Letter

Generalized maximum entropy methods as limits of the average spectrum method

Khaldoon Ghanem
Quantinuum, Leopoldstrasse 180, 80804 Munich, Germany

Erik Koch
Jülich Supercomputer Centre, Forschungszentrum Jülich, 52425 Jülich, Germany
and JARA High-Performance Computing, 52425 Jülich, Germany

(Received 21 July 2023; accepted 25 October 2023; published 8 November 2023)

We show that in the continuum limit, the average spectrum method (ASM) is equivalent to maximizing Rényi entropies of order η, of which Shannon entropy is the special case $\eta = 1$. The order of Rényi entropy is determined by the way the spectra are sampled. Our derivation also suggests a modification of Rényi entropy, giving it a nontrivial $\eta \to 0$ limit. We show that the sharper peaks generally obtained in ASM are associated with entropies of order $\eta < 1$. Our paper provides a generalization of the maximum entropy method that enables extracting more structure than the traditional method.

DOI: 10.1103/PhysRevB.108.L201107

Spectral reconstruction is an inverse problem where the goal is estimating a non-negative spectrum given incomplete and/or noisy data. It is a recurring problem in many branches of physics and has many important applications including, but not limited to, image reconstruction, x-ray diffraction, mass spectrometry, and analytic continuation of quantum Monte Carlo data [1–4]. Mathematically, the problem can be formulated as a Fredholm integral equation of the first kind,

$$g(y_j) = \int dx\, K(y_j, x) f(x), \tag{1}$$

where the left-hand side $g(y_j)$ represents m data points known numerically, while the integral kernel $K(y_j, x)$ is a continuous function known analytically and determined by the problem at hand. The spectrum $f(x)$ is a non-negative and integrable function $f(x)$ that needs to be estimated. In this Letter, we focus on problems where the norm of the spectrum is known *a priori*, so we can always rescale the above equation such that the spectrum is normalized to one.

The most widely used approach for solving this problem is the maximum entropy method (MaxEnt) [1,4,5]. In its most basic form, it aims to find the spectrum maximizing Shannon entropy under known constraints [6]

$$f_{\text{MaxEnt}} = \arg\max_{f \in \mathcal{C}} S[f],$$

where

$$S[f] := -\int dx\, f(x) \ln\left(\frac{f(x)}{D(x)}\right) \tag{2}$$

is the relative Shannon entropy or Kullback-Leibler divergence, and represents the expected amount of information in a spectrum f relative to another D, called the default model. The manifold $\mathcal{C}$ represents the non-negativity and normalization constraints as well as the m data constraints of Eq. (1). In case of noisy data, the latter is usually imposed via the discrepancy principle where the fit to the data is set equal to the expected level of noise [6–9]. Assuming Gaussain noise with zero mean and covariance matrix $\mathbf{L}$ and denoting the data residuals as $r_i := g(y_i) - \int dx\, K(y_i, x) f(x)$, the fit is defined as $\chi^2[f] := \mathbf{r}^{\mathsf{T}} \mathbf{L}^{-1} \mathbf{r}$, and the constrained manifold reads

$$\mathcal{C} := \left\{ f(x) \geqslant 0 : \int dx f(x) = 1,\ \chi^2(f) = m \right\}.$$

MaxEnt can be implemented efficiently [10] and produces good results in general, which contributed to its widespread adoption. Nevertheless, a common criticism, especially within the analytic continuation community, is that it produces spectra with overly smooth peaks. In addition, one needs to pick a default model whose choice may bias the result. Consequently, the average spectrum method (ASM) emerged as a promising alternative that addresses these issues [11–13]. It averages equally over all normalized non-negative spectra that fit the data constraints [14]

$$f_{\text{ASM}}(x) = \int_{\mathcal{C}} \mathcal{D}f\ f(x).$$

The averaging was assumed to wash out features not supported by the data and thus produce unbiased results without the need for a default model. It was shown later, however, that the above functional integral is not uniquely defined and that a default model is implied by the choice of the discretization grid [15]. Even a sampling of the grid points amounts to a change of the functional measure and thus produces a different result [16]. Nevertheless, ASM remains a viable alternative to MaxEnt due to its apparent ability of producing sharper spectra [12,13,17–20].

A connection between MaxEnt and ASM was first suggested by Ref. [21], which argued that the smoother spectra of MaxEnt are the result of it being a mean-field approximation of ASM. However, it was shown in Ref. [15] that such an approximation implies an entropy different from Shannon

entropy used by MaxEnt. The new relative entropy has the functional form of Eq. (2), but with the role of the default model and the spectrum reversed (reversed Kullbach-Leibler divergence),

$$S_{\mathrm{GK}}[f] \coloneqq -\int dx\, D(x) \ln\left(\frac{D(x)}{f(x)}\right), \tag{3}$$

where, following Ref. [22], the subscript GK is used to distinguish this specific form of entropy. GK entropy was derived for the case of sampling spectral weights on a fixed grid, and Ref. [22] recognized that sampling grid positions as well as spectral weights should lead to yet a different entropy and conjectured it to be a linear mixture of Shannon and GK entropies. One major contribution of our work here is deriving the correct entropic form for this case and generalizing it. The result turns out to be the family of Rényi entropies.

In his seminal work [23], Rényi relaxed the axioms of Shannon and derived the relative entropies

$$S_{\mathrm{R}}[f,\eta] \coloneqq \frac{1}{1-\eta} \ln \int dx\, \frac{f(x)^{\eta}}{D(x)^{\eta-1}}, \tag{4}$$

where the parameter η is called the order of the entropy. In the limit $\eta \to 1$, one recovers relative Shannon entropy as a special case. The relative form of Rényi entropy shown above is also called Rényi divergence. It was originally proposed as an alternative information measure and has since found important applications in various fields, including statistical mechanics [24] and quantum information [25]. Other generalized entropies have since been developed [26], the most prominent being Tsallis entropy [27]:

$$S_{\mathrm{T}}[f,\eta] \coloneqq \frac{1}{\eta-1}\left[\int dx\, \frac{f(x)^{\eta}}{D(x)^{\eta-1}} - 1\right]. \tag{5}$$

Rényi and Tsallis entropies are related by a monotonic mapping, so maximizing one is equivalent to maximizing the other [28]. The existence and usefulness of these entropies calls into question the uniqueness of Shannon entropy within MaxEnt, which has been the subject of extensive debate [29–34]. The results in this Letter supplement these arguments and give a systematic approach for deriving generalized entropies as limits of ASM. We also show how a simple rescaling of relative Rényi entropy changes the limit $\eta \to 0$ from simply vanishing to giving the GK entropy. Finally, we discuss the effect of maximizing generalized entropies of different orders, dubbed MaxGEnt, and exemplify it in an illustrative test case.

a. Limits of ASM.— To evaluate the functional integral of ASM, we discretize the spectrum by representing it as a linear combination of N delta functions

$$f(x) = \sum_{i=1}^{N} f_i \delta(x - x_i), \tag{6}$$

where x_i are their positions and f_i are their non-negative weights that sum up to one. The average spectrum is obtained by sampling the weights and/or positions, binning each spectrum on a discretized grid and averaging the binned spectra [21,22]. Let the binning grid have M intervals denoted as $\mathcal{I}_j$, then the amplitude of a binned sample on the jth bin is the integral of the sample over that bin $F_j \coloneqq \int_{\mathcal{I}_j} dx f(x) = \sum_{x_i \in \mathcal{I}_j} f_i$. Clearly, different samples can give rise to the same binned amplitudes. In physical terms, the unbinned sample $(x_1, \ldots, x_N; f_1, \ldots, f_N)$ can be thought of as a microstate, while the binned one $(F_1, \ldots, F_M)$ as the macrostate. The probability distribution of a macrostate is completely determined by how the microstates are sampled. As the number of delta functions increases, this distribution becomes singular but its normalized logarithm has a well-defined limit

$$\lim_{N\to\infty} \frac{\ln P(F_1, \ldots, F_M)}{N} =: H(F_1, \ldots, F_M),$$

which is defined as the entropy associated with that specific way of sampling [22]. In that limit, averaging binned spectra becomes equivalent to finding the spectrum with the maximum entropy. When the weights are sampled uniformly while the positions x_i are fixed and placed with a density $D(x)$, one obtains GK entropy [15]. On the other hand, when the weights are held fixed at $1/N$ while positions are sampled independently from a default model $D(x)$, it yields Shannon entropy [21,22]. In the following, we seek the entropy of sampling both the positions and weights simultaneously. Let n_j be the number of delta functions falling in the jth bin, then the desired probability can be decomposed as

$$P(\{F_j\}) = \sum_{\{n_j\}} P(\{n_j\}) \times P(\{F_j\}|\{n_j\}), \tag{7}$$

where $P(\{n_j\})$ is the probability that a configuration of N delta functions is distributed as $n_1, \ldots, n_M$ over the bins $\mathcal{I}_1, \ldots, \mathcal{I}_M$, and $P(\{F_j\}|\{n_j\})$ is the conditional probability that the weights of this configuration of delta functions adding up to $F_1, \ldots, F_M$ for the respective bins.

b. Configuration probability.— The positions are sampled independently from $D(x)$. The probability of a delta function falling inside the jth bin is $D_j \coloneqq \int_{\mathcal{I}_j} dx D(x)$, so the probability of having a specific configuration $\{n_j\}$ follows the multinomial distribution

$$P(\{n_j\}) = \frac{N!}{n_1! \ldots n_M!} D_1^{n_1} \ldots D_M^{n_M}.$$

We then use Stirling's formula $\ln x! \approx x \ln x - x$ to get the following approximation to the configuration probability:

$$P(\{n_j\}) \approx \exp\left[-\sum_j n_j \ln \frac{n_j}{N D_j}\right]. \tag{8}$$

c. Conditional probability of amplitudes.— The weights f_i are sampled under the condition of being non-negative and summing up to one. Therefore, they follow a Dirichlet distribution with unit shape parameters [15]. According to the aggregation property of the Dirichlet distribution, the partial sums of random variables following the Dirichlet distribution also follow a Dirichlet distribution where the corresponding shape parameters are summed [35]. Assuming n_j delta functions fall in the jth bin, the shape parameter of amplitude F_j is $\sum_{x_i \in \mathcal{I}_j} 1 = n_j$ and the desired Dirichlet distribution reads

$$P(\{F_j\}|\{n_j\}) = \frac{\Gamma(N)}{\Gamma(n_1)\ldots\Gamma(n_M)} F_1^{n_1-1} \ldots F_M^{n_M-1},$$

where $\Gamma(z)$ is the gamma function. In the limit of very large N, we can replace n_i by n_i+1 and use Stirling's formula again

to get an approximation to the conditional probability:

$$P(\{F_j\}|\{n_j\}) \approx \exp\left[-\sum_j n_j \ln \frac{n_j}{NF_j}\right]. \tag{9}$$

d. Total probability of amplitudes.— The total probability in Eq. (7) is a sum over the product of Eqs. (8) and (9) for all possible configurations. In the limit $N \to \infty$, the sum is dominated by the term with the largest exponent. Finding its value is simple when defining the auxiliary quantities

$$E \coloneqq \sum_j \sqrt{F_j D_j}, \quad \text{and} \quad E_j \coloneqq \sqrt{F_j D_j}/E,$$

and writing the total probability in terms of normalized configurations $\tilde{n}_j \coloneqq n_j/N$ as

$$P(\{F_j\}) = \exp[2N \ln E] \sum_{\tilde{n}_1,\ldots,\tilde{n}_M} \exp\left[-2N \sum_j \tilde{n}_j \ln \frac{\tilde{n}_j}{E_j}\right].$$

We recognize the exponent above as a multiple of the Shannon entropy of a normalized distribution, $\tilde{n}_j$, with respect to another, E_j. Its maximum value is zero and thus the whole sum of exponentials is dominated by one. In the limit $N \to \infty$, we can therefore write

$$P(F_1, \ldots, F_M) \approx \exp[2N \ln E], \tag{10}$$

from which we find the entropy of sampling both weights and positions:

$$H(F_1, \ldots, F_M) = 2 \ln \left(\sum_j \sqrt{D_j F_j}\right).$$

This is a discrete version of Rényi entropy Eq. (4) with $\eta = 1/2$. Taking the limit of fine binning grid $M \to \infty$, one can also recover the corresponding continuous version.

e. Generalized entropies.— It is rather interesting that sampling both weights and positions gives an entropy that is symmetric in both the spectrum and the default model, while sampling one of them gives the same entropic form but with the role of default model and spectrum reversed [compare Eq. (2) with (3)]. A natural question then is to ask whether other ways of sampling could give rise to a general form that connects the different entropies. This is indeed the case if we replace the uniform distribution of the weights with a symmetric Dirichlet distribution of arbitrary shape parameter $\alpha > 0$. The shape parameter controls how concentrated the weights are around their mean value $1/N$. The uniform distribution is the special case with $\alpha = 1$, and in the limit $\alpha \to \infty$, the sampling becomes equivalent to ASM with fixed weights. We can derive the entropy of a general shape parameter α in a similar way to the derivation above, which corresponds to $\alpha = 1$. In the general case, the configuration probability remains the same, while the conditional probability of amplitudes gets raised to the power α, so the total probability reads

$$P_\alpha(\{F_j\}) = \exp[(\alpha+1)N \ln E'] \times \sum_{\tilde{n}_1,\ldots,\tilde{n}_M} \exp\left[-(\alpha+1)N \sum_j \tilde{n}_j \ln \frac{\tilde{n}_j}{E'_j}\right],$$

where we defined the generalized auxiliary quantities

$$E' \coloneqq \sum_j \left(F_j^\alpha D_j\right)^{\frac{1}{1+\alpha}} \quad \text{and} \quad E'_j \coloneqq \left(F_j^\alpha D_j\right)^{\frac{1}{1+\alpha}}/E'.$$

Using a similar argument to the $\alpha = 1$ case, we find the entropy in the general α case to be

$$H_\alpha(F_1, \ldots, F_M) = (\alpha+1) \ln \left[\sum_j F_j \left(\frac{F_j}{D_j}\right)^{\frac{-1}{1+\alpha}}\right],$$

which is the Rényi entropy of order $\eta = \alpha/(1+\alpha)$.

f. Relating GK entropy to Rényi entropy.— In the limit of $\alpha \to \infty$, which gives fixed weights, the order of the entropy has the limit $\eta \to 1$ and we recover Shannon entropy as expected. One would then expect to recover GK entropy in the other limit $\eta \to 0$. Instead, Rényi entropy becomes trivial in that limit and vanishes when f has the same support as D. In fact, this limit corresponds to $\alpha \to 0$, where the earlier use of Stirling's approximation for the conditional probability breaks down. Nevertheless, we can still recover GK entropy by slightly modifying the definition of relative Rényi entropies as

$$S_{\tilde{\mathrm{R}}}[f, \eta] \coloneqq \frac{1}{\eta} S_{\mathrm{R}}[f, \eta] = \frac{1}{\eta(1-\eta)} \ln \int dx \, \frac{f(x)^\eta}{D(x)^{\eta-1}}.$$

In the limit $\eta \to 0$, this modification gives GK entropy (just as Shannon entropy is recovered for $\eta \to 1$), thus giving a richer structure than the original expression Eq. (4). Crucially, it does not change the properties of relative Rényi entropy at any nonvanishing order, merely rescaling it.

g. Maximizing generalized entropies (MaxGEnt).— It is well-known that maximizing Shannon entropy under linear constraints leads to distributions of exponential form, while maximizing Rényi (or equivalently Tsallis) entropies gives power-law distributions [27,36]. For convenience, instead of maximizing the modified Rényi entropy itself, we maximize its monotonic transformation

$$S_{\tilde{\mathrm{T}}}[f, \eta] \coloneqq \frac{e^{\eta(1-\eta)S_{\tilde{\mathrm{R}}}[f,\eta]} - 1}{\eta(1-\eta)}, \tag{11}$$

representing our modified version of Tsallis entropy with a nontrivial $\eta \to 0$ limit. We then define the Lagrangian

$$\mathcal{L}_\eta[f, \mu, \{\lambda_k\}] = S_{\tilde{T}}[f, \eta] - \mu\left[\int dx f(x) - 1\right] - \mu(\eta-1) \sum_k \lambda_k \left[\int dx f(x) c_k(x) - C_k\right], \tag{12}$$

where $\int dx f(x) c_k(x) = C_k$ are a set of linear constraints and μ, λ_k are Lagrange multipliers. Note that the λ_k are explicitly multiplied by $\mu(\eta - 1)$ for convenience. Rényi/Tsallis maximizers are saddle points of this Lagrangian and have the general form

$$f(x) = \frac{1}{Z} D(x) \left[1 + (1-\eta) \sum_i \lambda_k c_k(x)\right]^{\frac{1}{\eta-1}}, \tag{13}$$

where $Z(\{\lambda_i\}) \coloneqq \int dx D(x) [1 + (1-\eta) \sum_i \lambda_k c_k(x)]^{\frac{1}{\eta-1}}$ is a normalization constant. Using a flat default model and con-

straints on the moments, the above form has a power-law decay for any $\eta < 1$. In the limit $\eta \to 1$, we recover the familiar exponential form of Shannon maximizers. Spectra with fatter tails appear sharper than ones with the same moments, but which otherwise decay faster. Therefore, maximizing entropies with lower order η is generally expected to produce sharper spectra. For example, maximizing Rényi entropy under the constraints of zero mean and variance σ^2 gives Student's t-distributions [37,38]

$$f(x) \propto \left(1 + \frac{1-\eta}{3\eta-1}\frac{x^2}{\sigma^2}\right)^{\frac{1}{\eta-1}}, \tag{14}$$

which has the full width at half maximum:

$$\mathrm{FWHM} = 2\sigma\sqrt{\frac{3\eta-1}{1-\eta}(2^{1-\eta}-1)}. \tag{15}$$

As η decreases, the tail gets fatter but its FWHM shrinks, producing a sharper peak. This readily explains ASM's ability of producing sharper spectra than MaxEnt. The typical ways of performing ASM correspond to entropies with orders $\eta = 0$ (sampling only weights) and $\eta = 1/2$ (sampling both weights and positions). These are lower than the order $\eta = 1$ of MaxEnt's Shannon entropy and lead to peaks with narrower FWHM and fatter tails. The fat tails also elucidate the observation that ASM on a fixed grid is sensitive to the cutoff/width parameter of a default model, and also why this sensitivity is reduced when the grid positions are sampled [15,16].

h. Illustrative test case.— In practical applications, one does not have access to exact data values, only noisy ones. The constraint would then be imposed on the data fit χ^2, which is a quadratic function of the spectrum rather than linear. Nevertheless, the tendency of lower entropies to produce sharper spectra remains. As an illustrative example, we look at the analytic continuation problem of the density of states $D(\omega)$ from its Green's function $G(z) = 1/(2\pi)\int d\omega D(\omega)/(z-w)$ for a hole in the t-J model in infinite dimensions [39] at finite exchange $J/t = 1$. We evaluated the exact Green's function at eight Matsubara frequencies $z_j = i(2j+1)\pi/8$ and added relative Gaussian noise with standard deviation 10^{-3}. Different entropies are then optimized under the discrepancy principle constraint (see Fig. 1). As expected, using an entropy with lower order produces spectra of sharper peaks and fatter tails, with GK entropy ($\eta = 0$) producing the sharpest result. On the same plot, we also show ASM spectra using both fixed and free positions and under the same discrepancy constraint.

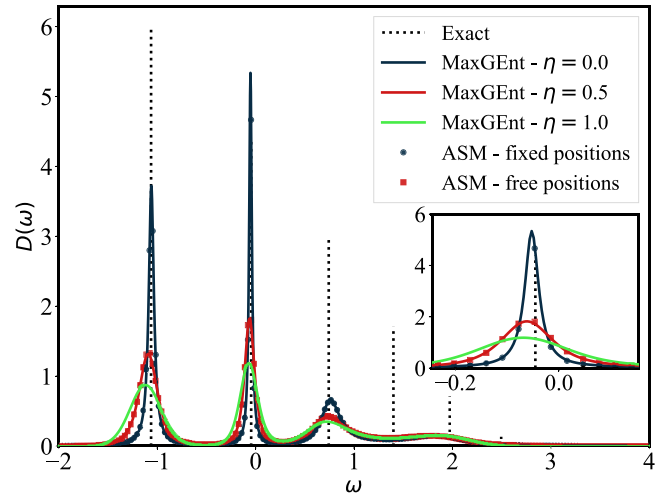

FIG. 1. MaxGEnt and ASM results for the analytic continuation of a density of states from the noisy Green's function on Matsubara frequencies. The exact density is a set of weighted delta peaks which are plotted here as vertical lines whose heights are proportional to their relative weights. ASM is performed using $N = 256$ delta functions. In all methods, a flat default model in the interval $\omega \in [-4, 4[$ is used.

Even using only $N = 256$ delta functions, the ASM spectra match the corresponding entropy maximization spectra providing a numerical confirmation of our derivation above.

i. Summary and discussion.— In this Letter, we provided a systematic approach for deriving the entropies associated with different ways of sampling in ASM. This resulted in a modified version of Rényi entropies with the order determined implicitly by the specific way of sampling. We further showed that the apparently sharp and desirable peaks of ASM can be attributed to the lower order of Rényi entropy associated with typical sampling methods. These results allow obtaining the same quality of ASM spectra more efficiently via maximization of the entropy instead of sampling. They also invite a revisit of the unique role of Shannon entropy in MaxEnt. Besides providing sharper spectra, maximizing Rényi/Tsallis entropies can now be justified by the underlying sampling process. In particular, sampling the weights with either fixed or free positions is no less justified than sampling the positions alone (which results in Shannon entropy). This generalization of MaxEnt gives an additional degree of freedom (the order of the entropy) that allows including prior knowledge of the expected profile of the spectral peaks.

[1] S. F. Gull and G. J. Daniell, Nature (London) **272**, 686 (1978).
[2] S. Wilkins, J. Varghese, and M. Lehmann, Acta Crystallogr. A Found Crystallogr. **39**, 594 (1983).
[3] Z. Zhang, S. Guan, and A. G. Marshall, J. Am. Soc. Mass Spectrom. **8**, 659 (1997).
[4] R. N. Silver, D. S. Sivia, and J. E. Gubernatis, Phys. Rev. B **41**, 2380 (1990).
[5] M. Jarrell, in *Correlated Electrons: From Models to Materials*, edited by E. Pavarini, E. Koch, F. Anders, and M. Jarrell (Forschungszentrum Jülich, Jülich, 2012).
[6] K. Ghanem and E. Koch, Phys. Rev. B **107**, 085129 (2023).
[7] V. A. Morozov, Criteria for selection of regularization parameter, in *Methods for Solving Incorrectly Posed Problems* (Springer, New York, 1984), pp. 32–64.
[8] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977).
[9] S. Gull and J. Skilling, IEE Proc. F Commun. Radar Signal Process. UK **131**, 646 (1984).
[10] R. K. Bryan, Eur. Biophys. J. **18**, 165 (1990).

[11] S. R. White, in *Computer Simulation Studies in Condensed Matter Physics III*, edited by D. P. Landau, K. K. Mon, and B.-B. Schüttler (Springer, Heidelberg, 1991), pp. 145–153.
[12] A. W. Sandvik, Phys. Rev. B **57**, 10287 (1998).
[13] A. W. Sandvik, Phys. Rev. E **94**, 063308 (2016).
[14] For noisy data, ASM usually weights the spectra with how well they fit the data. Restricting the averaging to a specific data fit makes the comparison with constrained MaxEnt more seamless, but it is straightforward to extend our arguments to that case.
[15] K. Ghanem and E. Koch, Phys. Rev. B **101**, 085111 (2020).
[16] K. Ghanem and E. Koch, Phys. Rev. B **102**, 035114 (2020).
[17] K. Vafayi and O. Gunnarsson, Phys. Rev. B **76**, 035115 (2007).
[18] O. F. Syljuåsen, Phys. Rev. B **78**, 174429 (2008).
[19] S. Fuchs, T. Pruschke, and M. Jarrell, Phys. Rev. E **81**, 056701 (2010).
[20] K. Ghanem, Stochastic analytic continuation: A Bayesian approach, Ph.D. thesis, RWTH Aachen University, 2017.
[21] K. S. D. Beach, arXiv:cond-mat/0403055.
[22] H. Shao and A. W. Sandvik, Phys. Rep. **1003**, 1 (2023).
[23] A. Rényi, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (University of California Press, Berkeley, California, 1961), Vol. 4, pp. 547–562.
[24] E. Lenzi, R. Mendes, and L. da Silva, Physica A **280**, 337 (2000).
[25] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, J. Math. Phys. **54**, 122203 (2013).
[26] J. M. Amigó, S. G. Balogh, and S. Hernández, Entropy **20**, 813 (2018).
[27] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[28] T.-K. L. Wong and J. Zhang, IEEE Trans. Inf. Theory **68**, 5353 (2022).
[29] J. Shore and R. Johnson, IEEE Trans. Inf. Theory **26**, 26 (1980).
[30] J. Skilling, The axioms of maximum entropy, in *Maximum-Entropy and Bayesian Methods in Science and Engineering: Foundations*, edited by G. J. Erickson and C. R. Smith (Springer Netherlands, Dordrecht, 1988), pp. 173–187.
[31] S. N. Karbelkar, Pramana **26**, 301 (1986).
[32] J. Uffink, Stud. Hist. Philos. Sci. Part B **26**, 223 (1995).
[33] Y. Burnier and A. Rothkopf, Phys. Rev. Lett. **111**, 182003 (2013).
[34] P. Jizba and J. Korbel, Phys. Rev. Lett. **122**, 120601 (2019).
[35] Dirichlet distribution, in *A Primer on Statistical Distributions* (John Wiley & Sons, Ltd, Hoboken, New Jersey, 2003), pp. 269–276.
[36] E. M. F. Curado and C. Tsallis, J. Phys. A: Math. Gen. **24**, L69 (1991).
[37] J. Costa, A. Hero, and C. Vignat, in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, edited by A. Rangarajan, M. Figueiredo, and J. Zerubia (Springer, Berlin, 2003), pp. 211–226.
[38] O. Johnson and C. Vignat, Ann. Inst. Henri Poincare, B´ **43**, 339 (2007).
[39] R. Strack and D. Vollhardt, Phys. Rev. B **46**, 13852 (1992).