

# Variable Muon- $a(t)$ , Boltzmann and HJB: The Control Layer Above the Paquette Volterra Equation

Anonymous

July 3, 2026

## Abstract

This companion studies the control layer in which the Muon exponent  $a$  is allowed to vary in time. The Paquette Volterra equation gives the state dynamics. AMP gives the instantaneous spectral denoiser. HJB adds the price of changing today's exponent on future risk and BBP margins. Boltzmann is the entropy-regularized version of that control problem. The theorem concerns this reduced spectral-control model. A comparison theorem against all online SQ algorithms is kept separate.

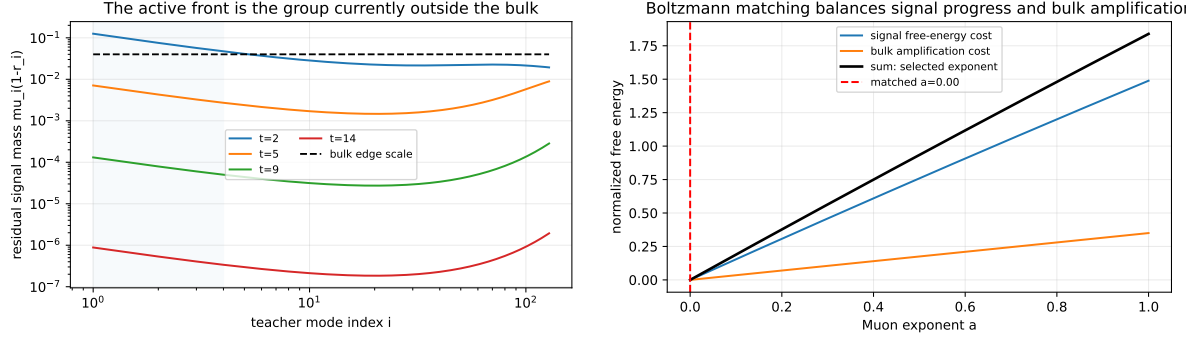
**Relation to the main paper.** This document is the variable-exponent companion to *Muon as Dynamic Spectral Denoising for Power-Law Phase Retrieval*. It keeps the Boltzmann, HJB and AMP-projection layer separate from the fixed- $a$  Volterra calculations. The controlled state is built on the Paquette Muon/Volterra dynamics [1]; the local spectral projection is the linearized multi-index AMP principle of [2], read through the VAMP denoising viewpoint [4]; and the dynamical reduced-state perspective is compatible with multi-index effective dynamics such as [3].

**How to read it.** The first section fixes the meaning of  $b$ ,  $c$  and the action  $A_t$ . Sections 2–3 explain why AMP suggests a local Muon power on a power-law front. Sections 4–6 add future BBP margins through HJB and explain the Boltzmann rule. The final section separates reduced optimality, which is proved inside the model, from global online/SQ optimality, which remains a separate comparison problem.

**Summary and outline.** This companion starts after the fixed- $a$  Volterra state is already known. It asks a different question: if the visible spectral front moves, how should the exponent  $a$  move with it? The local answer is spectral denoising: AMP/VAMP selects the best filter for the current gradient channel, and a power-law front projects that filter onto a Muon power. The nonlocal answer is control: HJB prices how today's exponent changes later risk and later BBP margins. Boltzmann is the entropy-regularized policy obtained by combining these two prices. The chain is:

$$\begin{aligned} \text{Volterra state} &\longrightarrow \text{local AMP cost } F_t(a) \longrightarrow \text{future action } A_t(a) \\ &\longrightarrow \text{Boltzmann/HJB rule for } a(t). \end{aligned}$$

The point is not that Boltzmann is automatically globally optimal. The point is that it is the reduced physical rule obtained when the optimizer is allowed to look at the current spectral front and at the future cost of uncaptured BBP events.



(a) The exponent is chosen from the active front: (b) Boltzmann converts the local denoising cost and the modes near the current BBP boundary carry the future HJB action into a smooth distribution over candidate exponents.

Figure 1: The variable- $a$  story before the formulas. AMP/VAMP gives the instantaneous spectral denoiser on the active front; HJB prices future BBP margins; Boltzmann is the entropy-regularized policy over  $a$ .

**What changes compared with fixed  $a$ .** The observables are the same as in the constant-exponent companion, but the phase boundary is now moving because the exponent is part of the state. At each time the active front determines which teacher modes are near visibility, the local AMP cost  $F_t(a)$  asks which Muon power best denoises that front, and the HJB action  $A_t(a)$  asks how the same choice changes later BBP contacts. A successful online rule is therefore judged by spectral predictivity: it should move  $a(t)$  so that the predicted gradient, weight and Hessian BBP times match the observed exits without requiring a posteriori branch switches.

Symbol	Literal meaning	Interpretation
$X_t$	reduced Paquette/Volterra state	contains masses, residuals, bulk transforms, Schur roots and BBP margins
$b(t, X, a)$	vector field of the controlled state	says how choosing exponent $a$ moves $X_t$ immediately
$F_t(a)$	instantaneous AMP spectral cost	measures how well Muon- $a$ denoises the current gradient channel
$A_t(a)$	HJB action or future shadow cost	measures how today's $a$ changes later risk and BBP margins
$c_t$	action gauge independent of $a$	can shift all actions without changing the policy
$\beta_t$	inverse temperature	large $\beta_t$ makes the policy near-greedy, small $\beta_t$ smooths it
$Z_t$	partition function normalizing the Gibbs policy	turns $F_t + A_t$ into a probability distribution over exponents

Table 1: Control-layer dictionary. Boltzmann is a Gibbs rule over  $a$ , not an additional optimizer primitive.

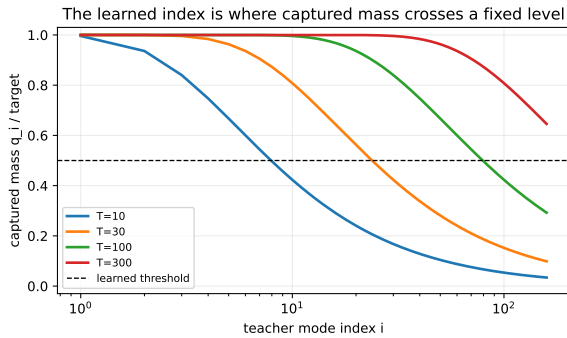
## The four online rules being compared

The variable- $a$  problem is easy to misread because several curves can look similar while meaning different things. The paper separates four rules.

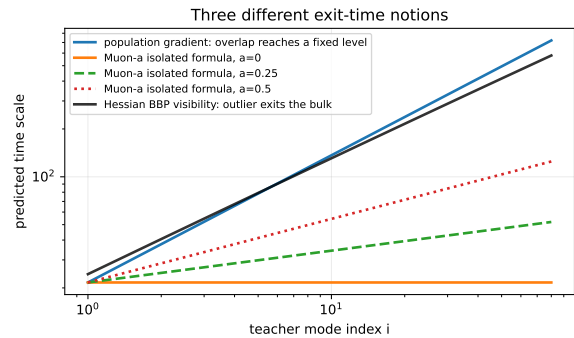
Rule	Definition	What it tests
Myopic AMP projection	minimize $F_t(a)$ at the current state	the best one-step spectral denoiser inside the Muon power family
HJB optimum	minimize $F_t(a) + A_t(a)$	the one-step choice after adding the future price of risk and BBP margins
Boltzmann	sample or average with density proportional to $\exp[-\beta_t(F_t(a) + A_t(a))]$	the entropy-regularized version of the HJB choice
Affine Boltzmann	replace $A_t(a)$ by $c_t + \lambda_t a$ on the Gibbs support	the practical low-dimensional closure tested in the figures

The important point is that these rules are not judged only by final loss. They are judged by the spectral story they produce. A good rule should expose the gradient front, keep the weight Schur roots outside the bulk at the predicted times, and avoid damaging the later Hessian margins. This is why the most informative plots are the  $a(t)$  trajectories together with the Schur spectra and BBP exit times.

Question	Theoretical answer in this companion	Experimental readout
What is optimized locally?	the AMP/VAMP spectral objective $F_t(a)$	gradient spectrum and signal-to-bulk Rayleigh quotient
What prices the future?	the HJB action $A_t(a) = \nabla_X V \cdot b - c_t$	replayed BBP margins and terminal risk curves
Why Boltzmann?	entropy-regularized minimization of $F_t(a) + A_t(a)$	smooth $a(t)$ curve and stable Schur branch timing
When is affine Boltzmann valid?	when $A_t(a)$ is nearly affine on the Gibbs support	small action residual on the selected exponent window
When is scalar Muon insufficient?	when the active spectral front has several incompatible slopes	mismatch between one $a(t)$ and blockwise Schur/BBP exits



(a) The online rule follows the learned front, not a fixed teacher index. As training proceeds, the relevant slope is the slope of the modes near the current spectral visibility boundary.



(b) The same  $a(t)$  affects several times: population progress, Muon front motion, and BBP exit from the bulk.

Figure 2: How to judge a variable exponent. The selected  $a(t)$  is useful when it improves the sequence of visible spectral exits, not merely when it decreases a one-step surrogate.

**Theorem 1** (Reduced online control layer). *Assume a Paquette/Volterra reduced state  $X_t$  is available and that the gradient channel has a bulk law  $\nu_{B,t}$  and a signal/front measure  $\mu_{S,t}$ . Write*

$$R_t = \frac{d\mu_{S,t}}{d\nu_{B,t}}$$

*for the local signal-to-bulk likelihood ratio. Then the variable-exponent layer has the following deterministic components.*

1. *The instantaneous spectral cost is  $F_t(a) = -\log \mathcal{H}_t(\phi_{a,\eta})$ , where  $\mathcal{H}_t$  is the Rayleigh quotient in (3).*
2. *If  $R_t$  is log-linear on the active front, the AMP/VAMP optimal spectral filter projects to a Muon power  $a_{\text{loc}}(t)$ .*
3. *If  $V(t, X)$  is a terminal value function penalizing risk and uncaptured BBP margins, the future price of choosing  $a$  is the HJB action  $A_t(a) = \nabla_X V(t, X_t) \cdot b(t, X_t, a) - c_t$ .*
4. *The entropy-regularized online rule is the Gibbs policy*

$$\pi_t(da) \propto \exp[-\beta_t(F_t(a) + A_t(a))]da.$$

*The theorem is a statement about the reduced control problem. It does not assert global optimality among all online algorithms, and it does not assert that a single scalar  $a(t)$  is optimal for every non-power-law spectrum.*

*Proof.* The first item is the Cauchy–Schwarz optimal spectral denoising problem, restricted to the Muon one-parameter family. The second item is the local regular-variation reduction: a log-linear likelihood ratio is a power on the active spectral window. The third item is the standard HJB linearization of future terminal cost along the controlled vector field  $b$ . Adding entropy to this one-step control problem gives the Gibbs policy in the fourth item.  $\square$

## 1 What are $b$ and $c$ ?

The equation

$$A_t(a) = \nabla_X V(t, X_t) \cdot b(t, X_t, a) - c_t \tag{1}$$

does not contain two universal numerical constants.

The symbol  $b(t, X, a)$  is the vector field of the Paquette/Volterra state. For the scalar phase-retrieval variables it contains, for each mode,

$$b_i(t, X, a) = -2\eta \delta_i^{(a)} Q^{(a-1)/2} q_i^2 + \eta^2 \nu_i^{(a)} Q^a, \tag{2}$$

and

$$b_Q(t, X, a) = \frac{1}{2} \sum_i \mu_i b_i(t, X, a).$$

The other entries of  $b$  are the deterministic updates of the resolvents, Schur kernels and BBP margins induced by the same Paquette state. In other words,  $b$  is the drift of the controlled ODE

$$\dot{X}_t = b(t, X_t, a_t).$$

The scalar  $c_t$  is not a physical parameter. It is a gauge. It can be chosen as 0, as a Gibbs average, or as the intercept of the best affine projection. It does not change the policy because it is independent of  $a$ :

$$\exp[-\beta(F_t(a) + A_t(a))] = e^{\beta c_t} \exp[-\beta(F_t(a) + \nabla_X V \cdot b(t, X_t, a))].$$

Thus the only meaningful quantities are differences in  $A_t(a)$ , its slope on the active Gibbs support, and its nonlinear residual. In the affine reduction

$$A_t(a) \simeq c_t + \lambda_t a,$$

$c_t$  is just the intercept, while  $\lambda_t$  is the shadow price. In the finite-dimensional checks below, the estimated median  $\lambda_t$  is about 16.8, 17.3, 17.5 for  $d = 40, 56, 72$ , close to the reference value 18. The intercept  $c_t$  itself is not invariant.

## 2 Instantaneous AMP free energy

At a fixed state  $X_t$ , write the gradient channel as bulk plus signal,

$$G_t^{\text{cav}} = B_t + S_t.$$

Let  $\nu_{B,t}$  be the bulk singular-value law and let  $\mu_{S,t}$  be the finite signal/front measure, including Schur residues. The AMP likelihood ratio is

$$R_t(\sigma) = \frac{d\mu_{S,t}(\sigma)}{d\nu_{B,t}}(\sigma).$$

For a spectral filter  $f$ , the local signal-to-noise Rayleigh quotient is

$$\mathcal{H}_t(f) = \frac{\langle f, R_t \rangle_{\nu_B}^2}{\langle f^2 \rangle_{\nu_B}}. \quad (3)$$

By Cauchy–Schwarz, the unconstrained local AMP filter is

$$f_{\text{AMP},t} \propto R_t.$$

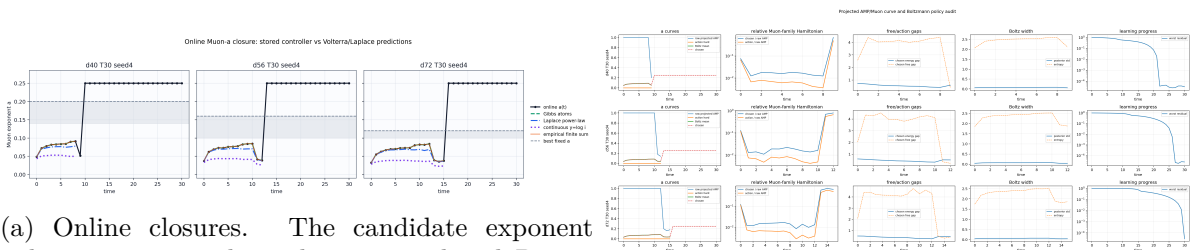
Muon restricts the filter to

$$\phi_{a,\eta}(\sigma) = \sigma(\sigma^2 + \eta^2)^{(a-1)/2}.$$

The instantaneous Muon projection is therefore

$$a_{\text{AMP}}(t) = \arg \max_a \mathcal{H}_t(\phi_{a,\eta}), \quad F_t(a) = -\log \mathcal{H}_t(\phi_{a,\eta}). \quad (4)$$

Thus  $F_t(a)$  has a concrete interpretation: it is the instantaneous AMP denoising cost restricted to Muon powers.



(a) Online closures. The candidate exponent rules are compared on the same reduced Paquette/Volterra state, so discrepancies come from the control layer rather than from a different training clock.

(b) AMP/Boltzmann comparison. The overlay checks whether the Boltzmann rule behaves like the projected AMP spectral denoiser along the active front.

Figure 3: The instantaneous layer:  $F_t(a)$  selects the spectral denoiser seen by the gradient channel.

### 3 Local power law and Muon

If, on the active front,

$$\log R_t(\sigma_t e^u) = \kappa_t + a_{\text{loc}}(t)u + o(1), \quad |u| \leq L, \quad (5)$$

then  $R_t$  is locally a power. The deterministic projection lemma gives

$$a_{\text{AMP}}(t) = a_{\text{loc}}(t) + o(1)$$

up to endpoint clipping. In Paquette's log- $x$  convention, the bulk slope is read from the Step-4 fixed point:

$$\alpha_{B,t}^{\text{Paq}} = 2 + 2x \frac{\Im m'_t(x + i0)}{\Im m_t(x + i0)}. \quad (6)$$

The local Muon exponent is the signal slope minus the bulk slope:

$$a_{\text{loc}}(t) = \alpha_{S,t}^{\text{Paq}} - \alpha_{B,t}^{\text{Paq}}. \quad (7)$$

For non power-law spectra this statement is blockwise: one scalar  $a(t)$  is optimal only if one front block dominates, or if all active blocks have the same local slope.

#### Where the AMP claim comes from

The exact backbone is not vanilla AMP for every possible spectrum. Vanilla AMP is the i.i.d. Gaussian/subgaussian story. For a nontrivial bulk spectrum we read the claim through VAMP/OAMP or, equivalently at first order, through optimal spectral denoising.

The relevant references are the VAMP/OAMP line Rangan–Schniter–Fletcher, Ma–Ping, and Takeuchi; the general spiked spectral calculus of Benaych–Georges–Nadakuditi and Nadakuditi's OptShrink; and the multi-index linearized-message-passing reference Defilippis–Dandi–Mergny–Krzakala–Loureiro.

Thus the mathematical statement is not

Muon- $a$  is AMP for all spectra.

The correct statement is

$$\boxed{\text{linearized AMP/VAMP optimal spectral filter} \longrightarrow \text{Muon-}a_{\text{loc}}}$$

under local regular variation of the active front. If

$$R_t = \frac{d\mu_{S,t}}{d\nu_{B,t}}$$

is locally log-linear, Muon is asymptotically the right one-parameter projection. If  $R_t$  is curved, oscillatory, or made of several active fronts, Loureiro's objection is correct: a scalar  $a(t)$  is not the exact optimal denoiser. The right object is either blockwise  $a_b(t)$  or the full spectral filter  $f_{\text{opt}} = R_t$ .

### 4 HJB control

The local choice (4) is only the instantaneous optimum, because today's  $a$  changes the future Volterra state. Let  $V(t, X)$  be the value function for a terminal objective such as risk plus uncaptured BBP margins. The entropy-regularized HJB policy is

$$\pi_t(da) = \frac{1}{Z_t} \exp[-\beta_t(F_t(a) + A_t(a))]da, \quad a_t = \int a \pi_t(da), \quad (8)$$

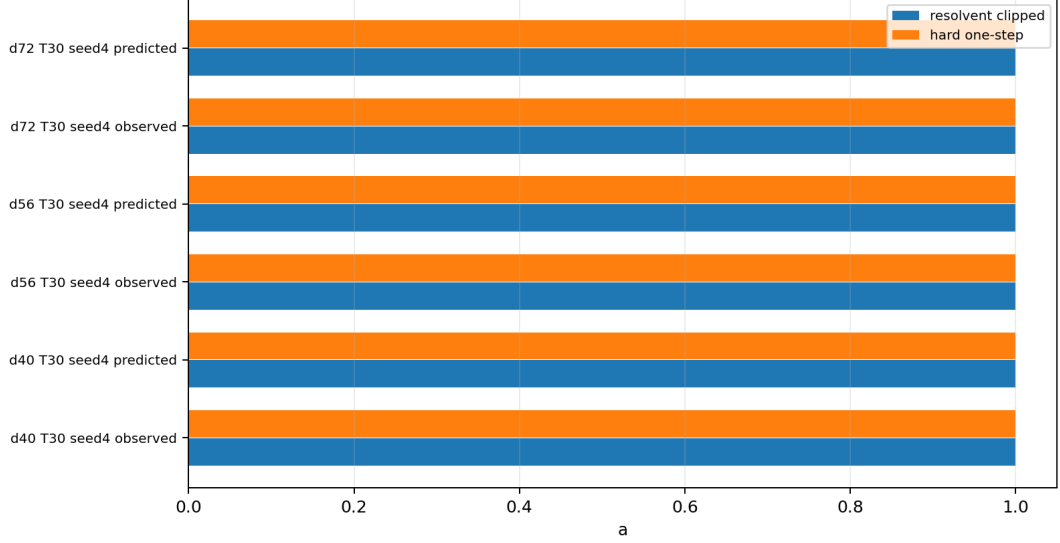


Figure 4: The bulk exponent used in (7) is a resolvent quantity, not a visual histogram comparison.

with

$$Z_t = \int \exp[-\beta_t(F_t(a) + A_t(a))]da$$

and

$$A_t(a) = \nabla_X V(t, X_t) \cdot b(t, X_t, a) - c_t. \quad (9)$$

This is the precise meaning of Boltzmann here. It is HJB with entropy. The temperature  $1/\beta_t$  says how sharply we trust the current reduced state.

If the terminal loss penalizes BBP margins, for example

$$\mathcal{L}_T(X_T) = \sum_{c,i} w_i^c \text{softplus}_\epsilon(-M_i^c(X_T)) + \omega Q(T),$$

then  $A_t(a)$  is a weighted price of future escapes. Away from contacts, if

$$t_j^c = \inf\{t : M_j^c(X_t) = 0\},$$

then a perturbation at time  $s < t_j^c$  gives

$$\frac{\delta t_j^c}{\delta a_s} = - \frac{\nabla_X M_j^c(X_{t_j^c}) \Phi(t_j^c, s) \partial_a b(s, X_s, a_s)}{\frac{d}{dt} M_j^c(X_t)|_{t=t_j^c}}. \quad (10)$$

This formula is the direct bridge between HJB and the spectral BBP plots.

## 5 Affine Boltzmann is a projection, not an automatic theorem

The practical Boltzmann rule used in the experiments replaces the full action by an affine shadow price,

$$A_t(a) \simeq c_t + \lambda_t a.$$

This is valid on the Gibbs support if the nonlinear residual is invisible:

$$\beta_t \sqrt{\text{Var}_{\pi_t}(A_t - c_t - \lambda_t a)} = o(1). \quad (11)$$

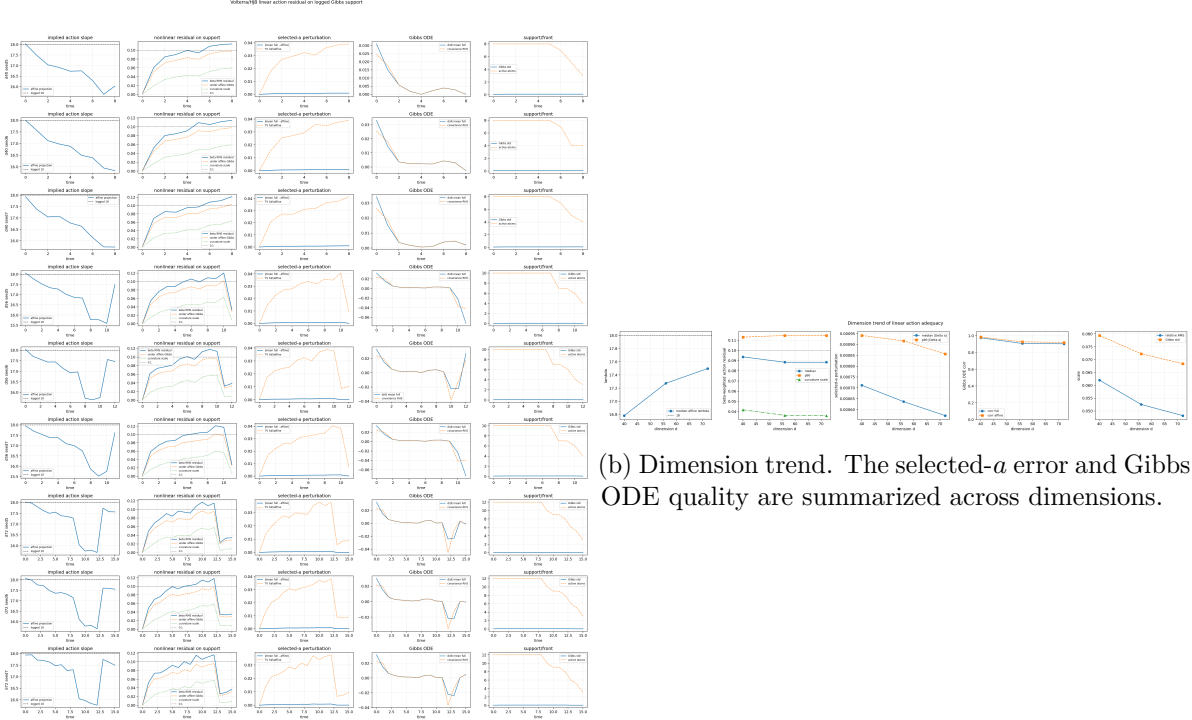
A simple sufficient condition is

$$\beta_t \text{Var}_{\pi_t}(a) \sup_a |\partial_{aa} A_t(a)| = o(1).$$

For the Muon filter,

$$\partial_{aa} \phi_{a,\eta}(\sigma) = \frac{1}{4} \log^2(\sigma^2 + \eta^2) \phi_{a,\eta}(\sigma).$$

So the real proof target is a signed second-log-moment bound on the active front. The numerical evidence shows that this affine projection is accurate in the observed regime, but it remains a condition rather than a theorem of global optimality.



(a) Time trace. The full HJB action is compared with its affine projection along the selected trajectory.

(b) Dimension trend. The selected- $a$  error and Gibbs ODE quality are summarized across dimensions.

Figure 5: Affine-action evidence. The selected- $a$  gap between full and affine action is below  $10^{-3}$  in the tested regimes; the Gibbs ODE correlation is above 0.90.

## 6 Variable $a(t)$ as a Volterra control

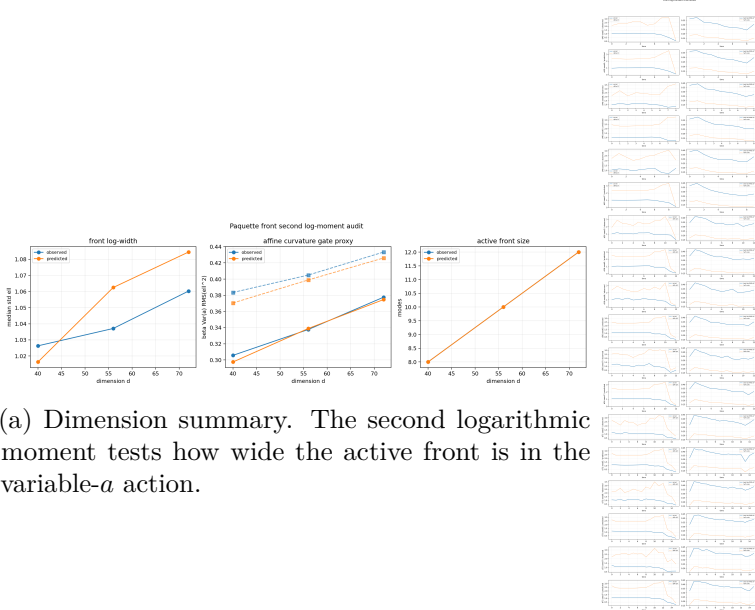
For a nonconstant exponent the primitive equation is not the ODE for  $a_*(t)$ . The primitive object is the non-autonomous Volterra dynamics

$$\dot{q}_i^2(t) = -2\eta \delta_i^{(a_t)} Q(t)^{(a_t-1)/2} q_i^2(t) + \eta^2 \nu_i^{(a_t)} Q(t)^{a_t}. \quad (12)$$

In a slowly moving power-law front, (12) reduces to the adiabatic formula

$$a_*^{PR}(t) = \frac{\alpha_t - 1}{3\alpha_t - 1}, \quad \dot{a}_*^{PR}(t) = \frac{2\dot{\alpha}_t}{(3\alpha_t - 1)^2}. \quad (13)$$

The derivative formula is therefore a consequence of the Volterra front, not the starting assumption.



(a) Dimension summary. The second logarithmic moment tests how wide the active front is in the variable- $a$  action.

(b) Time summary. The same log-moment observable identifies the times at which affine Boltzmann is most exposed to curvature of the action.

Figure 6: The front is not narrow enough by itself. The good affine behavior comes from the signed adjoint/Paquette normalization, not only from a small width.

## 7 Spectral consequences

The HJB state controls the same BBP margins as the constant- $a$  theory:

$$M_i^G(X_t), \quad M_i^W(X_t), \quad M_i^{H,-}(X_t), \quad M_i^{H,+}(X_t).$$

The only difference is that  $a_t$  now changes their future times through (10). This is why the online control rule must be checked on spectral plots, not only on final loss.

## 8 What is proved, and what is not

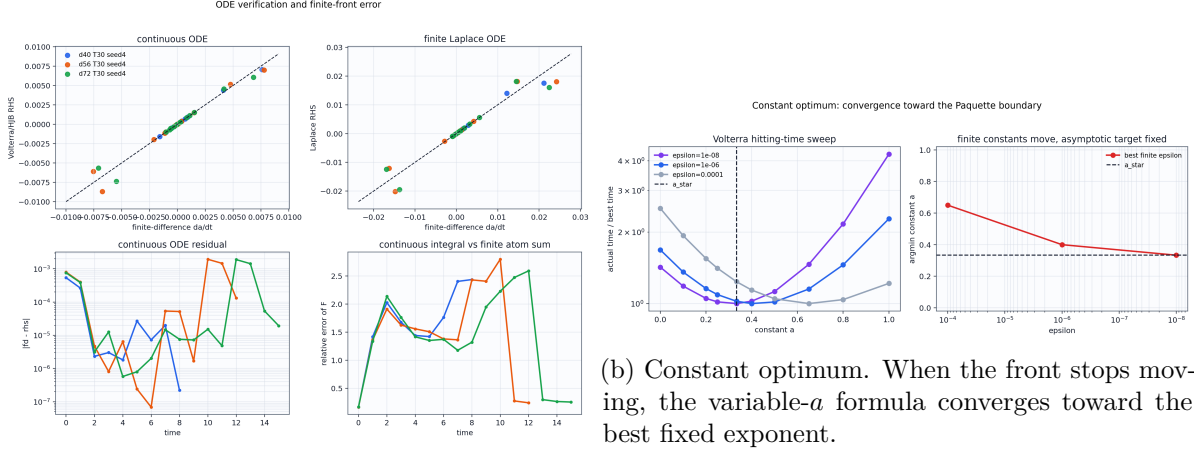
The precise comparison target is:

$$\boxed{\text{Paquette Volterra state} + \text{AMP free energy} + \text{HJB action} \Rightarrow \text{online } a_t \text{ and BBP margins.}}$$

The following pieces are deterministic once the Paquette fresh/frozen equivalents are accepted:

$$F_t(a) = -\log \mathcal{H}_t(\phi_{a,\eta}), \quad A_t(a) = \nabla_X V \cdot b - c_t, \quad \pi_t(da) \propto e^{-\beta_t(F_t(a)+A_t(a))} da.$$

Here “fresh/frozen” means the same thing as in the fixed- $a$  companion: the reduced state is held fixed before the spectral averages are evaluated, or the measuring sample is independent of the sample that generated the state. This convention isolates the control problem. The extra theorem needed for a fully reused online trajectory is the RFA transport estimate stated below.



(a) ODE quality. The reduced continuous rule for  $a_*(t)$  is compared with the discrete front extracted from the Volterra state.

(b) Constant optimum. When the front stops moving, the variable- $a$  formula converges toward the best fixed exponent.

Figure 7: The  $a_*(t)$  rule is read from the Volterra front. The constant optimum is recovered as the front stops moving.

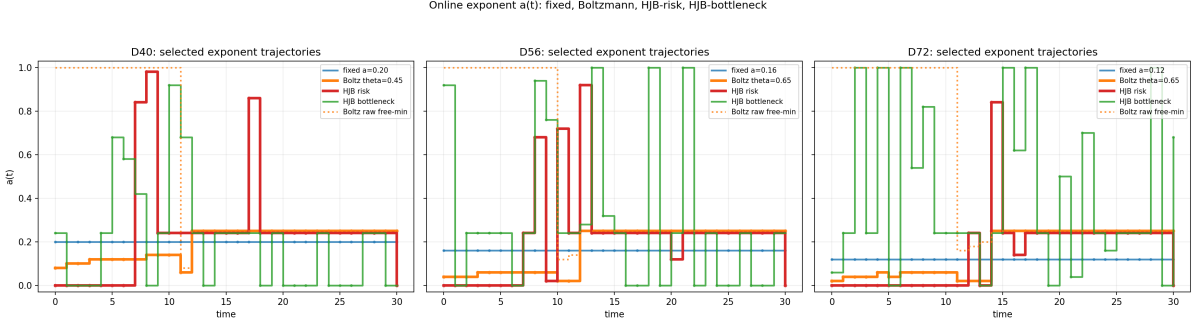
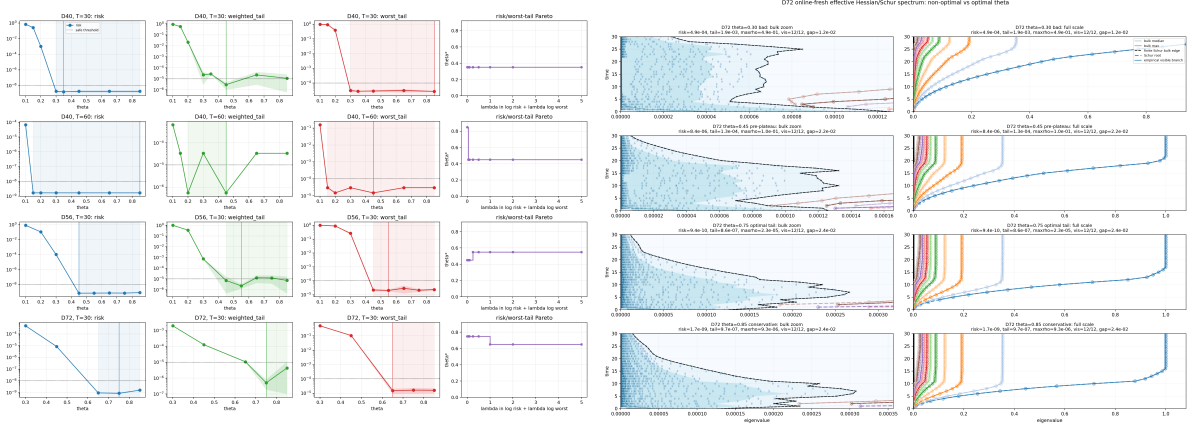


Figure 8: Visible trajectories of  $a_t$ . Boltzmann is stable because the policy averages the local spectral cost and the future action over a Gibbs window, instead of jumping to a single point minimizer.

Layer	Status in this companion	Interpretation
Instantaneous AMP cost $F_t(a)$	deterministic once the gradient-channel bulk and signal measures are known	local spectral denoising problem before future effects are priced
Power-law projection	exact for a log-linear active front, blockwise otherwise	explains why an AMP filter can reduce to a Muon power on a regular tail
HJB action $A_t(a)$	exact inside the chosen reduced control model	prices how today's exponent moves later risk and BBP margins
Boltzmann policy	exact entropy-regularized policy for the reduced cost $F_t + A_t$	smooths the selected exponent instead of taking an unstable hard minimum
Affine action reduction	finite-dimensional checks show a small residual on tested power-law fronts	controlled reduction when the Gibbs support sees an almost linear action
Global online/SQ optimality	separate comparison theorem	not inferred from the one-parameter reduced control model alone



(a) Reduced parameter lock. The scan identifies a stable plateau of parameters that preserve the desired spectral margins. (b) Schur consequence. The selected parameter is evaluated through the branches and margins it produces in the Schur spectrum.

Figure 9: Boltzmann/HJB is judged by the spectral margins it produces. The Schur-margin plots are the most direct check in this manuscript set.

The following pieces remain outside the reduced theorem.

- Boltzmann with affine  $A_t(a) \simeq c_t + \lambda_t a$  is not globally optimal among all online rules. It is the entropy-regularized HJB policy after projecting the full action to an affine shadow price on the Gibbs support.
- A single scalar  $a(t)$  is not AMP-optimal for every non-power-law spectrum. If the active front has several slopes, the appropriate theorem is blockwise Muon or a richer spectral filter.
- The reused-data trajectory still needs the local-law/RFA transport

$$\varepsilon_{\text{RFA},d} = \mathcal{O}_{\mathbb{P}}(1).$$

Thus the conclusion is precise. Boltzmann is the entropy-regularized HJB policy of the reduced spectral-control problem. Its affine version is a controlled reduction on tested power-law fronts because the measured action residual is small on the selected Gibbs support. A theorem of global online or SQ optimality would require an additional comparison result beyond this one-parameter reduced model.

## 9 Conclusion

The variable-exponent layer is a control problem above the fixed- $a$  Volterra state. AMP/VAMP supplies the local denoising objective: which spectral filter best separates signal from bulk now? HJB supplies the nonlocal correction: how does today's exponent change later risk and later BBP margins? Boltzmann is the entropy-regularized way to combine these two prices inside the reduced one-parameter Muon family. This explains why the observed  $a(t)$  trajectories are smooth and why the useful rule need not be the myopic loss minimizer. What remains beyond this companion is not the definition of the reduced policy, but a comparison theorem showing when this one-parameter spectral control is globally optimal among broader online algorithms.

Diagnostic	Reduced object	Interpretation
Online closure curves	$F_t(a) = -\log \mathcal{H}_t(\phi_{a,\eta})$	The instantaneous rule is a spectral denoising objective, not a loss-only calibration.
Bulk exponent panel	$\alpha_{B,t}^{\text{Paq}}$ in (6)	The bulk side of the local power is read from the resolvent.
Action residual panels	affine gate (11)	The affine Boltzmann reduction is justified only on the Gibbs support where the residual is small.
$a_t$ trajectory plot	Gibbs policy (8)	Boltzmann smooths the online rule by averaging over nearby exponents.
Schur spectrum comparison	sensitivity formula (10)	The correct online rule preserves or improves future BBP margins, rather than only reducing current risk.

Table 2: How to read the variable- $a$  figures. Each plot checks a part of the reduced HJB/AMP control layer, while global online optimality remains a separate comparison target.

## References

- [1] E. Paquette, N. Marshall, L. Benigni, G. Wang, A. Agarwala and C. Paquette, *Phases of Muon: When Muon Eclipses SignSGD*, arXiv:2605.09552.
- [2] L. Defilippis, Y. Dandi, P. Mergny, F. Krzakala and B. Loureiro, *Optimal Spectral Transitions in High-Dimensional Multi-Index Models*, arXiv:2502.02545.
- [3] Z. Fan and L. Wang, *High-dimensional learning dynamics of multi-pass Stochastic Gradient Descent in multi-index models*, arXiv:2601.21093.
- [4] S. Rangan, P. Schniter and A. Fletcher, *Vector approximate message passing*, arXiv:1610.03082.