



*Research article***Gibbs normalizability and maximum-entropy programs with bilinear moment constraints on orthants****Abdulaziz M. D. Aljohani* and Khaled M. Alhawiti**Faculty of Computers and Information Technology, University of Tabuk, Tabuk 47512, Saudi Arabia;
khalhawiti@ut.edu.sa* **Correspondence:** Email: a-aljohani@ut.edu.sa; Tel: +966508420903.

Abstract: Gibbs weights with attractive bilinear interactions need not define probability laws on unbounded orthants. This paper gives a maximum-entropy and partition-function classification of that failure. For densities on $[0, \infty)^2$ with prescribed marginal means and product moment $\mathbb{E}[XY]$, the entropy program is well posed exactly when $\mathbb{E}[XY] \leq \mathbb{E}[X]\mathbb{E}[Y]$. In the attractive case $\mathbb{E}[XY] > \mathbb{E}[X]\mathbb{E}[Y]$, the entropy supremum equals the means-only product-exponential value but is not attained: The attractive constraint is entropically invisible while destroying normalizability. We then identify the repairs. Self-moment damping of orders $p, q > 1$ normalizes every attractive bilinear exponent when $1/p + 1/q < 1$ and never normalizes it when $1/p + 1/q > 1$. In the critical case $1/p + 1/q = 1$, normalizability is decided by an explicit inequality between the damping coefficients, which at $p = q = 2$ is the Gaussian condition $\gamma^2 < 4\alpha\alpha'$; when that inequality holds with equality, the linear multipliers decide, and we classify this remaining boundary completely. Hard-wall confinement also repairs the model: Truncation to $[0, K]^2$ restores well-posedness and has sharp entropy deficit $\Delta(\mu_1^{-1} + \mu_2^{-1})/K + o(K^{-1})$. In dimension n , quadratic Gibbs normalizability is governed by copositivity of the negated coupling matrix together with the sign of the linear field on the zero cone, in every regime except a degenerate boundary of transverse codimension at least two, which we leave unresolved; the associated decision problem is co-NP-hard, where co-NP is the class of decision problems whose complements belong to nondeterministic polynomial time (NP). The results give checkable criteria for when bilinear maximum-entropy models on orthants are genuinely solvable.

Keywords: maximum entropy; Gibbs measures; partition function; normalizability; bilinear interactions; confinement; copositive matrices

Mathematics Subject Classification: 94A17, 62B10, 44A60, 15B48, 90C60

1. Introduction

A Gibbs weight is only a statistical-mechanical model when its partition function is finite. This elementary requirement becomes nontrivial on unbounded state spaces, especially when the exponent contains attractive bilinear interactions. The same issue appears in maximum-entropy modeling: Moment constraints formally produce a Gibbs density, but the formal density may fail to define a probability law, and the entropy supremum may fail to be attained.

The solvability question below is encountered, in different dress, by several communities: builders of exponential-family models with pairwise interaction terms on nonnegative data, users of moment methods who impose cross-moment targets on orthants, and statistical mechanics of unbounded continuous spins with attractive couplings, where finiteness of the partition function is the precondition for the model to exist. The classification below gives each of these settings a checkable criterion.

The maximum-entropy principle chooses, among densities satisfying prescribed moment constraints, the density of largest differential entropy. In regular cases, this density has Gibbs form $f \propto \exp(-\sum_i \lambda_i \varphi_i)$; this idea goes back to Jaynes [1] and Shannon [2], and its projection and attainment theory were developed by Csiszár [3] and Csiszár-Matúš [4]; see also [5]. Those results, and the broad sufficient conditions for existence of maximum-entropy distributions on unbounded supports given by Ishwar and Moulin [6], do not, by themselves, classify which constraint sets produce a normalizable Gibbs density. The maximum-entropy principle has also been extended to non-symmetric entropies for discrete families: Contreras-Reyes [7] maximizes a nonsymmetric entropy under Hurwitz-Lerch auxiliary information to obtain the Lerch distribution on the nonnegative integers. That setting is discrete, and the functional is not the Shannon differential entropy, so the failure studied here—divergence of the Gibbs partition function for an attractive bilinear exponent on an unbounded continuous orthant—is not addressed there. The present paper concerns Shannon differential entropy on continuous unbounded orthants, where that failure is exactly what has to be classified. The usual test is still ad hoc: Write the exponent, compute the partition function, and check whether the integral converges.

This paper studies that solvability question for bilinear cross-moment constraints on nonnegative orthants. The motivating example is simple. On $[0, \infty)^2$, prescribe the means of X and Y and the product moment $\mathbb{E}[XY]$, but do not prescribe variances. If the target product moment is below or equal to $\mathbb{E}[X]\mathbb{E}[Y]$, the maximum-entropy problem is well posed. If it is above $\mathbb{E}[X]\mathbb{E}[Y]$, the Gibbs partition function diverges and, more strongly, no maximizing density exists. The entropy supremum is nevertheless the same as in the means-only problem. Thus the attractive constraint is visible in the constraints but invisible in the entropy value.

Main contributions

The paper proves the following results.

- (i) A sharp two-dimensional dichotomy for means and one product moment on $[0, \infty)^2$: Non-attractive targets are attained, while attractive targets have the means-only entropy value but no maximizer (Theorem 3.1).
- (ii) A minimal damping trichotomy for attractive bilinear exponents with self-moment powers p and q : The regimes are determined by $1/p + 1/q < 1$, $= 1$, and > 1 (Theorem 4.1). The critical quadratic case gives the Gaussian condition $\gamma^2 < 4\alpha\alpha'$. On the critical equality manifold itself, finiteness is decided by the linear multipliers, and this boundary is classified completely (Proposition 4.3).

(iii) A truncation theorem: Support restriction to $[0, K]^2$ restores well-posedness, and the entropy deficit is $\Delta(\mu_1^{-1} + \mu_2^{-1})/K + o(K^{-1})$ (Theorem 5.3). The matched coupling satisfies $K\gamma_K \rightarrow \mu_1^{-1} + \mu_2^{-1}$ (Corollary 5.5).

(iv) A quadratic n -dimensional classification through copositivity of the negated coupling matrix, complete except on a degenerate zero-cone boundary of transverse codimension at least two, and including strict copositivity for unconditional normalizability and co-NP-hardness of the decision problem (Theorem 7.3 and Corollaries 7.8 and 7.9), where co-NP is the class of decision problems whose complements belong to nondeterministic polynomial time (NP).

The mechanism behind the main dichotomy is an escape of mass to infinity. A mass ε placed at distance $\varepsilon^{-1/2}$ can carry an $O(1)$ excess in $\mathbb{E}[XY]$, while changing the means by $o(1)$ and the entropy by $o(1)$. This works only on the attractive side because xy is unbounded above on the orthant but bounded below. The same mechanism explains why repulsive couplings cannot subsidize attractive ones in the variance-free n -dimensional case.

Scope

The two-dimensional dichotomy, the damping trichotomy—including its critical equality manifold, which is classified completely in Proposition 4.3—and the truncation asymptotics are complete in the stated regimes. In dimension n , Theorem 7.3 classifies quadratic normalizability in every regime except one: the degenerate zero-cone boundary at transverse codimension two or more. The codimension-zero and codimension-one boundary cases are proved divergent (Theorem 7.3 (d)); at codimension two, both alternatives occur, as shown by the two explicit three-dimensional examples of Remark 7.4, so copositivity together with the sign of β on the zero cone cannot decide that regime. We treat this as part of the structure rather than as a missing sign case.

From the viewpoint of statistical mechanics, the question is when an attractive interaction can be confined by the remaining terms in the energy; linear fields alone cannot confine the diagonal escape directions generated by a positive bilinear term.

The nearest classical relative of this question is the stability theory of unbounded spin systems: Ruelle's superstability condition [8] requires the energy to dominate a positive multiple of $\sum_i x_i^2$ up to an additive constant per site and is the standard confinement hypothesis under which the infinite-volume theory for unbounded spins exists; see also [9]. The present paper works one level down—normalizability of a single finite-dimensional Gibbs density rather than existence of infinite-volume states—and shows that on the orthant the exact confinement requirement is the cone-restricted form of that bound: Strict copositivity of the negated coupling matrix is necessary and sufficient for unconditional normalizability (Corollary 7.8), with the copositive classification, its remaining degenerate boundary, and its computational hardness developed in Section 7.

For background on entropy and large deviations, we use Kullback-Leibler, Sanov, and Donsker-Varadhan [10–12]. Further large-deviation treatments are provided by Dupuis-Ellis and Dembo-Zeitouni [13, 14]. For exponential-family theory, we use Barndorff-Nielsen and Brown [15, 16]. Convex duality for entropy-type problems is treated by Borwein-Lewis, Rockafellar, and Boyd-Vandenberghe [17–19], and Wainwright-Jordan [20] connect the two through the variational representation of exponential families. The classical moment problem is treated by Haviland, Shohat-Tamarkin, and Krein-Nudelman [21–23]. Truncated and semidefinite-moment formulations are represented by Curto-Fialkow and Lasserre [24, 25]. For copositive optimization and positivity, see [26–28]; the complexity result used later is due to Murty and Kabadi [29]. Recent contributions

on moment problems and copositive cones include [30, 31]. Copositivity has very recently been connected to integral convergence in another setting: Feynman integrals converge when their kinematic parameters lie in the copositive cone of the second Symanzik polynomial [32].

The Gibbs family in Section 3 is related to the bivariate exponential-conditionals distribution of Arnold and Strauss [33], with antecedents in Gumbel [34]; the conditional-specification literature [35, 36] records that this family supports only nonpositive correlation. Our result is not a rediscovery of that family. It shows that, at attractive targets, no density of any form attains the entropy supremum, and it computes that supremum exactly. The damped, truncated, and n -dimensional results have no analogue in that conditional-specification setting.

A separate strand uses entropy-derived criteria as learning objectives rather than as a variational principle for a density. Sun, Zhou, and Huang [37] obtain fast rates for an exponential cost function, and Jiang, Xu, Sun, and Huang [38] establish robustness of minimum-error-entropy regression under heavy-tailed noise. These are different problems from ours: There is no partition function to normalize, the object estimated is a regression function rather than a density on an orthant, and the analysis concerns generalization rates and noise robustness rather than existence of a maximizer. The damping and truncation of Sections 4 and 5 are, at most, conceptual analogues of tail control and regularization in that literature; we claim no formal equivalence, and none of our results transfers to those settings or follows from them.

Maximum-entropy existence theory remains active: Henrion [39] develops the maximum-entropy dual of the moment-body membership problem, and Liu and Ma [40] obtain maximum-entropy membership functions under expectation and variance constraints in the setting of uncertain sets.

What was known, what remained open, and what this paper settles

Known: The projection and attainment theory of Csizsár [3,4] and the sufficient existence conditions of Ishwar and Moulin [6] guarantee attainment once the program is well posed or their hypotheses are verified; the conditional-specification literature [33, 35, 36] records that one particular Gibbs family on the quadrant supports only nonpositive correlation; and copositivity was recently shown to govern integral convergence in a different setting [32].

Open: A classification of which bilinear cross-moment targets on an unbounded orthant yield a well-posed maximum-entropy program at all, the value and behavior of the entropy supremum when they do not, and the exact confinement required to repair the failure. This paper settles those questions in the regimes stated under Scope above; the degenerate zero-cone boundary in dimension n remains open and is stated as such.

The paper is organized as follows. Section 2 fixes notation. Section 3 proves the two-dimensional dichotomy. Section 4 proves the damping trichotomy and the augmented attainment statement. Section 5 proves the truncation results and gives a numerical illustration. Section 6 discusses the Gaussian and truncated-exponential autoregressive examples. Section 7 gives the n -dimensional copositive classification and delimits the degenerate boundary it leaves open. The final sections collect limitations, open problems, conclusions, and a dependency map for the main results.

Tables 1 and 2 summarize how the present work is positioned relative to the principal related lines of research: Table 1 compares the problems and assumptions, while Table 2 contrasts the corresponding normalizability and attainment results.

Table 1. Position of the paper relative to identifiable lines of work (part I: problem treated).

Work or literature	Problem and state space	Main assumptions	Attractive bilinear interaction treated?
Jaynes [1], Shannon [2]; Cover-Thomas [5] Csiszár [3], Csiszár-Matúš [4]	Maximum entropy under moment constraints, general spaces I -projections, generalized I -projections and duality, general measure spaces	Constraint set admits a Gibbs solution Convexity and closedness of the constraint set; regularity for the classical statements, relaxed in the generalized theory	No No
Ishwar-Moulin [6]	Existence of maxent laws on unbounded supports	Sufficient growth and moment-inequality conditions	Not as a separate case
Arnold-Strauss [33], Gumbel [34], Arnold-Castillo-Sarabia [36] Ruelle [8], Georgii [9]	Bivariate laws with exponential conditionals on Q	Conditional specification within a fixed family	Yes, at the level of the family
Copositive optimization [26–28]; complexity [29] Henrion [39]	Infinite-volume Gibbs states, unbounded continuous spins Positivity of quadratic forms on Q_n ; complexity	Superstability: energy dominates $c \sum_i x_i^2$ Symmetric matrix data; no integral is formed	Yes, via confinement Not applicable
Sun-Zhou-Huang [37], Jiang-Xu-Sun-Huang [38] Present manuscript	Maximum entropy dual of moment-body membership Entropy-related learning objectives, rates, and robustness Normalizability and maxent attainment for bilinear cross-moment constraints on Q and Q_n	Compact or otherwise regular moment body Hypothesis space, sampling model, noise tails Prescribed means and cross-moments; optional self-moments or a box	No No Yes, as the central case

Table 2. Position of the paper relative to identifiable lines of work (part II: result obtained and distinction from this paper).

Work or literature	Normalizability or attainment result	Distinction from the present manuscript
Jaynes, Shannon; Cover-Thomas Csiszár, Csiszár-Matúš	Gibbs form of the maximizer when one exists I -projection and duality theory, with generalized I -projections and attainment criteria that extend well beyond the regular case, including situations where the classical projection is not attained	Existence is presupposed; normalizability is not classified That theory does not say which bilinear cross-moment targets on an orthant admit a normalizable Gibbs density; our contribution is the explicit classification of the bilinear family and the exact value of the entropy supremum at the non-attained targets
Ishwar-Moulin	Sufficient conditions for a maxent law to exist	Sufficient, not necessary; our dichotomy is exact in two dimensions
Arnold-Strauss, Gumbel, Arnold-Castillo-Sarabia Ruelle, Georgii	The normalizable members of that family have nonpositive correlation Superstability gives infinite-volume existence	We rule out maximizers of any form at attractive targets and compute the entropy value We work one level down, on a single finite-dimensional density, and obtain the cone-restricted form: strict copositivity
Copositive optimization	Complexity of copositivity; NP-completeness of its complement	We identify copositivity as the exact normalizability criterion and inherit co-NP-hardness
Henrion	Maximum entropy over the moment body	Different question: we ask which constraint sets give a well-posed program at all
Sun-Zhou-Huang, Jiang-Xu-Sun-Huang Present manuscript	Fast rates and robustness for entropy-based losses Two-dimensional dichotomy; damping trichotomy with its equality manifold; exact truncation price; copositive classification in dimension n	Different problem: no partition function is at issue; see the discussion in Section 1 —

2. Setup and notation

Throughout, $Q = [0, \infty)^2$ denotes the closed nonnegative orthant equipped with Lebesgue measure, and $\ln$ denotes the natural logarithm. For a probability density g on Q , we write

$$h(g) = - \int_Q g \ln g$$

for its differential entropy (with the convention $0 \ln 0 = 0$).

We use the standard asymptotic notation $O(\cdot)$, $o(\cdot)$ and $\Theta(\cdot)$ for upper, strictly smaller, and matching two-sided orders of magnitude and write $a \asymp b$ when $a = \Theta(b)$. The symbol Θ carrying an argument always denotes an order of magnitude in this sense; the same letter without an argument is used from Section 3 onward for the natural parameter space of an exponential family, and the two usages are never combined.

Fix targets $\mu_1, \mu_2 > 0$ and $C > 0$ and define the feasible set

$$\mathcal{F}(\mu_1, \mu_2, C) = \left\{ g \text{ density on } Q : \mathbb{E}_g[X] = \mu_1, \mathbb{E}_g[Y] = \mu_2, \mathbb{E}_g[XY] = C \right\}.$$

Definition 2.1 (Well-posedness). The maximum-entropy program

$$\sup\{h(g) : g \in \mathcal{F}(\mu_1, \mu_2, C)\} \tag{2.1}$$

is well posed if the supremum is finite and attained by a unique density. We call the target (μ_1, μ_2, C) repulsive if $C < \mu_1\mu_2$, critical if $C = \mu_1\mu_2$, and attractive if $C > \mu_1\mu_2$.

These words are used at two related but distinct levels, and we keep them apart throughout.

At the level of moment targets, as in Definition 2.1, the benchmark is the product value $\mu_1\mu_2$ that independent coordinates with the prescribed means would give: A target is repulsive or attractive according as the prescribed covariance $C - \mu_1\mu_2$ is negative or positive.

At the level of Gibbs parameters, the same words describe the sign of the bilinear term in the exponent. In the convention (2.2), the exponent carries $-\beta_3xy$, so a coupling $\beta_3 = -\gamma < 0$ contributes $+\gamma xy$ and is called attractive: It raises the density where x and y are simultaneously large and so favors large values of one coordinate in the presence of large values of the other. A coupling $\beta_3 > 0$ is repulsive in the same sense.

That the two usages agree is a theorem, not a convention: Lemma 3.3 shows that a normalizable Gibbs density with $\beta_3 \geq 0$ realizes only non-attractive targets, and Proposition 4.9 (iv) gives the corresponding statement for the damped family. Where the level is not clear from context, we say “attractive target” or “attractive coupling” explicitly.

Any interior maximizer of (2.1), if one exists, has Gibbs form

$$f(x, y) = Z^{-1} \exp(-\beta_1x - \beta_2y - \beta_3xy), \quad Z = \int_Q e^{-\beta_1x - \beta_2y - \beta_3xy} \, dx \, dy, \tag{2.2}$$

but our main negative result will be proved without assuming the Gibbs form by working directly with the value of (2.1).

We record a classical fact used repeatedly.

Lemma 2.2 (Concavity of differential entropy). *For densities f, g on Q and $\varepsilon \in [0, 1]$,*

$$h((1 - \varepsilon)f + \varepsilon g) \geq (1 - \varepsilon)h(f) + \varepsilon h(g).$$

This is standard; see [5]. It is the integrated form of the pointwise concavity of $u \mapsto -u \ln u$ on $[0, \infty)$: Applying that concavity to $u = (1 - \varepsilon)f(x, y) + \varepsilon g(x, y)$ at each point, which is Jensen's inequality for the two-point mixture, and integrating over Q gives the stated bound.

3. The dichotomy theorem

Theorem 3.1 (Dichotomy for bilinear cross-moment programs). *Let $\mu_1, \mu_2 > 0$ and $C > 0$.*

(a) **Repulsive case** ($0 < C < \mu_1 \mu_2$): *The program (2.1) is well posed. The unique maximizer is of Gibbs form (2.2) with $\beta_1, \beta_2 > 0$ and $\beta_3 > 0$.*

(b) **Critical case** ($C = \mu_1 \mu_2$): *The program is well posed, the unique maximizer being the independent product of exponentials with means μ_1, μ_2 (i.e., $\beta_3 = 0$), and*

$$\sup_{g \in \mathcal{F}} h(g) = 2 + \ln(\mu_1 \mu_2).$$

(c) **Attractive case** ($C > \mu_1 \mu_2$): *The program is ill posed. More precisely, no maximizer exists, and*

$$\sup_{g \in \mathcal{F}(\mu_1, \mu_2, C)} h(g) = 2 + \ln(\mu_1 \mu_2), \quad (3.1)$$

the same value as in the critical case: The attractive cross-moment constraint is entropically invisible—it lowers the value of the program by nothing yet destroys attainment.

The proof occupies the remainder of this section. We first dispose of the Gibbs family (Lemmas 3.2 and 3.3), then prove part (c) by a direct variational argument, and finally construct the maximizer in parts (a) and (b) through a one-parameter exponential family.

3.1. The Gibbs family under attraction: divergence and the sign obstruction

Lemma 3.2 (Divergence of the attractive partition function). *Let $\beta_3 = -\gamma < 0$. Then for every $(\beta_1, \beta_2) \in \mathbb{R}^2$, the partition function in (2.2) is infinite.*

Proof. For fixed y , the inner integral is

$$\int_0^\infty e^{-(\beta_1 - \gamma y)x} dx,$$

which equals $+\infty$ whenever $\beta_1 - \gamma y \leq 0$, i.e., on the set $\{y \geq \beta_1/\gamma\}$ of infinite Lebesgue measure. Hence

$$Z = \int_0^\infty e^{-\beta_2 y} \left[\int_0^\infty e^{-(\beta_1 - \gamma y)x} dx \right] dy = +\infty,$$

since the outer integrand is $+\infty$ on a set of positive measure. Note that no choice of β_2 can repair this: The factor $e^{-\beta_2 y}$, however small, multiplies an inner integral that is already infinite. $\square$

Lemma 3.3 (Sign obstruction within the normalizable family). *Let f be a normalizable Gibbs density (2.2) with $\beta_3 = t \geq 0$ and finite means. Then $\text{Cov}_f(X, Y) \leq 0$, with strict inequality if $t > 0$. Consequently the normalizable Gibbs family realizes only non-attractive targets $C \leq \mu_1 \mu_2$.*

Proof. Under f , the conditional law of X given $Y = y$ is exponential with rate $\beta_1 + ty$, so

$$\mathbb{E}[X | Y] = (\beta_1 + tY)^{-1}$$

is a nonincreasing function of Y (strictly decreasing if $t > 0$). By the tower property,

$$\text{Cov}(X, Y) = \text{Cov}(\mathbb{E}[X | Y], Y),$$

the covariance of a nonincreasing and a (strictly) increasing function of the same random variable, which is nonpositive by Chebyshev's integral inequality [41] and strictly negative when $t > 0$ and Y is nondegenerate. $\square$

Lemmas 3.2 and 3.3 explain why the attractive program must fail if a Gibbs maximizer is sought: The attractive members of the family do not exist (Lemma 3.2), and the existing members cannot reach attractive targets (Lemma 3.3). Part (c) of Theorem 3.1 is stronger: It rules out maximizers of any form and identifies the value (3.1) exactly. This is the content of the next two subsections.

Remark 3.4 (Relation to the bivariate exponential-conditionals family). The normalizable members of (2.2) have $\beta_3 \geq 0$ together with the corresponding positive linear rates; they form the bivariate exponential-conditionals distribution of Arnold and Strauss [33]. Gumbel [34] provides an antecedent, while the later conditional-specification literature [35, 36] records the nonpositivity of the correlation within that family—the substance of Lemma 3.3. Theorem 3.1 supplies the variational counterpart: Part (a) shows that every member of the family is the unique maximum-entropy density for its own moment vector, and part (c) shows that the failure at attractive targets is a property of the program, not of the family—no density of any form attains the supremum, whose exact value the escape construction identifies.

3.2. Proof of Theorem 3.1 (c)

The proof has two steps: A strict entropy upper bound for every feasible density, obtained from Gibbs' inequality against a product-exponential reference (Step 1), and an escape-to-infinity mixture construction showing that the bound is the exact supremum yet is attained by no feasible density (Step 2).

Step 1: the strict upper bound

Let π be the product of exponential densities with means μ_1, μ_2 :

$$\pi(x, y) = \mu_1^{-1} \mu_2^{-1} \exp\left(-\frac{x}{\mu_1} - \frac{y}{\mu_2}\right).$$

For any $g \in \mathcal{F}(\mu_1, \mu_2, C)$ with $h(g) > -\infty$, Gibbs' inequality gives

$$0 \leq D(g \| \pi) = -h(g) + \ln(\mu_1 \mu_2) + \frac{\mathbb{E}_g[X]}{\mu_1} + \frac{\mathbb{E}_g[Y]}{\mu_2} = -h(g) + 2 + \ln(\mu_1 \mu_2).$$

Thus $h(g) \leq 2 + \ln(\mu_1 \mu_2)$. Equality would force $g = \pi$ a.e., hence $\mathbb{E}_g[XY] = \mu_1 \mu_2$. Therefore, whenever $C \neq \mu_1 \mu_2$,

$$h(g) < 2 + \ln(\mu_1 \mu_2) \quad \text{for every } g \in \mathcal{F}(\mu_1, \mu_2, C) \text{ with } h(g) > -\infty. \quad (3.2)$$

Densities with entropy $-\infty$ do not affect the supremum. Hence, if we show that the supremum equals $2 + \ln(\mu_1 \mu_2)$, no maximizer can exist.

Step 2: an escape-to-infinity sequence attaining the bound in the limit

Let $\Delta = C - \mu_1 \mu_2 > 0$. For $\varepsilon \in (0, 1)$ and a placement parameter $s \geq 0$, consider the two-component mixture

$$g_{\varepsilon, s} = (1 - \varepsilon)[\text{Exp}(\text{mean } m_1) \otimes \text{Exp}(\text{mean } m_2)] + \varepsilon \text{Unif}([s, s + 1]^2), \quad (3.3)$$

a product-exponential bulk carrying almost all the mass, plus a remote uniform blob of mass ε on a unit square along the diagonal. The blob has coordinatewise mean $s + 1/2$, cross-moment $(s + 1/2)^2$ (its coordinates are independent), and differential entropy 0.

Root existence for the matching equations (intermediate value theorem). Feasibility of (3.3) amounts to the three equations

$$(1 - \varepsilon)m_1 + \varepsilon(s + \frac{1}{2}) = \mu_1, \quad (3.4)$$

$$(1 - \varepsilon)m_2 + \varepsilon(s + \frac{1}{2}) = \mu_2, \quad (3.5)$$

$$(1 - \varepsilon)m_1 m_2 + \varepsilon(s + \frac{1}{2})^2 = C. \quad (3.6)$$

Solve (3.4) and (3.5) for

$$m_i(s) = \frac{\mu_i - \varepsilon(s + \frac{1}{2})}{1 - \varepsilon},$$

admissible (i.e., $m_1, m_2 > 0$) for

$$s < s_{\max} := \frac{\mu_*}{\varepsilon} - \frac{1}{2},$$

where $\mu_* := \mu_1 \wedge \mu_2$, and substitute into (3.6) to obtain a single continuous equation $F_\varepsilon(s) = 0$ on $[0, s_{\max})$, where

$$F_\varepsilon(s) = (1 - \varepsilon)m_1(s)m_2(s) + \varepsilon(s + \frac{1}{2})^2 - C.$$

Restrict from now on to

$$0 < \varepsilon \leq \varepsilon_0 := \min \left\{ \frac{\Delta}{2\mu_1 \mu_2 + \Delta}, \Delta, \mu_*, \frac{1}{2}, \frac{\mu_*^2}{2C} \right\}. \quad (3.7)$$

The first entry of (3.7) is exactly the condition

$$\frac{\mu_1 \mu_2}{1 - \varepsilon} \leq \mu_1 \mu_2 + \frac{\Delta}{2},$$

so, since $m_i(s) \leq \mu_i/(1 - \varepsilon)$ for every admissible s ,

$$(1 - \varepsilon)m_1(s)m_2(s) \leq \frac{\mu_1 \mu_2}{1 - \varepsilon} \leq \mu_1 \mu_2 + \frac{\Delta}{2} = C - \frac{\Delta}{2},$$

and therefore, using also $(1 - \varepsilon)m_1 m_2 > 0$,

$$\frac{\Delta}{2} \leq C - (1 - \varepsilon)m_1(s)m_2(s) \leq C \quad \text{for every admissible } s. \quad (3.8)$$

The entries $\varepsilon \leq \mu_*$ and $\varepsilon \leq 1/2$ of (3.7) make $s = 0$ admissible ($m_i(0) > 0$ and $s_{\max} \geq 1/2$), and $\varepsilon \leq \Delta$ gives, with (3.8), the strict sign at the left endpoint:

$$F_\varepsilon(0) = [(1 - \varepsilon)m_1(0)m_2(0) - C] + \frac{\varepsilon}{4} \leq -\frac{\Delta}{2} + \frac{\Delta}{4} = -\frac{\Delta}{4} < 0. \quad (3.9)$$

At the right endpoint, the value of F_ε is finite and must be computed exactly. Since

$$\varepsilon(s_{\max} + \tfrac{1}{2}) = \mu_*,$$

we have

$$m_i(s_{\max}) = \frac{\mu_i - \mu_*}{1 - \varepsilon},$$

one of which vanishes, so $m_1 m_2 \rightarrow 0$, while

$$\varepsilon(s_{\max} + \tfrac{1}{2})^2 = \frac{\mu_*^2}{\varepsilon}.$$

Hence F_ε extends continuously to $s = s_{\max}$ with

$$\lim_{s \uparrow s_{\max}} F_\varepsilon(s) = \frac{\mu_*^2}{\varepsilon} - C \geq C > 0, \quad (3.10)$$

the inequality being exactly the last entry $\varepsilon \leq \mu_*^2/(2C)$ of (3.7). (The value μ_*^2/ε is finite for each fixed ε ; it is only as $\varepsilon \downarrow 0$ that it diverges, which is why the threshold in (3.7) is needed.) Combining (3.9) and (3.10) with the continuity of F_ε on $[0, s_{\max}]$, the intermediate value theorem gives a root $s_\varepsilon \in (0, s_{\max})$, and we fix this choice, writing $g_\varepsilon := g_{\varepsilon, s_\varepsilon} \in \mathcal{F}(\mu_1, \mu_2, C)$.

Localization of the root. At the root,

$$\varepsilon(s_\varepsilon + \tfrac{1}{2})^2 = C - (1 - \varepsilon)m_1 m_2,$$

so (3.8) gives the two-sided bound

$$\frac{\Delta}{2} \leq \varepsilon(s_\varepsilon + \tfrac{1}{2})^2 \leq C,$$

that is,

$$\sqrt{\frac{\Delta}{2\varepsilon}} \leq s_\varepsilon + \tfrac{1}{2} \leq \sqrt{\frac{C}{\varepsilon}}, \quad \text{so} \quad s_\varepsilon + \tfrac{1}{2} = \Theta(\varepsilon^{-1/2}). \quad (3.11)$$

Multiplying (3.11) by ε ,

$$\sqrt{\frac{\varepsilon\Delta}{2}} \leq \varepsilon(s_\varepsilon + \tfrac{1}{2}) \leq \sqrt{\varepsilon C}, \quad \text{so} \quad \varepsilon(s_\varepsilon + \tfrac{1}{2}) = \Theta(\sqrt{\varepsilon}) = O(\sqrt{\varepsilon}),$$

and consequently from

$$m_i = \frac{\mu_i - \varepsilon(s_\varepsilon + \tfrac{1}{2})}{1 - \varepsilon},$$

$$m_i = \mu_i + O(\sqrt{\varepsilon}), \quad i = 1, 2. \quad (3.12)$$

The vanishing blob, placed at distance $\Theta(\varepsilon^{-1/2})$, thus carries the entire excess cross-moment Δ while perturbing the means by $O(\sqrt{\varepsilon})$. We record that the rate in (3.12) is $\sqrt{\varepsilon}$ and not ε : It comes from the displacement of the blob, not from its mass, and it is this rate that governs the entropy estimate below.

Entropy of the sequence. By Lemma 2.2 and $h(\text{Unif}([s, s+1]^2)) = 0$,

$$\begin{aligned} h(g_\varepsilon) &\geq (1 - \varepsilon) h(\text{Exp}(m_1) \otimes \text{Exp}(m_2)) \\ &= (1 - \varepsilon)(2 + \ln(m_1 m_2)) \\ &= 2 + \ln(\mu_1 \mu_2) - O(\sqrt{\varepsilon}) \xrightarrow{\varepsilon \downarrow 0} 2 + \ln(\mu_1 \mu_2), \end{aligned}$$

where the error rate is that of (3.12), the $O(\varepsilon)$ loss from the mixture weight $1 - \varepsilon$ being of smaller order. Combining with the strict upper bound (3.2) proves (3.1), and non-attainment follows because (3.2) is strict for every feasible g . $\square$

Remark 3.5 (The mechanism and its one-sidedness). The escape construction succeeds because the statistic xy is unbounded above on the orthant: A mass ε placed at distance $\Theta(\varepsilon^{-1/2})$ contributes $\Theta(1)$ to $\mathbb{E}[XY]$ but only $O(\sqrt{\varepsilon})$ to each mean, and hence—through that mean displacement, which dominates the $O(\varepsilon)$ loss carried by the mixture weight—only $O(\sqrt{\varepsilon})$ to the entropy gap furnished by concavity. The mirror-image trick for lowering $\mathbb{E}[XY]$ below $\mu_1 \mu_2$ is impossible because $xy \geq 0$ is bounded below; this asymmetry of the bilinear statistic on the orthant is precisely the source of the dichotomy, and it is this mechanism—not any two-dimensional accident—that we will generalize to n dimensions. The n -dimensional generalization is carried out in Section 7.

3.3. Proof of Theorem 3.1 (a) and (b)

Part (b) follows from the Step 1 bound; part (a) requires constructing feasible densities for every repulsive or product target, which occupies the remainder of the subsection.

Part (b) is immediate: The product of exponential densities with means μ_1, μ_2 lies in $\mathcal{F}(\mu_1, \mu_2, \mu_1 \mu_2)$ and attains the bound of Step 1. Equality in Gibbs' inequality with respect to this product law forces equality a.e., giving uniqueness.

For part (a), we construct the maximizer inside the Gibbs family by a one-parameter continuation in the coupling $t = \beta_3 \geq 0$.

Lemma 3.6 (Mean matching and strict profile monotonicity). *Let $\mu_1, \mu_2 > 0$.*

- (i) *For every $t \geq 0$, there exists a unique pair $(\beta_1(t), \beta_2(t)) \in (0, \infty)^2$ such that the Gibbs density f_t of (2.2) with parameters $(\beta_1(t), \beta_2(t), t)$ satisfies $\mathbb{E}_{f_t}[X] = \mu_1$ and $\mathbb{E}_{f_t}[Y] = \mu_2$.*
- (ii) *$t \mapsto (\beta_1(t), \beta_2(t))$ is continuously differentiable on $[0, \infty)$, one-sidedly at $t = 0$.*
- (iii) *The profile $c(t) = \mathbb{E}_{f_t}[XY]$ is continuously differentiable with*

$$c'(t) = -(d(t) - b(t)^\top V(t)^{-1} b(t)) < 0, \quad (3.13)$$

where $V = \text{Cov}_{f_t}((X, Y))$, $b = \text{Cov}_{f_t}((X, Y), XY)$, and $d = \text{Var}_{f_t}(XY)$. Consequently c is a strictly decreasing bijection from $[0, \infty)$ onto $(0, \mu_1 \mu_2]$, and its inverse $C \mapsto t(C)$ is continuously differentiable on $(0, \mu_1 \mu_2)$.

Proof. **The natural domain.** Fix $t \geq 0$ and consider the base measure

$$\nu_t(dx, dy) = e^{-txy} dx dy$$

on Q , with natural parameter $\theta = (\theta_1, \theta_2)$, sufficient statistic (x, y) , and log-partition $A_t(\theta) = \ln Z_t(\theta)$,

$$Z_t(\theta) = \int_Q e^{\theta_1 x + \theta_2 y} d\nu_t.$$

Since $e^{-txy} \leq 1$, $Z_t(\theta) < \infty$ whenever $\theta_1, \theta_2 < 0$. Conversely, if $\theta_1 \geq 0$, then for $t > 0$ the inner x -integral equals $(ty - \theta_1)^{-1}$ for $ty > \theta_1$ and $+\infty$ otherwise; in either subcase the double integral diverges (for $\theta_1 > 0$ on the set $\{ty \leq \theta_1\}$ of positive measure, and for $\theta_1 = 0$ from the non-integrable singularity $(ty)^{-1}$ at $y = 0$); the case $t = 0$ is classical. The set on which Z_t is finite is therefore exactly the open quadrant $\Theta = (-\infty, 0)^2$.

(i) **Existence and uniqueness by direct minimization.** Put

$$\Phi_t(\theta) = A_t(\theta) - \mu_1 \theta_1 - \mu_2 \theta_2, \quad \theta \in \Theta.$$

On compact subsets of Θ , the integrands defining A_t and its θ -derivatives are dominated by integrable envelopes

$$P(x, y) e^{\theta'_1 x + \theta'_2 y}$$

with $\theta' \in \Theta$ and P polynomial, so $A_t \in C^\infty(\Theta)$ with $\nabla A_t(\theta) = \mathbb{E}_\theta[(X, Y)]$ and $\nabla^2 A_t(\theta) = \text{Cov}_\theta((X, Y))$. The latter is strictly positive definite: For $(a_1, a_2) \neq (0, 0)$, the function $a_1 x + a_2 y$ is not a.e. constant on Q , and the density is positive on Q , so $\text{Var}_\theta(a_1 X + a_2 Y) > 0$. Hence Φ_t is strictly convex on the open convex set Θ .

Growth as $\|\theta\| \rightarrow \infty$ within Θ . Let

$$\delta = \frac{1}{2} (\mu_1 \wedge \mu_2) > 0.$$

Restricting the integral defining Z_t to $[0, \delta]^2$, where $e^{-txy} \geq e^{-t\delta^2}$ and $\theta_1 x + \theta_2 y \geq -\delta(|\theta_1| + |\theta_2|)$, gives

$$A_t(\theta) \geq 2 \ln \delta - t\delta^2 - \delta(|\theta_1| + |\theta_2|).$$

Since $-\mu_i \theta_i = \mu_i |\theta_i|$ on Θ and $\mu_i - \delta \geq \delta$,

$$\Phi_t(\theta) \geq 2 \ln \delta - t\delta^2 + \delta(|\theta_1| + |\theta_2|) \xrightarrow{\|\theta\| \rightarrow \infty} +\infty. \quad (3.14)$$

Growth at the finite boundary. Performing the x -integration first gives, for every $\theta \in \Theta$,

$$Z_t(\theta) = \int_0^\infty \frac{e^{\theta_2 y}}{ty + |\theta_1|} dy.$$

Fix $M > 0$ and let $|\theta_2| \leq M$. Then $e^{\theta_2 y} \geq e^{-1}$ for $0 \leq y \leq 1/M$, so

$$Z_t(\theta) \geq e^{-1} \int_0^{1/M} \frac{dy}{ty + |\theta_1|} = \begin{cases} \frac{e^{-1}}{t} \ln \left(1 + \frac{t}{M|\theta_1|} \right), & t > 0, \\ \frac{e^{-1}}{M|\theta_1|}, & t = 0, \end{cases}$$

and in both cases $Z_t(\theta) \rightarrow \infty$ as $|\theta_1| \downarrow 0$ uniformly over $|\theta_2| \leq M$. Exchanging the roles of the two coordinates gives the same conclusion as $|\theta_2| \downarrow 0$ uniformly over $|\theta_1| \leq M$. Since $\Phi_t \geq A_t$ on Θ , the same holds for Φ_t .

Conclusion. Fix $\theta^0 \in \Theta$ and let $S = \{\theta \in \Theta : \Phi_t(\theta) \leq \Phi_t(\theta^0)\}$, a nonempty set closed in Θ . By (3.14), S is bounded, say $|\theta_1|, |\theta_2| \leq M$ on S ; by the boundary estimate applied with this M , there is $\eta > 0$ with $|\theta_1|, |\theta_2| \geq \eta$ on S . Hence S is a compact subset of Θ , and the continuous function Φ_t attains its minimum over Θ at some $\theta(t) \in S \subset \Theta$, an interior point. There $\nabla \Phi_t(\theta(t)) = 0$, that is,

$$\nabla A_t(\theta(t)) = (\mu_1, \mu_2).$$

Strict convexity makes the minimizer unique and ∇A_t injective on Θ , so $\theta(t)$ is the only parameter in Θ satisfying the two mean equations. Setting $(\beta_1(t), \beta_2(t)) = -\theta(t) \in (0, \infty)^2$ proves (i). For a regular exponential family, the same conclusion is available from the general mean-map theory of [15, 16], which we cite as context; the argument above is given in full so that nothing here depends on which convention for steepness is adopted.

(ii) Regularity in t . On compact subsets of $\Theta \times [0, \infty)$, the integrands defining A_t and its θ -derivatives admit the same envelopes as above, so all these integrals are continuous in (θ, t) there and may be differentiated under the integral sign, and

$$\nabla_{\theta}^2 A_t = \text{Cov}_{\theta}((X, Y)) > 0$$

is nonsingular.

For $t > 0$, the map $(\theta, t) \mapsto \nabla_{\theta} A_t(\theta)$ is C^{∞} on the open set $\Theta \times (0, \infty)$, so the implicit function theorem applied to

$$\nabla_{\theta} A_t(\theta) = (\mu_1, \mu_2)$$

yields C^1 (indeed C^{∞}) dependence of $\theta(t)$, and hence of $\beta(t)$, on $(0, \infty)$.

At $t = 0$, the parameter ranges only over $[0, \infty)$: For $t < 0$, the coupling is attractive and the partition function is infinite by Lemma 3.2, so A_t is not defined on any two-sided neighborhood of $t = 0$, and the implicit function theorem is not available in that form. Instead, the envelopes above give continuity of $\nabla_{\theta} A_t(\theta)$ and of $\nabla_{\theta}^2 A_t(\theta)$ up to $t = 0$ and permit differentiating the moment equations

$$\nabla_{\theta} A_t(\theta(t)) = (\mu_1, \mu_2)$$

one-sidedly in t ; solving the resulting linear system, whose matrix $\nabla_{\theta}^2 A_0 > 0$ is nonsingular, produces the right derivative $\theta'(0^+)$. Thus $t \mapsto \beta(t)$ is continuous on $[0, \infty)$ and continuously differentiable there, one-sidedly at the origin, which is the assertion of (ii).

(iii) Strict profile monotonicity. Extend the family to the full statistic $T = (X, Y, XY)$ with natural parameter $\vartheta = (\theta_1, \theta_2, \theta_3)$ and log-partition

$$A(\vartheta) = \ln \int_{\mathcal{Q}} e^{\theta_1 x + \theta_2 y + \theta_3 xy} \, dx \, dy,$$

finite exactly on $\{\theta_1 < 0, \theta_2 < 0, \theta_3 \leq 0\}$ (Lemma 3.2 excludes $\theta_3 > 0$). The mean-matched family of (i) is the curve $t \mapsto (\theta(t), -t)$, and $c(t) = \partial_{\theta_3} A(\theta(t), -t)$. Write

$$V = (\partial_{\theta_i \theta_j}^2 A)_{i,j \leq 2} = \text{Cov}((X, Y)),$$

$$b = (\partial_{\theta_1 \theta_3}^2 A)_{i \leq 2} = \text{Cov}((X, Y), XY),$$

$$d = \partial_{\theta_3 \theta_3}^2 A = \text{Var}(XY),$$

all finite. Indeed, under f_t the conditional law of X given $Y = y$ is exponential with rate

$$\beta_1(t) + ty \geq \beta_1(t) > 0,$$

so

$$\mathbb{E}[X^i | Y] = i! (\beta_1(t) + tY)^{-i} \leq i! \beta_1(t)^{-i},$$

while the Y -marginal obtained by integrating out x satisfies

$$f_Y(y) = \frac{1}{Z} \frac{e^{-\beta_2(t)y}}{\beta_1(t) + ty} \leq \frac{1}{Z\beta_1(t)} e^{-\beta_2(t)y}, \quad (3.15)$$

so Y has exponentially decaying tails and all its polynomial moments are finite. Combining the two,

$$\mathbb{E}[X^i Y^j] = \mathbb{E}[Y^j \mathbb{E}[X^i | Y]] \leq i! \beta_1(t)^{-i} \mathbb{E}[Y^j] < \infty$$

for all $i, j \geq 0$, which covers every entry of V , b , and d . The same envelopes justify differentiating under the integral sign (one-sidedly in θ_3 at $\theta_3 = 0$). Differentiating the two mean equations $\partial_{\theta_i} A(\theta(t), -t) = \mu_i$ with respect to t gives $V\theta'(t) - b = 0$, i.e., $\theta'(t) = V^{-1}b$, and then differentiating $c(t) = \partial_{\theta_3} A(\theta(t), -t)$ gives

$$c'(t) = b^\top \theta'(t) - d = -(d - b^\top V^{-1}b),$$

which is (3.13). The quantity $d - b^\top V^{-1}b$ is the Schur complement of V in $\text{Cov}(T)$, equivalently the variance of the residual of XY after projection in $L^2(f_t)$ onto $\text{span}\{1, X, Y\}$. It is strictly positive: Were it zero, XY would agree a.e. under f_t with some affine $a_0 + a_1X + a_2Y$, so the real-analytic function $xy - a_0 - a_1x - a_2y$ would vanish on a subset of Q of full Lebesgue measure and hence identically, which it does not. Therefore $c'(t) < 0$ for every $t \geq 0$.

Finally $c(0) = \mu_1\mu_2$, the mean-matched density at $t = 0$ being the product of exponentials with means μ_1, μ_2 ; and for $t > 0$, conditioning as in Lemma 3.3,

$$c(t) = \mathbb{E}[Y \mathbb{E}[X | Y]] = \mathbb{E}\left[\frac{Y}{\beta_1(t) + tY}\right] \leq \mathbb{E}\left[\frac{Y}{tY}\right] = \frac{1}{t} \xrightarrow{t \rightarrow \infty} 0 \quad (3.16)$$

using only $\beta_1(t) > 0$. Thus c is continuous and strictly decreasing on $[0, \infty)$, with $c(0) = \mu_1\mu_2$ and $c(t) \downarrow 0$; it is therefore a bijection onto $(0, \mu_1\mu_2]$, and since $c' < 0$ the inverse function theorem makes $t(\cdot)$ continuously differentiable on $(0, \mu_1\mu_2)$. $\square$

Proof of Theorem 3.1 (a). Let $C \in (0, \mu_1\mu_2)$. By Lemma 3.6 (iii) the profile c is a strictly decreasing bijection from $[0, \infty)$ onto $(0, \mu_1\mu_2]$, so there is exactly one $t^* > 0$ with $c(t^*) = C$. The corresponding mean-matched density $f^* := f_{t^*}$ then lies in $\mathcal{F}(\mu_1, \mu_2, C)$, with $\beta_1(t^*), \beta_2(t^*) > 0$ and $\beta_3 = t^* > 0$.

Maximality and uniqueness. For any $g \in \mathcal{F}(\mu_1, \mu_2, C)$, nonnegativity of relative entropy gives the Gibbs inequality

$$h(g) \leq - \int_Q g \ln f^* = \beta_1\mu_1 + \beta_2\mu_2 + t^*C + \ln Z = - \int_Q f^* \ln f^* = h(f^*),$$

where the middle equality uses only the shared moments of g and f^* . Equality holds if and only if $D(g \| f^*) = 0$, i.e., $g = f^*$ a.e. Hence f^* is the unique maximizer, and the program is well posed. $\square$

Remark 3.7. By Lemma 3.6 (iii) the profile c is strictly decreasing, so the mean-matched continuation sweeps the repulsive range $(0, \mu_1\mu_2)$ exactly once and sweeps nothing else—in exact agreement with the dichotomy. Lemma 3.3 gives $c(t) < \mu_1\mu_2$ for $t > 0$ on independent grounds by a conditioning argument that does not use (3.13).

Remark 3.8 (The value function). Write

$$H(C) = \sup\{h(g) : g \in \mathcal{F}(\mu_1, \mu_2, C)\}.$$

Theorem 3.1 says H is finite everywhere, equals $2 + \ln(\mu_1\mu_2)$ on the entire closed attractive ray $C \geq \mu_1\mu_2$, is attained only at its left endpoint, and is attained (by a Gibbs density with $\beta_3 > 0$) for every repulsive C . The flatness of H on the attractive side is the precise sense in which an attractive bilinear constraint is entropically inactive on the orthant. Moreover H is strictly increasing on the repulsive side, and this is now a consequence of Lemma 3.6 rather than an assumption. For $0 < C < \mu_1\mu_2$, let $t(C)$ be the unique matching coupling of Lemma 3.6 (iii) and $f_{t(C)}$ the maximizer of part (a), so that, evaluated along the matched family,

$$H(C) = \beta_1\mu_1 + \beta_2\mu_2 + t(C)C + \ln Z.$$

By Lemma 3.6 (ii) and (iii) the map $C \mapsto (\beta_1, \beta_2, t)$ is C^1 on $(0, \mu_1\mu_2)$, so H is differentiable there. Writing $\ln Z = A(\vartheta)$ with $\vartheta = (-\beta_1, -\beta_2, -t)$ and using $\nabla A(\vartheta) = (\mu_1, \mu_2, C)$ along the family, the chain rule gives

$$\frac{d}{dC} \ln Z = -\mu_1 \frac{d\beta_1}{dC} - \mu_2 \frac{d\beta_2}{dC} - C \frac{dt}{dC},$$

so that in

$$H'(C) = \mu_1 \frac{d\beta_1}{dC} + \mu_2 \frac{d\beta_2}{dC} + t(C) + C \frac{dt}{dC} + \frac{d}{dC} \ln Z,$$

all parameter derivatives cancel—the same envelope computation used again in Corollary 5.5—leaving $H'(C) = \beta_3(C) > 0$ for $0 < C < \mu_1\mu_2$. The value function thus rises strictly on $(0, \mu_1\mu_2]$ to its maximum $2 + \ln(\mu_1\mu_2)$ and is flat there on the entire attractive ray, attained only at the junction.

4. Minimal damping: the trichotomy for augmented programs

Theorem 3.1 (c) shows that attractive bilinear coupling cannot stand alone. The classical escape is Gaussian: Adding self second-moment constraints contributes quadratic damping $-\alpha x^2 - \alpha' y^2$ to the exponent, and the Gibbs weight becomes normalizable precisely when the quadratic form is positive definite. This section determines exactly how much damping is needed, and in what form, by classifying normalizability of the augmented Gibbs class

$$Z_{p,q}(\gamma, \alpha, \alpha', \beta_1, \beta_2) = \int_Q \exp(\gamma xy - \alpha x^p - \alpha' y^q - \beta_1 x - \beta_2 y) dx dy, \quad (4.1)$$

where $\gamma > 0$ (attraction), $\alpha, \alpha' > 0$ (damping multipliers, arising from added constraints on $\mathbb{E}[X^p]$ and $\mathbb{E}[Y^q]$), $p, q > 1$, and $\beta_1, \beta_2 \in \mathbb{R}$ are unrestricted. Throughout, $p' = p/(p-1)$ denotes the conjugate exponent and

$$K_p = (p-1) p^{-p/(p-1)}, \quad \text{so that in particular } K_2 = \frac{1}{4}. \quad (4.2)$$

The constant K_p is exactly the sharp constant in Young's inequality on the orthant:

$$\sup_{x \geq 0} (\gamma y x - \alpha x^p) = K_p \alpha^{1-p'} (\gamma y)^{p'}, \quad y \geq 0, \quad (4.3)$$

attained at

$$x^*(y) = \left(\frac{\gamma y}{p\alpha} \right)^{1/(p-1)}$$

as a one-line calculus computation shows.

Theorem 4.1 (Damping trichotomy). *Fix $p, q > 1$ and let $Z = Z_{p,q}(\gamma, \alpha, \alpha', \beta_1, \beta_2)$ as in (4.1).*

(i) **Subcritical damping** ($\frac{1}{p} + \frac{1}{q} < 1$): $Z < \infty$ for all $\gamma, \alpha, \alpha' > 0$ and all $\beta_1, \beta_2 \in \mathbb{R}$. Normalizability is unconditional.

(ii) **Critical damping** ($\frac{1}{p} + \frac{1}{q} = 1$, i.e., $q = p'$): With K_p as in (4.2),

$$Z < \infty \quad \text{if} \quad \alpha' > K_p \alpha^{1-p'} \gamma^{p'}, \quad Z = \infty \quad \text{if} \quad \alpha' < K_p \alpha^{1-p'} \gamma^{p'}$$

for all $\beta_1, \beta_2 \in \mathbb{R}$; on the equality manifold

$$\alpha' = K_p \alpha^{1-p'} \gamma^{p'},$$

finiteness is decided by the linear multipliers β_1, β_2 and is classified completely in Proposition 4.3. At $p = q = 2$, the strict condition reads $\gamma^2 < 4\alpha\alpha'$: The Gaussian positive-definiteness condition is recovered as the critical case.

(iii) **Supercritical coupling** ($\frac{1}{p} + \frac{1}{q} > 1$): $Z = \infty$ for all $\gamma, \alpha, \alpha' > 0$ and all $\beta_1, \beta_2 \in \mathbb{R}$. No choice of coefficients restores normalizability.

The critical case of Theorem 4.1 rests on the following sharp two-sided estimate for the inner integral, which also yields the complete classification of the equality manifold in Proposition 4.3.

Lemma 4.2 (Sharp two-sided envelope estimate). *Let $p > 1$, $\alpha > 0$ and set*

$$G(s) = \int_0^\infty e^{sx - \alpha x^p} dx, \quad A = K_p \alpha^{1-p'}, \quad \rho = \frac{p-2}{2(p-1)}.$$

Then G is finite, continuous, and nondecreasing on $\mathbb{R}$, and there exist constants $0 < c_- \leq c_+ < \infty$ and $s_0 \geq 1$, depending only on p and α , such that

$$c_- s^{-\rho} e^{A s^{p'}} \leq G(s) \leq c_+ s^{-\rho} e^{A s^{p'}} \quad (s \geq s_0). \quad (4.4)$$

Moreover $\rho < 1/2$ for every $p > 1$.

Proof. Finiteness, continuity, and monotonicity are immediate from $p > 1$. Fix $s > 0$ and substitute $x = x^*(s)v$, where $x^*(s) = (s/(p\alpha))^{1/(p-1)}$ is the maximizer in (4.3). Since $\alpha(x^*)^p = sx^*/p$,

$$sx^*v - \alpha(x^*)^p v^p = sx^* \left(v - \frac{v^p}{p} \right), \quad sx^* = p' A s^{p'} =: \lambda,$$

the second identity because

$$\sup_{x \geq 0} (sx - \alpha x^p) = sx^* \left(1 - \frac{1}{p} \right) = s x^* / p'$$

equals $As^{p'}$ by (4.3). This gives the exact factorization

$$G(s) = x^*(s) e^{As^{p'}} J(\lambda), \quad J(\lambda) = \int_0^\infty e^{-\lambda\psi(v)} dv, \quad \psi(v) = \frac{1}{p'} + \frac{v^p}{p} - v. \quad (4.5)$$

By (4.3), $\psi \geq 0$; moreover $\psi(1) = \psi'(1) = 0$, $\psi''(1) = p - 1 > 0$, $\psi(0) = 1/p' > 0$, and $v^{-p}\psi(v) \rightarrow 1/p$ as $v \rightarrow \infty$. Hence there are $v_1 > 2$ and $0 < a_1 \leq a_2 < \infty$, depending only on p , with

$$\psi \geq a_1 \min\{(v-1)^2, 1\} \quad \text{on } [0, v_1],$$

$$\psi \geq a_1 v^p \quad \text{on } [v_1, \infty),$$

$$\psi \leq a_2 (v-1)^2 \quad \text{on } [\frac{1}{2}, \frac{3}{2}],$$

the last by Taylor expansion at the double zero $v = 1$. For $\lambda \geq 1$, the first two bounds give

$$J(\lambda) \leq \int_{|v-1| \leq 1} e^{-\lambda a_1 (v-1)^2} dv + v_1 e^{-\lambda a_1} + e^{-\lambda a_1 v_1^p/2} \int_{v_1}^\infty e^{-a_1 v^p/2} dv \leq c'_+ \lambda^{-1/2},$$

and for $\lambda \geq 4$, the third bound, restricted to

$$|v-1| \leq \lambda^{-1/2} \leq \frac{1}{2},$$

gives

$$J(\lambda) \geq 2\lambda^{-1/2} e^{-a_2} =: c'_- \lambda^{-1/2}.$$

Thus $J(\lambda) \asymp \lambda^{-1/2}$ for $\lambda \geq 4$. Since $x^*(s) \asymp s^{1/(p-1)}$, $\lambda^{-1/2} \asymp s^{-p'/2}$, and

$$\frac{1}{p-1} - \frac{p'}{2} = \frac{2-p}{2(p-1)} = -\rho,$$

estimate (4.4) follows, with s_0 chosen so that $\lambda = p'As^{p'} \geq 4$ for $s \geq s_0$. Finally $\rho < 1/2$ is equivalent to $p-2 < p-1$. $\square$

The estimate (4.4) is sharp in both the exponential and the polynomial factor, and it is exactly this polynomial factor that decides the most degenerate boundary cases below.

Proof of Theorem 4.1. Part (i): subcritical case. Splitting the damping and applying (4.3) with $\alpha/2$ in place of α ,

$$\gamma xy \leq \frac{\alpha}{2} x^p + K_p \left(\frac{\alpha}{2}\right)^{1-p'} \gamma^{p'} y^{p'}.$$

Since

$$\frac{1}{p} + \frac{1}{q} < 1$$

means $q > p'$, for every $\delta > 0$, there is $C_\delta < \infty$ with $y^{p'} \leq \delta y^q + C_\delta$ on $[0, \infty)$. Choosing δ so small that

$$K_p (\alpha/2)^{1-p'} \gamma^{p'} \delta \leq \alpha'/2,$$

the exponent in (4.1) is bounded above by

$$-\frac{\alpha}{2}x^p - \frac{\alpha'}{2}y^q - \beta_1x - \beta_2y + C',$$

which is integrable over Q for arbitrary $\beta_1, \beta_2 \in \mathbb{R}$ because the polynomial damping of order > 1 dominates the linear terms. Hence $Z < \infty$.

Part (ii): critical convergence. Suppose $\alpha' > K_p \alpha^{1-p'} \gamma^{p'}$. Since $\alpha \mapsto K_p \alpha^{1-p'} \gamma^{p'}$ is continuous and increasing as α decreases (the exponent $1 - p'$ is negative), we may pick $\delta \in (0, 1)$ so small that

$$\alpha' > K_p ((1 - \delta)\alpha)^{1-p'} \gamma^{p'} =: \alpha' - 2\varepsilon'$$

for some $\varepsilon' > 0$. Splitting $\alpha = (1 - \delta)\alpha + \delta\alpha$ and applying (4.3) to the first part,

$$\gamma xy - \alpha x^p - \alpha' y^{p'} \leq -\delta\alpha x^p - (\alpha' - K_p ((1 - \delta)\alpha)^{1-p'} \gamma^{p'}) y^{p'} \leq -\delta\alpha x^p - 2\varepsilon' y^{p'},$$

and, as in (i), the residual damped exponent is integrable for all β_1, β_2 . Hence $Z < \infty$.

Part (ii): critical divergence. Suppose $\alpha' < K_p \alpha^{1-p'} \gamma^{p'} = A\gamma^{p'}$ and put $2\eta := A\gamma^{p'} - \alpha' > 0$. Integrating x out and applying the lower bound in (4.4) at $s = \gamma y - \beta_1$ gives, for all large y ,

$$\int_0^\infty e^{\gamma xy - \alpha x^p - \beta_1 x} dx = G(\gamma y - \beta_1) \geq c_- (\gamma y - \beta_1)^{-\rho} \exp(A(\gamma y - \beta_1)^{p'}),$$

and

$$A(\gamma y - \beta_1)^{p'} = A\gamma^{p'} y^{p'} + O(y^{p'-1}) = (\alpha' + 2\eta) y^{p'} + O(y^{p'-1}).$$

The outer integrand is therefore at least

$$c_- (\gamma y - \beta_1)^{-\rho} \exp(2\eta y^{p'} - \beta_2 y + O(y^{p'-1})) \longrightarrow \infty,$$

since $p' > 1$ makes $y^{p'}$ dominate both $\beta_2 y$ and the polynomial factor. Hence $Z = \infty$.

Part (iii): supercritical case, by a growing-cell estimate. Since

$$\frac{1}{p} + \frac{1}{q} > 1$$

is equivalent to $(p - 1)(q - 1) < 1$, the interval $(p - 1, 1/(q - 1))$ is nonempty; fix b in it, so that with the parametrization $x \asymp t, y \asymp t^b$,

$$1 + b > p \cdot 1, \quad 1 + b > qb, \quad 1 + b > \max\{1, b\}.$$

On the unit cells $R_t = [t, t + 1] \times [t^b, t^b + 1]$, $t \in \mathbb{N}$, the exponent satisfies

$$\begin{aligned} \gamma xy - \alpha x^p - \alpha' y^q - \beta_1 x - \beta_2 y &\geq \gamma t^{1+b} - \alpha(t + 1)^p - \alpha'(t^b + 1)^q - O(t^{\max\{1, b\}}) \\ &= \gamma t^{1+b}(1 - o(1)), \end{aligned}$$

so

$$Z \geq \sum_{t \geq t_0} e^{\gamma t^{1+b}/2} = \infty.$$

□

Proposition 4.3 (Complete classification on the critical equality manifold). *Let $p > 1$, $q = p'$, $\gamma, \alpha > 0$, and $\beta_1, \beta_2 \in \mathbb{R}$, and suppose the coefficients lie on the equality manifold*

$$\alpha' = K_p \alpha^{1-p'} \gamma^{p'}, \quad (4.6)$$

put

$$c_p = p' K_p \alpha^{1-p'} \gamma^{p'-1} > 0, \quad L(y) = c_p \beta_1 y^{p'-1} + \beta_2 y \quad (y > 0).$$

Then

$$Z_{p,p'}(\gamma, \alpha, \alpha', \beta_1, \beta_2) < \infty \iff L(y) \longrightarrow +\infty \quad (y \rightarrow \infty). \quad (4.7)$$

Explicitly:

(a) **Case $1 < p < 2$ (so that the conjugate damping exponent satisfies $q = p' > 2$):** $Z < \infty$ if and only if either $\beta_1 > 0$, or $\beta_1 = 0$ and $\beta_2 > 0$.

(b) **Case $p = 2$ (so that $q = p' = 2$ and $\gamma = 2\sqrt{\alpha\alpha'}$):** $Z < \infty$ if and only if

$$\frac{\beta_1}{\sqrt{\alpha}} + \frac{\beta_2}{\sqrt{\alpha'}} > 0.$$

(c) **Case $p > 2$ (so that $q = p' < 2$):** $Z < \infty$ if and only if either $\beta_2 > 0$, or $\beta_2 = 0$ and $\beta_1 > 0$.

Here p is the damping exponent of x throughout, and $q = p'$ that of y ; the three cases are cases of p alone, not of the pair (p, q) . In particular the equality manifold carries both convergent and divergent coefficient vectors for every $p > 1$: The deciding multiplier is β_1 when the x -damping exponent lies below the Gaussian value and β_2 when it lies above, with the tie broken by the other multiplier and, when both vanish, always in favor of divergence.

Proof. Write $A = K_p \alpha^{1-p'}$, so that (4.6) reads $\alpha' = A\gamma^{p'}$ and $c_p = p'A\gamma^{p'-1}$. By Tonelli and the definition of G in Lemma 4.2,

$$Z = \int_0^\infty e^{-\alpha' y^{p'} - \beta_2 y} G(\gamma y - \beta_1) dy. \quad (4.8)$$

Since G is finite and continuous, the contribution of any bounded y -interval to (4.8) is finite; fix y_0 with $\gamma y_0 - \beta_1 \geq s_0$ and $y_0 > 2|\beta_1|/\gamma$. On $[y_0, \infty)$ we have $s := \gamma y - \beta_1 \asymp y$, so (4.4) gives

$$Z < \infty \iff \int_{y_0}^\infty y^{-\rho} e^{-R(y)} dy < \infty, \quad R(y) := R_0(y) + \beta_2 y,$$

where, with $u = \beta_1/\gamma$ and using $\alpha' = A\gamma^{p'}$,

$$R_0(y) = A\gamma^{p'} [y^{p'} - (y - u)^{p'}] = -(A(\gamma y - \beta_1)^{p'} - \alpha' y^{p'}).$$

The two-sided constants in (4.4) and the comparison $s \asymp y$ affect only multiplicative constants, hence not convergence. If $\beta_1 = 0$, then $R_0 \equiv 0$ and $R = L$ exactly. If $\beta_1 \neq 0$, the mean value theorem applied to $t \mapsto t^{p'}$ on the interval with endpoints $y - u$ and y yields, for $y > 2|u|$, $R_0(y) = A\gamma^{p'} p' u \xi_y^{p'-1}$ with $|\xi_y - y| \leq |u|$, so

$$R_0(y) = c_p \beta_1 y^{p'-1} (1 + O(y^{-1})), \quad y \rightarrow \infty. \quad (4.9)$$

Throughout we use $\rho < 1/2 < 1$ (Lemma 4.2): The polynomial factor $y^{-\rho}$ alone never produces convergence.

(a) $p < 2$. Here $p' - 1 = 1/(p - 1) > 1$, so the term $c_p \beta_1 y^{p'-1}$ in (4.9) dominates both $\beta_2 y$ and the error term. If $\beta_1 > 0$, then $R(y) \geq \kappa y^{p'-1}$ for large y with $\kappa > 0$, and

$$\int_{-\infty}^{\infty} y^{-p} e^{-\kappa y^{p'-1}} dy < \infty.$$

If $\beta_1 < 0$, then $R(y) \leq -\kappa y^{p'-1} \rightarrow -\infty$, and the integrand tends to $+\infty$. If $\beta_1 = 0$, then $R(y) = \beta_2 y$, and the integral converges precisely for $\beta_2 > 0$; for $\beta_2 = 0$, the integrand is $\asymp y^{-p}$ with $p < 1$, and for $\beta_2 < 0$, it grows.

(b) $p = 2$. Then $p' = 2$ and $\rho = 0$, and (4.9) holds with no error term: $R_0(y) = A\gamma^2(2uy - u^2)$, so

$$R(y) = (c_2 \beta_1 + \beta_2) y - A\gamma^2 u^2, \quad c_2 = 2K_2 \alpha^{-1} \gamma = \frac{\gamma}{2\alpha}.$$

The integrand is $\asymp e^{-(c_2 \beta_1 + \beta_2)y}$, so the integral converges if and only if $c_2 \beta_1 + \beta_2 > 0$; when $c_2 \beta_1 + \beta_2 = 0$, it is a positive constant, and the integral diverges. On the manifold $\gamma = 2\sqrt{\alpha\alpha'}$, $c_2 = \sqrt{\alpha'}/\alpha$ and

$$c_2 \beta_1 + \beta_2 > 0 \iff \frac{\beta_1}{\sqrt{\alpha}} + \frac{\beta_2}{\sqrt{\alpha'}} > 0.$$

(c) $p > 2$. Here $p' - 1 < 1$, so $\beta_2 y$ dominates R whenever $\beta_2 \neq 0$, and convergence holds exactly for $\beta_2 > 0$, as above. If $\beta_2 = 0$, then

$$R(y) = c_p \beta_1 y^{p'-1} (1 + O(y^{-1}))$$

with $p' - 1 > 0$: For $\beta_1 > 0$, the stretched-exponential factor $e^{-\kappa y^{p'-1}}$ dominates y^{-p} , and the integral converges; for $\beta_1 < 0$, the integrand tends to $+\infty$; for $\beta_1 = 0$, it is $\asymp y^{-p}$, and the integral diverges.

Finally, the stated conditions are exactly $L(y) \rightarrow +\infty$: For $p < 2$, the term $c_p \beta_1 y^{p'-1}$ dominates L , for $p > 2$, the term $\beta_2 y$ does, and for $p = 2$, one has $L(y) = (c_2 \beta_1 + \beta_2)y$. This proves (4.7). $\square$

Remark 4.4 (Consistency of the classification). The integral (4.1) is invariant under the exchange $(x, p, \alpha, \beta_1) \leftrightarrow (y, q, \alpha', \beta_2)$, so any correct criterion must be invariant as well; Proposition 4.3 is as follows: The manifold (4.6) is symmetric, since solving it for α gives $\alpha = K_p^{p-1} \gamma^p \alpha'^{1-p}$ and

$$K_{p'} = (p' - 1) p'^{-p} = (p - 1)^{p-1} p^{-p} = K_p^{p-1}.$$

The exchange sends $p < 2$ to $p' > 2$, carrying case (a) to case (c) with the roles of β_1 and β_2 interchanged, including the tie-breaking clauses.

Remark 4.5 (Geometry of the boundary and a numerical warning). At $p = q = 2$, part (b) has a transparent geometric proof: Completing the square, the exponent equals

$$-(\sqrt{\alpha} x - \sqrt{\alpha'} y)^2 - \beta_1 x - \beta_2 y,$$

which is flat along the ray $\sqrt{\alpha} x = \sqrt{\alpha'} y$; parameterizing that ray by $x = t/\sqrt{\alpha}$, $y = t/\sqrt{\alpha'}$ turns the exponent into

$$-\left(\frac{\beta_1}{\sqrt{\alpha}} + \frac{\beta_2}{\sqrt{\alpha'}}\right)t,$$

and a change of variables along and across the ray recovers the criterion. For $p \neq 2$, the maximizing curve $x^*(y) \asymp y^{1/(p-1)}$ is not a ray: The flat direction is curved, and the residual along it is the nonlinear L rather than a linear form, and this is precisely why the deciding multiplier switches from β_1 to β_2 as p crosses the Gaussian exponent.

Numerical integration over a fixed range can be misleading here because divergence may set in only past an arbitrarily large threshold. For $p = 3$, $\alpha = \gamma = 1$, $\beta_1 = 50$, and $\beta_2 = -1/100$, one has $c_3 = 3^{-1/2}$ and $R(y) = 50 \cdot 3^{-1/2} \sqrt{y} - y/100$, which rises to about 2.1×10^4 before crossing zero at $y \approx 8.3 \times 10^6$: The outer integrand is suppressed by a factor of order $e^{-2 \times 10^4}$ over four decades, so partial integrals computed below that scale can appear numerically stabilized, although $Z = \infty$ by case $p > 2$.

Corollary 4.6 (Quadratic damping is the minimal symmetric order). *For symmetric damping $p = q$: If $p > 2$, $Z < \infty$ unconditionally; if $p = 2$, $Z < \infty$ if and only if either $\gamma^2 < 4\alpha\alpha'$ or $\gamma^2 = 4\alpha\alpha'$ and $\beta_1/\sqrt{\alpha} + \beta_2/\sqrt{\alpha'} > 0$ (Proposition 4.3 (b)); if $p < 2$, $Z = \infty$ for all coefficient values. In particular, the classical Gaussian program sits exactly on the critical manifold: Self second moments are the weakest symmetric constraints capable of supporting attractive bilinear coupling on the orthant.*

Corollary 4.7 (Every coordinate must be damped). *One-sided damping never suffices: If $\alpha' = 0$ (no self-moment constraint on Y), then for every $p > 1$, $\alpha > 0$, and $\beta_1, \beta_2 \in \mathbb{R}$, $Z = \infty$.*

Proof. With $\alpha' = 0$, integrating x out and applying the lower bound in (4.4) leaves an outer integrand bounded below by

$$c_-(\gamma y - \beta_1)^{-p} \exp(A(\gamma y - \beta_1)^{p'} - \beta_2 y)$$

for large y , which tends to $+\infty$ because $p' > 1$. Hence $Z = \infty$. $\square$

Remark 4.8 (Interpretation at the level of constraints and scope). In the language of maximum-entropy programs, Theorem 4.1 classifies which augmentations of the ill-posed attractive program of Section 3 yield a normalizable Gibbs weight: Adding self-moment constraints $\mathbb{E}[X^p]$, $\mathbb{E}[Y^q]$ (with multipliers necessarily acting as damping, $\alpha, \alpha' > 0$) succeeds unconditionally if and only if

$$\frac{1}{p} + \frac{1}{q} < 1,$$

conditionally on the explicit coefficient inequality if and only if

$$\frac{1}{p} + \frac{1}{q} = 1,$$

and never otherwise. By Corollary 4.7, the augmentation must touch both coordinates of the coupled pair. The scope of the theorem is finiteness of the partition function of the augmented Gibbs class. The conversion into attainment for prescribed moment targets $(\mu_1, \mu_2, C, m_p, m_q)$ is carried out in Proposition 4.9 below—but only for targets lying in the region M° actually realized by the family. We do not claim that every feasible attractive target lies in that region.

Theorem 4.1 settles when the damped Gibbs family exists, which is a statement about coefficients. A modeler who holds prescribed moment targets needs more: That the maximum-entropy program at those targets is attained, by a member of the family, with multipliers of the expected signs. The next proposition supplies this on the region the family actually realizes, extending the continuation argument of Theorem 3.1 (a) to the augmented setting. Which prescribed targets belong to that region is a separate question and is not settled here.

Proposition 4.9 (Attainment on the realized moment region). *Fix $p, q > 1$ with*

$$\frac{1}{p} + \frac{1}{q} \leq 1,$$

let $T(x, y) = (x, y, xy, x^p, y^q)$ with natural parameter $\theta = (\theta_1, \theta_2, \gamma, -\alpha, -\alpha')$ and log-partition

$$A(\theta) = \ln \int_{\mathcal{Q}} e^{\theta \cdot T(x, y)} dx dy,$$

and set

$$\Theta^\circ = \{\theta : \alpha, \alpha' > 0, \text{ and, if } \frac{1}{p} + \frac{1}{q} = 1, \gamma < K_p^{-1/p'} \alpha^{1/p} \alpha'^{1/p'}\},$$

with no restriction on γ in the subcritical case

$$\frac{1}{p} + \frac{1}{q} < 1.$$

Then the following hold. Items (i)–(iii) distinguish three logically separate levels, and item (iv) records the sign relation between the coupling and the realized correlation.

(i) **Normalizability for specified multipliers.** Θ° is an open convex subset of the natural domain; in particular, $A(\theta) < \infty$ for every $\theta \in \Theta^\circ$, so each such θ determines a Gibbs density $f_\theta \propto e^{\theta \cdot T}$.

(ii) **Realization of a prescribed moment vector.** The mean map $\theta \mapsto \nabla A(\theta) = \mathbb{E}_\theta[T]$ is a C^∞ diffeomorphism from Θ° onto the open set

$$M^\circ := \nabla A(\Theta^\circ) \subset \mathbb{R}^5,$$

the realized moment region. A prescribed moment vector $m^ \in \mathbb{R}^5$ is represented by a member of this family if and only if $m^* \in M^\circ$, and in that case the representing parameter $\theta^* \in \Theta^\circ$, defined by $\nabla A(\theta^*) = m^*$, is unique.*

(iii) **Maximum-entropy attainment.** For every $m^* \in M^\circ$, with θ^* as in (ii), the maximum-entropy program with the five prescribed moments m^* is well posed, and its unique maximizer is f_{θ^*} .

(iv) **Sign trichotomy.** The realized correlation carries the sign of the coupling— $C > \mu_1 \mu_2$, $C = \mu_1 \mu_2$, or $C < \mu_1 \mu_2$ according as $\gamma > 0$, $\gamma = 0$, or $\gamma < 0$. In particular, M° contains attractive targets (any $\theta \in \Theta^\circ$ with $\gamma > 0$ realizes one), and every attractive target in M° is attained with strictly positive damping in both coordinates, in accordance with Corollary 4.7.

Scope of the proposition. Parts (ii)–(iv) are statements about M° and about nothing else. The proposition does not characterize the full feasible moment set of the five statistics; it does not show that every feasible attractive target lies in M° ; and it does not establish attainment for every feasible interior moment vector. Because the augmented family is not steep (Remark 4.11), M° need not exhaust the interior of the moment set. What happens at a target outside M° is not settled here. Such a target may be realized by a normalizable boundary parameter or attained only in the generalized I -projection sense of Csiszár and Matúš [3,4], or not attained at all. We do not classify these possibilities. An exact description of M° , together with that classification, remains open; both are listed as Problem (1) of Section 8.

Proof. (i) and (ii) Convexity and openness. In the subcritical case Θ° is a product of open convex sets. In the critical case the constraint reads

$$\gamma - K_p^{-\frac{1}{p'}} \alpha^{\frac{1}{p}} \alpha'^{\frac{1}{p'}} < 0;$$

since

$$\frac{1}{p} + \frac{1}{p'} = 1,$$

the map $(\alpha, \alpha') \mapsto \alpha^{1/p} \alpha'^{1/p'}$ is a weighted geometric mean, hence concave on $(0, \infty)^2$, so the left-hand side is convex, and its strict sublevel set is open and convex. Membership $\Theta^\circ \subseteq \Theta$ holds by Theorem 4.1 (i)–(ii) when $\gamma > 0$ and by the domination $\gamma xy \leq 0$ when $\gamma \leq 0$.

On Θ° the components of T are affinely independent a.e. on $\mathcal{Q}$, so the family is minimal: $\nabla^2 A(\theta) = \text{Cov}_\theta(T) > 0$, and A is smooth on Θ° by dominated convergence with locally uniform integrable envelopes. Strict convexity of A on the convex set Θ° makes ∇A injective there, and the inverse function theorem makes it a local diffeomorphism; hence it is a diffeomorphism onto the open image $M^\circ = \nabla A(\Theta^\circ)$.

(iii) For feasible g , nonnegativity of relative entropy gives

$$h(g) \leq - \int_{\mathcal{Q}} g \ln f_{\theta^*} = -\theta^* \cdot m^* + \ln Z(\theta^*) = h(f_{\theta^*}),$$

the middle equality using that g and f_{θ^*} share all five moments; equality holds if and only if $g = f_{\theta^*}$ a.e.

(iv) Under f_θ the conditional density of X given $Y = y$ is proportional to $e^{sx - \alpha x^p}$ with $s = \theta_1 + \gamma y$: a one-parameter exponential family with carrier $e^{-\alpha x^p} dx$, normalizable for every $s \in \mathbb{R}$ since $\alpha > 0$ and $p > 1$, whose mean $m(s)$ satisfies $m'(s) = \text{Var} > 0$. Hence $\mathbb{E}[X | Y] = m(\theta_1 + \gamma Y)$ is strictly increasing, constant, or strictly decreasing in Y according to the sign of γ , and

$$\text{Cov}(X, Y) = \text{Cov}(\mathbb{E}[X | Y], Y)$$

inherits that sign by Chebyshev's correlation inequality [41], strictly when $\gamma \neq 0$ (as Y is nondegenerate). $\square$

Remark 4.10 (Three levels of statement). Three assertions are easily conflated, and the paper keeps them apart.

(1) Coefficient normalizability. If the coefficients satisfy the conditions of Theorem 4.1, then $Z < \infty$ and a Gibbs family exists.

(2) Realization. If $m^* \in M^\circ$, that family contains a member whose moment vector is exactly m^* .

(3) Attainment. If $m^* \in M^\circ$, the maximum-entropy program at m^* is uniquely attained by that member.

Statement (1) alone implies neither (2) nor (3). It says that certain multiplier vectors produce integrable Gibbs weights, not that a given moment target is produced by any of them; the passage from (1) to (2) is precisely the question of whether m^* lies in M° , which Proposition 4.9 does not answer for a general feasible target.

Remark 4.11 (Non-steepness: the boundary of the method). The augmented family is not steep, so one cannot conclude that M° exhausts the interior of the full moment set. Indeed the natural domain contains boundary points at which all five moments remain finite. In the subcritical case, take

$$\theta_b = (\theta_1, \theta_2, \gamma, 0, -\alpha')$$

with $\theta_1 < 0$, $\gamma < 0$, $\alpha' > 0$. The conditional rate $|\theta_1| + |\gamma|y$ is bounded below by $|\theta_1| > 0$, so $Z(\theta_b) < \infty$ and

$$\mathbb{E}[X^p] \leq \frac{\Gamma(p+1)}{|\theta_1|^p} < \infty,$$

while every neighborhood of θ_b meets $\{\theta_4 > 0\}$, where $Z = \infty$; approaching θ_b from Θ° , the mean map converges to a finite limit—precisely the failure of steepness. The convergence is by dominated convergence, with the envelopes above; the x^p -coordinate is in addition monotone in θ_4 , since

$$\partial_{\theta_4} \mathbb{E}_\theta[X^p] = \text{Var}_\theta(X^p) > 0,$$

but we make no monotonicity claim for the mean map as a whole. What becomes of an interior target outside M° is not settled here. It may be realized by a normalizable boundary parameter such as θ_b or attained only in the generalized I -projection sense of Csiszár and Matúš [3, 4], or not attained at all; we do not classify these cases. In the subcritical case these boundary regimes are repulsively coupled ($\gamma > 0$ with a vanishing damping multiplier is excluded by Corollary 4.7), whereas on the critical manifold the equality boundary of Theorem 4.1 (ii) contributes attractively coupled boundary points as well (Remark 4.5). An exact description of M° is deferred to Section 8.

5. Restoration by truncation and the exact entropic price

Theorem 3.1 (c) leaves the attractive program with a supremum it cannot attain. This section shows that restriction of the support to the box $Q_K = [0, K]^2$ always restores well-posedness—with an exact characterization of feasibility—and then computes the price of confinement: The truncated value lies below the unattainable supremum by exactly $\Delta(\mu_1^{-1} + \mu_2^{-1})/K$ to leading order.

Throughout, fix an attractive target $\mu_1, \mu_2 > 0$, $C > \mu_1\mu_2$, write $\Delta = C - \mu_1\mu_2 > 0$, and assume without loss of generality $\mu_2 \leq \mu_1$. Let $\mathcal{F}_K$ denote the densities supported on Q_K with the prescribed moments, $H_K = \sup_{\mathcal{F}_K} h$, and let $\Pi = \text{Exp}(\mu_1) \otimes \text{Exp}(\mu_2)$ be the product-exponential density on the quadrant.

Theorem 5.1 (Well-posedness on the box). (i) $\mathcal{F}_K \neq \emptyset$ if and only if

$$\max\{\mu_1, \mu_2\} < K \quad \text{and} \quad C < K \min\{\mu_1, \mu_2\}.$$

(ii) In that case the truncated program is well posed. The unique maximizer is a Gibbs density on the box,

$$f_K(x, y) \propto \exp(-\beta_1^{(K)}x - \beta_2^{(K)}y + \gamma_K xy) \quad \text{on } Q_K, \quad \text{with } \gamma_K > 0.$$

(iii) H_K is strictly increasing in K on the feasible range, and $H_K < 2 + \ln(\mu_1\mu_2)$.

Proof. Part (i): necessity. A density on Q_K has both means strictly below K , so $\max\{\mu_1, \mu_2\} < K$. Pointwise $xy \leq Ky$ and $xy \leq Kx$, so $C \leq K\mu_2$ and $C \leq K\mu_1$, that is, $C \leq K \min\{\mu_1, \mu_2\}$; equality in the first would force $(K - X)Y = 0$ almost surely and in the second $(K - Y)X = 0$, both impossible for an absolutely continuous law. Under the standing convention $\mu_2 \leq \mu_1$, the two conditions read $\mu_1 < K$ and $C < K\mu_2$, and it is in this form that they are used below.

Part (i): sufficiency. Let ν_1 be the product of uniform densities on $[\mu_i - \rho, \mu_i + \rho]$ (ρ small): means exact, cross-moment $\mu_1\mu_2$. For $\delta \in (0, 1)$ let

$$\nu_2 = p \text{Unif}([K - \delta, K]^2) + (1 - p) \text{Unif}([x_0, x_0 + \delta] \times [0, \delta]),$$

with

$$p = \frac{\mu_2 - \delta/2}{K - \delta}$$

and x_0 chosen so that both means are exactly (μ_1, μ_2) ; then $x_0 \geq 0$ and

$$x_0 \rightarrow \frac{K(\mu_1 - \mu_2)}{K - \mu_2} < K$$

as $\delta \downarrow 0$ (using $\mu_1 < K$), so ν_2 is supported in Q_K for small δ . Its cross-moment satisfies

$$c(\delta) = p(K - \delta/2)^2 + O(\delta) \rightarrow K\mu_2$$

as $\delta \downarrow 0$; since $C < K\mu_2$, fix δ with $c(\delta) > C$. The convex combinations $\lambda\nu_1 + (1 - \lambda)\nu_2$ have exact means and cross-moment sweeping $[\mu_1\mu_2, c(\delta)] \ni C$ continuously, so the intermediate value theorem produces a feasible density of finite entropy.

Part (ii). The exponential family on Q_K with statistic $T(x, y) = (x, y, xy)$ and Lebesgue base measure has finite log-partition for every $\theta \in \mathbb{R}^3$ (bounded statistic, compact support), so it is regular; it is minimal because $1, x, y$, and xy are linearly independent on Q_K . By [16] the mean map is a bijection from $\mathbb{R}^3$ onto the interior of the moment set $M_K = \text{conv } T(Q_K)$. We claim any target $m^* = (\mu_1, \mu_2, C)$ achieved by some density $g \in \mathcal{F}_K$ is automatically interior. If not, a supporting functional $(a, b, c) \neq 0$ would give

$$\mathbb{E}_g[aX + bY + cXY] = V := \max_{Q_K}(ax + by + cxy),$$

forcing $ax + by + cxy = V$ for g -a.e. (x, y) , hence on a set of positive Lebesgue measure; a nonzero real-analytic function cannot vanish on such a set, so $a = b = c = 0$, a contradiction. Thus the Gibbs member f_K with moments m^* exists and is unique in the family, and the Gibbs inequality (exactly as in the proof of Theorem 3.1 (a)) shows it uniquely maximizes h over $\mathcal{F}_K$.

To determine the sign of γ_K , profile out the means: For each $\gamma \in \mathbb{R}$ choose $(\beta_1, \beta_2)(\gamma)$ matching $\mathbb{E}[X] = \mu_1$, $\mathbb{E}[Y] = \mu_2$ (possible for every γ : The fixed- γ subfamily is again regular and minimal, with two-dimensional moment set of interior $(0, K)^2 \ni (\mu_1, \mu_2)$), and set $c(\gamma) = \mathbb{E}[XY]$ along this profile. By the standard formula for profiled exponential families, $c'(\gamma)$ equals the variance of the $L^2(f)$ -residual of XY after projection onto $\text{span}\{1, X, Y\}$, which is strictly positive because xy is not a.e. affine in (x, y) on the full-support box. Hence c is strictly increasing, and $c(0) = \mu_1\mu_2 < C$ (at $\gamma = 0$ the mean-matched density is a product), forcing $\gamma_K > 0$.

Part (iii). If $K < K'$ are both feasible, then $\mathcal{F}_K \subset \mathcal{F}_{K'}$ gives $H_K \leq H_{K'}$, f_K vanishes a.e. off Q_K while $f_{K'}$ is a.e. positive on $Q_{K'}$, so $f_K \neq f_{K'}$ and uniqueness of the maximizer on $Q_{K'}$ gives strictness. The upper bound is (3.2) applied to box-supported feasible densities of finite entropy; densities with entropy $-\infty$ are irrelevant to the supremum. $\square$

Lemma 5.2 (I-projection identity). *For every $g \in \mathcal{F}_K$,*

$$D(g \parallel \Pi) = (2 + \ln(\mu_1 \mu_2)) - h(g).$$

Consequently the entropic deficit

$$\mathcal{E}(K) := 2 + \ln(\mu_1 \mu_2) - H_K = \min_{g \in \mathcal{F}_K} D(g \parallel \Pi) \in (0, \infty)$$

is strictly decreasing in K , attained at f_K : The truncated maximum-entropy problem is exactly the I-projection of the product-exponential law onto the box-supported constraint set.

Proof. Since g is supported in Q_K and has matched means,

$$D(g \parallel \Pi) = -h(g) - \int g \ln \Pi = -h(g) + \ln(\mu_1 \mu_2) + \frac{\mu_1}{\mu_1} + \frac{\mu_2}{\mu_2},$$

which is the stated identity. Minimizing $D(\cdot \parallel \Pi)$ is therefore the same as maximizing h ; positivity is (3.2), finiteness and attainment come from Theorem 5.1, and strict monotonicity from Theorem 5.1 (iii). $\square$

Theorem 5.3 (The exact entropic price). As $K \rightarrow \infty$, $\mathcal{E}(K) \downarrow 0$ and

$$\lim_{K \rightarrow \infty} K \mathcal{E}(K) = \Delta \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right).$$

Proof. Write $L_K = K - 1/2$ throughout.

Step 1: an exactly feasible corner-blob mixture. Let $T_{m,K}$ denote the exponential density on $[0, K]$ with mean m , and consider

$$g_K = (1 - \varepsilon_K) T_{m_{1,K},K} \otimes T_{m_{2,K},K} + \varepsilon_K \text{Unif}([K - 1, K]^2).$$

The uniform factor has independent coordinates, both with mean L_K , hence cross-moment L_K^2 . Feasibility of g_K is therefore the system

$$(1 - \varepsilon) m_i + \varepsilon L_K = \mu_i, \quad (i = 1, 2), \quad (5.1)$$

$$(1 - \varepsilon) m_1 m_2 + \varepsilon L_K^2 = C. \quad (5.2)$$

This system is solvable in closed form; no fixed-point argument is needed. Solving (5.1),

$$m_{i,K} = \frac{\mu_i - \varepsilon_K L_K}{1 - \varepsilon_K}, \quad (5.3)$$

and substituting into (5.2) and clearing the denominator,

$$(\mu_1 - \varepsilon L_K)(\mu_2 - \varepsilon L_K) + \varepsilon L_K^2 (1 - \varepsilon) = C(1 - \varepsilon).$$

The two terms in ε^2 are $\varepsilon^2 L_K^2$ and $-\varepsilon^2 L_K^2$ and cancel identically, leaving the linear equation

$$\varepsilon [L_K^2 - L_K(\mu_1 + \mu_2) + C] = \Delta.$$

Hence

$$\varepsilon_K = \frac{\Delta}{L_K^2 - L_K(\mu_1 + \mu_2) + C} = \frac{\Delta}{(L_K - \mu_1)(L_K - \mu_2) + \Delta}, \quad (5.4)$$

the second expression because

$$(L - \mu_1)(L - \mu_2) + \Delta = L^2 - L(\mu_1 + \mu_2) + \mu_1 \mu_2 + \Delta = L^2 - L(\mu_1 + \mu_2) + C.$$

Admissibility. Assume

$$K > \max\{\mu_1, \mu_2\} + \frac{1}{2},$$

so that $L_K > \max\{\mu_1, \mu_2\}$ and $(L_K - \mu_1)(L_K - \mu_2) > 0$. The second form in (5.4) then gives at once

$$0 < \varepsilon_K < 1.$$

Furthermore

$$\varepsilon_K L_K = \frac{\Delta L_K}{(L_K - \mu_1)(L_K - \mu_2) + \Delta} \longrightarrow 0 \quad \text{as } K \rightarrow \infty,$$

so $\varepsilon_K L_K < \min\{\mu_1, \mu_2\}$ for all large K , and (5.3) gives $m_{i,K} > 0$; since $m_{i,K} \rightarrow \mu_i < K$, we also have $m_{i,K} < K$, so each $T_{m_{i,K},K}$ is well defined. Thus $g_K \in \mathcal{F}_K$ exactly for every sufficiently large K .

Rates. Since

$$L_K^2 - L_K(\mu_1 + \mu_2) + C = K^2(1 + O(K^{-1}))$$

and $L_K^{-1} = K^{-1} + O(K^{-2})$, (5.3) and (5.4) give

$$\varepsilon_K = \frac{\Delta}{K^2} + O(K^{-3}), \quad \varepsilon_K L_K = \frac{\Delta}{K} + O(K^{-2}), \quad m_{i,K} = \mu_i - \frac{\Delta}{K} + O(K^{-2}), \quad (5.5)$$

the last using also $(1 - \varepsilon_K)^{-1} = 1 + O(K^{-2})$. All implied constants depend only on μ_1, μ_2, C .

Step 2: the exact divergence of a truncated exponential. Parameterize the truncated exponential on $[0, K]$ by its rate,

$$q_{r,K}(x) = \frac{r e^{-rx}}{1 - e^{-rK}}, \quad 0 \leq x \leq K,$$

whose mean is

$$m = \frac{1}{r} - \frac{K}{e^{rK} - 1}. \quad (5.6)$$

Since

$$\ln q_{r,K}(x) = \ln r - rx - \ln(1 - e^{-rK}) \quad \text{and} \quad \ln \text{Exp}(\mu)(x) = -\ln \mu - \frac{x}{\mu},$$

integrating $\ln(q_{r,K}/\text{Exp}(\mu))$ against $q_{r,K}$ gives the exact identity

$$D(q_{r,K} \parallel \text{Exp}(\mu)) = \ln(r\mu) - \ln(1 - e^{-rK}) - rm + \frac{m}{\mu}. \quad (5.7)$$

The rate is not the reciprocal of the truncated mean, but it is exponentially close to it, uniformly on compacta. Fix $0 < a < b < \infty$ and let $m \in [a, b]$. Substituting $s = rK$ in (5.6) gives $m = K\phi(s)$ with

$$\phi(s) = \frac{1}{s} - \frac{1}{e^s - 1},$$

a strictly decreasing bijection from $(0, \infty)$ onto $(0, 1/2)$. Strict monotonicity is elementary:

$$\phi'(s) = -s^{-2} + e^s (e^s - 1)^{-2}$$

is negative precisely when $s^2 e^s < (e^s - 1)^2$, which, after taking positive square roots, is equivalent to

$$s < e^{s/2} - e^{-s/2} = 2 \sinh(s/2),$$

which holds for every $s > 0$, and $\phi(0^+) = 1/2$, $\phi(+\infty) = 0$.

Assume from now on $K > 3b$. Then $m/K < 1/2$, so s is determined by $\phi(s) = m/K$. From $\phi(s) \leq 1/s$ we get the upper bound $s \leq K/m \leq K/a$. For the lower bound, first note that

$$\phi(2) = \frac{1}{2} - (e^2 - 1)^{-1} > \frac{1}{3},$$

so

$$\phi(s) = \frac{m}{K} \leq \frac{b}{K} < \frac{1}{3} < \phi(2),$$

and strict monotonicity of ϕ gives $s > 2$. Now $e^s - 1 \geq 2s$ for $s \geq 2$, which is exactly $\phi(s) \geq 1/(2s)$; hence $m/K \geq 1/(2s)$, that is, $s \geq K/(2m) \geq K/(2b)$. Also

$$\left| \phi(s) - \frac{1}{s} \right| = (e^s - 1)^{-1} \leq 2e^{-s} \quad \text{for } s \geq 1.$$

Therefore, using

$$\begin{aligned} r - \frac{1}{m} &= \frac{s}{m} \left(\frac{m}{K} - \frac{1}{s} \right), \\ \left| r - \frac{1}{m} \right| &\leq \frac{s}{m} \cdot 2e^{-s} \leq \frac{2K}{a^2} e^{-K/(2b)} = O(Ke^{-cK}), \quad c = \frac{1}{2b}, \end{aligned} \quad (5.8)$$

uniformly for $m \in [a, b]$. Feeding (5.8) into (5.7) term by term— $\ln(r\mu) = \ln(\mu/m) + \ln(rm)$ with $rm = 1 + O(Ke^{-cK})$; $-\ln(1 - e^{-rK}) = O(e^{-cK})$ since $rK = s \geq K/(2b)$; and $rm = 1 + O(Ke^{-cK})$ again in the third term—gives, uniformly for $m \in [a, b]$,

$$D(q_{r,K} \parallel \text{Exp}(\mu)) = \ln \frac{\mu}{m} + \frac{m}{\mu} - 1 + O(Ke^{-cK}). \quad (5.9)$$

Finally,

$$\ln \left(\frac{\mu}{m} \right) + \frac{m}{\mu} - 1 = \frac{(m - \mu)^2}{2\mu^2} + O(|m - \mu|^3),$$

so by (5.5), applied with $m = m_{i,K} \in [\mu_i/2, 2\mu_i]$ for large K ,

$$D(T_{m_{i,K},K} \parallel \text{Exp}(\mu_i)) = O(K^{-2}), \quad i = 1, 2. \quad (5.10)$$

Step 3: the upper bound. By convexity of $D(\cdot \parallel \Pi)$ and tensorization of relative entropy across the product reference $\Pi = \text{Exp}(\mu_1) \otimes \text{Exp}(\mu_2)$,

$$D(g_K \parallel \Pi) \leq (1 - \varepsilon_K) \sum_{i=1,2} D(T_{m_{i,K},K} \parallel \text{Exp}(\mu_i)) + \varepsilon_K D(\text{Unif}([K-1, K]^2) \parallel \Pi).$$

The uniform density equals 1 on its support, so its differential entropy vanishes and

$$D(\text{Unif}([K-1, K]^2) \parallel \Pi) = \ln(\mu_1\mu_2) + \left(K - \frac{1}{2}\right) \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right), \quad (5.11)$$

each coordinate of the uniform law having mean $L_K = K - 1/2$. Combining (5.10), (5.11), and the rates (5.5) and using $\varepsilon_K \ln(\mu_1\mu_2) = O(K^{-2})$,

$$\mathcal{E}(K) \leq D(g_K \parallel \Pi) \leq \frac{\Delta}{K} \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right) + O(K^{-2}), \quad (5.12)$$

with constants depending only on μ_1, μ_2, C . In particular $\mathcal{E}(K) \rightarrow 0$.

Step 4: the lower bound. Fix θ with $0 < \theta < \mu_1^{-1} + \mu_2^{-1}$ and put

$$\varphi_K(x, y) = \frac{\theta}{K} xy \mathbf{1}_{Q_K}(x, y),$$

a bounded measurable function, $0 \leq \varphi_K \leq \theta K$. For every $g \in \mathcal{F}_K$ the Donsker-Varadhan inequality [13] gives

$$D(g \parallel \Pi) \geq \mathbb{E}_g[\varphi_K] - \ln \mathbb{E}_\Pi[e^{\varphi_K}],$$

and since g is supported on Q_K and has cross-moment C , $\mathbb{E}_g[\varphi_K] = \theta C/K$. The remaining term does not depend on g . Decompose

$$\mathbb{E}_\Pi[e^{\varphi_K}] = \Pi(Q_K^c) + \int_{Q_K} e^{\theta xy/K} d\Pi = \Pi(Q_K^c) + \mathbf{I}_K + \mathbf{II}_K, \quad (5.13)$$

where $\mathbf{I}_K$ and $\mathbf{II}_K$ are the parts of the box integral with $0 \leq y \leq \eta K$ and $\eta K < y \leq K$, respectively, and

$$\eta := \min\{1, \frac{1}{2\theta\mu_1}\} \in (0, 1].$$

A union bound gives

$$\Pi(Q_K^c) \leq e^{-K/\mu_1} + e^{-K/\mu_2}.$$

Region I ($0 \leq y \leq \eta K$). Put $u = \theta\mu_1 y/K$; then $u \leq \theta\mu_1\eta \leq \frac{1}{2}$ by the choice of η . Extending the inner integral to $[0, \infty)$,

$$\begin{aligned} \int_0^K e^{\theta xy/K} d\text{Exp}(\mu_1)(x) &\leq \int_0^\infty \frac{1}{\mu_1} e^{-x(\mu_1^{-1} - \theta y/K)} dx \\ &= \frac{1}{1-u} \leq 1 + u + 2u^2, \end{aligned}$$

the last step because

$$\frac{1}{1-u} = 1 + u + \frac{u^2}{1-u} \quad \text{and} \quad u \leq \frac{1}{2}.$$

Integrating over $y \in [0, \infty)$ against $\text{Exp}(\mu_2)$, whose first two moments are μ_2 and $2\mu_2^2$,

$$\mathbf{I}_K \leq 1 + \frac{\theta\mu_1\mu_2}{K} + \frac{4\theta^2\mu_1^2\mu_2^2}{K^2}. \quad (5.14)$$

Region II ($\eta K < y \leq K$). If $\eta = 1$ this region is empty and $\mathbf{II}_K = 0$; assume therefore $\eta = (2\theta\mu_1)^{-1} < 1$. For $0 \leq x \leq K$ the integrand $e^{x(\theta y/K - \mu_1^{-1})}$ is monotone in x , so it is bounded by its value at one of the endpoints, namely by $e^{(\theta y - K/\mu_1)_+}$; hence

$$\int_0^K e^{\theta xy/K} d\text{Exp}(\mu_1)(x) \leq \frac{K}{\mu_1} \exp[(\theta y - K/\mu_1)_+]. \quad (5.15)$$

Writing $\psi_\theta(y) = (\theta y - K/\mu_1)_+ - y/\mu_2$ and bounding the length of the y -interval by K ,

$$\mathbf{II}_K \leq \frac{K}{\mu_1\mu_2} \int_{\eta K}^K e^{\psi_\theta(y)} dy \leq \frac{K^2}{\mu_1\mu_2} \exp\left[\sup_{\eta K \leq y \leq K} \psi_\theta(y)\right].$$

The function ψ_θ is piecewise linear with the single kink $y_0 = K/(\theta\mu_1)$: For $y \leq y_0$ it equals $-y/\mu_2$, strictly decreasing; for $y \geq y_0$ it equals $(\theta - \mu_2^{-1})y - K/\mu_1$, affine. Its supremum over $[\eta K, K]$ is therefore attained at $y = \eta K$, at $y = K$, or at $y = y_0$ when $y_0 \in [\eta K, K]$. Note first that $\eta K < y_0$ always, since $\eta \leq (2\theta\mu_1)^{-1}$ gives $\eta K \leq \frac{1}{2}y_0$; hence the left endpoint lies on the decreasing piece and

$$\psi_\theta(\eta K) = -\frac{\eta K}{\mu_2}, \quad \psi_\theta(y_0) = -\frac{y_0}{\mu_2} = -\frac{K}{\theta\mu_1\mu_2}.$$

At the right endpoint, there are two cases. If $y_0 \leq K$, equivalently $\theta \geq \mu_1^{-1}$, then

$$\psi_\theta(K) = -K(\mu_1^{-1} + \mu_2^{-1} - \theta),$$

negative because $\theta < \mu_1^{-1} + \mu_2^{-1}$. If $y_0 > K$ the kink lies outside the interval, the whole of $[\eta K, K]$ lies on the decreasing piece, and the supremum is the left-endpoint value $-\eta K/\mu_2$ already listed. In every case,

$$\sup_{\eta K \leq y \leq K} \psi_\theta(y) \leq -c_2(\theta)K, \quad c_2(\theta) = \min\left\{\frac{\eta}{\mu_2}, \frac{1}{\theta\mu_1\mu_2}, \frac{1}{\mu_1} + \frac{1}{\mu_2} - \theta\right\} > 0, \quad (5.16)$$

each of the three entries being strictly positive, the last by the standing restriction on θ . Hence

$$\Pi_K \leq \frac{K^2}{\mu_1\mu_2} e^{-c_2(\theta)K}.$$

Combination. Substituting the three bounds into (5.13),

$$\mathbb{E}_\Pi[e^{\varphi_K}] \leq 1 + \frac{\theta\mu_1\mu_2}{K} + \frac{4\theta^2\mu_1^2\mu_2^2}{K^2} + e^{-K/\mu_1} + e^{-K/\mu_2} + \frac{K^2}{\mu_1\mu_2} e^{-c_2(\theta)K}.$$

The last three terms are $o(K^{-2})$, so there is $c_3(\theta) < \infty$, depending only on θ, μ_1 , and μ_2 , with

$$\mathbb{E}_\Pi[e^{\varphi_K}] \leq 1 + \frac{\theta\mu_1\mu_2}{K} + \frac{c_3(\theta)}{K^2}$$

for all large K ; by $\ln(1+z) \leq z$,

$$\ln \mathbb{E}_\Pi[e^{\varphi_K}] \leq \frac{\theta\mu_1\mu_2}{K} + \frac{c_3(\theta)}{K^2},$$

uniformly in $g \in \mathcal{F}_K$. Therefore

$$\mathcal{E}(K) = \min_{g \in \mathcal{F}_K} D(g \parallel \Pi) \geq \frac{\theta C}{K} - \frac{\theta\mu_1\mu_2}{K} - \frac{c_3(\theta)}{K^2} = \frac{\theta\Delta}{K} - \frac{c_3(\theta)}{K^2}.$$

Multiplying by K and letting $K \rightarrow \infty$ gives $\liminf_K K\mathcal{E}(K) \geq \theta\Delta$ for each fixed admissible θ ; only then letting $\theta \uparrow \mu_1^{-1} + \mu_2^{-1}$ yields

$$\liminf_K K\mathcal{E}(K) \geq \Delta(\mu_1^{-1} + \mu_2^{-1}).$$

This proof fixes θ , lets $K \rightarrow \infty$, and only then moves θ because the Region II estimate is not uniform in θ . For each fixed θ , the Region II bound

$$\Pi_K \leq \frac{K^2}{\mu_1\mu_2} e^{-c_2(\theta)K}$$

is $o(K^{-2})$, but $c_2(\theta) \downarrow 0$ as θ approaches $\mu_1^{-1} + \mu_2^{-1}$, and the value of K beyond which that bound falls below a prescribed multiple of K^{-2} therefore grows without bound as θ approaches the endpoint. The estimate is thus not uniform in θ , and the argument accordingly fixes θ , lets $K \rightarrow \infty$, and only then lets θ approach the endpoint. Together with (5.12), this proves the theorem. $\square$

Remark 5.4 (Interpretation and the multiplier scaling). The constant has a direct probabilistic interpretation. To carry the covariance excess Δ inside $[0, K]^2$, the asymptotically sharp upper-bound construction places mass of order Δ/K^2 near the corner (K, K) . Under the product-exponential reference law, that corner costs the tail rate $\mu_1^{-1} + \mu_2^{-1}$ per unit distance, giving the first-order cost $\Delta(\mu_1^{-1} + \mu_2^{-1})/K$. The Donsker-Varadhan lower bound uses the same rate from the dual side: The threshold $\theta \uparrow \mu_1^{-1} + \mu_2^{-1}$ is where the tilted expectation $\mathbb{E}_\Pi[e^{\theta XY/K}]$ stops being controlled by the bulk and starts being controlled by the corner. Corollary 5.5 then identifies the multiplier scale: The attractive coupling vanishes at rate $1/K$, with $\gamma_K K$ converging to the same threshold. Both limits are illustrated numerically in Section 5.1.

Corollary 5.5 (Convergence of the maximizers and the multiplier law). *As $K \rightarrow \infty$ along the feasible range:*

- (i) $\|f_K - \Pi\|_{\text{TV}} \leq \sqrt{\mathcal{E}(K)/2} = O(K^{-1/2})$: *The truncated maximizers converge in total variation to the product law that ignores the attractive constraint—the dynamical form of entropic invisibility;*
- (ii) $K \gamma_K \rightarrow \mu_1^{-1} + \mu_2^{-1}$.

Proof. (i) By Lemma 5.2, $D(f_K \| \Pi) = \mathcal{E}(K)$; Theorem 5.3 and Pinsker's inequality [5] conclude.

(ii) Regard the cross-moment target as a variable: For c in the feasible interior, let $V_K(c)$ denote the truncated value at target (μ_1, μ_2, c) and $\gamma_K(c)$ the coupling of its maximizer, so that $\gamma_K = \gamma_K(C)$. At matched parameters,

$$V_K(c) = \beta_1 \mu_1 + \beta_2 \mu_2 - \gamma_K(c) c + \ln Z_K;$$

differentiating in c , the parameter derivatives vanish by moment matching (∂_{β_i} produces $\mu_i - \mathbb{E}[X_i] = 0$, and ∂_γ produces $-c + \mathbb{E}[XY] = 0$; the matched parameters are C^1 in c by the implicit function theorem, the Jacobian being the nonsingular covariance of the minimal statistics). This leaves the envelope identity $V'_K(c) = -\gamma_K(c)$. Writing $\mathcal{E}_t(K)$ for the deficit at covariance excess $t > 0$, the critical-target constant cancels in differences:

$$\mathcal{E}_{t_2}(K) - \mathcal{E}_{t_1}(K) = \int_{t_1}^{t_2} \gamma_K(\mu_1 \mu_2 + u) \, du, \quad 0 < t_1 < t_2.$$

Since $c \mapsto \gamma_K(c)$ is strictly increasing—it is the inverse of the strictly increasing profile $\gamma \mapsto c(\gamma)$ from the proof of Theorem 5.1 (ii)—we obtain, for $0 < \delta < \Delta$,

$$\frac{\mathcal{E}_\Delta(K) - \mathcal{E}_{\Delta-\delta}(K)}{\delta} \leq \gamma_K(C) \leq \frac{\mathcal{E}_{\Delta+\delta}(K) - \mathcal{E}_\Delta(K)}{\delta}.$$

Multiplying by K and applying Theorem 5.3 at the attractive targets with excesses $\Delta - \delta$, Δ , and $\Delta + \delta$ (all feasible for large K), both bounds converge to

$$(\mu_1^{-1} + \mu_2^{-1}) \frac{\delta}{\delta} = \mu_1^{-1} + \mu_2^{-1}.$$

□

5.1. A numerical consistency check

The limits in Theorem 5.3 and Corollary 5.5 are analytic results and are used nowhere else in the paper. The computation below is included only as a consistency check on the scale and on the implementation of the moment-matching equations.

Each tabulated K is solved independently: The quadrature grid is rebuilt and the solver restarted, so no continuation across values of K is used. Within a single K , by contrast, successive coupling trials are warm started, each taking as its initial multipliers those matched at the previous trial; this changes only the starting point of the inner iteration, not the multipliers it converges to.

For each K , the coupling γ is located by bisection on a bracket that is verified before any bisection is performed, using the strictly monotone profile $\gamma \mapsto \mathbb{E}_\gamma[XY]$ of Theorem 5.1 (ii). The lower end is $\gamma = 0$, where $\mathbb{E}_\gamma[XY]$ is checked to lie strictly below the target C . The upper end begins at $(\mu_1^{-1} + \mu_2^{-1})/K$ and is enlarged adaptively by the factor 1.5, at most 40 times, until mean matching succeeds there and $\mathbb{E}_\gamma[XY]$ exceeds C ; failure to bracket within those 40 enlargements is raised as an error rather than passed on. The verified bracket is then bisected 60 times.

For each trial γ , the linear multipliers (β_1, β_2) are matched to the prescribed means using Newton iteration, with the covariance matrix of (X, Y) as the Jacobian—the derivative of $(\mathbb{E}[X], \mathbb{E}[Y])$ with respect to (β_1, β_2) is $-\text{Cov}(X, Y)$ —and are initialized at (μ_1^{-1}, μ_2^{-1}) when no warm start is available. The Newton step is globalized by backtracking rather than by a fixed step cap: The step length is halved from 1, through at most 50 trial lengths, until the maximum absolute mean residual strictly decreases, and if no trial length decreases it, the iteration stops. At most 80 Newton iterations are taken, with early exit once that residual falls below 10^{-12} ; a trial is accepted as solved when it ends below 10^{-10} . The multipliers are unrestricted in sign, since on a compact box they need not be positive. After the last bisection, all three moments are recomputed, and the largest residual over the whole grid is required to be below the tolerance 10^{-8} ; the value actually attained is 7.7×10^{-13} .

Integrals are evaluated by composite Gauss-Legendre quadrature with three adaptively placed panels in each coordinate. With

$$a = \min\{10, K/3\}, \quad b = \max\{a, K - \min(8, K/3)\},$$

the panels are $[0, a]$, $[a, b]$, and $[b, K]$, carrying 160 nodes each, so 480 nodes per coordinate in total. Clipping at $K/3$ keeps the panels ordered and non-overlapping at every K , however small. For $K \geq 30$, hence for every box in Table 3 except the smallest, the definition reduces to a bulk panel $[0, 10]$, a middle panel $[10, K-8]$, and a corner layer $[K-8, K]$ of unit scale, so that both the product-exponential bulk near the origin and the thin corner layer near (K, K) are resolved; at $K = 20$ it returns the three equal panels

$$\left[0, \frac{20}{3}\right], \quad \left[\frac{20}{3}, \frac{40}{3}\right], \quad \text{and} \quad \left[\frac{40}{3}, 20\right].$$

The same deterministic procedure reproduces Table 3; no random sampling is used.

Table 3. Normalized entropy deficit $R_E(K) = K\mathcal{E}(K)/[\Delta(\mu_1^{-1} + \mu_2^{-1})]$ and normalized multiplier $R_\gamma(K) = K\gamma_K/(\mu_1^{-1} + \mu_2^{-1})$, both of which tend to 1 by Theorem 5.3 and Corollary 5.5. Six scenarios (μ_1, μ_2, Δ) spanning balanced and unequal means and weak, baseline, and strong excess interaction. The complete grid, including $K = 120$ and $K = 320$ and the matched multipliers, is reported in Supplementary Tables S1–S3 after the References.

(μ_1, μ_2, Δ)	$R_E(K)$				$R_\gamma(K)$			
	20	40	80	160	20	40	80	160
$(1, 2, \frac{1}{2})$	0.545	0.767	0.876	0.933	0.822	0.930	0.959	0.975
$(1, 1, \frac{1}{4})$	0.676	0.829	0.907	0.949	0.883	0.937	0.962	0.977
$(\frac{1}{2}, 2, \frac{1}{4})$	0.642	0.833	0.914	0.954	0.931	0.976	0.984	0.989
$(2, 3, 1)$	0.316	0.495	0.732	0.857	0.507	0.786	0.917	0.953
$(1, 2, \frac{1}{10})$	0.162	0.311	0.586	0.782	0.301	0.598	0.929	0.964
$(1, 2, 1)$	0.718	0.857	0.921	0.956	0.942	0.961	0.972	0.981

In every scenario, both normalized quantities increase toward 1 from below (Figure 1). The approach is slower when the excess interaction Δ is small relative to $\mu_1\mu_2$, or when the means are large: At $(\mu_1, \mu_2, \Delta) = (1, 2, 1/10)$ the ratio R_E has reached only 0.78 at $K = 160$, whereas at $(1, 2, 1)$ it has reached 0.96. This ordering across scenarios is compatible with scenario-dependent remainder constants: The remainder in Theorem 5.3 is $O(K^{-2})$ with constants depending on μ_1, μ_2, C , and different constants move the scale at which the limit becomes visible. It is not, however, something the theorem predicts. Theorem 5.3 and Corollary 5.5 assert the limits for each scenario separately; they do not order the convergence speeds of different scenarios against one another, and nothing in them implies that the ordering observed here would persist at larger K or under other parameter choices. Four values of K per scenario cannot establish an asymptotic rate in any case, and we make no claim about the order of the remaining error from these data. What the grid does show is numerical consistency of both limits across substantially different parameter settings.

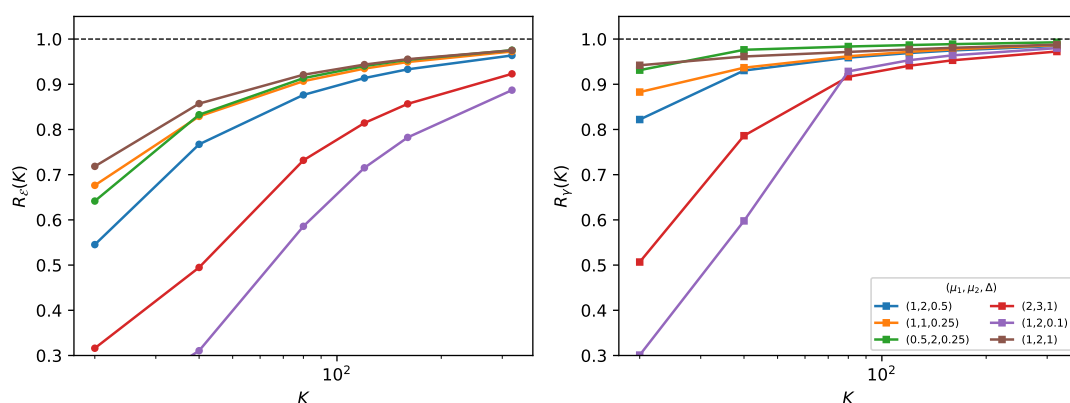


Figure 1. Normalized entropy deficit $R_E(K)$ (left) and normalized multiplier $R_\gamma(K)$ (right) for the six scenarios of Table 3, at $K = 20, 40, 80, 120, 160, 320$. The dashed line is the common limit 1 of Theorem 5.3 and Corollary 5.5.

Figure 2 displays the matched density at $K = 20$ in the baseline scenario. Most of the mass sits in the bulk near the origin, and the density at the capacity corner remains far below its value near the origin, so it is not the case that the density rises toward that corner globally. What does happen is local: Along the high- x , high- y diagonal, the density falls, reaches a minimum, and then turns upward again as (K, K) is approached, which is the finite-box trace of the attractive term γxy . The cross-moment is correspondingly drawn from the upper-right region out of all proportion to the mass there: The region $\{x > 5, y > 5\}$ carries 0.36% of the mass but 14% of $\mathbb{E}[XY]$, while the unit corner square $[K - 1, K]^2$ carries 2.1×10^{-5} of the mass and 0.32% of $\mathbb{E}[XY]$. Both percentages are reproduced in Supplementary Table S5. This is the finite-box counterpart of the escape mechanism of Theorem 3.1 (c), in which a vanishing mass placed far out carries a fixed contribution to the cross-moment.

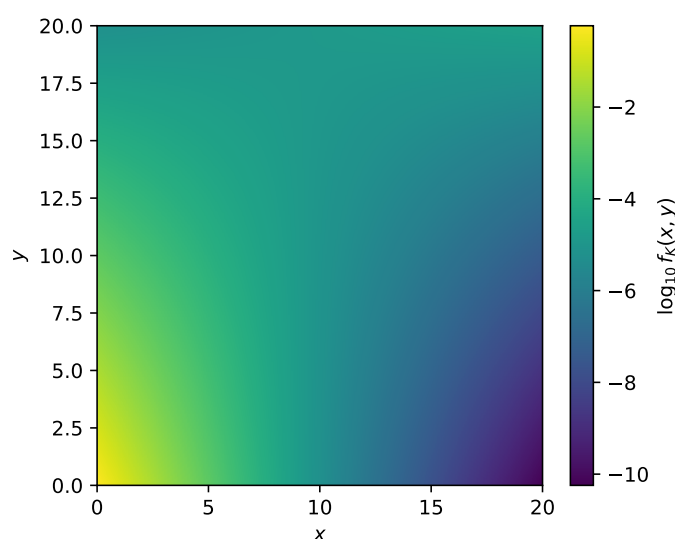


Figure 2. The truncated maximizer f_K at $K = 20$ in the baseline scenario ($\beta_1 = 1.1532$, $\beta_2 = 0.5741$, $\gamma_K = 0.06165$) plotted on a $\log_{10}$ scale. The density does not rise toward the capacity corner globally—its value there is far below the bulk near the origin—but along the high- x , high- y diagonal, it passes through a minimum and turns upward again as (K, K) is approached. The cross-moment is drawn from the upper-right region out of all proportion to the mass: The region $\{x > 5, y > 5\}$ holds 0.36% of the mass and contributes 14% of $\mathbb{E}[XY]$, and the unit corner square $[K - 1, K]^2$ holds 2.1×10^{-5} of the mass and contributes 0.32%. We do not decompose $\mathbb{E}[XY]$ into “bulk” and “excess” parts; the figure shows the density itself.

Remark 5.6 (Reproducibility of the numerical check). The quadrature order, panel locations, coupling bisection bracket, and nonlinear solver tolerances were fixed before producing the table. The complete reproducibility specification, software versions, full 36-row numerical grid, quadrature-sensitivity check, and region diagnostics are incorporated directly into the Supplementary section after the References (Supplementary Tables S1–S5). The computation is deterministic and uses no random sampling. The largest moment residual over the whole grid is 7.7×10^{-13} . The sensitivity check resolves representative extreme cases at 120 and 220 nodes per panel instead of 160; this is numerical validation that the reported digits are not artifacts of the quadrature order, not a proof of quadrature convergence. No statement in the paper depends on this numerical check.

Remark 5.7 (Choosing the capacity bound in practice). The results of this section bear on the modeling choice of K in three ways.

If K is too small, the target is simply infeasible: By Theorem 5.1 (i) feasibility requires $\max\{\mu_1, \mu_2\} < K$ and $C < K \min\{\mu_1, \mu_2\}$, so a capacity at or below $C/\min\{\mu_1, \mu_2\}$ admits no density with the prescribed moments at all. Just above that threshold, the entropy deficit is at its largest, since $\mathcal{E}$ is strictly decreasing in K on the feasible range (Theorem 5.1 (iii)). We do not quantify that regime: Theorem 5.3 and Corollary 5.5 are large- K statements and describe neither the deficit nor the multiplier near the feasibility boundary.

Increasing K lowers the entropy deficit, at the rate $\mathcal{E}(K) \sim \Delta(\mu_1^{-1} + \mu_2^{-1})/K$ of Theorem 5.3, and weakens the attractive coupling, at the rate $\gamma_K \sim (\mu_1^{-1} + \mu_2^{-1})/K$ of Corollary 5.5; by Corollary 5.5 (i), the confined maximizer converges in total variation to the product law that ignores the attractive constraint. The cost is numerical. Two separate indications, neither of them a theorem about the exact maximizer, point the same way. In the mixture used for the upper bound in Theorem 5.3, the excess cross-moment is carried by a blob of unit width and mass $\varepsilon_K = \Theta(K^{-2})$ in a box of side K ; in the computed example of Section 5.1, the corresponding feature of the exact maximizer is a thin ridge near the capacity corner (Figure 2, at $K = 20$). The results of this section do not characterize the width of that layer for the exact maximizer. What they do make plain is that the moment equations stiffen as K grows: The quadrature panels of Section 5.1 were placed for exactly this reason.

Neither rate is a rule for selecting K from data. Both are asymptotic statements with unquantified $o(K^{-1})$ remainders, and they describe how the model changes with the capacity bound, not which capacity bound is correct. In applications K should come from the modeling context—a physical or contractual capacity, a saturation level, or an instrument range—with Theorems 5.1 and 5.3 then serving to quantify what that choice costs in entropy and how far the resulting law sits from the unconfined ideal.

6. Examples

Example 6.1 (Gaussian AR(1) as the exactly critical case). Let $\{X_t\}$ be the stationary Gaussian first-order autoregression (AR(1))

$$X_t - \mu = \phi(X_{t-1} - \mu) + \epsilon_t, \quad \epsilon_t \text{ i.i.d. } N(0, \sigma_\epsilon^2), \quad |\phi| < 1, \quad (6.1)$$

for which

$$\text{Var}(X_t) = \frac{\sigma_\epsilon^2}{1 - \phi^2}, \quad \text{Corr}(X_t, X_{t-1}) = \phi.$$

On $\mathbb{R}^2$ (rather than $\mathcal{Q}$), the joint law of $(X_{t-1}, X_t) = (x, y)$ is the maximum-entropy law under the marginal means, the marginal variances, and the lag-one covariance. Its covariance matrix is

$$\frac{\sigma_\epsilon^2}{1 - \phi^2} \begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix},$$

whose inverse is

$$\sigma_\epsilon^{-2} \begin{pmatrix} 1 & -\phi \\ -\phi & 1 \end{pmatrix},$$

so the exponent equals

$$-\frac{(x-\mu)^2 + (y-\mu)^2 - 2\phi(x-\mu)(y-\mu)}{2\sigma_\epsilon^2} \quad (6.2)$$

up to an additive constant. Comparing (6.2) with $-\alpha x^2 - \alpha' y^2 + \gamma xy$ plus linear terms identifies

$$\alpha = \alpha' = \frac{1}{2\sigma_\epsilon^2}, \quad \gamma = \frac{\phi}{\sigma_\epsilon^2}, \quad (6.3)$$

and hence $4\alpha\alpha' = \sigma_\epsilon^{-4}$ and $\gamma^2 = \phi^2\sigma_\epsilon^{-4}$, so that

$$\frac{\gamma^2}{4\alpha\alpha'} = \phi^2, \quad \text{and therefore} \quad \gamma^2 < 4\alpha\alpha' \iff |\phi| < 1. \quad (6.4)$$

The Gaussian positive-definiteness condition of Corollary 4.6 and the stationarity condition for (6.1) are thus the same condition, with normalizability margin $1 - \gamma^2/(4\alpha\alpha') = 1 - \phi^2$; Table 4 traces it across the range of ϕ . Note that γ carries the sign of ϕ , so positive autocorrelation is an attractive coupling in the sense of Section 2, while the margin depends on $|\phi|$ alone: $\phi = \pm 0.7$ sit at the same distance from the boundary but on opposite sides of the coupling trichotomy. Recent information-theoretic work on autocorrelation for weakly stationary processes, including explicit formulas for autoregressive moving average (ARMA) models, is given by Contreras-Reyes [42].

Table 4. The Gaussian AR(1) of (6.1) across the range of ϕ , through the identification (6.3).

ϕ	coupling	$\gamma^2/(4\alpha\alpha') = \phi^2$	margin $1 - \phi^2$	behavior
-0.7	repulsive ($\gamma < 0$)	0.49	0.51	Comfortably inside the normalizable region; negative lag-one correlation.
0	uncoupled ($\gamma = 0$); critical moment target	0	1	Independent coordinates; the joint law is a product.
0.7	attractive ($\gamma > 0$)	0.49	0.51	Same margin as $\phi = -0.7$, opposite coupling sign.
0.95	attractive ($\gamma > 0$)	0.9025	0.0975	Near the boundary: $\text{Var}(X_t) = \sigma_\epsilon^2/(1 - \phi^2)$ is about ten times σ_ϵ^2 , and the density flattens along the diagonal. As $ \phi \uparrow 1$ the margin vanishes, the quadratic form degenerates, and the law ceases to be normalizable.

Corollary 4.6 shows this classical fact is not an accident of Gaussianity but the critical case $p = q = 2$ of a general trichotomy: Variance constraints supply the minimal admissible damping order, and weakening them to any sub-quadratic self-moment destroys the program regardless of coefficients. We also record the relation between the plane and the orthant. With damping statistics $|x|^p, |y|^q$, the trichotomy of Theorem 4.1 transfers to $\mathbb{R}^2$ verbatim by quadrant decomposition: The two quadrants with $xy \leq 0$ are dominated outright, while the two with $xy \geq 0$ reduce to the orthant integral under reflection, which flips only the signs of the linear coefficients—and parts (i) and (iii), together with the strict inequality of part (ii), are independent of β_1, β_2 . One sharpening appears on the equality manifold, and Proposition 4.3 now gives it for every critical pair (p, p') rather than only at $p = q = 2$: Convergence on $\mathbb{R}^2$ would require the criterion (4.7) to hold for the reflected escape direction as well, that is, both $L(y) \rightarrow +\infty$ and $-L(y) \rightarrow +\infty$, which is impossible. Hence on the plane, the equality case always diverges—recovering the classical fact that the degenerate Gaussian ($\gamma^2 = 4\alpha\alpha'$) is not normalizable on $\mathbb{R}^2$ for any choice of means. The boundary escape of Remark 4.5 is thus an orthant-only phenomenon.

Example 6.2 (A variance-free attractive program and its conditional resolution: The truncated-exponential autoregressive chain). The results of this paper originate in the following modeling problem, which we record here in full since it is resolved by the machinery developed above. One seeks a stationary Markov chain $\{X_n\}$ of nonnegative states with a prescribed marginal mean and positive lag-1 correlation (“friction”: large states breed large states), constructed in the variance-free, first-moment-only spirit of the exponential family. The natural joint maximum-entropy ansatz—marginal means and the lag-1 product moment $\mathbb{E}[X_n X_{n-1}]$, with no self second moments—is precisely the attractive program that Theorem 3.1 (c) shows to be ill posed, the divergence mechanism of Lemma 3.2 being that the conditional rate $\beta_1 - \gamma x_{n-1}$ turns nonpositive once the prior state exceeds β_1/γ .

The conditional object survives the failure of the joint one. Retaining the conditional density of the divergent ansatz and confining the state to $[0, K]$ with $K < \beta_1/\gamma$ defines the Markov chain

$$X_n \mid X_{n-1} = x \sim \text{TruncExp}(r(x), K), \quad f(u \mid x) = \frac{r(x)e^{-r(x)u}}{1 - e^{-r(x)K}} \quad \text{on } [0, K], \quad r(x) = \beta_1 - \gamma x,$$

whose rate satisfies $r(x) \geq \beta_1 - \gamma K > 0$ for all $x \in [0, K]$, so the chain is well defined by induction. Each conditional law is the one-dimensional instance of the box maximizers of Section 5: the maximum-entropy density on $[0, K]$ under a single (state-dependent) mean constraint. The chain is therefore a conditional resolution of the ill-posed joint program that preserves its information-theoretic principle, with the capacity wall $K < \beta_1/\gamma$ appearing on both sides: From the conditional side it keeps the rate positive, and it is exactly the divergence threshold of the joint program. Corollary 5.5 explains its scale from the joint side as well. The attraction sustainable on a box of side K is $\Theta(1/K)$, so the conditional requirement $\gamma < \beta_1/K$ and the entropy-matched coupling $\gamma_K \sim (\mu_1^{-1} + \mu_2^{-1})/K$ are the same scaling law expressed from the two viewpoints, at the exact entropic price $\Delta(\mu_1^{-1} + \mu_2^{-1})/K$ of Theorem 5.3. The process-level theory of the chain—stationarity, ergodicity, and queueing applications—lies outside the scope of the present paper. General sufficient conditions for geometric ergodicity of nonlinear autoregressions, which provide the natural background for such a process-level analysis, are given by An and Huang [43]; the chain displayed above is ours and is not one of the models treated there.

Remark 6.3 (Connection with energy-based models). The same obstruction appears in continuous energy-based models. A model of the form $p_\theta(x) \propto \exp(-E_\theta(x))$ is meaningful only when the partition function is finite. If, on an unbounded nonnegative domain, the energy contains an attractive bilinear term

$$E_\theta(x, y) = \beta_1 x + \beta_2 y - \gamma xy, \quad \gamma > 0,$$

then Lemma 3.2 shows that no choice of the linear coefficients normalizes the model. This does not explain all training instabilities in energy-based learning, but it gives a simple analytic test for one of them: Attractive cross-interactions on unbounded continuous state spaces require damping or truncation. The damping thresholds of Theorem 4.1 and the finite-box law $K\gamma_K \rightarrow \mu_1^{-1} + \mu_2^{-1}$ quantify those two repairs. Nonlinear time-series modeling has a long tradition of departing from linear specifications; De Gooijer [44] is cited here for threshold moving-average context only. That work does not treat product nonlinearities, and the normalizability question studied here is not addressed there.

7. The n -dimensional classification and its degenerate boundary

We now classify normalizability in arbitrary dimension. Let $Q_n = [0, \infty)^n$ and, for a symmetric matrix $\Gamma \in \mathbb{R}^{n \times n}$ and $\beta \in \mathbb{R}^n$, define

$$Z(\beta, \Gamma) = \int_{Q_n} \exp(-\beta^\top x + \tfrac{1}{2} x^\top \Gamma x) dx. \quad (7.1)$$

In the maximum-entropy interpretation, off-diagonal entries Γ_{ij} arise from pairwise cross-moment constraints ($\Gamma_{ij} > 0$ meaning attraction), while diagonal entries $\Gamma_{ii} = -2\alpha_i \leq 0$ arise from self second-moment (variance) constraints. The variance-free programs of Section 3 correspond to zero diagonal; the quadratically damped programs of Section 4 (case $p = q = 2$) to strictly negative diagonal.

Definition 7.1. A symmetric matrix M is copositive if $x^\top Mx \geq 0$ for all $x \in Q_n$ and strictly copositive if $x^\top Mx > 0$ for all $x \in Q_n \setminus \{0\}$. For Γ with $-\Gamma$ copositive, the zero cone is

$$\mathcal{Z}(\Gamma) = \{v \in Q_n \setminus \{0\} : v^\top \Gamma v = 0\}.$$

A contemporaneous parallel from a different field is worth noting: Sturmfels and Telek [32] show that Feynman integrals converge when their kinematic parameters lie in a copositive cone, so the pattern “copositivity governs integral convergence” appears independently there; Theorem 7.3 identifies the exact form this pattern takes for Gibbs densities on the orthant.

The following first-order information at zero directions is the key structural fact.

Lemma 7.2 (First-order conditions on the zero cone). *Let $-\Gamma$ be copositive and $v \in \mathcal{Z}(\Gamma)$. Then $(\Gamma v)_i \leq 0$ for every i , and $(\Gamma v)_i = 0$ for every $i \in \text{supp}(v)$.*

Proof. The function $q(x) = x^\top \Gamma x$ satisfies $q \leq 0$ on Q_n and $q(v) = 0$, so v is a global maximizer of q over the cone Q_n . For any i and $t > 0$, $q(v + te_i) = 2t(\Gamma v)_i + t^2 \Gamma_{ii} \leq 0$ forces $(\Gamma v)_i \leq 0$ on dividing by t and letting $t \downarrow 0$. If $i \in \text{supp}(v)$, then $v - te_i \in Q_n$ for small $t > 0$, and $q(v - te_i) = -2t(\Gamma v)_i + t^2 \Gamma_{ii} \leq 0$ forces $(\Gamma v)_i \geq 0$ likewise; hence $(\Gamma v)_i = 0$. $\square$

Theorem 7.3 (Copositive classification away from the higher-codimension degenerate boundary). *Let $\Gamma \in \mathbb{R}^{n \times n}$ be symmetric and $\beta \in \mathbb{R}^n$, and for $v \in Q_n \setminus \{0\}$ write*

$$d(v) = n - |\text{supp}(v)|$$

for the transverse codimension of v .

- (a) *If $-\Gamma$ is not copositive, then $Z(\beta, \Gamma) = \infty$ for every $\beta \in \mathbb{R}^n$.*
- (b) *If $-\Gamma$ is copositive and $\beta^\top v > 0$ for every $v \in \mathcal{Z}(\Gamma)$, then $Z(\beta, \Gamma) < \infty$.*
- (c) *If $-\Gamma$ is copositive and $\beta^\top v < 0$ for some $v \in \mathcal{Z}(\Gamma)$, then $Z(\beta, \Gamma) = \infty$.*
- (d) *If $-\Gamma$ is copositive and there exists $v \in \mathcal{Z}(\Gamma)$ with $\beta^\top v = 0$ and $d(v) \leq 1$, then $Z(\beta, \Gamma) = \infty$.*
- (e) *There remains one regime: $-\Gamma$ copositive, $\beta^\top v \geq 0$ throughout $\mathcal{Z}(\Gamma)$, and equality attained only at zero directions with $d(v) \geq 2$. There the pair (copositivity of $-\Gamma$, the sign of $\beta^\top v$ on $\mathcal{Z}(\Gamma)$) is not a sufficient criterion: Both $Z < \infty$ and $Z = \infty$ occur (Remark 7.4), and deciding between them requires finer information about the transverse vanishing behavior of $q(x) = x^\top \Gamma x$ relative to the zero cone—the two examples of Remark 7.4 have the same zero cone and opposite answers. The full pair (Γ, β) of course determines $Z(\beta, \Gamma)$; what is insufficient is this first-order data.*

Proof. Each part isolates one geometric mechanism: a single quadratically growing cell for (a), a radial foliation of the orthant for (b), a shrinking tube along a zero direction for (c), and a disjoint family of shrinking cells along a zero direction for (d).

(a) Choose $v \in Q_n$, $|v| = 1$, with $v^\top \Gamma v = 2c > 0$, and consider the single translated cell

$$R_T = Tv + [0, 1]^n \subset Q_n.$$

For $x = Tv + u$, $u \in [0, 1]^n$,

$$\frac{1}{2}x^\top \Gamma x - \beta^\top x = cT^2 + T v^\top \Gamma u + \frac{1}{2}u^\top \Gamma u - \beta^\top (Tv + u) = cT^2 + O(T),$$

so

$$Z \geq \int_{R_T} e^{cT^2 + O(T)} dx \geq e^{cT^2/2} \quad \text{for all large } T;$$

letting $T \rightarrow \infty$ gives $Z = \infty$.

(b) Foliate Q_n by the ℓ_1 -radius: $x = r\omega$ with $r = |x|_1$ and ω in the unit simplex $S = \{\omega \geq 0 : |\omega|_1 = 1\}$, so that $dx = r^{n-1} dr d\zeta(\omega)$ for a finite measure ζ on S . Write $q(\omega) = \omega^\top \Gamma \omega \leq 0$ and $\ell(\omega) = \beta^\top \omega$. The compact set $\mathcal{Z}_S = \{q = 0\} \cap S$ satisfies $\ell > 0$ on it by hypothesis, so there are an open $U \supseteq \mathcal{Z}_S$ in S and $\epsilon > 0$ with $\ell \geq \epsilon$ on U ; on the compact complement $S \setminus U$, we have $q \leq -\delta$ for some $\delta > 0$. Hence for $\omega \in U$, the exponent is at most

$$-\epsilon r + \frac{1}{2}r^2 q(\omega) \leq -\epsilon r,$$

while for $\omega \notin U$ it is at most

$$\|\beta\|_\infty r - \frac{\delta}{2}r^2.$$

Both radial profiles are integrable against $r^{n-1} dr$, uniformly in ω , so $Z < \infty$.

(c) Let $v \in \mathcal{Z}(\Gamma)$, $|v| = 1$, with $\beta^\top v = -\kappa < 0$, and set $S = \text{supp}(v)$. Consider the shrinking cell

$$R_T = \{Tv + u + w : u \in [0, 1]^S, w \in [0, \frac{1}{T}]^{S^c}\} \subset Q_n.$$

For $x = Tv + u + w$, Lemma 7.2 gives

$$T v^\top \Gamma(u + w) = T \sum_{i \in S^c} (\Gamma v)_i w_i = O(1),$$

since $(\Gamma v)_i = 0$ on S and $|w|_\infty \leq 1/T$; also

$$\frac{1}{2}(u + w)^\top \Gamma(u + w) = O(1)$$

and $\beta^\top(u + w) = O(1)$. Hence the exponent on R_T equals $\kappa T + O(1)$, while $\text{vol}(R_T) \asymp T^{-(n-|S|)}$. Therefore

$$Z \geq e^{\kappa T + O(1)} T^{-(n-|S|)} \rightarrow \infty \quad \text{as } T \rightarrow \infty.$$

(d) Here the exponential gain of (c) is absent, and divergence must instead be assembled from infinitely many cells of individually bounded contribution. Let $v \in \mathcal{Z}(\Gamma)$ with $|v| = 1$, $\beta^\top v = 0$, and

$$d := d(v) = |S^c| \leq 1,$$

where $S = \text{supp}(v)$. Fix $\delta \in (0, 1]$, set $L = \max\{1, 3\delta\sqrt{n}\}$ and $T_k = kL$ for $k = 1, 2, \dots$, and consider the disjoint cells

$$R_k = \left\{ T_k v + u + w : u \in [0, \delta]^S, w \in [0, \delta/T_k]^{S^c} \right\} \subset Q_n.$$

For $x = T_k v + u + w$, expanding and using $v^\top \Gamma v = 0$ and $\beta^\top v = 0$,

$$\frac{1}{2} x^\top \Gamma x - \beta^\top x = T_k v^\top \Gamma (u + w) + \frac{1}{2} (u + w)^\top \Gamma (u + w) - \beta^\top (u + w).$$

By Lemma 7.2, $(\Gamma v)_i = 0$ for $i \in S$ and $(\Gamma v)_i \leq 0$ for $i \in S^c$, so the first term is

$$T_k \sum_{i \in S^c} (\Gamma v)_i w_i \geq T_k \cdot \frac{\delta}{T_k} \sum_{i \in S^c} (\Gamma v)_i = \delta \sum_{i \in S^c} (\Gamma v)_i,$$

a bound independent of k ; since $T_k \geq L \geq 1$ we have $|u + w|_\infty \leq \delta$, so the remaining two terms are bounded below by a constant depending only on Γ, β, δ , and n . Hence there is $M < \infty$, independent of k , with exponent $\geq -M$ on every R_k , whence

$$\int_{R_k} e^{-\beta^\top x + \frac{1}{2} x^\top \Gamma x} dx \geq e^{-M} \text{vol}(R_k) = e^{-M} \delta^n T_k^{-d}.$$

The cells are pairwise disjoint: Each R_k lies within Euclidean distance $\delta\sqrt{n}$ of $T_k v$, while

$$|T_k v - T_l v| = |k - l| L \geq L \geq 3\delta\sqrt{n} \quad \text{for } k \neq l.$$

Summing,

$$Z(\beta, \Gamma) \geq e^{-M} \delta^n \sum_{k \geq 1} (kL)^{-d} = \infty,$$

the series being $\sum_k 1$ when $d = 0$ and $L^{-1} \sum_k k^{-1}$ when $d = 1$. For $d \geq 2$ this series converges, so the present cell argument is inconclusive; Remark 7.4 shows that both alternatives already occur at $d = 2$. $\square$

The scope of the classification is summarized in Table 5.

Table 5. Scope of the quadratic n -dimensional normalizability classification.

Regime	Status
$-\Gamma$ not copositive	Settled: $Z = \infty$ (Theorem 7.3 (a))
$-\Gamma$ copositive, $\beta^\top v > 0$ on $\mathcal{Z}(\Gamma)$	Settled: $Z < \infty$ (Theorem 7.3 (b))
$-\Gamma$ copositive, $\beta^\top v < 0$ somewhere	Settled: $Z = \infty$ (Theorem 7.3 (c))
Boundary, some zero direction with $d(v) \leq 1$	Settled: $Z = \infty$ (Theorem 7.3 (d))
Boundary, equality only where $d(v) \geq 2$	Open: both alternatives occur (Remark 7.4; Problem (3), Section 8)

Remark 7.4 (The degenerate boundary beyond codimension one). Theorem 7.3 (d) settles the degenerate boundary whenever some zero direction with $\beta^\top v = 0$ has transverse codimension $d(v) \leq 1$. The simplest instance is $n = 2$,

$$\Gamma = \begin{pmatrix} 0 & -\lambda \\ -\lambda & -2\alpha \end{pmatrix}$$

with $\lambda, \alpha > 0$ and $\beta = (0, \beta_2), \beta_2 > 0$: Integrating out x_1 leaves the integrand $e^{-\beta_2 x_2 - \alpha x_2^2} / (\lambda x_2)$, whose singularity x_2^{-1} in the single transverse dimension is non-integrable at the origin, reproducing $Z = \infty$ directly.

What remains open is the regime $d(v) \geq 2$, where both alternatives genuinely occur. The two examples below both have $n = 3$, with the zero cone equal to the e_1 -axis—hence $d(v) = 2$ —and $\beta_1 = 0$.

A convergent example. Take $\Gamma_{11} = 0, \Gamma_{12} = \Gamma_{13} = -\lambda, \Gamma_{22} = \Gamma_{33} = -2\alpha$, and $\Gamma_{23} = 0$, with $\lambda, \alpha > 0$ and $\beta = (0, \beta_2, \beta_3)$ with $\beta_2, \beta_3 > 0$. Then

$$x^\top \Gamma x = -2\lambda x_1(x_2 + x_3) - 2\alpha(x_2^2 + x_3^2), \quad (7.2)$$

which is ≤ 0 on Q_3 and vanishes there exactly when $x_2 = x_3 = 0$; therefore, $-\Gamma$ is copositive and $\mathcal{Z}(\Gamma)$ is the e_1 -axis. Integrating out x_1 ,

$$Z = \frac{1}{\lambda} \int_0^\infty \int_0^\infty \frac{e^{-\beta_2 x_2 - \beta_3 x_3 - \alpha(x_2^2 + x_3^2)}}{x_2 + x_3} dx_2 dx_3.$$

The substitution $u = x_2 + x_3, v = x_2/(x_2 + x_3)$, under which $(x_2, x_3) = (uv, u(1-v))$ and $dx_2 dx_3 = u du dv$ on $(u, v) \in (0, \infty) \times (0, 1)$, cancels the singular factor exactly:

$$Z = \frac{1}{\lambda} \int_0^\infty \int_0^1 \exp[-u\{\beta_2 v + \beta_3(1-v)\} - \alpha u^2\{v^2 + (1-v)^2\}] dv du. \quad (7.3)$$

Since $\beta_2 v + \beta_3(1-v) \geq \min\{\beta_2, \beta_3\}$ and $v^2 + (1-v)^2 \geq 1/2$ for $v \in [0, 1]$,

$$Z \leq \frac{1}{\lambda} \int_0^\infty e^{-\min\{\beta_2, \beta_3\}u - \alpha u^2/2} du < \infty. \quad (7.4)$$

A single estimate thus controls the origin and infinity together; no separate local-integrability argument is involved.

A divergent example. Take $\Gamma_{11} = 0, \Gamma_{12} = -\lambda, \Gamma_{13} = 0, \Gamma_{22} = \Gamma_{33} = -2\alpha, \Gamma_{23} = 0$, and $\beta = (0, \beta_2, \beta_3)$ with $\beta_2, \beta_3 > 0$. The zero cone is again the e_1 -axis with $d(v) = 2$, but after integrating out x_1 the integral factorizes:

$$Z = \left(\int_0^\infty \frac{e^{-\beta_2 x_2 - \alpha x_2^2}}{\lambda x_2} dx_2 \right) \left(\int_0^\infty e^{-\beta_3 x_3 - \alpha x_3^2} dx_3 \right) = \infty,$$

the first factor diverging logarithmically at the origin because x_3 does not couple to the zero direction: $(\Gamma e_1)_3 = 0$.

Together the two examples show that the transverse codimension alone does not decide the boundary: Both have $d(v) = 2$ and opposite answers. They differ in how many transverse coordinates couple to the zero direction at first order—two in the convergent example, one in the divergent one—but we do not claim that this count is the complete invariant, and we have not proved that it is. What the examples do establish is that the first-order data used in Theorem 7.3 (a)–(d) must be supplemented by finer information about how $q(x) = x^\top \Gamma x$ vanishes transversally to the zero cone. A complete criterion is listed as open Problem (3) of Section 8.

Corollary 7.5 (Variance-free case: no cross-subsidization). *Suppose $\Gamma_{ii} = 0$ for all i (no self second-moment constraints). Then the following are equivalent: (i) $Z(\beta, \Gamma) < \infty$ for some $\beta \in \mathbb{R}^n$; (ii) $-\Gamma$ is copositive; (iii) $\Gamma_{ij} \leq 0$ for all $i \neq j$. Moreover, under (iii), $Z(\beta, \Gamma) < \infty$ for every $\beta \in (0, \infty)^n$. In particular, a single attractive pair $\Gamma_{ij} > 0$ forces $Z = \infty$ for every β , no matter how strongly repulsive all other couplings are. On the orthant, repulsion elsewhere cannot subsidize attraction anywhere.*

Proof. (iii) $\Rightarrow$ (ii). With zero diagonal, $x^\top \Gamma x = 2 \sum_{i < j} \Gamma_{ij} x_i x_j \leq 0$ on Q_n . (ii) $\Rightarrow$ (iii). Evaluate at $x = e_i + e_j$ to get $2\Gamma_{ij} \leq 0$. (iii) $\Rightarrow$ finiteness for $\beta \in (0, \infty)^n$: The exponent is at most $-\beta^\top x$, which is integrable. Finally, $\neg$ (iii) $\Rightarrow \neg$ (i) is Theorem 7.3 (a) with $v = e_i + e_j$, and this also proves the no-cross-subsidization statement, since the divergent cell argument along $e_i + e_j$ is insensitive to all other entries of Γ . $\square$

Remark 7.6. The contrast with the full space $\mathbb{R}^n$ is instructive: There, a positive off-diagonal entry can always be absorbed by sufficiently strong variance damping (positive definiteness is an eigenvalue condition mixing all entries). On the orthant with zero diagonal, the condition is entrywise—the co-NP-hard copositivity condition trivializes—because the escape mechanism of Remark 3.5 runs along the two-dimensional face spanned by the attractive pair, where no other coordinate can intervene. The full copositive (hard) regime appears only once damping is present, as the next two corollaries show.

Remark 7.7 (Relation to exponential graphical models). The variance-free sign condition of Corollary 7.5 is known at the family level in the graphical-models literature: Besag’s auto-exponential models [45] and their modern successors [46] are exactly the densities $\propto \exp(\theta^\top x + \sum_{i < j} \Gamma_{ij} x_i x_j)$ on positive variables, and it is recorded there that normalizability restricts such models to nonpositive dependencies; the square-root graphical models of Inouye et al. [47] repair this by changing the sufficient statistics, a third route beside the damping and confinement studied here. Relative to that literature, the contributions of this paper are the variational upgrade—at attractive targets, no density of any form attains the entropy supremum, whose value Theorem 7.10 computes exactly—and the damped classification of Theorem 7.3, with its zero-cone boundary and complexity consequences, which has no graphical-models counterpart.

Corollary 7.8 (Strict copositivity characterizes unconditional normalizability). $Z(\beta, \Gamma) < \infty$ for every $\beta \in \mathbb{R}^n$ if and only if $-\Gamma$ is strictly copositive.

Proof. If $-\Gamma$ is strictly copositive, then $\mathcal{Z}(\Gamma) = \emptyset$ and Theorem 7.3 (b) applies vacuously for every β . Conversely, if $-\Gamma$ is not copositive, Theorem 7.3 (a) gives divergence; if it is copositive but not strictly, pick $v \in \mathcal{Z}(\Gamma)$ and any β with $\beta^\top v < 0$, and apply Theorem 7.3 (c). $\square$

Corollary 7.9 (Deciding normalizability is co-NP-hard). *The decision problem “given rational (β, Γ) , is $Z(\beta, \Gamma) < \infty$?” is co-NP-hard.*

Proof. Reduce from copositivity, whose complement is NP-complete [29]. Given a symmetric rational M , set $\Gamma = -M$ and $\beta = \mathbf{1}$. Since $\mathbf{1}^\top v = |v|_1 > 0$ for every $v \in Q_n \setminus \{0\}$, the zero-cone condition of Theorem 7.3 (b) is automatic, and the degenerate boundary never arises. By parts (a) and (b), $Z(\mathbf{1}, -M) < \infty$ if and only if M is copositive. $\square$

We conclude the section by lifting the entropic-invisibility phenomenon of Theorem 3.1 (c) to n dimensions for constraint sets in which every constrained pair is attractive.

Theorem 7.10 (Entropic invisibility in n dimensions). *Let $P \neq \emptyset$ be a set of pairs $\{i, j\}$, let $\mu \in (0, \infty)^n$, and let $C_{ij} > \mu_i \mu_j$ for every $\{i, j\} \in P$. Consider the program of maximizing $h(g)$ over densities g on $\mathcal{Q}_n$ with $\mathbb{E}[X_i] = \mu_i$ for all i and $\mathbb{E}[X_i X_j] = C_{ij}$ for all $\{i, j\} \in P$ (other pairs unconstrained). Then*

$$\sup h = n + \sum_{i=1}^n \ln \mu_i,$$

the value of the means-only program, and the supremum is not attained.

Proof. Upper bound. Let

$$\pi_\mu = \bigotimes_{i=1}^n \text{Exp}(\text{mean } \mu_i).$$

For every feasible g with $h(g) > -\infty$, Gibbs' inequality gives

$$0 \leq D(g \parallel \pi_\mu) = -h(g) + \sum_i \ln \mu_i + \sum_i \frac{\mathbb{E}_g[X_i]}{\mu_i} = -h(g) + n + \sum_i \ln \mu_i.$$

Hence $h(g) \leq n + \sum_i \ln \mu_i$. Equality would force $g = \pi_\mu$ a.e., which would give $\mathbb{E}[X_i X_j] = \mu_i \mu_j < C_{ij}$ for every constrained attractive pair, contradicting feasibility. Thus the bound is strict for every finite-entropy feasible density; densities with entropy $-\infty$ do not affect the supremum.

Escape sequence. For $e = \{i, j\} \in P$ and $s \geq 0$, let $B_s^{(e)}$ denote the uniform density on the box with coordinates i and j ranging over $[s, s+1]$ and all other coordinates over $[0, 1]$; it has differential entropy 0, means $s + \frac{1}{2}$ in coordinates i, j and $1/2$ elsewhere, and cross-moments $(s + 1/2)^2$ for the pair e itself, $(s + 1/2)/2$ for pairs sharing exactly one index with e , and $1/4$ for pairs disjoint from e . Write $M = |P|$, $\Delta_e = C_{ij} - \mu_i \mu_j > 0$, let d_i denote the degree of i in the constrained graph $([n], P)$, and set

$$A(e) = \{f \in P \setminus \{e\} : |e \cap f| = 1\}, \quad N_e^0 = \#\{f \in P : e \cap f = \emptyset\}.$$

For $\varepsilon > 0$, put $\tau = \sqrt{\varepsilon}$ and

$$g_\varepsilon = (1 - M\varepsilon) \bigotimes_{i=1}^n \text{Exp}(\text{mean } m_i) + \varepsilon \sum_{e \in P} B_{s_e}^{(e)}, \quad \sigma_e = \varepsilon \left(s_e + \frac{1}{2}\right)^2,$$

so that

$$\varepsilon \left(s_e + \frac{1}{2}\right) = \tau \sqrt{\sigma_e}.$$

Collecting the contributions listed above, feasibility of g_ε is exactly the system

$$\mu_i = (1 - M\tau^2) m_i + \tau \sum_{e \ni i} \sqrt{\sigma_e} + \frac{\tau^2}{2} (M - d_i), \quad i = 1, \dots, n, \quad (7.5)$$

$$C_{ij} = (1 - M\tau^2) m_i m_j + \sigma_e + \frac{\tau}{2} \sum_{f \in A(e)} \sqrt{\sigma_f} + \frac{\tau^2}{4} N_e^0, \quad e = \{i, j\} \in P. \quad (7.6)$$

Every term is exact and accounted for: In (7.5) a blob on an edge through i contributes

$$\varepsilon \left(s_e + \frac{1}{2}\right) = \tau \sqrt{\sigma_e}$$

to the mean of X_i , while each of the $M - d_i$ blobs not incident to i contributes $\varepsilon/2 = \tau^2/2$; in (7.6) the edge's own blob contributes

$$\varepsilon \left(s_e + \frac{1}{2}\right)^2 = \sigma_e,$$

a blob on an adjacent edge f contributes

$$\frac{1}{2}\varepsilon \left(s_f + \frac{1}{2}\right) = \frac{1}{2}\tau \sqrt{\sigma_f},$$

and each of the N_e^0 disjoint blobs contributes $\varepsilon/4 = \tau^2/4$.

Let $F(m, \sigma, \tau) \in \mathbb{R}^{n+M}$ collect the residuals of (7.5) and (7.6). On the open set $\{\sigma_e > 0 \text{ for all } e\}$, the map F is C^∞ in (m, σ, τ) , and

$$F(\mu, \Delta, 0) = 0,$$

since at $\tau = 0$ the system reduces to $\mu_i = m_i$ and $C_{ij} = m_i m_j + \sigma_e$. Ordering the variables as (m, σ) , the Jacobian at $(\mu, \Delta, 0)$ is block triangular:

$$\left. \frac{\partial F}{\partial(m, \sigma)} \right|_{(\mu, \Delta, 0)} = - \begin{pmatrix} I_n & 0 \\ B & I_M \end{pmatrix}, \quad B_{e,k} = \mu_j \delta_{ik} + \mu_i \delta_{jk} \quad (e = \{i, j\}), \quad (7.7)$$

because every σ -derivative in (7.5) and every adjacent-edge term in (7.6) carries a factor τ and so vanishes at $\tau = 0$. Its determinant is $(-1)^{n+M}$, so (7.7) is nonsingular, and the implicit function theorem supplies $\tau_0 > 0$ and unique C^∞ functions $\tau \mapsto (m(\tau), \sigma(\tau))$ on the two-sided interval $|\tau| < \tau_0$ solving $F = 0$; since $\tau = \sqrt{\varepsilon}$ we use only $0 \leq \tau < \tau_0$ from here on. They satisfy

$$m_i(\tau) = \mu_i + O(\tau), \quad \sigma_e(\tau) = \Delta_e + O(\tau).$$

Shrinking τ_0 if necessary, all the requirements hold simultaneously: $1 - M\tau^2 > 0$, $m_i(\tau) > 0$, $\sigma_e(\tau) > 0$, and

$$s_e(\tau) = \frac{\sqrt{\sigma_e(\tau)}}{\tau} - \frac{1}{2} \geq 0,$$

the last because $\sqrt{\sigma_e(\tau)} \rightarrow \sqrt{\Delta_e} > 0$ while $\tau \downarrow 0$. Hence for every $\varepsilon = \tau^2$ with $0 < \tau < \tau_0$, the density g_ε is exactly feasible, with positive bulk means and positive mixture weights, whatever the structure of the constrained graph. Shared vertices, stars, cycles, and complete subgraphs are all covered because (7.5) and (7.6) account for the contribution of every blob to every prescribed moment. Moreover $m(\tau) \rightarrow \mu$ as $\tau \downarrow 0$.

Entropy. By concavity of h over the finite mixture (Lemma 2.2),

$$h(g_\varepsilon) \geq (1 - M\tau^2) \left(n + \sum_i \ln m_i(\tau) \right) + \sum_{e \in P} \tau^2 \cdot 0 \xrightarrow{\tau \downarrow 0} n + \sum_i \ln \mu_i.$$

Combining this with the strict upper bound, the supremum equals $n + \sum_i \ln \mu_i$ and is not attained. $\square$

Remark 7.11 (Mixed constraint sets). When P also contains repulsive or critical pairs (prescribed $C_{ij} \leq \mu_i \mu_j$), the blobs pollute those equality constraints upward by $O(\sqrt{\varepsilon})$, and a product bulk cannot compensate. The natural repair is to take as bulk the maximizer of the reduced (attractive-constraints-deleted) program at targets perturbed by the pollution—which stay strictly repulsive—so

that the conjectured statement reads: The value of the mixed program equals the value of its reduced program and is attained if and only if no constrained pair is attractive. Proving this requires (i) the n -dimensional attainment theorem for purely repulsive programs (the analogue of Theorem 3.1 (a), where the intermediate value theorem must be replaced by a degree-theoretic or dual-attainment argument over $\binom{n}{2}$ coupling parameters) and (ii) continuity of the reduced value function in its targets. Both ingredients are deferred; we list them as open Problem (2) of Section 8.

Remark 7.12 (Diagnostic guide for model checking). The results can be used as a short diagnostic for pairwise-coupled maximum-entropy models on Q_n .

Step 1: Look for undamped attractive pairs. If a target has $C_{ij} > \mu_i \mu_j$ and the corresponding exponent contains an attractive bilinear term without self-moment damping on both coordinates, the unbounded model is at risk of non-attainment. In the variance-free case, one attractive pair is enough to force divergence and entropy invisibility (Corollary 7.5 and Theorem 7.10).

Step 2: Check the damping orders. For an attractive pair regularized by powers $p_i, p_j > 1$, the two-dimensional test is sharp: $1/p_i + 1/p_j < 1$ is safely normalizing, equality is coefficient-dependent, and $1/p_i + 1/p_j > 1$ cannot be repaired by coefficients (Theorem 4.1).

Step 3: Check the critical coefficients. On the critical boundary, use the coefficient inequality in Theorem 4.1 (ii); in the quadratic case, this is $\gamma^2 < 4\alpha\alpha'$. If the coefficients lie on the equality manifold itself, Proposition 4.3 decides the question through the linear multipliers.

Step 4: For quadratic n -dimensional exponents, check copositivity. For

$$-\beta^\top x + \frac{1}{2}x^\top \Gamma x,$$

Theorem 7.3 reduces normalizability to copositivity of $-\Gamma$ and the sign of β on the zero cone. Strict copositivity gives normalizability for all linear terms, while failure of copositivity gives divergence. These two data decide every nondegenerate regime, and also the degenerate one when some zero direction v with $\beta^\top v = 0$ has transverse codimension $d(v) \leq 1$, where the answer is divergence. They do not decide the remaining case, equality attained only where $d(v) \geq 2$: That boundary is unresolved and requires additional transverse analysis (Remark 7.4).

Step 5: If the unbounded model fails, use a wall. Truncation to $[0, K]^2$ restores well-posedness in the feasible range of Theorem 5.1, and Theorem 5.3 gives the leading entropy cost of that repair.

The first three checks are elementary in dimension two. The final copositivity check is computationally hard in general, in line with Corollary 7.9. All five checks concern normalizability for given coefficients. Whether a prescribed moment target is then realized by the resulting family, and the associated entropy program attained, is a further question, answered by Proposition 4.9 only on the realized region M° .

8. Discussion, limitations, and open problems

The results above identify a simple obstruction that is easy to miss in maximum-entropy modeling and in Gibbs modeling on unbounded orthants. An attractive bilinear term may produce a feasible moment constraint while the corresponding Gibbs weight has infinite partition function. The failure is not a numerical artifact and not merely a bad choice of parameters. It is a loss of confinement: A small amount of mass can move outward and carry the excess product moment at negligible entropy cost.

From the statistical-mechanics viewpoint, the main dichotomy says that linear fields do not confine attractive bilinear interactions on the orthant. Damping terms play the role of confining self-energies. Theorem 4.1 gives the exact power threshold at which those self-energies dominate the interaction, and Proposition 4.3 settles the threshold itself, where the linear multipliers decide; in two dimensions the damping classification is therefore complete, in contrast with the degenerate zero-cone boundary in dimension n , which remains open (item (3) below). Truncation gives the hard-wall alternative. The constant in Theorem 5.3 is then the leading entropy cost of imposing that wall as it recedes to infinity.

There are two repairs. The first is damping. Self-moment constraints must dominate the bilinear interaction in the sense of Theorem 4.1; in particular, symmetric damping must be at least quadratic. This restores normalizability at the level of coefficients; whether a prescribed moment target is then realized and attained is a separate question, settled by Proposition 4.9 only on the realized region M° . The second is truncation. A finite capacity bound restores attainment, and the entropy price of the wall has an explicit first-order constant. These two repairs correspond to two modeling choices: Either add real tail information to the constraint set or state explicitly that the model has a finite operating range. Table 6 sets the two side by side. Neither dominates: They require different information and deliver different guarantees, and the choice should follow from which of the two kinds of information the application actually supplies.

Table 6. Self-moment damping and hard-wall truncation compared. Neither repair is universally preferable.

Issue	Self-moment damping	Hard-wall truncation
Modeling change	Adds tail-moment constraints	Restricts the support
Information required	Prescribed self-moments of orders p, q	A physical or operational capacity K
Mathematical guarantee	Coefficient normalizability (Theorem 4.1); attainment only on the realized region M° (Proposition 4.9)	Attainment for every feasible box target (Theorem 5.1)
Threshold	$1/p + 1/q$, and on the critical manifold the coefficient boundary of Theorem 4.1 (ii)	$\max\{\mu_1, \mu_2\} < K$ and $C < K \min\{\mu_1, \mu_2\}$
Main limitation	The realized moment region is not characterized	Sensitivity to the capacity, and numerical stiffness as K grows (Remark 5.7)
Appropriate use	Genuine tail information exists	A defensible finite operating range exists

The n -dimensional results show where the problem becomes computational. For quadratic exponents, the correct cone is the copositive cone rather than the positive-semidefinite cone. This distinction matters on the orthant: In the variance-free case, no amount of repulsion elsewhere can compensate for a single attractive undamped pair. Once diagonal damping is present, the full copositive problem appears, and the decision problem is co-NP-hard.

The paper leaves four points open.

- (1) An exact description of the realized moment region M° in Proposition 4.9, together with a classification of what happens at targets outside it. Non-steepness (Remark 4.11) shows that M° need not exhaust the interior of the moment set, but does not by itself decide whether such a target is realized by a normalizable boundary parameter, attained in the generalized I-projection sense or not attained at all.
- (2) The n -dimensional repulsive attainment theorem and the corresponding mixed attractive–repulsive variational theorem (Remark 7.11).

- (3) A full boundary classification in Theorem 7.3. Part (d) settles the boundary at transverse codimension 0 and 1; what remains is codimension 2 and above, where the examples of Remark 7.4 show that neither the first-order sign data nor the transverse codimension alone can decide the answer. Those examples differ in how many transverse coordinates couple to the zero direction at first order, which suggests where to look, but we do not claim that this count is the invariant. A complete criterion presumably involves the order of vanishing of $q(x) = x^\top \Gamma x$ transverse to each stratum of the zero cone.
- (4) A finite-box analogue in dimension n . We expect the leading price to be additive over constrained attractive pairs, $\sum_{\{i,j\}} \Delta_{ij}(\mu_i^{-1} + \mu_j^{-1})/K$, but this requires a separate multi-corner analysis.

9. Conclusions

A Gibbs formula does not by itself define a probability model. For bilinear interactions on nonnegative unbounded domains, the sign and damping of the interaction determine whether the partition function is finite and whether the associated entropy program is attained. The basic two-dimensional case is sharp: Non-attractive targets are attained, while attractive targets have the same entropy value as the means-only program but no maximizer. Damping and truncation repair the problem in different ways, and both repairs admit explicit thresholds or constants; in two dimensions the damping threshold is resolved down to and including its critical equality manifold, where finiteness is decided by the linear multipliers. In higher dimensions, quadratic normalizability is controlled by copositivity together with the sign of the linear field on the zero cone, which explains both the geometric boundary cases and the computational hardness of the general decision problem. That description is complete except on the degenerate boundary at transverse codimension two or more, where both convergence and divergence occur and the criterion remains open.

The resulting rule is direct: Before using a Gibbs or maximum-entropy model with cross-moment interactions on an unbounded orthant, one must check normalizability and attainment—and these are different checks, the first about coefficients and the second about a prescribed moment target. Attractive interactions require genuine tail damping or an explicit capacity bound.

Author contributions

Abdulaziz M. D. Aljohani: Conceptualization, methodology, formal analysis, writing—original draft preparation, writing—review and editing; Khaled M. Alhawiti: Conceptualization, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Data availability

No external datasets were analyzed in this study. The numerical values used for the illustrative table and figures are generated deterministically from the moment-matching procedure described in Section 5.1. The complete numerical grid and reproducibility details are incorporated directly into the Supplementary section after the References; no separate supplementary files are required.

Use of Generative-AI tools declaration

During the preparation of this manuscript, the authors used an AI assistant (Claude, Anthropic) to assist in the development and checking of proofs and theorem statements, the drafting and editing of the text, and the preparation of the code for the numerical illustration of Section 5.1. The authors have reviewed, verified, and edited all results and text, and take full responsibility for the content of this publication.

Acknowledgments

The authors acknowledge the institutional support of the University of Tabuk. This research received no external funding.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. E. T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.*, **106** (1957), 620–630. <https://doi.org/10.1103/PhysRev.106.620>
2. C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.*, **27** (1948), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
3. I. Csiszár, I -divergence geometry of probability distributions and minimization problems, *Ann. Probab.*, **3** (1975), 146–158. <https://doi.org/10.1214/aop/1176996454>
4. I. Csiszár, F. Matúš, Information projections revisited, *IEEE Trans. Inform. Theory*, **49** (2003), 1474–1490. <https://doi.org/10.1109/TIT.2003.810633>
5. T. M. Cover, J. A. Thomas, *Elements of information theory*, Hoboken: Wiley, 2006. <https://doi.org/10.1002/047174882X>
6. P. Ishwar, P. Moulin, On the existence and characterization of the maxent distribution under general moment inequality constraints, *IEEE Trans. Inform. Theory*, **51** (2005), 3322–3333. <https://doi.org/10.1109/TIT.2005.853317>
7. J. E. Contreras-Reyes, Lerch distribution based on maximum nonsymmetric entropy principle: application to Conway’s game of life cellular automaton, *Chaos Solitons Fract.*, **151** (2021), 111272. <https://doi.org/10.1016/j.chaos.2021.111272>
8. D. Ruelle, Superstable interactions in classical statistical mechanics, *Comm. Math. Phys.*, **18** (1970), 127–159. <https://doi.org/10.1007/BF01646091>
9. H. O. Georgii, *Gibbs measures and phase transitions*, Berlin: De Gruyter, 2011. <https://doi.org/10.1515/9783110250329>
10. S. Kullback, R. A. Leibler, On information and sufficiency, *Ann. Math. Statist.*, **22** (1951), 79–86. <https://doi.org/10.1214/aoms/1177729694>

11. I. N. Sanov, On the probability of large deviations of random variables, *Mat. Sb.*, **42** (1957), 11–44.
12. M. D. Donsker, S. R. S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time. I, *Comm. Pure Appl. Math.*, **28** (1975), 1–47. <https://doi.org/10.1002/cpa.3160280102>
13. P. Dupuis, R. S. Ellis, *A weak convergence approach to the theory of large deviations*, New York: Wiley, 1997. <https://doi.org/10.1002/9781118165904>
14. A. Dembo, O. Zeitouni, *Large deviations techniques and applications*, New York: Springer, 1998. <https://doi.org/10.1007/978-3-642-03311-7>
15. O. E. Barndorff-Nielsen, *Information and exponential families in statistical theory*, Chichester: Wiley, 1978. <https://doi.org/10.1002/9781118857281>
16. L. D. Brown, *Fundamentals of statistical exponential families with applications in statistical decision theory*, Hayward: Institute of Mathematical Statistics, 1986.
17. J. M. Borwein, A. S. Lewis, Duality relationships for entropy-like minimization problems, *SIAM J. Control Optim.*, **29** (1991), 325–338. <https://doi.org/10.1137/0329017>
18. R. T. Rockafellar, *Convex analysis*, Princeton: Princeton University Press, 1970. <https://doi.org/10.1515/9781400873173>
19. S. Boyd, L. Vandenberghe, *Convex optimization*, Cambridge: Cambridge University Press, 2004.
20. M. J. Wainwright, M. I. Jordan, Graphical models, exponential families, and variational inference, *Found. Trends Mach. Learn.*, **1** (2008), 1–305. <https://doi.org/10.1561/22000000001>
21. E. K. Haviland, On the momentum problem for distributions in more than one dimension, *Amer. J. Math.*, **57** (1935), 562–568. <https://doi.org/10.2307/2371187>
22. J. A. Shohat, J. D. Tamarkin, *The problem of moments*, New York: American Mathematical Society, 1943. <https://doi.org/10.1090/SURV/001>
23. M. G. Krein, A. A. Nudelman, *The Markov moment problem and extremal problems*, Providence: American Mathematical Society, 1977. <https://doi.org/10.1090/mmono/050>
24. R. E. Curto, L. A. Fialkow, Solution of the truncated complex moment problem for flat data, *Mem. Amer. Math. Soc.*, **119** (1996), 568. <https://doi.org/10.1090/memo/0568>
25. J. B. Lasserre, *Moments, positive polynomials and their applications*, London: Imperial College Press, 2009. <https://doi.org/10.1142/p665>
26. A. Berman, N. Shaked-Monderer, *Completely positive matrices*, Singapore: World Scientific, 2003. <https://doi.org/10.1142/5273>
27. I. M. Bomze, Copositive optimization—recent developments and applications, *European J. Oper. Res.*, **216** (2012), 509–520. <https://doi.org/10.1016/j.ejor.2011.04.026>
28. P. J. C. Dickinson, L. Gijben, On the computational complexity of membership problems for the completely positive cone and its dual, *Comput. Optim. Appl.*, **57** (2014), 403–415. <https://doi.org/10.1007/s10589-013-9594-z>
29. K. G. Murty, S. N. Kabadi, Some NP-complete problems in quadratic and nonlinear programming, *Math. Program.*, **39** (1987), 117–129. <https://doi.org/10.1007/BF02592948>

30. O. Olteanu, Applications of the Hahn-Banach theorem, a solution of the moment problem and the related approximation, *Mathematics*, **12** (2024), 2878. <https://doi.org/10.3390/math12182878>
31. M. F. Iqbal, F. Ahmed, Approximation hierarchies for the copositive tensor cone and their application to the polynomial optimization over the simplex, *Mathematics*, **10** (2022), 1683. <https://doi.org/10.3390/math10101683>
32. B. Sturmfels, M. L. Telek, Copositive geometry of Feynman integrals, *Lett. Math. Phys.*, **115** (2025), 74. <https://doi.org/10.1007/s11005-025-01961-w>
33. B. C. Arnold, D. Strauss, Bivariate distributions with exponential conditionals, *J. Amer. Statist. Assoc.*, **83** (1988), 522–527. <https://doi.org/10.1080/01621459.1988.10478627>
34. E. J. Gumbel, Bivariate exponential distributions, *J. Amer. Statist. Assoc.*, **55** (1960), 698–707. <https://doi.org/10.1080/01621459.1960.10483368>
35. B. C. Arnold, D. J. Strauss, Bivariate distributions with conditionals in prescribed exponential families, *J. Roy. Statist. Soc. Ser. B*, **53** (1991), 365–375. <https://doi.org/10.1111/j.2517-6161.1991.tb01829.x>
36. B. C. Arnold, E. Castillo, J. M. Sarabia, *Conditional specification of statistical models*, New York: Springer, 1999. <https://doi.org/10.1007/b97592>
37. Q. Sun, Y. Zhou, S. Huang, Fast rates of exponential cost function, *Stat. Papers*, **66** (2025), 51. <https://doi.org/10.1007/s00362-025-01672-3>
38. S. Jiang, H. Xu, Q. Sun, S. Huang, Robust learning of minimum error entropy under heavy-tailed noise, *J. Comput. Appl. Math.*, **487** (2026), 117686. <https://doi.org/10.1016/j.cam.2026.117686>
39. D. Henrion, Maximal entropy in the moment body, *arXiv Preprint*, 2025. <https://doi.org/10.48550/arXiv.2507.02461>
40. C. Liu, G. Ma, Maximum entropy principle for uncertain sets with expectation and variance constraints, *AIMS Math.*, **11** (2026), 9303–9318. <https://doi.org/10.3934/math.2026384>
41. G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities*, Cambridge: Cambridge University Press, 1934.
42. J. E. Contreras-Reyes, Jensen-autocorrelation function for weakly stationary processes and applications, *Phys. D*, **470** (2024), 134424. <https://doi.org/10.1016/j.physd.2024.134424>
43. H. Z. An, F. C. Huang, The geometrical ergodicity of nonlinear autoregressive models, *Statist. Sinica*, **6** (1996), 943–956.
44. J. G. De Gooijer, On threshold moving-average models, *J. Time Ser. Anal.*, **19** (1998), 1–18. <https://doi.org/10.1111/1467-9892.00074>
45. J. Besag, Spatial interaction and the statistical analysis of lattice systems, *J. Roy. Statist. Soc. Ser. B*, **36** (1974), 192–236. <https://doi.org/10.1111/j.2517-6161.1974.tb00999.x>
46. E. Yang, P. Ravikumar, G. I. Allen, Z. Liu, Graphical models via univariate exponential family distributions, *J. Mach. Learn. Res.*, **16** (2015), 3813–3847.
47. D. I. Inouye, P. Ravikumar, I. S. Dhillon, Square root graphical models: multivariate generalizations of univariate exponential families that permit positive dependencies, *ICML 2016*, **48** (2016), 2445–2453.

Supplementary

S1. Numerical reproducibility specification

The numerical material supporting Section 5.1 is incorporated here directly, rather than supplied as separate files. It is a deterministic consistency check and is not used in any proof. The computation contains no random sampling. The reported results were produced with Python 3.12.3, NumPy 2.4.4, and Matplotlib 3.10.8.

For each scenario (μ_1, μ_2, Δ) and each box size K , the calculation solves

$$\mathbb{E}[X] = \mu_1, \quad \mathbb{E}[Y] = \mu_2, \quad \mathbb{E}[XY] = C = \mu_1\mu_2 + \Delta$$

for the truncated Gibbs density

$$f_K(x, y) \propto \exp(-\beta_1 x - \beta_2 y + \gamma xy), \quad (x, y) \in [0, K]^2.$$

The six scenarios are $(1, 2, 1/2)$, $(1, 1, 1/4)$, $(1/2, 2, 1/4)$, $(2, 3, 1)$, $(1, 2, 1/10)$, and $(1, 2, 1)$, with $K \in \{20, 40, 80, 120, 160, 320\}$.

Quadrature. Composite Gauss-Legendre quadrature uses three adaptively placed panels per coordinate. With

$$a = \min\{10, K/3\}, \quad b = \max\{a, K - \min(8, K/3)\},$$

the panels are $[0, a]$, $[a, b]$, and $[b, K]$, with 160 nodes per panel (480 nodes per coordinate). For $K \geq 30$ these are $[0, 10]$, $[10, K - 8]$, and $[K - 8, K]$; at $K = 20$ they are three equal panels.

Coupling search. The lower end of the coupling bracket is $\gamma = 0$. The upper end starts at $(\mu_1^{-1} + \mu_2^{-1})/K$ and is enlarged by a factor of 1.5, at most 40 times, until mean matching succeeds and $\mathbb{E}[XY] > C$. Failure to bracket within 40 enlargements is treated as an error. The verified bracket is bisected 60 times.

Mean matching. For fixed γ , Newton iteration matches (β_1, β_2) using the covariance matrix of (X, Y) as Jacobian, since the derivative of $(\mathbb{E}[X], \mathbb{E}[Y])$ with respect to (β_1, β_2) is $-\text{Cov}(X, Y)$. Backtracking halves the step length from 1 through at most 50 trial lengths until the maximum absolute mean residual strictly decreases. At most 80 Newton iterations are taken, with early exit below 10^{-12} ; a trial is accepted below 10^{-10} . The linear multipliers are unrestricted in sign.

Warm starts and verification. Each value of K is solved independently. Within a fixed K , successive bisection trials warm-start from the previously matched multipliers. After the final bisection, all three moments are recomputed. The largest residual over the full 36-point grid is 7.65×10^{-13} , below the required 10^{-8} tolerance.

S2. Complete numerical grid

The following three tables together reproduce all 19 columns of the numerical grid used for Section 5.1. The scenario column denotes (μ_1, μ_2, Δ) .

Table S1. Matched parameters and maximum moment residual for the complete numerical grid.

Scenario	C	K	β_1	β_2	γ	max. residual
$(1, 2, \frac{1}{2})$	2.500	20	1.153210	0.574054	0.06164954	3.28×10^{-13}
$(1, 2, \frac{1}{2})$	2.500	40	1.084508	0.539875	0.03489023	7.23×10^{-13}
$(1, 2, \frac{1}{2})$	2.500	80	1.042784	0.519962	0.01797865	6.45×10^{-13}
$(1, 2, \frac{1}{2})$	2.500	120	1.028654	0.513339	0.01211340	3.20×10^{-14}
$(1, 2, \frac{1}{2})$	2.500	160	1.021550	0.510023	0.00913864	6.82×10^{-13}
$(1, 2, \frac{1}{2})$	2.500	320	1.010833	0.505033	0.00461630	5.51×10^{-14}
$(1, 1, \frac{1}{4})$	1.250	20	1.105790	1.105790	0.08827457	1.28×10^{-13}
$(1, 1, \frac{1}{4})$	1.250	40	1.054305	1.054305	0.04683191	7.65×10^{-13}
$(1, 1, \frac{1}{4})$	1.250	80	1.027482	1.027482	0.02404923	4.91×10^{-13}
$(1, 1, \frac{1}{4})$	1.250	120	1.018418	1.018418	0.01619652	6.56×10^{-13}
$(1, 1, \frac{1}{4})$	1.250	160	1.013856	1.013856	0.01221564	9.39×10^{-14}
$(1, 1, \frac{1}{4})$	1.250	320	1.006967	1.006967	0.00616641	6.65×10^{-13}
$(\frac{1}{2}, 2, \frac{1}{4})$	1.250	20	2.290370	0.568740	0.11640827	1.02×10^{-13}
$(\frac{1}{2}, 2, \frac{1}{4})$	1.250	40	2.149791	0.533631	0.06101516	8.22×10^{-15}
$(\frac{1}{2}, 2, \frac{1}{4})$	1.250	80	2.074686	0.516490	0.03073466	5.23×10^{-13}
$(\frac{1}{2}, 2, \frac{1}{4})$	1.250	120	2.049771	0.510947	0.02055739	3.74×10^{-13}
$(\frac{1}{2}, 2, \frac{1}{4})$	1.250	160	2.037333	0.508197	0.01545001	3.76×10^{-13}
$(\frac{1}{2}, 2, \frac{1}{4})$	1.250	320	2.018684	0.504094	0.00775630	3.20×10^{-13}
$(2, 3, 1)$	7.000	20	0.573255	0.378412	0.02111528	2.21×10^{-13}
$(2, 3, 1)$	7.000	40	0.557088	0.371098	0.01637931	4.60×10^{-13}
$(2, 3, 1)$	7.000	80	0.532503	0.354533	0.00954774	4.62×10^{-13}
$(2, 3, 1)$	7.000	120	0.522030	0.347646	0.00653501	4.83×10^{-13}
$(2, 3, 1)$	7.000	160	0.516650	0.344133	0.00496364	9.99×10^{-14}
$(2, 3, 1)$	7.000	320	0.508429	0.338790	0.00253237	1.85×10^{-13}
$(1, 2, \frac{1}{10})$	2.100	20	1.047392	0.523446	0.02256748	3.62×10^{-13}
$(1, 2, \frac{1}{10})$	2.100	40	1.047073	0.523536	0.02241562	1.85×10^{-13}
$(1, 2, \frac{1}{10})$	2.100	80	1.036446	0.518141	0.01740924	7.38×10^{-13}
$(1, 2, \frac{1}{10})$	2.100	120	1.024851	0.512317	0.01191577	2.77×10^{-13}
$(1, 2, \frac{1}{10})$	2.100	160	1.018808	0.509303	0.00903710	5.17×10^{-13}
$(1, 2, \frac{1}{10})$	2.100	320	1.009529	0.504700	0.00459364	4.10×10^{-13}
$(1, 2, 1)$	3.000	20	1.204872	0.592230	0.07065185	2.93×10^{-13}
$(1, 2, 1)$	3.000	40	1.100644	0.544197	0.03605433	3.64×10^{-13}
$(1, 2, 1)$	3.000	80	1.049782	0.521738	0.01821641	2.48×10^{-13}
$(1, 2, 1)$	3.000	120	1.033137	0.514467	0.01221369	1.87×10^{-14}
$(1, 2, 1)$	3.000	160	1.024849	0.510851	0.00919370	3.88×10^{-13}
$(1, 2, 1)$	3.000	320	1.012437	0.505434	0.00462959	3.90×10^{-13}

Table S2. Achieved moments and entropy quantities for the complete numerical grid.

Scenario	K	$\mathbb{E}[X]$	$\mathbb{E}[Y]$	$\mathbb{E}[XY]$	$h(f_K)$	$\log Z_K$
$(1, 2, \frac{1}{2})$	20	1.000000000	2.000000000	2.500000000	2.672699011	0.525504368
$(1, 2, \frac{1}{2})$	40	1.000000000	2.000000000	2.500000000	2.678763017	0.601730069
$(1, 2, \frac{1}{2})$	80	1.000000000	2.000000000	2.500000000	2.684930572	0.647169338
$(1, 2, \frac{1}{2})$	120	1.000000000	2.000000000	2.500000000	2.687436396	0.662387444
$(1, 2, \frac{1}{2})$	160	1.000000000	2.000000000	2.500000000	2.688773403	0.670023723
$(1, 2, \frac{1}{2})$	320	1.000000000	2.000000000	2.500000000	2.690888420	0.681531085
$(1, 1, \frac{1}{4})$	20	1.000000000	1.000000000	1.250000000	1.983089261	-0.118147577
$(1, 1, \frac{1}{4})$	40	1.000000000	1.000000000	1.250000000	1.989637906	-0.060431299
$(1, 1, \frac{1}{4})$	80	1.000000000	1.000000000	1.250000000	1.994332285	-0.030570046
$(1, 1, \frac{1}{4})$	120	1.000000000	1.000000000	1.250000000	1.996105685	-0.020484051
$(1, 1, \frac{1}{4})$	160	1.000000000	1.000000000	1.250000000	1.997033732	-0.015408931
$(1, 1, \frac{1}{4})$	320	1.000000000	1.000000000	1.250000000	1.998480565	-0.007746178
$(\frac{1}{2}, 2, \frac{1}{4})$	20	0.500000000	2.000000000	1.250000000	1.979949487	-0.157204871
$(\frac{1}{2}, 2, \frac{1}{4})$	40	0.500000000	2.000000000	1.250000000	1.986990251	-0.078898485
$(\frac{1}{2}, 2, \frac{1}{4})$	80	0.500000000	2.000000000	1.250000000	1.992863571	-0.039041102
$(\frac{1}{2}, 2, \frac{1}{4})$	120	0.500000000	2.000000000	1.250000000	1.995102843	-0.025979336
$(\frac{1}{2}, 2, \frac{1}{4})$	160	0.500000000	2.000000000	1.250000000	1.996273589	-0.019475123
$(\frac{1}{2}, 2, \frac{1}{4})$	320	0.500000000	2.000000000	1.250000000	1.998094963	-0.009738793
$(2, 3, 1)$	20	2.000000000	3.000000000	7.000000000	3.778589620	1.644650095
$(2, 3, 1)$	40	2.000000000	3.000000000	7.000000000	3.781451910	1.668639389
$(2, 3, 1)$	80	2.000000000	3.000000000	7.000000000	3.784136591	1.722365694
$(2, 3, 1)$	120	2.000000000	3.000000000	7.000000000	3.786104473	1.744852703
$(2, 3, 1)$	160	2.000000000	3.000000000	7.000000000	3.787297811	1.756343534
$(2, 3, 1)$	320	2.000000000	3.000000000	7.000000000	3.789355626	1.773852638
$(1, 2, \frac{1}{10})$	20	1.000000000	2.000000000	2.100000000	2.691930774	0.645039214
$(1, 2, \frac{1}{10})$	40	1.000000000	2.000000000	2.100000000	2.691982460	0.644909767
$(1, 2, \frac{1}{10})$	80	1.000000000	2.000000000	2.100000000	2.692048858	0.655879999
$(1, 2, \frac{1}{10})$	120	1.000000000	2.000000000	2.100000000	2.692253196	0.667792103
$(1, 2, \frac{1}{10})$	160	1.000000000	2.000000000	2.100000000	2.692413649	0.673977652
$(1, 2, \frac{1}{10})$	320	1.000000000	2.000000000	2.100000000	2.692731404	0.683450243
$(1, 2, 1)$	20	1.000000000	2.000000000	3.000000000	2.639266852	0.461891019
$(1, 2, 1)$	40	1.000000000	2.000000000	3.000000000	2.661000466	0.580125054
$(1, 2, 1)$	80	1.000000000	2.000000000	3.000000000	2.675877568	0.637269732
$(1, 2, 1)$	120	1.000000000	2.000000000	3.000000000	2.681352943	0.655923055
$(1, 2, 1)$	160	1.000000000	2.000000000	3.000000000	2.684189422	0.665220448
$(1, 2, 1)$	320	1.000000000	2.000000000	3.000000000	2.688576740	0.679160112

Table S3. Entropy-deficit and multiplier diagnostics for the complete numerical grid.

Scenario	K	deficit	K deficit	$K\gamma$	R_E	R_γ
$(1, 2, \frac{1}{2})$	20	0.020448170	0.408963	1.232991	0.545285	0.821994
$(1, 2, \frac{1}{2})$	40	0.014384163	0.575367	1.395609	0.767155	0.930406
$(1, 2, \frac{1}{2})$	80	0.008216609	0.657329	1.438292	0.876438	0.958861
$(1, 2, \frac{1}{2})$	120	0.005710784	0.685294	1.453608	0.913726	0.969072
$(1, 2, \frac{1}{2})$	160	0.004373777	0.699804	1.462182	0.933073	0.974788
$(1, 2, \frac{1}{2})$	320	0.002258761	0.722803	1.477215	0.963738	0.984810
$(1, 1, \frac{1}{4})$	20	0.016910739	0.338215	1.765491	0.676430	0.882746
$(1, 1, \frac{1}{4})$	40	0.010362094	0.414484	1.873276	0.828968	0.936638
$(1, 1, \frac{1}{4})$	80	0.005667715	0.453417	1.923939	0.906834	0.961969
$(1, 1, \frac{1}{4})$	120	0.003894315	0.467318	1.943583	0.934636	0.971791
$(1, 1, \frac{1}{4})$	160	0.002966268	0.474603	1.954503	0.949206	0.977251
$(1, 1, \frac{1}{4})$	320	0.001519435	0.486219	1.973252	0.972439	0.986626
$(\frac{1}{2}, 2, \frac{1}{4})$	20	0.020050513	0.401010	2.328165	0.641616	0.931266
$(\frac{1}{2}, 2, \frac{1}{4})$	40	0.013009749	0.520390	2.440606	0.832624	0.976243
$(\frac{1}{2}, 2, \frac{1}{4})$	80	0.007136429	0.570914	2.458773	0.913463	0.983509
$(\frac{1}{2}, 2, \frac{1}{4})$	120	0.004897157	0.587659	2.466887	0.940254	0.986755
$(\frac{1}{2}, 2, \frac{1}{4})$	160	0.003726411	0.596226	2.472001	0.953961	0.988800
$(\frac{1}{2}, 2, \frac{1}{4})$	320	0.001905037	0.609612	2.482015	0.975379	0.992806
$(2, 3, 1)$	20	0.013169850	0.263397	0.422306	0.316076	0.506767
$(2, 3, 1)$	40	0.010307560	0.412302	0.655172	0.494763	0.786207
$(2, 3, 1)$	80	0.007622878	0.609830	0.763819	0.731796	0.916583
$(2, 3, 1)$	120	0.005654997	0.678600	0.784201	0.814320	0.941041
$(2, 3, 1)$	160	0.004461658	0.713865	0.794183	0.856638	0.953019
$(2, 3, 1)$	320	0.002403844	0.769230	0.810359	0.923076	0.972430
$(1, 2, \frac{1}{10})$	20	0.001216407	0.024328	0.451350	0.162188	0.300900
$(1, 2, \frac{1}{10})$	40	0.001164721	0.046589	0.896625	0.310592	0.597750
$(1, 2, \frac{1}{10})$	80	0.001098322	0.087866	1.392739	0.585772	0.928493
$(1, 2, \frac{1}{10})$	120	0.000893985	0.107278	1.429892	0.715188	0.953261
$(1, 2, \frac{1}{10})$	160	0.000733532	0.117365	1.445936	0.782434	0.963957
$(1, 2, \frac{1}{10})$	320	0.000415776	0.133048	1.469963	0.886990	0.979975
$(1, 2, 1)$	20	0.053880328	1.077607	1.413037	0.718404	0.942025
$(1, 2, 1)$	40	0.032146714	1.285869	1.442173	0.857246	0.961449
$(1, 2, 1)$	80	0.017269613	1.381569	1.457312	0.921046	0.971542
$(1, 2, 1)$	120	0.011794237	1.415308	1.465642	0.943539	0.977095
$(1, 2, 1)$	160	0.008957759	1.433241	1.470991	0.955494	0.980661
$(1, 2, 1)$	320	0.004570441	1.462541	1.481470	0.975027	0.987646

S3. Quadrature-sensitivity check

Three representative extreme cases were re-solved at 120 and 220 nodes per panel, compared with the reported 160-node setting. The largest absolute change in either normalized diagnostic was 4.02×10^{-10} , more than five orders of magnitude below the 10^{-4} reporting resolution of Table 3. This is numerical validation only; it is not a proof of quadrature convergence and provides no analytic error bound.

Table S4 reports the normalized diagnostics at 120, 160, and 220 nodes per panel and the changes relative to the 160-node calculation.

Table S4. Quadrature-sensitivity check. Differences are relative to the 160-node-per-panel calculation.

Scenario, K	nodes	$R_{\mathcal{E}}$	$\Delta R_{\mathcal{E}}$	R_{γ}	ΔR_{γ}
(1, 2, 0.5), 320	160	0.963738	–	0.984810	–
	120	0.963738	-4.02×10^{-10}	0.984810	-5.32×10^{-13}
	220	0.963738	3.52×10^{-11}	0.984810	4.74×10^{-14}
(1, 2, 0.1), 160	160	0.782434	–	0.963957	–
	120	0.782434	-1.56×10^{-11}	0.963957	-7.99×10^{-15}
	220	0.782434	1.22×10^{-10}	0.963957	7.82×10^{-14}
(2, 3, 1), 320	160	0.923076	–	0.972430	–
	120	0.923076	-4.26×10^{-12}	0.972430	-8.88×10^{-16}
	220	0.923076	4.42×10^{-11}	0.972430	-1.62×10^{-14}

S4. Region diagnostics for the baseline case

For $(\mu_1, \mu_2, \Delta) = (1, 2, 1/2)$ at $K = 20$, the upper-right region carries a disproportionate share of the cross-moment despite its small probability mass, consistent with the finite-box trace of the escape mechanism discussed in Section 5.1.

Table S5 reports, for the baseline case, the probability mass and share of $\mathbb{E}[XY]$ carried by the upper-right region and the corner square.

Table S5. Region diagnostics for the baseline scenario at $K = 20$.

Region	Probability mass	Share of $\mathbb{E}[XY]$
$\{x > 5, y > 5\}$	0.00363979 (0.36%)	0.14031 (14%)
$[K - 1, K]^2$	2.08×10^{-5}	0.00318024 (0.32%)

S5. Dependency map for the main results

This table records how the main statements depend on earlier material. It is included to make the proof structure easier to audit.

Table S6. Logical dependencies of the main results.

Result	Main inputs	Role in the paper
Theorem 3.1 (c)	Product-exponential KL bound, escape mixture	Non-attainment and entropy invisibility
Theorem 3.1 (a)	Lemma 3.6, Gibbs inequality	Attainment in the repulsive case
Theorem 4.1	Sharp Young inequality and the envelope estimate of Lemma 4.2	Minimal damping threshold
Proposition 4.3	Lemma 4.2 and the residual expansion on the manifold	Complete classification of the critical equality manifold
Proposition 4.9	Convexity of the natural domain and covariance non-singularity	Attainment for damped targets on the realized region M°
Theorem 5.1	Compact exponential-family theory and moment-set interiority	Well-posedness on $[0, K]^2$
Theorem 5.3	I-projection identity and Donsker–Varadhan bound	Sharp entropy cost of confinement
Theorem 7.3	Copositivity, zero-cone KKT conditions, radial foliation, disjoint shrinking cells	n -dimensional normalizability away from the codimension- ≥ 2 boundary
Corollary 7.9	Theorem 7.3 and NP-completeness of non-copositivity	Computational hardness
Theorem 7.10	Product-exponential KL bound and multi-blob escape sequence	n -dimensional entropy invisibility



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)