Supplemental Material:
Unification of Field Theory and Maximum Entropy Methods for Learning Probability Densities

Justin B. Kinney*
Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

This document contains supplemental material for the preprint titled "Unification of Field Theory and Maximum Entropy Methods for Learning Probability Densities." Derivations of Eqs. (4), (10), and (12) of the main text are provided. A description is also given of the predictor-corrector homotopy algorithm used to produce Figs. 1 and 2.

CONTENTS

Derivation of the action $S_\ell[\phi]$ (Eq. 4) 1

Derivation of the evidence ratio E (Eq. 10) 3

Derivation of the K coefficient (Eq. 12) 4
- Step 1: Expansion of ϕ_ℓ to first order in η. 5
- Step 2: Expansion of $S_\ell[\phi_\ell]$ to first order in η. 6
- Step 3: Expansion of $\ln\det[L^{2\alpha}\Delta^\alpha + \eta e^{-\phi_\ell}]$ to first order in η. 7
- Putting it all together 9

Homotopy algorithm 9
- Computing the MaxEnt density 10
- Predictor step 10
- Corrector step 11

References 11

DERIVATION OF THE ACTION $S_\ell[\phi]$ (EQ. 4)

Our derivation of the action $S_\ell[\phi]$ in Eq. 4 of the main text is essentially that used in [1]. This derivation is not entirely straight-forward, however, and the details of it have yet to be published.

Our adopted prior is

$$p(\phi|\ell) = \frac{1}{Z_\ell^0} e^{-S_\ell^0}[\phi], \qquad S_\ell^0 = \int \frac{dx}{L}\frac{\ell^{2\alpha}}{2}\phi\Delta^\alpha\phi, \qquad Z_\ell^0 = \int \mathcal{D}\phi\, e^{-S_\ell^0[\phi]}. \tag{1}$$

As written, this prior is improper due to the α-dimensional kernel of Δ^α. To avoid unnecessary mathematical difficulties, we can render $p(\phi|\ell)$ proper by considering a regularized form of the action

$$S_\ell^0 = \int \frac{dx}{L}\frac{1}{2}\phi\left[\ell^{2\alpha}\Delta^\alpha + \epsilon\right]\phi. \tag{2}$$

where ϵ is an infinitesimal positive number. All quantities of interest, of course, should be evaluated in the $\epsilon \to 0$ limit.

The prior over Q that results from Eq. (1) is

$$
\begin{aligned}
p(Q|\ell) &= \int_{-\infty}^{\infty} d\phi_c \, p(\phi|\ell) && (3)\\
&= \frac{1}{Z_\ell^0} \int_{-\infty}^{\infty} d\phi_c \, \exp\left(-\frac{1}{2}\int \frac{dx}{L}\phi\left[\ell^{2\alpha}\Delta^\alpha + \epsilon\right]\phi\right) && (4)\\
&= \frac{1}{Z_\ell^0}\left\{\int_{-\infty}^{\infty} d\phi_c \, \exp\left(-\frac{\epsilon}{2}\phi_c^2\right)\right\}\exp\left(-\frac{1}{2}\int \frac{dx}{L}\phi_{nc}\left[\ell^{2\alpha}\Delta^\alpha + \epsilon\right]\phi_{nc}\right) && (5)\\
&= \frac{1}{Z_\ell^0}\sqrt{\frac{2\pi}{\epsilon}}\exp\left(-\frac{1}{2}\int \frac{dx}{L}\phi_{nc}\left[\ell^{2\alpha}\Delta^\alpha + \epsilon\right]\phi_{nc}\right), && (6)
\end{aligned}
$$

where ϕ_c is the constant Fourier component of ϕ and $\phi_{nc}(x)$ is the non-constant component.

The likelihood of Q given the data,

$$
p(\text{data}|Q) = \prod_i Q(x_i), \tag{7}
$$

can be expressed in a somewhat non-obvious but nevertheless useful form. Using the identity

$$
\frac{1}{a^N} = \frac{N^N}{\Gamma(N)}\int_{-\infty}^{\infty} du \, e^{-N\left(u + a e^{-u}\right)}, \tag{8}
$$

which holds for any positive number a and positive integer N, we find that

$$
\begin{aligned}
p(\text{data}|Q) &= \exp\left(-\sum_i \phi_{nc}(x_i)\right)\frac{1}{L^N}\left[\int \frac{dx}{L}\, e^{-\phi_{nc}(x)}\right]^{-N} && (9)\\
&= \exp\left(-\int dx\, NR\phi_{nc}\right)\frac{N^N}{L^N\Gamma(N)}\int_{-\infty}^{\infty} d\phi_c \exp\left(-N\left[\phi_c + \int \frac{dx}{L}\, e^{-\phi_{nc}(x)-\phi_c}\right]\right) && (10)\\
&= \frac{N^N}{L^N\Gamma(N)}\int_{-\infty}^{\infty} d\phi_c \exp\left(-\int \frac{dx}{L}\left\{NLR\phi + Ne^{-\phi}\right\}\right). && (11)
\end{aligned}
$$

The prior probability of Q and the data together is therefore given by

$$
\begin{aligned}
p(\text{data}, Q|\ell) &= p(\text{data}|Q)\, p(Q|\ell) && (12)\\
&= \frac{N^N}{L^N\Gamma(N)}\sqrt{\frac{2\pi}{\epsilon}}\int_{-\infty}^{\infty} d\phi_c \exp\left(-\int \frac{dx}{L}\left\{NLR\phi - Ne^{-\phi}\right\}\right)\frac{1}{Z_\ell^0}\exp\left(-\int \frac{dx}{L}\frac{\ell^{2\alpha}}{2}(\partial^\alpha\phi_{nc})^2\right) && (13)\\
&= \frac{N^N}{L^N\Gamma(N)}\sqrt{\frac{2\pi}{\epsilon}}\frac{1}{Z_\ell^0}\int_{-\infty}^{\infty} d\phi_c \, \exp\left(-\int \frac{dx}{L}\left\{\frac{\ell^{2\alpha}}{2}(\partial^\alpha\phi)^2 + NLR\phi + Ne^{-\phi}\right\}\right) && (14)\\
&= \frac{N^N}{L^N\Gamma(N)}\sqrt{\frac{2\pi}{\epsilon}}\frac{1}{Z_\ell^0}\int_{-\infty}^{\infty} d\phi_c \, e^{-S_\ell[\phi]}, && (15)
\end{aligned}
$$

where

$$
\boxed{S_\ell[\phi] = \int \frac{dx}{L}\left\{\frac{\ell^{2\alpha}}{2}(\partial^\alpha\phi)^2 + NLR\phi + Ne^{-\phi}\right\}.} \tag{16}
$$

is the action in Eq. 4 of the main text.

The "evidence" for ℓ – i.e., the probability of the data given ℓ – is therefore given by,

$$
\begin{aligned}
p(\text{data}|\ell) &= \int \mathcal{D}Q\, p(\text{data}, Q|\ell) && (17)\\
&= \frac{N^N}{L^N\Gamma(N)}\sqrt{\frac{2\pi}{\epsilon}}\frac{1}{Z_\ell^0}\int \mathcal{D}\phi\, e^{-S_\ell[\phi]} && (18)\\
&= \frac{N^N}{L^N\Gamma(N)}\sqrt{\frac{2\pi}{\epsilon}}\frac{Z_\ell}{Z_\ell^0}, && (19)
\end{aligned}
$$

where

$$Z_\ell = \int \mathcal{D}\phi \, e^{-S_\ell[\phi]}. \tag{20}$$

is the partition function corresponding to the action S_ℓ. The posterior distribution over Q is then given by Bayes's rule:

$$p(Q|\text{data}, \ell) = \frac{p(\text{data}, Q|\ell)}{p(\text{data}|\ell)} = \int_{-\infty}^{\infty} d\phi_c \, \frac{e^{-S_\ell[\phi]}}{Z_\ell}. \tag{21}$$

This motivates us to *define* the posterior distribution over ϕ as

$$p(\phi|\text{data}, \ell) \equiv \frac{e^{-S_\ell[\phi]}}{Z_\ell}. \tag{22}$$

This definition of $p(\phi|\text{data}, \ell)$ is consistent with the formula for $p(Q|\text{data}, \ell)$ obtained in Eq. (21), in that

$$p(Q|\text{data}, \ell) = \int_{-\infty}^{\infty} d\phi_c \, p(\phi|\text{data}, \ell). \tag{23}$$

However, Eq. (22) appears to violate Bayes's rule, since

$$p(\phi|\text{data}, \ell) \neq \frac{p(\text{data}, \phi|\ell)}{p(\text{data}|\ell)}. \tag{24}$$

This is not a problem, however, since ϕ itself is not directly observable; only Q is observable, and so Bayes's rule only requires Eq. (21). Put another way, Eq. (22) violates Bayes's rule only in the way that it specifies the behavior of ϕ_c. This constant component of ϕ, however, has no effect on Q.

Note that all of the above calculations have been exact; no large N approximation was used. This contrasts with prior work [2, 3], which used a large N Laplace approximation to derive the formula for $S_\ell[\phi]$. Also note that the regularization parameter ϵ has vanished in the formulas for the posterior distributions $p(Q|\text{data}, \ell)$ and $p(\phi|\text{data}, \ell)$. However, this parameter still appears in the formula for the evidence $p(\text{data}|\ell)$, both explicitly and implicitly through the value of Z_ℓ^0.

DERIVATION OF THE EVIDENCE RATIO E (EQ. 10)

In Eq. (19), we found that the evidence for a length scale ℓ is given by

$$p(\text{data}|\ell) = \frac{N^N}{L^N \Gamma(N)} \sqrt{\frac{2\pi}{\epsilon}} \frac{Z_\ell}{Z_\ell^0}. \tag{25}$$

We now turn to the task of evaluating the partition functions Z_ℓ and Z_ℓ^0, so that we can compute this quantity. Defining $\Lambda = L^{2\alpha}\Delta$ and $\eta = N(L/\ell)^{2\alpha}$, and using a Laplace approximation to compute the path integral, we get,

$$\begin{aligned}
Z_\ell &= \int \mathcal{D}\phi \, e^{-S_\ell[\phi]} &(26)\\
&\approx e^{-S_\ell[\phi_\ell]} \int \mathcal{D}\delta\phi \, \exp\left(-\frac{1}{2} \int dx dx' \, \left.\frac{\delta^2 S_\ell}{\delta\phi(x)\delta\phi(x')}\right|_{\phi_\ell} \delta\phi(x)\delta\phi(x') \right) &(27)\\
&= e^{-S_\ell[\phi_\ell]} \int \mathcal{D}\delta\phi \, \exp\left(-\frac{1}{2} \int \frac{dx}{L} \, \delta\phi \left[\ell^{2\alpha}\Delta^\alpha + N e^{-\phi_\ell} \right] \delta\phi \right) &(28)\\
&= e^{-S_\ell[\phi_\ell]} \int \mathcal{D}\delta\phi \, \exp\left(-\frac{1}{2} \int \frac{dx}{L} \, \delta\phi \left(\frac{\ell^{2\alpha}}{L^{2\alpha}} \right) \left[\Lambda + \eta e^{-\phi_\ell} \right] \delta\phi \right) &(29)\\
&= e^{-S_\ell[\phi_\ell]} \int \mathcal{D}\delta\phi \, \exp\left(-\frac{1}{2} \sum_{i,j} \delta\phi_i \left(\frac{\ell^{2\alpha}}{G L^{2\alpha}} \right) \left[\Lambda + \eta e^{-\phi_\ell} \right]_{ij} \delta\phi_j \right) &(30)\\
&= e^{-S_\ell[\phi_\ell]} \left\{ \left(\frac{\ell^{2\alpha}}{2\pi G L^{2\alpha}} \right)^G \det\left[\Lambda + \eta e^{-\phi_\ell} \right] \right\}^{-1/2}. &(31)
\end{aligned}$$

In Eq. (30) we switched to the grid representation, in which $\delta\phi_i = \delta\phi(x_i), \Lambda_{ij} = L^{2\alpha}\left[(\partial_G^\alpha)^\intercal \partial_G^\alpha\right]_{ij}$, $[e^{-\phi_\ell}]_{ij} = \delta_{ij} e^{-\phi_\ell(x_i)}$, etc.. Note that the operator Λ is unitless and has eigenvalues that, if ranked in ascending order, approach limiting values as $G \to \infty$. Also note that η is unitless. For these reasons, η will emerge as a natural perturbation parameter in the $\ell \to \infty$ limit.

Evaluating the partition function Z_ℓ^0 requires more care because the regularized form of the action S_ℓ^0, given in Eq. (2), has to be used in order for the equations we derive to make sense. We get

$$Z_\ell^0 = \int \mathcal{D}\phi\, e^{-S_\ell^0[\phi]} \tag{32}$$

$$= \int \mathcal{D}\phi\, \exp\left(-\frac{1}{2}\int \frac{dx}{L}\phi\left[\ell^{2\alpha}\Delta^\alpha + \epsilon\right]\phi\right) \tag{33}$$

$$= \int \mathcal{D}\phi\, \exp\left(-\frac{1}{2}\int \frac{dx}{L}\phi\left(\frac{\ell^{2\alpha}}{L^{2\alpha}}\right)\left[\Lambda + \frac{\eta\epsilon}{N}\right]\phi\right) \tag{34}$$

$$= \int \mathcal{D}\phi\, \exp\left(-\frac{1}{2}\sum_{i,j}\phi_i\left(\frac{\ell^{2\alpha}}{GL^{2\alpha}}\right)\left[\Lambda + \frac{\eta\epsilon}{N}\right]_{ij}\phi_j\right) \tag{35}$$

$$= \left\{\left(\frac{\ell^{2\alpha}}{2\pi G L^{2\alpha}}\right)^G \det\left[\Lambda + \frac{\eta\epsilon}{N}\right]\right\}^{-1/2} \tag{36}$$

$$= \left\{\left(\frac{\ell^{2\alpha}}{2\pi G L^{2\alpha}}\right)^G N^{-\alpha}\eta^\alpha\epsilon^\alpha \det_{\text{row}}[\Lambda]\right\}^{-1/2}. \tag{37}$$

As in the main text, the subscript "row" on the determinant denotes restriction to the row space of Λ. Putting these values for Z_ℓ and Z_ℓ^0 back into Eq. (25), we get

$$p(\text{data}|\ell) = \frac{N^N}{L^N\Gamma(N)} e^{-S_\ell[\phi_\ell]}\sqrt{\frac{2\pi}{\epsilon}}\sqrt{\frac{N^{-\alpha}\epsilon^\alpha\eta^\alpha \det_{\text{row}}[\Lambda]}{\det[\Lambda + \eta e^{-\phi_\ell}]}} \tag{38}$$

$$= \epsilon^{\frac{\alpha-1}{2}}\frac{\sqrt{2\pi}N^{N-\frac{\alpha}{2}}}{L^N\Gamma(N)} e^{-S_\ell[\phi_\ell]}\sqrt{\frac{\det_{\text{row}}[\Lambda]}{\eta^{-\alpha}\det[\Lambda + \eta e^{-\phi_\ell}]}}. \tag{39}$$

Although both Z_ℓ and Z_ℓ^0 depend strongly on the number of grid points G, the value for the evidence does not. The evidence does, however, depend on the regularization parameter ϵ; specifically, it is proportional to $\epsilon^{\frac{\alpha-1}{2}}$. This is the basis for the claim in the main text that the evidence vanishes for $\alpha > 1$.

In the large ℓ limit, $\eta \to 0$, and so

$$\det\left[\Lambda + \eta e^{-\phi_\ell}\right] \to \eta^\alpha \det_{\text{ker}}\left[e^{-\phi_\infty}\right]\det_{\text{row}}[\Lambda] \tag{40}$$

where "ker" denotes restriction to the kernel of Λ. As a result,

$$p(\text{data}|\infty) = \epsilon^{\frac{\alpha-1}{2}}\frac{\sqrt{2\pi}N^{N-\frac{\alpha}{2}}}{L^N\Gamma(N)}\frac{e^{-S_\infty[\phi_\infty]}}{\sqrt{\det_{\text{ker}}[e^{-\phi_\infty}]}}. \tag{41}$$

The evidence ratio is therefore given by

$$\boxed{E = \frac{p(\text{data}|\ell)}{p(\text{data}|\infty)} = e^{S_\infty[\phi_\infty] - S_\ell[\phi_\ell]}\sqrt{\frac{\det_{\text{ker}}[e^{-\phi_\infty}]\det_{\text{row}}[\Lambda]}{\eta^{-\alpha}\det[\Lambda + \eta e^{-\phi_\ell}]}}.} \tag{42}$$

Note that E, unlike the evidence itself, does not depend on the regularization parameter ϵ.

DERIVATION OF THE K COEFFICIENT (EQ. 12)

The goal of this section is to clarify the large length scale behavior of

$$\ln E = S_\infty[\phi_\infty] - S_\ell[\phi_\ell] + \frac{1}{2}\ln\left\{\frac{\det_{\text{ker}}\left[e^{-\phi_\infty}\right]\det_{\text{row}}[\Lambda]}{\eta^{-\alpha}\det[\Lambda + \eta e^{-\phi_\ell}]}\right\}. \tag{43}$$

To do this we expand $\ln E$ as a power series in η about $\eta = 0$. We will find that

$$\ln E = K\eta + O(\eta^2), \tag{44}$$

where K is a nontrivial coefficient that can be either positive or negative, depending on the data. We carry out this expansion in three steps:

1. Expand ϕ_ℓ to first order in η.
2. Expand $S_\ell[\phi_\ell]$ to first order in η.
3. Expand $\ln\det[\Lambda + \eta e^{-\phi_\ell}]$ to first order in η.

In what follows we will sometimes use the ket notion of quantum mechanics, primarily as a notational convenience. For any two functions f and g and any operator H, we define

$$\langle f|H|g\rangle = \int \frac{dx}{L} f^* H g. \tag{45}$$

Note the convention of dividing by L in the inner product integral. This allows us to take inner products without altering units. The eigenvalues λ_i and corresponding eigenfunctions $\psi_i(x)$ defined in the main text satisfy

$$\langle \psi_i|\psi_j\rangle = \delta_{ij} \quad \text{for all } i, j. \tag{46}$$

and

$$\langle \psi_i|\Lambda|\psi_j\rangle = \delta_{ij}\zeta_j \quad \text{for } i, j \le \alpha \tag{47}$$

We will also make use of the following indexed quantities,

$$v_i = L\,\langle \psi_i|Q_\infty - R\rangle = \int dx\,(Q_\infty - R)\psi_i \tag{48}$$

$$z_{ij} = L\,\langle \psi_i|Q_\infty|\psi_j\rangle = \int dx\,Q_\infty \psi_i\psi_j \tag{49}$$

$$z_{ijk} = L\,\langle \psi_i|Q_\infty\psi_j|\psi_k\rangle = \int dx\,Q_\infty \psi_i\psi_j\psi_k. \tag{50}$$

Step 1: Expansion of ϕ_ℓ to first order in η.

We begin by representing ϕ_ℓ as a power series in η. Abusing notation somewhat, we write

$$\phi_\ell = \phi_\infty + \eta\phi_1 + \eta^2\phi_2 + \cdots. \tag{51}$$

Plugging this expansion into the equation of motion,

$$\Lambda\phi_\ell + \eta(LR - e^{-\phi_\ell}) = 0, \tag{52}$$

then collecting terms of equal order in η, we get,

$$\Lambda\phi_\infty + \eta\left(\phi_1 + LR - e^{-\phi_\infty}\right) + \eta^2\left(\Lambda\phi_2 + e^{-\phi_\infty}\phi_1\right) + \cdots = 0. \tag{53}$$

At lowest order in η, we recover

$$\Lambda\phi_\infty = 0. \tag{54}$$

This just reflects the restriction of ϕ_∞ to the kernel of Λ. At first order we get

$$\Lambda\phi_1 + LR - e^{-\phi_\infty} = 0, \tag{55}$$

which we will proceed to investigate. To compute ϕ_1, we first write it in terms of the eigenfunctions of Λ.

$$\phi_1(x) = \sum_i A_i\psi_i(x). \tag{56}$$

Taking the inner product of Eq. (55) with ψ_i, we get

$$\lambda_i A_i + L\,\langle \psi_i|R - Q_\infty\rangle = 0. \tag{57}$$

Since $\lambda_i > 0$ for $i > \alpha$,

$$A_i = \frac{v_i}{\lambda_i}, \quad i > \alpha. \tag{58}$$

As yet we have no information about the value of A_i for $i \le \alpha$. For this we need to consider the second order terms in Eq. (53). At second order in η, we obtain,

$$\Lambda\phi_2 + e^{-\phi_\infty}\phi_1 = 0. \tag{59}$$

Dotting this with ψ_j for $j \le \alpha$, we obtain

$$0 = \left\langle \psi_j | e^{-\phi_\infty} | \phi_1 \right\rangle = \sum_i A_i \left\langle \psi_i | e^{-\phi_\infty} | \psi_j \right\rangle = A_j \zeta_j + \sum_{i>\alpha} \frac{v_i z_{ij}}{\lambda_i}. \tag{60}$$

The A coefficients left unspecified in Eq. (58) are therefore given by,

$$A_j = -\sum_{i>\alpha} \frac{v_i z_{ij}}{\lambda_i \zeta_j}, \quad j \le \alpha. \tag{61}$$

This completes our computation of ϕ_ℓ to first order in η:

$$\boxed{\phi_\ell = \phi_\infty + \eta\phi_1 + O(\eta^2) \quad \text{where} \quad \phi_1(x) = \sum_{i>\alpha} \frac{v_i}{\lambda_i}\psi_i(x) - \sum_{\substack{i>\alpha \\ j\le\alpha}} \frac{v_i z_{ij}}{\lambda_i \zeta_j}\psi_j(x).} \tag{62}$$

Step 2: Expansion of $S_\ell[\phi_\ell]$ to first order in η.

The value of the action S_ℓ, evaluated at its minimum ϕ_ℓ, is

$$S_\ell[\phi_\ell] = \int \frac{dx}{L} \left\{ \frac{\ell^{2\alpha}}{2} \phi_\ell \Delta^\alpha \phi_\ell + NLR\phi_\ell + Ne^{-\phi_\ell} \right\} \tag{63}$$

$$= N \int \frac{dx}{L} \left\{ \frac{1}{2\eta} \phi_\ell \Lambda \phi_\ell + LR\phi_\ell - e^{-\phi_\ell} \right\} \tag{64}$$

$$= N \int \frac{dx}{L} \left\{ \frac{1}{2\eta} (\phi_\infty + \eta\phi_1) \Lambda (\phi_\infty + \eta\phi_1) + LR(\phi_\infty + \eta\phi_1) + LQ_\infty - \eta LQ_\infty \phi_1 \right\} + O(\eta^2) \tag{65}$$

$$= S_\infty[\phi_\infty] + N\eta \left\{ \frac{1}{2} \langle \phi_1 | \Lambda | \phi_1 \rangle - \langle \phi_1 | Q_\infty - R \rangle \right\} + O(\eta^2). \tag{66}$$

The first inner product term in Eq. (66) is

$$\frac{1}{2} \langle \phi_1 | \Lambda | \phi_1 \rangle = \frac{1}{2} \sum_{i,j} A_i A_j \lambda_j \langle \psi_i | \psi_j \rangle \tag{67}$$

$$= \frac{1}{2} \sum_{i,j>\alpha} \frac{v_i}{\lambda_i} \frac{v_j}{\lambda_j} \lambda_j \delta_{ij} \tag{68}$$

$$= \frac{1}{2} \sum_{i>\alpha} \frac{v_i^2}{\lambda_i}. \tag{69}$$

Here we have used the fact that $\lambda_k = 0$ for $k \le \alpha$ to restrict the sum to $i, j > \alpha$ in Eq. (68). The second inner product term in Eq. (66) is

$$-\langle \phi_1 | Q_\infty - R \rangle = -\sum_i A_i \langle \psi_i | Q_\infty - R \rangle \tag{70}$$

$$= -\sum_{i>\alpha} \frac{v_i}{\lambda_i} v_i \tag{71}$$

$$= -\sum_{i>\alpha} \frac{v_i^2}{\lambda_i}. \tag{72}$$

Here, the sum over i is restricted to $i > \alpha$ due to the moment-matching condition $\langle\psi_k|Q_\infty - R\rangle = 0$ for $k \le \alpha$. To first order in η, we therefore find the rather simple result,

$$\boxed{S_\ell[\phi_\ell] - S_\infty[\phi_\infty] = -\eta \sum_{i>\alpha} \frac{N v_i^2}{2\lambda_i}.} \tag{73}$$

We pause to note that this moment matching condition has interesting implications in the context of perturbation theory. For $k \le \alpha$,

$$0 = \langle\psi_k|R - Q_\ell\rangle = \left\langle\psi_k|R - Q_\infty + \eta Q_\infty\phi_1 + O(\eta^2)\right\rangle. \tag{74}$$

This expression must vanish at each order in η, and so

$$0 = \langle\psi_k|Q_\infty|\phi_1\rangle = \sum_{\substack{i>\alpha \\ j\le\alpha}} \frac{v_i z_{ij} z_{jk}}{\lambda_i \zeta_j}. \tag{75}$$

In a similar manner, considering the higher-order terms of Eq. (74) leads to nontrivial constraints on the higher-order corrections to ϕ_ℓ. We don't make use of this finding in what follows, but it seems worth noting.

Step 3: Expansion of $\ln\det[L^{2\alpha}\Delta^\alpha + \eta e^{-\phi_\ell}]$ to first order in η.

We first outline how we will go about computing

$$\ln\det\Gamma \quad \text{where} \quad \Gamma = \Lambda + \eta e^{-\phi_\ell}. \tag{76}$$

Computing this quantity requires computing the eigenvalues of Γ. We will use γ_i and ρ_i to denote the eigenvalues and corresponding orthonormal eigenfunctions of Γ, i.e.,

$$\Gamma\rho_i = \gamma_i\rho_i. \tag{77}$$

Our primary task is to compute each eigenvalue γ_i as a power series in η:

$$\gamma_i = \lambda_i + \eta\gamma_i^1 + \eta^2\gamma_i^2 + \cdots. \tag{78}$$

This task, as we will see, also requires computing the eigenfunctions ρ_i as power series in η:

$$\rho_i = \psi_i + \eta\rho_i^1 + \eta^2\rho_i^2 + \cdots. \tag{79}$$

As usual, it will help to expand ρ_i^1 in the eigenfunctions of Λ:

$$\rho_i^1(x) = \sum_j B_j^i \psi_j(x). \tag{80}$$

Keeping in mind that $\lambda_i > 0$ for $i > \alpha$, and $\lambda_j = 0$ for $j \le \alpha$, we see that

$$\ln\det\Gamma = \ln\prod_i \gamma_i \tag{81}$$

$$= \ln\left\{\prod_{j\le\alpha} \eta\gamma_j^1 \left(1 + \eta\frac{\gamma_j^2}{\gamma_j^1} + O(\eta^2)\right)\right\} + \ln\left\{\prod_{i>\alpha} \lambda_i \left(1 + \eta\frac{\gamma_i^1}{\lambda_i} + O(\eta^2)\right)\right\} \tag{82}$$

$$= \ln\left\{\eta^\alpha \left(\prod_{j\le\alpha} \gamma_i^1\right)\left(\prod_{i>\alpha} \lambda_i\right)\right\} + \eta\left\{\sum_{i>\alpha} \frac{\gamma_i^1}{\lambda_i} + \sum_{j\le\alpha} \frac{\gamma_j^2}{\gamma_j^1}\right\} + O(\eta^2). \tag{83}$$

So while the larger eigenvalues of Γ need only be computed to first order in η, the α lowest eigenvalues must actually be computed to second order in η. Performing this second order calculation will require that we also compute the eigenfunctions ρ_i to first order in η.

Plugging Eq. (78) and Eq. (79) into Eq. (77) and collecting terms by order in η, we get

$$0 = (\Lambda + \eta e^{-\phi_\infty} - \eta^2 e^{-\phi_\infty}\phi_1 + \cdots)(\psi_i + \eta\rho_i^1 + \eta^2\rho_i^2 + \cdots) - (\lambda_i + \eta\gamma_i^1 + \eta^2\gamma_i^2 + \cdots)(\psi_i + \eta\rho_i^1 + \eta^2\rho_i^2 + \cdots) \tag{84}$$

$$\begin{aligned} &= (\Lambda\psi_i - \lambda_i\psi_i) + \\ &\quad \eta\left(\Lambda\rho_i^1 + e^{-\phi_\infty}\psi_i - \lambda_i\rho_i^1 - \gamma_i^1\psi_i\right) + \\ &\quad \eta^2\left(\Lambda\rho_i^2 + e^{-\phi_\infty}\rho_i^1 - e^{-\phi_\infty}\phi_1\psi_i - \lambda_i\rho_i^2 - \gamma_i^1\rho_i^1 - \gamma_i^2\psi_i\right) + O(\eta^3). \end{aligned} \tag{85}$$

From the zeroth order term, we recover the eigenvalue equation $\Lambda\psi_i = \lambda_i\psi_i$, which we already knew. From the first order term, we get,

$$0 = \Lambda\rho_i^1 + e^{-\phi_\infty}\psi_i - \lambda_i\rho_i^1 - \gamma_i^1\psi_i. \tag{86}$$

Dotting this with ψ_k gives

$$\begin{aligned}
0 &= \left\langle\psi_k|\Lambda|\rho_i^1\right\rangle + \left\langle\psi_k|e^{-\phi_\infty}|\psi_i\right\rangle - \lambda_i\left\langle\psi_k|\rho_i^1\right\rangle - \gamma_i^1\left\langle\psi_k|\psi_i\right\rangle &(87)\\
&= \sum_j \lambda_k B_j^i\left\langle\psi_k|\psi_j\right\rangle + z_{ki} - \lambda_i\sum_j B_j^i\left\langle\psi_k|\psi_j\right\rangle - \gamma_i^1\delta_{ik} &(88)\\
&= \lambda_k B_k^i + z_{ik} - \lambda_i B_k^i - \gamma_i^1\delta_{ik} &(89)\\
&= (\lambda_k - \lambda_i)B_k^i + z_{ik} - \gamma_i^1\delta_{ik}. &(90)
\end{aligned}$$

Choosing $k = i$, we recover the standard first order correction to the eigenvalues,

$$\gamma_i^1 = z_{ii} \quad \text{for all} \quad i, \tag{91}$$

in particular,

$$\gamma_j^1 = \zeta_j, \quad j \le \alpha. \tag{92}$$

Moreover, for $i \le \alpha$, Eq. (90) gives

$$\rho_i^1(x) = \sum_{j>\alpha} B_j^i\psi_j(x) = -\sum_{j>\alpha}\frac{z_{ij}}{\lambda_j}\psi_j(x). \tag{93}$$

From the second order term in Eq. (85), we get,

$$0 = \Lambda\rho_i^2 + e^{-\phi_\infty}\rho_i^1 - e^{-\phi_\infty}\phi_1\psi_i - \lambda_i\rho_i^2 - \gamma_i^1\rho_i^1 - \gamma_i^2\psi_i. \tag{94}$$

Focusing on $i \le \alpha$ and dotting with ψ_i,

$$\begin{aligned}
0 &= \left\langle\psi_i|\Lambda|\rho_i^2\right\rangle + \left\langle\psi_i|e^{-\phi_\infty}|\rho_i^1\right\rangle - \left\langle\psi_i|e^{-\phi_\infty}\phi_1|\psi_i\right\rangle - \lambda_i\left\langle\psi_i|\rho_i^2\right\rangle - \gamma_i^1\left\langle\psi_i|\rho_i^1\right\rangle - \gamma_i^2\left\langle\psi_i|\psi_i\right\rangle &(95)\\
&= \left\langle\psi_i|e^{-\phi_\infty}|\rho_i^1\right\rangle - \left\langle\psi_i|e^{-\phi_\infty}\phi_1|\psi_i\right\rangle - \gamma_i^1\left\langle\psi_i|\rho_i^1\right\rangle - \gamma_i^2. &(96)
\end{aligned}$$

Using Eq. (93),

$$\begin{aligned}
\gamma_i^2 &= \left\langle\psi_i|e^{-\phi_\infty}|\rho_i^1\right\rangle - \left\langle\psi_i|e^{-\phi_\infty}\phi_1|\psi_i\right\rangle &(97)\\
&= -\sum_{j>\alpha}\frac{z_{ij}}{\lambda_j}\left\langle\psi_i|e^{-\phi_\infty}|\psi_i\right\rangle - \sum_{j>\alpha}\frac{v_j}{\lambda_j}\left\langle\psi_i|e^{-\phi_\infty}\psi_j|\psi_i\right\rangle + \sum_{\substack{j>\alpha\\ k\le\alpha}}\frac{v_j z_{jk}}{\lambda_j\zeta_k}\left\langle\psi_i|e^{-\phi_\infty}\psi_k|\psi_i\right\rangle &(98)\\
&= -\sum_{j>\alpha}\frac{z_{ij}^2}{\lambda_j} - \sum_{j>\alpha}\frac{v_j z_{iij}}{\lambda_j} + \sum_{\substack{j>\alpha\\ k\le\alpha}}\frac{v_j z_{jk} z_{iik}}{\lambda_j\zeta_k}. &(99)
\end{aligned}$$

Switching indices from $i \leftrightarrow j$, so that $i > \alpha$ and $j, k \le \alpha$, we find the required second order correction to γ_j:

$$\gamma_j^2 = -\sum_{i>\alpha}\frac{z_{ij}^2 + v_i z_{ijj}}{\lambda_i} + \sum_{\substack{i>\alpha\\ k\le\alpha}}\frac{v_i z_{ik} z_{jjk}}{\lambda_i\zeta_k}, \quad j \le \alpha. \tag{100}$$

We can now compute $\ln\det\Gamma$. Plugging the values for γ_i^1 and γ_i^2 into Eq. (83), and using

$$\prod_{j\le\alpha}\zeta_j = \det_{\text{ker}}[e^{-\phi_\infty}], \qquad \prod_{i>\alpha}\lambda_i = \det_{\text{row}}[\Lambda], \tag{101}$$

we get what we set out to find:

$$\boxed{\ln\det[\Lambda + \eta^{-\phi_\infty}] = \ln\left\{\eta^\alpha\det_{\text{ker}}[e^{-\phi_\infty}]\det_{\text{row}}[\Lambda]\right\} + \eta\left\{\sum_{i>\alpha}\frac{z_{ii}}{\lambda_i} - \sum_{\substack{i>\alpha\\ j\le\alpha}}\frac{z_{ij}^2 + v_i z_{ijj}}{\lambda_i\zeta_j} + \sum_{\substack{i>\alpha\\ j,k\le\alpha}}\frac{v_i z_{ik} z_{jjk}}{\lambda_j\zeta_j\zeta_k}\right\} + O(\eta^2).} \tag{102}$$

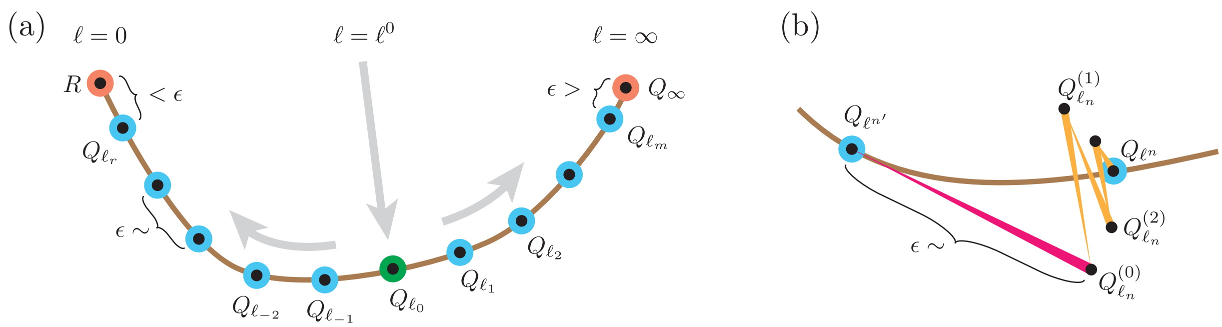

FIG. S1. Illustration of the predictor-corrector homotopy algorithm used for computations in the main text. (a) The MAP curve (brown) is approximated using finite set of densities $\{R, Q_{\ell_r}, \cdots, Q_{\ell_{-2}}, Q_{\ell_{-1}}, Q_{\ell_0}, Q_{\ell_1}, Q_{\ell_2}, \ldots, Q_{\ell_m}, Q_\infty\}$. First the MAP density at an intermediate length scale ℓ_0 is computed. A predictor-corrector algorithm is then used to extend the set of MAP densities outward to larger and to smaller values of ℓ. These ℓ values are chosen so that neighboring MAP densities lie within a geodesic distance of $\lesssim \epsilon$ of each other. (b) Each step $Q_{\ell_{n'}} \to Q_{\ell_n}$ has two parts. First, a predictor step (magenta) computes a new length scale ℓ_n and an approximation $Q^{(0)}_{\ell_n}$ of Q_{ℓ_n}. A series of corrector steps $Q^{(0)}_{\ell_n} \to Q^{(1)}_{\ell_n} \to Q^{(2)}_{\ell_n} \cdots$ (orange) then converges to Q_{ℓ_n}.

Putting it all together

Putting together our results from Eq. (73) and Eq. (102), and switching indices $j \leftrightarrow k$ in the last term of Eq. (102) (just to smooth out the notation), we get

$$\ln E = -\left(S_\ell[\phi_\ell] - S_\infty[\phi_\infty]\right) - \frac{1}{2}\ln\left\{\frac{\det[\Lambda + \eta e^{-\phi_\infty}]}{\eta^\alpha \det_{\mathrm{ker}}[e^{-\phi_\infty}]\det_{\mathrm{row}}[\Lambda]}\right\} \tag{103}$$

$$= \eta\left\{\sum_{i>\alpha}\frac{N v_i^2 - z_{ii}}{2\lambda_i} + \sum_{\substack{i>\alpha\\ j\leq\alpha}}\frac{z_{ij}^2 + v_i z_{ijj}}{2\lambda_i\zeta_j} - \sum_{\substack{i>\alpha\\ j,k\leq\alpha}}\frac{v_i z_{ij} z_{jkk}}{2\lambda_i\zeta_j\zeta_k}\right\} + O(\eta^2) \tag{104}$$

$$= K\eta + O(\eta^2), \tag{105}$$

where the K coefficient is

$$\boxed{K = \sum_{i>\alpha}\frac{N v_i^2 - z_{ii}}{2\lambda_i} + \sum_{\substack{i>\alpha\\ j\leq\alpha}}\frac{z_{ij}^2 + v_i z_{ijj}}{2\lambda_i\zeta_j} - \sum_{\substack{i>\alpha\\ j,k\leq\alpha}}\frac{v_i z_{ij} z_{jkk}}{2\lambda_i\zeta_j\zeta_k}.} \tag{106}$$

HOMOTOPY ALGORITHM

In the space of probability densities, the set of MAP densities forms a one-parameter curve extending from the MaxEnt density at $\ell = \infty$ to the data histogram at $\ell = 0$. The algorithm used to perform the computations shown in Figs. 1 and 2 of the main text approximates this "MAP curve" by computing the MAP densities at a finite set of length scales. This set of length scales extends from $\ell = 0$ to $\ell = \infty$ (inclusively), and are chosen so that, at any two neighboring length scales ℓ and ℓ', the MAP densities Q_ℓ and $Q_{\ell'}$ are sufficiently similar. By computing this "string of beads" approximation to the MAP curve, the homotopy algorithm allows one to compute the evidence ratio E over all length scales.

This algorithm proceeds as follows. First, the MaxEnt density Q_∞ is computed essentially as described by Ormoneit and White [4]. The details of this computation are reported below.

Next, the MAP density Q_{ℓ_0} is computed at an intermediate length scale $\ell_0 = L/\sqrt{G}$. This density, Q_{ℓ_0}, serves as the starting point for a predictor-corrector homotopy algorithm [5] that traces the MAP curve from ℓ_0 towards both larger and smaller length scales (Fig. S1a). The result is a finite set of length scales $\{0, \ell_r, \cdots, \ell_{-2}, \ell_{-1}, \ell_0, \ell_1, \ell_2, \cdots, \ell_m, \infty\}$ chosen so that MAP densities at any two neighboring length scales ℓ and ℓ' satisfy

$$D_{\mathrm{geo}}(Q_\ell, Q_{\ell'}) \lesssim \epsilon, \tag{107}$$

where D_{geo} is the geodesic distance measure [1, 6] and ϵ is a small user-specified distance. The value $\epsilon = 10^{-2}$ was used for the calculations reported in the main text.

Each step $Q_{\ell_{n'}} \to Q_{\ell_n}$ is accomplished in two parts (Fig. S1b). First, a "predictor step" is used to compute the new length scale ℓ_n as well as an approximation $Q^{(0)}_{\ell_n}$ to the corresponding MAP density Q_{ℓ_n}. A series of "corrector steps" is then used to compute a corresponding series of densities $Q^{(1)}_{\ell_n}, Q^{(2)}_{\ell_n}, \cdots$ that converges to Q_{ℓ_n}. These predictor and corrector steps are described in more detail below.

Computing the MaxEnt density

We saw in the main text that ϕ_∞ is a polynomial of order $\alpha - 1$, i.e.,

$$\phi_\infty(x) = \sum_{i=0}^{\alpha-1} a_i x^i. \tag{108}$$

The problem of computing the MaxEnt density Q_∞ therefore reduces to finding the vector of coefficients $\boldsymbol{a} = (a_0, a_1, \ldots, a_{\alpha-1})$ that minimizes the action

$$S_\infty(\boldsymbol{a}) = N \int \frac{dx}{L} \left\{ LR \sum_{i=0}^{\alpha-1} a_i x^i + \exp\left[- \sum_{i=0}^{\alpha-1} a_i x^i \right] \right\}. \tag{109}$$

As suggested by Ormoneit and White [4], we solve this optimization problem using the Newton-Raphson algorithm with backtracking. Starting at $\boldsymbol{a}^0 = \boldsymbol{0}$, we iterate

$$\boldsymbol{a}^n \to \boldsymbol{a}^{n+1} = \boldsymbol{a}^n + \gamma_n \delta \boldsymbol{a}^n \tag{110}$$

where $0 < \gamma_n \leq 1$ and the vector $\delta \boldsymbol{a}^n$ is the solution to

$$\sum_{j=0}^{\alpha-1} \left. \frac{\partial^2 S}{\partial a_i \partial a_j} \right|_{\boldsymbol{a}^n} \delta a_j^n = - \left. \frac{\partial S}{\partial a_i} \right|_{\boldsymbol{a}^n}. \tag{111}$$

From Eq. (109),

$$\frac{\partial S}{\partial a_i} = N\mu_i - N \int \frac{dx}{L} x^i \exp\left[- \sum_{k=1}^{\alpha-1} a_k x^k \right] \quad \text{and} \quad \frac{\partial^2 S}{\partial a_i \partial a_j} = N \int \frac{dx}{L} x^{i+j} \exp\left[- \sum_{k=1}^{\alpha-1} a_k x^k \right], \tag{112}$$

where $\mu_i = \int dx\, R x^i$ is the i'th moment of the data. The Hessian matrix $\frac{\partial^2 S}{\partial a_i \partial a_j}$ is positive definite at all $\boldsymbol{a}$, so a unique solution for $\delta \boldsymbol{a}^n$ can always be found [7]. The scalar γ_n is then chosen so that

$$S_\infty(\boldsymbol{a}^n + \gamma_n \delta \boldsymbol{a}^n) - S_\infty(\boldsymbol{a}^n) < \frac{\gamma_n}{2} \sum_{i=0}^{\alpha-1} \left. \frac{\partial S}{\partial a_i} \right|_{\boldsymbol{a}^n} \delta a_i^n. \tag{113}$$

Specifically, γ_n is first set to 1, then is reduced by factors of $\frac{1}{2}$ until Eq. (113) is satisfied. This "dampening" of the Newton-Raphson method is sufficient to guarantee convergence [4, 8]. This algorithm is terminated when the magnitude of the change in the action $|S_\infty(\boldsymbol{a}^{n+1}) - S_\infty(\boldsymbol{a}^n)|$ falls below a specified tolerance.

Predictor step

The predictor step computes a change $\ell \to \ell'$ in the length scale, as well as an *approximation* to the corresponding change $\phi_\ell \to \phi_{\ell'}$ in the MAP field. Specifically, the predictor step returns a scalar δt and a function $\rho(x)$ such that,

$$\boxed{t' = t + \delta t \quad \text{and} \quad \phi_{\ell'}(x) \approx \phi_\ell(x) + \rho(x)\delta t}, \tag{114}$$

where $t = \ln \eta$ is a numerically convenient reparametrization of the length scale ℓ. To determine the function ρ, we examine the equation of motion at ℓ':

$$\begin{aligned} 0 &= \Lambda \phi_{\ell'} + \eta'(LR - e^{-\phi_{\ell'}}) && (115)\\ &= \Lambda(\phi_\ell + \rho \delta t) + \eta(1 + \delta t)(LR - e^{-\phi_\ell} + e^{-\phi_\ell} \rho \delta t + O[(\delta t)^2]) && (116)\\ &= \Lambda \phi_\ell + \eta(LR - e^{-\phi_\ell}) + \delta t \left\{ \left[\Lambda + \eta e^{-\phi_\ell} \right] \rho + \eta(LR - e^{-\phi_\ell}) \right\} + O[(\delta t)^2]. && (117) \end{aligned}$$

The $O(\delta t)$ term must vanish, and thus we obtain a linear equation for ρ:

$$\boxed{\left[\Lambda + \eta e^{-\phi_\ell}\right]\rho = \eta(e^{-\phi_\ell} - LR)}. \tag{118}$$

The scalar δt is then chosen to satisfy the distance criterion,

$$\epsilon^2 \approx D^2_{\text{geo}}(Q_\ell, Q_{\ell'}) \approx \int dx \frac{\delta Q^2}{Q_\ell} \approx (\delta t)^2 \int dx\, Q_\ell\, \rho^2. \tag{119}$$

We therefore adopt

$$\boxed{\delta t = \pm \frac{\epsilon}{\sqrt{\int dx\, Q_\ell\, \rho^2}}}, \tag{120}$$

with the sign of δt chosen according to the direction we wish to traverse the MAP curve.

Corrector step

The purpose of the corrector step is to accurately solve the equation of motion,

$$\Lambda\phi_\ell + \eta(LR - e^{-\phi_\ell}) = 0, \tag{121}$$

at fixed length scale ℓ. This step is used initially to compute Q_{ℓ_0} at the starting length scale ℓ_0. It is also used to hone in on the MAP density at each new length scale chosen by the predictor step of the homotopy algorithm.

As with the computation of the MaxEnt density, this corrector step uses the Newton-Raphson algorithm with backtracking. Starting from a hypothesized field $\phi^{(0)}$ (e.g., returned by the predictor algorithm), the iterative step

$$\boxed{\phi^{(n)} \to \phi^{(n+1)} = \phi^{(n)} + \gamma_n \delta\phi^{(n)}} \tag{122}$$

is repeatedly performed. The function $\delta\phi^{(n)}$ and scalar γ_n are chosen at each step so that this iteration converges to the true field ϕ_ℓ. To derive the field perturbation $\delta\phi^{(n)}$, we plug this formula for $\phi^{(n+1)}$ into the equation of motion. Keeping only terms of order $\delta\phi^{(n)}$ or less, we see that

$$\Lambda\phi^{(n)} + \Lambda\delta\phi^{(n)} + \eta(LR - e^{-\phi^{(n)}} + e^{-\phi^{(n)}}\delta\phi^{(n)}) = 0. \tag{123}$$

The field perturbation $\delta\phi^{(n)}$ is therefore given by the linear equation

$$\boxed{\left[\Lambda + \eta e^{-\phi^{(n)}}\right]\delta\phi^{(n)} = \eta(e^{-\phi} - LR) - \Lambda\phi^{(n)}.} \tag{124}$$

As before, γ_n is chosen so that so that the action decreases by a respectable amount compared to what is expected from the linear approximation, i.e.,

$$S_\ell[\phi^{(n)} + \gamma_n\delta\phi^{(n)}] - S_\ell[\phi^{(n)}] < \frac{\gamma_n}{2} \int dx\ \left.\frac{\delta S_\ell}{\delta\phi(x)}\right|_{\phi^{(n)}} \delta\phi^{(n)}. \tag{125}$$

This iterative process is terminated when the magnitude of the change in the action, $|S_\ell[\phi^{(n+1)}] - S_\ell[\phi^{(n)}]|$ falls below a specified tolerance.

* Email correspondence to jkinney@cshl.edu

[1] J. B. Kinney, Phys. Rev. E **90**, 011301(R) (2014).
[2] W. Bialek, C. G. Callan, and S. P. Strong, Phys. Rev. Lett. **77**, 4693 (1996).
[3] I. Nemenman and W. Bialek, Phys. Rev. E **65**, 026137 (2002).
[4] D. Ormoneit and H. White, Economet. Rev. **18**, 127 (1999).
[5] E. L. Allgower and K. Georg, *Numerical Continuation Methods: An Introduction* (Springer, 1990).
[6] J. Skilling, AIP Conf. Proc. **954**, 39 (2007).
[7] L. R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984).
[8] S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge University Press, 2009).