - \eta\delta_t\nabla\mathcal{H}(\theta_t)$, and secondly projecting back on the sphere $\theta_{t_{k+1}} = \sqrt{n}\theta_{t_k+\frac{\delta t}{2}}\|\theta_{t_k+\frac{\delta t}{2}}\|^{-1}$. These steps are implemented using Tensorflow in Python and run seamlessly on a standard single computer configuration. The initial vectors θ_0 and θ^* are chosen deterministically as $\sqrt{n}\theta_0 = \alpha \mathrm{e}_1 + \sqrt{1-\alpha^2}\mathrm{e}_2$ and $\sqrt{n}\theta^* = \mathrm{e}_1$ with $(e_i)_{1\leq i\leq n}$ the canonical basis of $\mathbb{R}^n$, while the noise matrix H is generated randomly. To account for the randomness of H at each execution, we perform 100 runs and give the quantiles for quantities of interest.

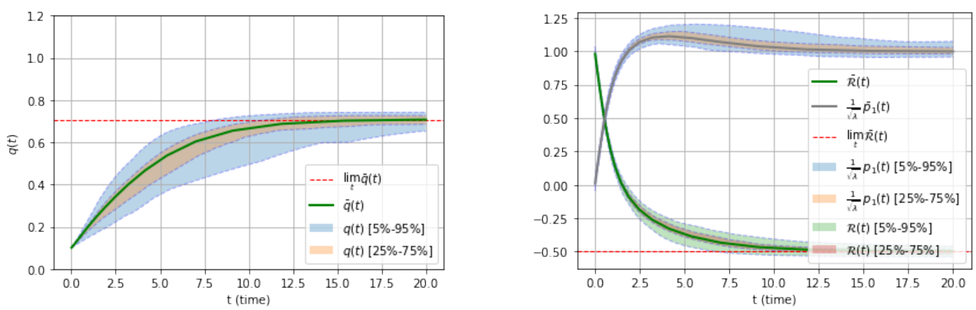
As shown in figure 6, the learning curve matches the theoretical limiting curve with some fluctuations. As illustrated below, these fluctuations diminish as n is increased. Noticeably, in the regime where $\lambda > 1$, smaller values of λ require higher values of n to keep the same concentration. Therefore, the formula from theorem 1 provides a good theoretical framework to predict the behavior of the experimental learning algorithm. Such formulas potentially allow to benchmark the time-evolution of gradient descent techniques and provide guidelines for early-stopping commonly used in machine learning.

We provide a range of further different experiments for different values of λ, α, n.



(a) Overlap (b) Cost

Figure 7: $\lambda = 10, n = 1000, \alpha = 0.1, \delta_t = 0.1$

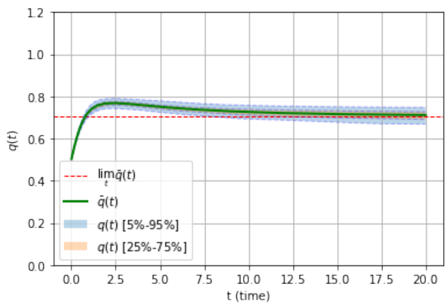


(a) Overlap (b) Cost

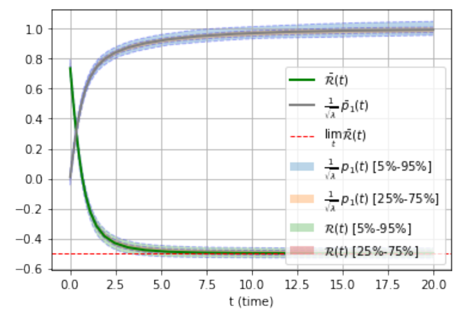
Figure 8: $\lambda = 2, n = 1000, \alpha = 0.1, \delta_t = 0.1$

Let us first comment the regime $\lambda > 1$ illustrated on figures 7, 8, 9. Figure 7 clearly shows that increasing n up to 1000 concentrates the experimental curves around the expect limiting overlap and cost $\bar{q}, \bar{\mathcal{H}}$. We also see even more clearly the characteristic change of p_1 with a "self-healing" process at some specific point in the dynamics of the learning algorithm (recall that p_1 is a similarity measure between the reconstructed matrix $\theta_t \theta_t^T$ and the noise matrix H). This is also seen in Figures 8 and 9 for different values of λ and α. Figures 7 and 8 only differ in the value λ: we observe that decreasing this parameter closer to 1 not only decreases the overlap, but also increases the deviation from the limiting theoretical overlap $\bar{q}$ - and thus as λ decreases higher values of n would thus be needed to match closely $\bar{q}$.

Finally, in the regime $\lambda < 1$, we observe on Figure 10 that similarity measure p_1 explodes and overtakes the risk.

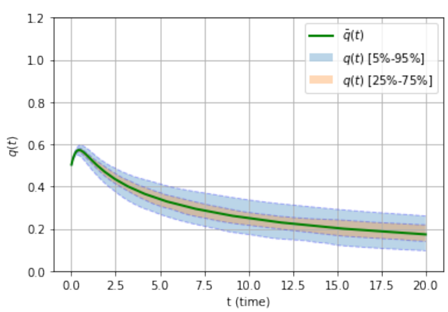


(*a*) Overlap

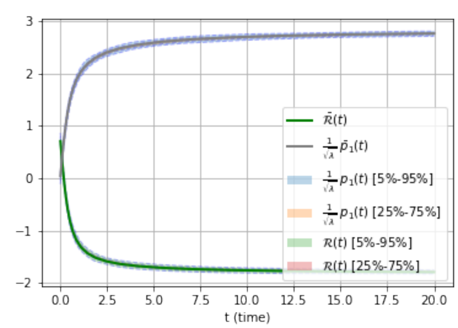


(*b*) Cost

Figure 9: $\lambda = 2, n = 1000, \alpha = 0.5, \delta_t = 0.1$



(*a*) Overlap



(*b*) Cost

Figure 10: $\lambda = 0.5, n = 1000, \alpha = 0.5, \delta_t = 0.1$. Note the different scale for the cost.