Extremizing The Expected Energy: Fisher Information, and Shrodinger

Carlos A. Gómez-Uribe
CGOMEZ@ALUM.MIT.EDU

Editor:

Abstract

We consider a quantum-mechanical spinless particle in a one dimensional space under the influence of a smooth but otherwise arbitrary potential, and study its expected energy. The expected energy is a sum of terms, one of which is the variance of momentum. We find that this variance in momentum is directly proportional to the Fisher information matrix of a hypothetical location parameter. We then turn to the problem of extremizing the expected energy, a problem in variational calculus, e.g., to find the ground state and other stationary states. To our surprise, the result is the time-independent Schrodinger equation. In addition, the expected energy extremization problem resembles the problem of variational learning in the context of machine learning, widely used to learn the parameters of useful but complex models (e.g., with hidden variables). Although the two problems are not equivalent, they are related, and further study of this relation could provide useful to both fields.

1. Introduction

Consider a spin-less particle with mass m in a one dimensional space under the influence of a smooth potential. The Hamiltonian operator is

$$\widetilde{H} = \widetilde{V}(\widetilde{X}) + \frac{\widetilde{P^2}}{2m}, \tag{1}$$

where the operator $\widetilde{X}$ is the position, $\widetilde{V}(\widetilde{X})$ is the potential energy, and $\widetilde{P}$ is the particle's momentum.

Let $\rho(\widetilde{x})$ and $\gamma(\widetilde{p})$ denote the probability density functions for the particle's position and momentum. The expected energy is then

$$\langle\widetilde{H}\rangle = \langle\widetilde{V}(\widetilde{X})\rangle + \frac{\langle\widetilde{P^2}\rangle}{2m} = \int \rho(\widetilde{x})V(\widetilde{x})d\widetilde{x} + \int \gamma(\widetilde{p})\widetilde{p}^2 d\widetilde{p}, \tag{2}$$

where $\langle\rangle$ denotes expectation, and the integrals run from $-\infty$ to ∞.

We consider the problem of finding the states that extremize the energy, that is, states for which small perturbations away from it do not change the expected energy up to first order. These states are often called stationary states. Extremizing the expected energy is obviously an optimization problem in the space of functions, i.e., a problem in variational calculus. The motivation to study this problem is to find the state of lowest energy, the so-called ground state, since often other stationary states can be found mechanistically once the ground state is known, and knowing the full set of stationary states allows for a full solution of the system, including its dynamics.

2. Formulating The Optimization Problem

Our first goal is to formulate the optimization problem of extremizing the expected energy in simpler terms and in a way that makes explicit the connection to other problems in machine learning and information theory. We accomplish this by (i) simplifying the equations above by choosing appropriate coordinates, and (ii) expressing the expectations of functions of momentum in terms of position using the well-known relationship. We find that the variance in momentum is the Fisher information matrix in disguise, which leaves much to think about.

Our second goal is to show that solutions of the resulting optimization problem are exactly the solutions to the stationary Schrodinger equation. In other words, the stationary Schrodinger equation is derived as a consequence of extremizing the expected energy. I think this is really cool, but don't know what to make of it. More initial thoughts about this in the discussion.

2.1 Changing Coordinates

Consider the function $\widetilde{V}(\widetilde{x})$ corresponding to the operator $\widetilde{V}(\widetilde{X})$ above. Let $\widetilde{x_o}$ be a (potentially local) minimum of $\widetilde{V}(\widetilde{x})$, and $\widetilde{V_o} = \widetilde{V}(\widetilde{x_o})$ be the value of the potential at this minimum. Then $\widetilde{V}'(\widetilde{x_o}) = 0$, where the prime denotes the first spatial derivative, and the second spatial derivative satisfies $\widetilde{V}''(\widetilde{x_o}) > 0$.

Now we make the change of coordinates to make $\widetilde{x_o}$ the new origin, i.e., $x = \widetilde{x} - \widetilde{x_o}$. We also redefine the potential in terms of this new coordinate, and also subtract $\widetilde{V_o}$ from it, i.e., $U(x) = \widetilde{V}(x + \widetilde{x_o}) - \widetilde{V_o}$. Of course, $U(0) = U'(0) = 0$ and $U''(0) > 0$ by construction. The Hamiltonian then becomes $U(X) + \frac{\widetilde{P^2}}{2m} + \widetilde{V_o}$.

In quantum mechanics (QM), position and momentum are dependent quantities, but as is well-known and may be shown below, shift position by a constant, as we have, does not change the probability density of momentum. Similarly, shifting the momentum coordinates by a constant leaves $\rho(x)$ unchanged. So let's do that. Let $\mu_{\widetilde{p}}$ denote the expected value of $\widetilde{p}$, and define $p = \widetilde{p} - \mu_{\widetilde{p}}$. Note that $\langle P \rangle = 0$, and $\sigma_p^2 = \sigma_{\widetilde{p}}^2$, where σ_a^2 denotes the variance of quantity a.

Lastly, we express the Hamiltonian in our new coordinates, and scale it by $8m/\hbar^2$, where $\hbar$ is Planck's constant. The factor of 8 will make things look pretty later on:

$$H(X,P) = \frac{8m}{\hbar^2}\widetilde{H}(X + \widetilde{x_o}, P + \mu_{\widetilde{p}}) = \frac{8m}{\hbar^2}U(X) + \frac{4}{\hbar^2}(P^2 + 2\mu_{\widetilde{p}}P + \mu_{\widetilde{p}}^2) + \frac{8m}{\hbar^2}\widetilde{V_o}. \tag{3}$$

We now let $V(X) = \frac{8m}{\hbar^2}U(X)$ denote the effective potential in our new Hamiltonian. The expected energy for H is

$$\langle H \rangle = \langle V(x) \rangle + \frac{4}{\hbar^2}\sigma_p^2 + \frac{4}{\hbar^2}\mu_{\widetilde{p}}^2 + \frac{8m}{\hbar^2}\widetilde{V_o}. \tag{4}$$

Now, since changing $\mu_{\widetilde{p}}^2$ does not affect σ_p^2 nor $\rho(x)$, then all extrema of $\langle H \rangle$ must have $\mu_{\widetilde{p}}^2 = 0$. Also, $\frac{8m}{\hbar^2}\widetilde{V_o}$ is just a constant. Taking this into account, then our optimization problem is extremizing $\langle V(x)\rangle + \frac{4}{\hbar^2}\sigma_p^2$. As will become evident below, the ground state is the global minimizer of $\langle V(x)\rangle + \frac{4}{\hbar^2}\sigma_p^2$, while other stationary states are just extrema of the same expression (I suppose we may only be interested in the local minima and not maxima, since the latter will be unstable?).

Finally, letting $\rho_n(x)$ denote the distribution of the $n-th$ stationary state, with $\rho_o(x)$ the ground state, our optimization problem is:

$$\rho_o(x) = \underset{\rho\in\mathcal{D}}{\operatorname{argmin}}\{\frac{4}{\hbar^2}\sigma_p^2 + \langle V(x)\rangle\}, \tag{5}$$

where $\mathcal{D}$ is the space of probability distributions. While this already looks cleaner, it is still not a well posed problem because we are optimizing over distributions for x but the optimization objective has a term proportional to the variance in momentum. Our next goal is to re-express this variance in terms of x.

2.2 The Fisher Information And Momentum

The Fisher information turns out to be a world in itself, defining the efficiency limits of unbiased estimators in Statistics, and showing up often in algorithms to learn models from data, including the so-called natural gradient, and a variety of Kalman filters. However, it is lesser known and used than the Shannon information, probably because it is the Shannon entropy that underpins the communication technology of our times. The latter is obviously related to the entropy in Thermodynamics. That said, much of information theory can be and to some extent has been analogously developed for the Fisher information. At the same time, there are several known relations between the Fisher information and the Shannon entropy. It turns out there is also a whole subfield focused on deriving much of Physics from Fisher information, e.g., Frieden (2004). As far as I can tell, however, this work has not yet studied the problem of extremizing the expected energy in QM. But they have studied other problems that end up having similar attributes that will help us below.

Anyway, the Fisher information $J(\theta)$ of a continuous probability distribution $\rho(x)$ that depends on parameter θ is

$$J(\theta) = \int \rho(x)|\frac{\partial}{\partial\theta}\log\rho(x)|^2 dx = \int \frac{1}{\rho(x)}\left(\frac{\partial\rho(x)}{\partial\theta}\right)^2 dx. \tag{6}$$

It is the average of a function of the smoothness of the distribution, where smoothness is measured based on the magnitude (squared) of the gradient with respect to the parameter.

Now consider a distribution that admits a location parameter, i.e., of the form $\rho(x-\theta)$, Then the gradient with respect to θ is just the negative gradient with respect to x evaluated at $\theta = 0$, so the integrand above can be re-expressed purely in terms of x, making $J(\theta)$ a function of $\rho(x)$ alone. Specifically, for such distributions the Fisher information becomes

$$J(x) = \int \rho(x)|\nabla\log\rho(x)|^2 dx = \int \frac{1}{\rho(x)}\left(\nabla\rho(x)\right)^2 dx, \tag{7}$$

where $\nabla = \frac{\partial}{\partial x}$. Check out Zegers (2015) for a helpful introduction and summary of results around the Fisher information, including some information theory results based on it.

Back to Quantum Mechanics now. It turns out that the variance in momentum is a linear function of the Fisher information $J(x)$. There are at least two ways to see this. The first one writes $\gamma(p)$ in terms of the probability amplitudes $\psi(p)$, i.e., $\gamma(p) = \psi^2(p) = \psi^*(p)\psi(p)$, and then uses the fact that $\psi(p)$ is the Fourier transform of $\psi(x)$. We use * to denote complex conjugation. Some calculus including wrestling with delta functions gives the desired result. However, there is a much shorter path that makes use of another basic relationship in QM that is unfortunately less intuitive for those new to QM. We'll follow this less intuitive shorter path here to keep this note brief.

In QM, the momentum operator P in the position basis is proportional to the spatial gradient, i.e., $P = -i\hbar\frac{\partial}{\partial x}$, where $i = \sqrt{-1}$. Writing $\rho(x) = \psi^2(x)$ in terms of the probability amplitude $\psi(x)$, also called the wave function, we find that

$$\begin{aligned}
\langle P^2\rangle &= \int \rho(x) P^2 dx = \int \psi^*(x)(-i\hbar\frac{\partial}{\partial x})^2\psi(x)dx \\
&= -\hbar^2\int \psi^*(x)\frac{\partial^2}{\partial x^2}\psi(x)dx = \hbar^2\int \frac{\partial}{\partial x}\psi^*(x)\frac{\partial}{\partial x}\psi(x)dx \\
&= \hbar^2\int \rho(x)\big(\nabla\log\psi(x)\big)^2 dx = \langle\big|\nabla\log\psi(x)\big|^2\rangle \qquad (8)\\
&= \frac{\hbar^2}{4}\int \rho(x)\big[\big(\nabla\log\rho(x)\big)^2 + 4\mathrm{Im}\{\nabla\log\psi(x)\}^2\big]dx \\
&= \frac{\hbar^2}{4}J(x) + \hbar^2\left\langle|\mathrm{Im}\{\nabla\log\psi(x)\}|^2\right\rangle. \qquad (9)
\end{aligned}$$

The second equality in the second line follows from integration by parts and assuming that the gradient of the wavefunctions vanish at plus and minus infinity. The equality in the fourth line follows from $\big(\nabla\log\rho(x)\big)^2 = 2[\big(\nabla\log\psi(x)\big)^2 + \mathrm{Re}\{\nabla\log\psi(x)\}^2 - \mathrm{Im}\{\nabla\log\psi(x)\}^2]$. Notice that $\langle|\mathrm{Im}\{\nabla\log\psi(x)\}|^2\rangle \geq 0$, so $\sigma_p^2 = \langle P^2\rangle \geq \frac{\hbar^2}{4}J(x)$, i.e., the Fisher information provides a lower bound for σ_p^2 that is tight for real wave functions $\psi(x)$.

Substituting Eq. 9 into Eq.5 we find that

$$\rho_o(x) = \underset{\rho\in\mathcal{D}}{\mathrm{argmin}}\{J(x) + \langle V(x)\rangle + 4\left\langle|\mathrm{Im}\{\nabla\log\psi(x)\}|^2\right\rangle\}. \qquad (10)$$

Now, let's Taylor expand the potential about zero

$$V(x) = \sum_{j=2}^{\infty} v_j x^j, \qquad (11)$$

where $v_j = \frac{1}{j!}\frac{\partial^j V(x)}{\partial x^j}\big|_{x=0}$ are just the Taylor coefficients, and $v_0 = v_1 = 0$ by construction. Letting $\mathcal{L}(x) = J(x) + \langle V(x)\rangle$, the optimization becomes

$$\begin{aligned}
\rho_o(x) &= \underset{\rho\in\mathcal{D}}{\mathrm{argmin}}\{\mathcal{L}(x) + 4\left\langle|\mathrm{Im}\{\nabla\log\psi(x)\}|^2\right\rangle\} \\
&= \underset{\rho\in\mathcal{D}}{\mathrm{argmin}}\{J(x) + \sum_{j=2}^{\infty} v_j\left\langle x^j\right\rangle + 4\left\langle|\mathrm{Im}\{\nabla\log\psi(x)\}|^2\right\rangle\}, \qquad (12)
\end{aligned}$$

where the $\langle x^j \rangle$ are the central moments of $\rho(x)$.

The appearance of the term in the optimization objective that depends on $\psi(x)$ rather than $\rho(x)$ is unfortunate. However, for at least some systems, like the simple harmonic oscillator ($V(x) = x^2$), the wave functions $\psi_n(x)$ are real, so $\langle |\mathrm{Im}\{\nabla \log \psi(x)\}|^2 \rangle = 0$, making the above optimization problem well posed. But in general, for what systems are the wave functions $\psi_n(x)$ real?

In any case, it is still worth thinking about the optimization problem below, which extremizes a lower bound of the expected energy.

$$\rho_o(x) = \underset{\rho \in \mathcal{D}}{\mathrm{argmin}} \{J(x) + \langle V(x) \rangle\} = \underset{\rho \in \mathcal{D}}{\mathrm{argmin}}\, \mathcal{L}(x). \tag{13}$$

In words, the optimization problem in Equation 13 seeks the distribution that minimizes the sum of the Fisher information and the expected potential energy. This problem is very similar in spirit to problems encountered in statistics and machine learning, particularly the problem of variational inference in probabilisitic graphical models. The main difference between those problems and the one above is that while we want to minimize the Fisher information, they need to maximize the Shannon differential entropy; see, e.g., chapters 3 and 4 in Wainwright et al. (2008). Both optimization problems, at a high level, then aim to find the distribution that maximizes the uncertainty in x given constraints that depend on the potential function. And in at least one problem, the case of the SHO or a quadratic potential, give the same result. We'll try to paint a fuller picture about this in the Discussion. First, let's try to solve the optimization problem in 13.

2.3 Example

To gain some intuition, let's look at an approximation of the optimization problem in Eq. 13 by restricting the space of distributions we optimize over to be the space of Gaussian distributions. It can be easily shown that $J(x) = \frac{1}{\sigma^2}$ for a Gaussian with variance σ^2, so $\mathcal{L}(x) = \frac{1}{\sigma^2} + \langle V(x) \rangle$. So the two terms in the objective are in tension with each other: minimizing $J(x)$ means making σ^2 large, while minimizing $\langle V(x) \rangle$ means making σ^2 small. The latter follows directly from $x = 0$ being a minima of $V(x)$, so $\langle V(x) \rangle \approx \frac{\sigma^2}{2} V''(0)$ by Taylor expansion, and the approximation being exact for the quadratic potential of the simple harmonic oscillator. To extremize $\mathcal{L}(x)$, we can then set $\frac{d\mathcal{L}}{d\sigma^2} = 0$ which gives $\sigma_o^2 = \sqrt{\frac{2}{V''(0)}} = 1/J_o(x)$, and $\mathcal{L}_o(x) = \frac{1}{2}\sqrt{8V''(0)}$. This corresponds to an rescaled energy (multiply by $\hbar^2/8m$) of $\langle H_o \rangle = \frac{\hbar^2}{2m}\sqrt{\frac{V''(0)}{8}}$.

In the general problem, $\langle V(x) \rangle \geq V(0)$ for distributions "centered" close to $x = 0$. Because of the local convexity at $x = 0$, a local minimum of the potential, starting with a variance in x of zero, $\langle V(x) \rangle$ at least initially grows as the variance in x increases. So we expect that minimizing $J(x)$ and $\langle V(x) \rangle$ to be in tension with each other, and the minimizer of $\mathcal{L}(x)$ to always have a positive variance.

3. Solving The Optimization Problem

Let's massage the objective to make it amenable to variational calculus.

$$\begin{aligned}\mathcal{L}(x) &= J(x) + \langle V(x)\rangle \\ &= \int\Big(\frac{1}{\rho(x)}(\rho'(x))^2 + \rho(x)\sum_{j=2}^{\infty} v_j x^j + \alpha\big(\rho(x)-1\big)\Big)dx \qquad (14)\\ &= \int \ell\big(\rho(x),\rho'(x)\big)dx. \qquad (15)\end{aligned}$$

where we introduced the Largange multiplier α that guarantees $\rho(x)$ is normalized. Extremizing the above is a standard variational calculus problem, solved by setting the first variation to zero. This amounts to consider small perturbations $\delta(x)$ about $\rho(x)$ that are zero at plus and minus infinity, and that integrate to zero, so that $\rho(x)+\delta(x)$ is also a distribution. The result is the Euler-Lagrange (EL) equation for $\rho(x)$, namely $\frac{d}{dx}\big(\frac{\partial \ell}{\partial \rho'}\big) = \frac{\partial \ell}{\partial \rho}$. See, e.g., Gelfand et al. (2000) for more about calculus of variations. Simple algebra gives:

$$\begin{aligned}\frac{\partial \ell}{\partial \rho} &= \alpha + \sum_{j=2}^{\infty} v_j x^j - (\nabla \log \rho)^2\\ \frac{\partial \ell}{\partial \rho'} &= 2\nabla \log \rho\\ \frac{d}{dx}\frac{\partial \ell}{\partial \rho'} &= 2\nabla^2 \log \rho,\end{aligned}$$

Plugging these into the EL equation finally gives a differential equation for our solution

$$2\nabla^2 \log \rho + \big(\nabla \log \rho\big)^2 = \alpha + \sum_{j=2}^{\infty} v_j x^j. \qquad (16)$$

This Equation is identical to Eq. (16) in Frieden et al. (1999), a study of Thermodynamics and Fisher information. Equation 16 turns out to be a Ricatti equation [1], but more importantly, can easily be expressed in terms of the wave function. First note that

$$\nabla^2 \log \rho(x) = 2\Big(\frac{\nabla^2\psi(x)}{\psi(x)} - \frac{(\nabla\psi(x))^2}{\psi^2(x)}\Big).$$

Substituting this into 16 and a bit of algebra yields

$$\Big(\frac{\partial^2}{\partial x^2} - \frac{1}{4}\sum_{j=2}^{\infty} v_j x^j\Big)\psi(x) = \frac{\alpha}{4}\psi,$$

1. The change of variables $\nu(x) = \nabla \log \psi(x)$ puts it in the standard Ricatti equation form $\nu'(x) + \nu^2(x) = 1/4(\alpha + \sum_{j=2}^{\infty} v_j x^j)$.

or rescaling by $\frac{-\hbar^2}{2m}$ to go back to our original energy scale:

$$\Big(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \frac{\hbar^2}{8m}\sum_{j=2}^{\infty} v_j x^j \Big)\psi(x) = -\frac{\alpha\hbar^2}{8m}\psi(x) \text{ or} \tag{17}$$

$$\Big(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + U(X) \Big)\psi(x) = E\psi(x). \tag{18}$$

I find this connection, despite its remaining gaps, both interesting and surprising. If correct, it means that extremizing a lower bound of the expected energy, using the Fourier relation between position and momentum space, is enough to derive Schrodinger. Alternatively, one can argue that interpreting the variance of momentum as the Fisher information already essentially assumes Shrodinger, so the result above should be obvious. That said, it seems plausible to directly motivate the optimization objective: extremize the expected potential energy while maximizing uncertainty. But why measure uncertainty in terms of the Fisher information?

4. Maximizing Shannon Entropy

One can imagine conceptually solving Eq. 13 by keeping $\langle V(x) \rangle = V$ fixed and minimizing $J(x)$ given $\langle V(x) \rangle = V$ using variational calculus, and then evaluating $J(x)$ for the resulting distribution, making $\mathcal{L}$ a function of the scalar parameter V, which can then be optimized through regular calculus. The first part of the problem is very similar to the problem often encountered in statistics of maximizing the Shannon differential entropy of x given the moment-based constraint $\langle V(x) \rangle = \sum_{j=2}^{\infty} v_j \langle x_j \rangle = V$. This turns out to be an easy variational calculus problem, as shown next, solved by an exponential family distribution.

The (Shannon differential) entropy is defined as

$$h(x) = -\int \rho(x) \log \rho(x) dx. \tag{19}$$

There are several relations between the Fisher information and the Shannon entropy, including

$$\lim_{t\to 0} \frac{\partial}{\partial t} h\big(X + \mathcal{N}(0,t)\big) = \frac{1}{2} J(x). \tag{20}$$

In words, the change in entropy when adding a zero-mean Gaussian with small variance t is directly proportional to the Fisher information. See Chapter 16.6 in Cover and Thomas (2012).

Let's consider what happens when we replace minimization of $J(x)$ with maximization of $h(x)$. These problems are solved by the same distribution in at least a few cases. We first consider these problems in the space of distributions that are only non-zero in an interval, say when $x \in [-1, 1]$ with loss of generality. Minimizing $J(x)$ is evidently accomplished by the uniform distribution in $[-1, 1]$, since then $\nabla \log \rho(x) = 0$ for all x in the interval, and $J(x) = 0$. Since we know $J(x) \geq 0$, then the uniform distribution is a minimizer. Similarly, it is well known that the uniform distribution maximizes $h(x)$.

Now consider the actual problem we care about

$$\rho_o(x) = \underset{\rho\in\mathcal{D} \text{ s.t. } \langle V(x)\rangle = V}{\operatorname{argmax}} h(x). \tag{21}$$

This problem is identical to the variational inference optimization problem in machine learning. See chapters 3 and 4 in Wainwright et al. (2008) for on variational inference in machine learning.

We set up the Lagrangian objective

$$\mathcal{L}_\langle(x) = \int\Big(-\rho(x)\log\rho(x) - \lambda\big(\rho(x)-1\big) - \gamma\big(V(x)-V\big)\Big)dx = \int \ell_h(\rho)dx, \tag{22}$$

and use it to get the corresponding Euler Lagrange equation, which simplifies to

$$\begin{aligned}\frac{\partial \ell_h}{\partial\rho} &= -\log\rho - 1 - \lambda - \gamma V(x) = 0\\ \rho_o(x) &= \frac{1}{Z}e^{-\gamma V(x)},\end{aligned} \tag{23}$$

since ℓ_h is independent of $\rho'(x)$. The optimizing distribution is the Boltzmann or Gibbs distribution, a member of the so-called exponential family in statistics.

Note that when $V(x) = x^2$, $\rho_o(x)$ is a zero-mean Gaussian distribution.

Now, consider

$$\rho_o(x) = \underset{\rho\in\mathcal{D} \text{ s.t. } \langle V(x)\rangle = V}{\operatorname{argmax}} J(x), \tag{24}$$

with corresponding objective

$$\mathcal{L}_\{(x) = \int\Big(\frac{\rho'^2(x)}{\rho} - \alpha\big(\rho(x)-1\big) - \beta\big(V(x)-V\big)\Big)dx = \int \ell_f(\rho)dx, \tag{25}$$

The Euler-Lagrange equation for the extremizer, as we saw above, is just

$$2\nabla^2\log\rho + \big(\nabla\log\rho\big)^2 = \alpha + \beta V(x). \tag{26}$$

In particular, when $V(x) = x^2$, the extremizer is a zero-mean Gaussian distribution. So this is another example when the two problems have the same solution. In general, however, it seems like the two problems have different solutions. For example, we know that Equation 23 solves the maximum entropy problem. For that distribution, $\nabla\log\rho_o = -\gamma V'(x)$, so substitution into Eq. 26 yields

$$-2\gamma V''(x) + \gamma^2\big(V'(x)\big)^2 = \alpha + \beta V(x). \tag{27}$$

which is obviously not generally satisfied for arbitrary potentials $V(x)$. But how far are the optimizers of the two problems, in general? If close enough, one might then search for solutions for Eq. 13 in the space of Boltzmann distributions, for which $J(x) = \gamma^2\left\langle V'^2(x)\right\rangle$, i.e., replacing the problem in Eq. 13 by the simpler problem

$$\rho_o(x) = \underset{\rho \in \mathcal{D}_\mathcal{B}}{\text{argmin}} \{\gamma^2 \langle V'^2(x)\rangle + \langle V(x)\rangle\}, \tag{28}$$

where $\mathcal{D}_\mathcal{B}$ is the space of Boltzmann distributions of the form in Eq. 23, which can be indexed by the single parameter γ, i.e., this problem is a simpler search in a one dimensional space. E.g., for a given γ, the expectations in 28 can be computed via Monte Carlo methods and then the optimal γ found through gradient climbing.

5. Discussion

- For which systems is the lower bound used in Equation 13 tight, i.e., for which systems do we expect the stationary wave functions to be real? If not generally true, then let's reformulate our optimization problem directly in terms of ψ and approach using calculus of variations again.
- More generally, it seems worthwhile revisiting variational inference in machine learning but adding or replacing the Shannon-derived concepts there with Fisher information derived ones.

Acknowledgments

CGU wants to thank Meltem Tolunay and Yanjun Han for their encouragement to pursue this direction, and some useful initial information theory pointers and ideas.

References

Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

B Roy Frieden. *Science from Fisher information: a unification*. Cambridge University Press, 2004.

BR Frieden, A Plastino, AR Plastino, and BH Soffer. Fisher-based thermodynamics: Its legendre transform and concavity properties. *Physical Review E*, 60(1):48, 1999.

Izrail Moiseevitch Gelfand, Richard A Silverman, et al. *Calculus of variations*. Courier Corporation, 2000.

Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

Pablo Zegers. Fisher information properties. *Entropy*, 17(7):4918–4939, 2015.