A Probabilistic Approach to Pattern Matching in the Continuous Domain

Daniel Keren[*]

Computer Science Department

University of Haifa

Haifa, Israel 32000

dkeren@cs.haifa.ac.il

Michael Werman

School of Engineering

and Computer Science

Hebrew University of

Jerusalem

Jerusalem 91904 Israel

werman@cs.huji.ac.il

Joshua Feinberg

Physics Department

University of Haifa at Oranim

Tivon 36006, Israel

and

Physics Department, Technion

Haifa 32000, Israel

joshua@physics.technion.ac.il

Abstract

The goal of this paper is to solve the following basic problem: given discrete noisy samples from a continuous signal, compute the probability distribution of its distance from a fixed template. As opposed to the typical restoration problem, which considers a single optimal signal, the computation of the entire probability distribution necessitates integrating over the entire signal space. To achieve this, we apply path integration techniques. The problem is studied in one and two dimension, and an accurate solution as well as an efficient approximation scheme are provided.

1 Introduction

A fundamental question in pattern matching is: given a fixed template $y(t)$ and a signal $x(t)$ (usually, for the sake of notational simplicity, t will be suppressed, as will dt when the integration is over t), we wish to compare x with y. Typically, the information on x is sparse and noisy. Given this information, and for every threshold β, we wish to calculate the probability that the distance between x and y is larger than β – that is, the cumulative

---

[*]Corresponding author.

distribution of the distance $d(x, y)$ between x and y. Here we use the standard L^2 norm, so the distance is defined as

$$d(x, y) \triangleq \int (x - y)^2.$$

A special and very important case is $y = 0$; the question then reduces to computing the distribution of $\int x^2$, or the *energy* of x, which is arguably the most fundamental property of a signal.

If the entire signal x can be measured with absolute accuracy, or is given in closed form, one can exactly compute $d(x, y)$. Alas, typically the information on the signal is sparse and noisy, and $d(x, y)$ cannot be computed. Since usually the ground truth signal cannot be reconstructed, it is clear that the correct reply is probabilistic in nature, and allows to answer questions such as: given some measurements of x, what is the probability that its distance from y is larger than a certain threshold? Generally we would like to compute, given the sparse and noisy measurements, the probability that $d(x, y)$ lies within a certain range. Stated otherwise, a probability on signals x induces a probability on $d(x, y)$. We propose a method for computing this probability. First, a probability distribution is defined over the space of possible signals, and then the probability volume of all signals x for which $d(x, y)$ is larger than a given threshold is computed. To this end we adapt a technique known as *path integration* or *functional integration*.

The paper proceeds as follows. First, the well-known regularization method is briefly reviewed, as well as the probability measure it induces on the space of signals. In Section 2 the problem is formally defined and previous work reviewed. Section 3 reviews path integration, which allows to integrate over the infinite-dimensional signal space. In Sections 4-6 we study three problems of increasing difficulty, where Sections 4 and 5 discuss problems which are both important on their own right, and also include preparatory calculations for their preceding sections: Section 4 discusses the distribution of the signal's value at a point, and in Section 5 the distribution of the signal's energy is derived. In Section 6 the distribution of $d(x, y)$ is calculated. Section 7 discusses a very general solution to the problem, and extends it to handle functions in two variables. We show that for functions with more than one variable a divergence problem occurs, and offer how to remedy it. In Section 8 we present an approximation scheme and analyze its accuracy. Conclusions are offered in Section 9, followed by two appendices in which some technical issues are elaborated.

1.1 Regularization and Priors on Functions

The problem of restoring a continuous signal $x(t)$, given discrete samples $x(t_i)$, is one of the most well-studied in signal processing. A very popular restoration technique is *regularization,* which seeks a tradeoff between data fidelity and smoothness. Work on regularization abounds; some references are [25, 24, 5, 16, 12]. The regularization paradigm proceeds as follows: first, a class of "admissible functions" is defined, the members of which are the possible solutions (we shall abide here with the common terminology and use "functions" instead of "signals"). In order to allow as much flexibility as possible, typically the class of admissible functions is very large (e.g. an infinite-dimensional function space). Then, a *cost functional $M(x)$* is defined for every admissible function x by $M(x) = D(x) + \lambda S(x)$, where $D(x)$ measures the deviation of x from the samples, $S(x)$ measures the roughness of x, and λ is a positive parameter. The function x minimizing M can be found by variational methods; denote it x_0. D is defined as follows: if the measured values of the function at a set of points $\{t_i\}_{i=1}^l$ (from now on called *sampling points*) are denoted by $\{x_i\}_{i=1}^l$, then $D(x) = \sum_{i=1}^{l} \frac{1}{2\sigma_i^2} [x(t_i) - x_i]^2$ where σ_i is the magnitude of the measurement noise at t_i (for simplicity it will be assumed hereafter that $2\sigma_i^2 = 1$), and $S(x) = \int \dot{x}^2$ (the *first order smoothness term*), where $\dot{x}$ denotes the derivative of x by t^1. The *total variation* smoothness term, which is also often used, is defined by $\int |\dot{x}|$ [2, 3].

It will be assumed hereafter that the functions are always defined on the unit interval, so all the integrals on $x, \dot{x}$ etc. are from 0 to 1. We shall assume so-called *natural (Dirichlet) boundary conditions,* $x(0) = x(1) = 0$ (these conditions are not crucial to the following derivations, although they make them a little simpler. In practice, the boundary conditions can either be dropped, or imposed at the endpoints of a very long interval, thus having negligible effect).

Quite a while after this variational approach was introduced in the context of regularization [25], the seminal work [6] anchored it in a probabilistic setting. According to Bayes' rule –

$$\Pr(x|\mathbf{Data}) = \frac{\Pr(\mathbf{Data}|x)\Pr(x)}{\Pr(\mathbf{Data})}$$

where x is the function in question and **Data** are the samples. When Gaussian error is

---

[1] Sometimes the second derivative $\ddot{x}$ is used instead of $\dot{x}$, leading to the so-called *second order smoothness term.*

assumed in the measurement process, the following holds (up to the obvious normalization)

$$\Pr(\mathbf{Data}|x) = \exp\left(-\sum_{i=1}^{l}[x(t_i) - x_i]^2\right). \tag{1}$$

A crucial question is how to define the a-priori probability of a function ("prior on the functions" in the Bayesian terminology). [10, 27] address this question for images, [19, 18] for acoustic signals, and [13] for video. Here the following definition is used:

$$\Pr(x) = \exp\left(-\lambda \int \dot{x}^2\right) \tag{2}$$

where λ is a positive constant. This means that a-priori, the smoother the function, the more probable it is. This probability distribution resembles the well-known *Boltzmann distribution* in statistical physics, and its application to signals was rigorously justified in [6]. In [7, 14, 15, 13, 18], an empirical analysis of such prior probabilities on signals demonstrated a very good match between them and reality.

Combining Eqs. 1,2 yields

$$\Pr(x|\mathbf{Data}) \propto \exp(-M(x)), \quad M(x) = \sum_{i=1}^{l}[x(t_i) - x_i]^2 + \lambda \int \dot{x}^2. \tag{3}$$

This justifies – in hindsight – the choice of the solution x_0 as the function which minimizes M. This probabilistic approach will be adopted here, and the a-priori probability of x will be assumed to equal $\exp(-M(x))$.

The advantages of this Bayesian approach are far more important than justifying previous work. Typically, one is not interested only in the value of the optimal solution, but also in an estimate of how reliable it is – e.g. it is very desirable to place confidence intervals around the solution's values. Assigning a probability to every admissible function enables a global view in which *every* function, not just the optimal one, plays a role, and allows to assign a measure of uncertainty to various solution parameters. Conceptually, the optimal solution x_0 can be viewed as the expectation, and the uncertainty measure as the variance, of the distribution of some signal characteristic.

The goal here is to compute the probability distribution of the signal's distance from a template. This problem is next rigorously defined and solved.

2 The Problem

Given discrete and noisy measurements of a signal x, and a fixed signal y, we seek to compute $\Pr(\int (x - y)^2 \geq \beta)$, for every β. Hereafter this probability will be denoted $\Pr(y, \beta)$. This is equivalent to computing the probability density function (p.d.f) of $\int (x-y)^2$. The case $y = 0$ is of considerable interest, as it measures the distribution of the signal's energy. With some minimal abuse of notation, we shall denote the probability in this case by $Pr(E, \beta)$.

2.1 Previous Work

Previously the uncertainty of linear functionals was computed, given a prior on the function space [23, 26, 7]. It turns out [9, 11, 7] that under the assumption that $\Pr(x) = \exp(-Q(x))$, where $Q(x)$ is a positive definite quadratic functional, then for every linear functional L defined on the space of admissible functions, $f \to L(f)$ is a normal random variable, and therefore it is possible to compute $\Pr(L(f) \geq \beta)$ (the Q's used very much resemble the cost functional M of Eq. 3). Typical L's were the value of x or its derivative at some point. Since $L(f)$ is normal, its entire p.d.f is characterized by its average and variance, which can be calculated as described in [7]. Here we propose to tackle the considerably more difficult problem of the *quadratic* functional $\int (x - y)^2$, whose distribution is not normal. The method presented here can be applied to other quadratic functionals.

2.1.1 Computing the p.d.f vs. the Restoration Problem

There is a great deal of work on restoring signals from sparse, noisy samples (see Section 1.1). Restoration can be viewed as choosing the "best" (i.e. maximum likelihood) function from the entire function space. Here, we seek to compute the p.d.f of a non-linear functional on the function space, which necessitates considering the *entire* space. The restored function is just one point in this immense space, and it is useless as far as computing the p.d.f – or even only the expectation of the non-linear functional – is considered. To see this, assume we want to estimate the expectation of $\int x^2$, given sparse measurements of x. Assume further that all these measurements are zero. Clearly, the restored function is the zero function, and it satisfied $\int x^2 = 0$. Alas, it is also clear that the expectation of $\int x^2$ is non-zero (this expectation is explicitly computed in Section 7.1). Hence, even for the simpler problem of computing only the expectation of the random variable $\int x^2$, the

5

restored function is useless – not to mention the far more general problem of computing the entire p.d.f of this random variable.

2.1.2 Gaussian Processes

The theory of Gaussian processes [17] considers a family of random variables, such that each sample of a finite number of them obeys a Gaussian distribution. Indeed, the point values of a function with a Gaussian prior form a Gaussian process [7], but here we seek to study the p.d.f of a functional which is 1) non-linear (and non-Gaussian), and 2) defined by the entire set of the function's values, not a finite sample.

3 A (very) Brief Review of Path Integrals

The tool we apply builds on the theory of *path integrals*, extensively used in theoretical physics, for example statistical mechanics, quantum mechanics, and quantum field theory. For an excellent introduction, see [20]. The probability we seek to compute is

$$\Pr(y, \beta) = \Pr\left(\int (x-y)^2 \geq \beta\right) = \frac{\int \mathcal{D}x \,\theta[\int (x-y)^2 - \beta] \Pr(x)}{\int \mathcal{D}x \Pr(x)} \tag{4}$$

where θ is the Heaviside step function, equal to 1 when the argument is positive and zero otherwise. This will exactly capture the probability of the functions x for which $\int (x-y)^2 \geq \beta$. The functional measure $\mathcal{D}x$ denotes that the integral is computed over the entire space of admissible functions.

3.1 Why is the Path Integral Necessary?

Before proceeding with the introduction and calculation of path integrals, we offer some motivation as to why their usage here is necessary. Let us look at the problem mentioned in Section 2.1.1, which resembles the general one we seek to solve, but is quite simpler: all the measured x_i are equal to zero. This means that the probability of a function x is, up to some normalizing factor, equal to $\exp\left(-\lambda \int \dot{x}^2\right)$. As in Section 2.1.1, we restrict ourselves to the problem of computing the expectation of $\int x^2$. Assume that our function space consists only of second degree polynomials $p_2(t)$ (a linear function has to be identically zero because of the boundary conditions). Since $p_2(0) = p_2(1) = 0$, we must have $p_2(t) = \alpha t(1-t)$, which means that our function space is indexed by a single

parameter α. Since for $x = \alpha t(1 - t)$ we have $\int \dot{x}^2 = \frac{\alpha^2}{3}$ and $\int x^2 = \frac{\alpha^2}{30}$, the expectation of $\int x^2$ equals

$$\frac{\int d\alpha \, \left(\frac{\alpha^2}{30}\right) \exp\left(-\frac{\alpha^2}{3}\right)}{\int d\alpha \exp\left(-\frac{\alpha^2}{3}\right)}$$

which can be computed in closed form. Naturally, the space of parabolas is far too restricted for real application; so let's move on to cubics. The space of cubics which satisfy the boundary conditions is of dimension two, and can be written as $\alpha c_1(t) + \beta c_2(t)$, where $c_1(t), c_2(t)$ span the space of cubics which satisfy the boundary conditions (e.g. $c_1(t) = t(1-t)^2, c_2(t) = t^2(1-t)$). Proceeding as for the quadratic case, the expectation of $\int x^2$ can be written as the quotient of two double integrals over α and β. As we continue to increase the class of allowable functions we are led to expressions which can be written as a quotient of integrals over domains whose dimension tends to infinity – and it is exactly this type of integrals which path integration techniques enable to compute.

3.2 Some Intuition

The path integral may appear formidable, as the domain of integration is infinite-dimensional. Perhaps the easiest way to intuitively comprehend its meaning is to view it as an extension of integrals over $\mathcal{R}^n$, in which vectors are replaced by functions, and matrices by linear operators. The similarity is especially evident if one considers a function as a vector of infinite length, whose components are the function's values at every point, and linear operators as "$\infty \times \infty$" matrices which can be "multiplied" – in the usual manner – with the infinitely long vectors. In Section 3.3 this similarity will be further elucidated for the case of Gaussian integrals.

3.3 Calculating Path Integrals

In most cases, it is not possible to exactly calculate the path integral (one famous exception being the *harmonic oscillator path integral* [20]). One type of path integral which can in some cases be explicitly evaluated is

$$\int \mathcal{D}x \exp\left(-\int L(x, \dot{x})\right) = \int \mathcal{D}x \exp\left(-\int \left(\lambda \dot{x}^2 + b(t)x^2 + c(t)x\right)\right). \qquad (5)$$

The quadratic functional $L(x, \dot{x}) = \lambda\dot{x}^2 + b(t)x^2 + c(t)x$ is known in physics as the *Euclidean time Lagrangian* (hereafter simply the Lagrangian). To calculate this path integral, first

the following *Euler-Lagrange* differential equation should be solved

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 2\lambda\ddot{x} + 2b(t)x + c(t) = 0 \tag{6}$$

call the solution x_{cl} (in physics, this is the *classical solution* to the dynamics of the system, or the "classical path"). The integral in Eq. 5 then equals

$$\exp\left(-\int\left(\lambda\dot{x}_{cl}^2 + b(t)x_{cl}^2 + c(t)x_{cl}\right)\right)\int\mathcal{D}x\exp\left(-\int\left(\lambda\dot{x}^2 + b(t)x^2\right)\right) \tag{7}$$

This is best understood by making an analogy with the ordinary Gaussian integral, which is computed by completing the square in the exponent (just as the ordinary quadratic form is expanded around its minimum to obtain a pure quadratic, the quadratic functional is expanded around its minimum (the classical solution) and the integral is reduced to a simpler, purely quadratic functional). For a more formal derivation, see Chapter 6 in [20]. This integral can sometimes be solved by applying the *Gelfand-Yaglom method* [20], however that is not straightforward in our case, since (as will be shown later) the Lagrangian contains both real and complex parts. Since to further discuss this here will require a deviation from the paper's main theme, the Gelfand-Yaglom method and its adaptation to the integrals in this paper are discussed in the appendix.

The integral of Eq. 7 is computed as follows. The integral of a Gaussian $\int_{\mathcal{R}^n} du \exp(-uAu^t)$, for a positive definite matrix A, equals (up to a normalizing factor) the inverse of the square root of the product of A's eigenvalues. Similarly, the infinite-dimensional integral is the inverse of the (infinite) product of the square roots of the eigenvalues of the function space operator T which satisfies $(x, Tx) = \int(\lambda\dot{x}^2 + b(t)x^2)$, where the inner product is the usual one in L^2, $(x, y) = \int xy$ (see summary in Section 3.4). For operators on functions defined on a finite interval, which are constrained to be zero at the endpoints, the set of eigenvalues in known to be discrete [21] (physically, this corresponds to the fact that a particle in a box has a discrete set of energy levels). In Feynman's celebrated formulation of quantum mechanics by means of path integrals [4], this infinite product converges due to the presence of an (infinite) number of normalizing factors; here, convergence follows from the normalization of the probability on the entire space to equal 1. This normalization results in the quotient of two infinite products, which converges to a finite result.

3.4 Overview and Terminology

Following the physics nomenclature, we refer to x_{cl} as the classical path and to $\int(\lambda\dot{x}_{cl}^2 + b(t)x_{cl}^2 + c(t)x_{cl})$ as the classical action. The integrand $\lambda\dot{x}^2 + b(t)x^2 + c(t)x$ will

be referred to as the Lagrangian. Thus, to compute the path integral we need to:

1. Solve the differential equation (Eq. 6) defining x_{cl} and evaluate the classical action, denoted $A(x_{cl})$, which equals $\int (\lambda \dot{x}_{cl}^2 + b(t)x_{cl}^2 + c(t)x_{cl})$.

2. Compute the eigenvalues $\alpha_k, k = 1 \ldots \infty$, of the operator T which satisfies $\int xT(x) = \int (\lambda \dot{x}^2 + b(t)x^2)$.

3. The path integral then equals $\exp(-A(x_{cl})) \left(\prod_{k=1}^{\infty} \alpha_k \right)^{-\frac{1}{2}}$.

This paradigm and terminology will next be applied to our path integrals.

4 Warm-up: Path Integral for the Value at a Point

Let x_i be the noisy samples at t_i, $i = 1 \ldots n$. A natural question is: given the samples, what is the probability that $x(t_0) \geq \beta$? (note that t_0 is independent of t_i, $i = 1 \ldots n$, although it may equal one of them). We address this problem not only because it is important in its own right, but also since its solution will serve to introduce some of the methods and notations used later. Also, some of the calculations required for computing the path integrals of Sections 5 and 6 will be carried out here.

Assuming for the while $n = 1$, that is, a single sample point (extension to more than one point is discussed in Sections 5.3.1 and 7), then the probability, denoted $\Pr(t_0, \beta) \triangleq \Pr(x(t_0) \geq \beta)$, equals

$$\frac{\int \mathcal{D}x\ \theta[x(t_0) - \beta] \Pr(x)}{\int \mathcal{D}x \Pr(x)} = \frac{\int \mathcal{D}x\ \theta[x(t_0) - \beta] \exp\left(-\left(\lambda \int \dot{x}^2 + x^2(t_1) - 2x_1 x(t_1)\right)\right)}{\int \mathcal{D}x\ \exp\left(-\left(\lambda \int \dot{x}^2 + x^2(t_1) - 2x_1 x(t_1)\right)\right)} \quad (8)$$

(the $[x(t_1) - x_1]^2$ term in the exponent of $\Pr(x)$ contains x_1^2, but after exponentiation it cancels out in the numerator and denominator).

The main difficulty is handling the Heaviside function θ, which is not linear or quadratic. θ has many representations; aiming to tailor it into the Gaussian integrals in Eq. 8, we choose the following one:

$$\theta(u) = \frac{1}{2\pi i} \lim_{\epsilon \to 0^+} \int_{-\infty}^{\infty} d\omega \frac{\exp(u\omega i)}{\omega - \epsilon i} \quad (9)$$

The correctness of this formula for $\theta(u)$ follows from the identity

$$\frac{1}{2\pi i} \int_{-\infty}^{\infty} d\omega \frac{\exp(u\omega i)}{\omega - \epsilon i} = \begin{cases} 0 & \text{if } u < 0 \\ \exp(-\epsilon u) & \text{if } u \geq 0 \end{cases}$$

9

which can be derived by contour integration technique [22].

Next, we obtain (after interchanging the order of integration between $\mathcal{D}x$ and $d\omega$):

$$\text{Pr}(t_0, \beta) = \frac{1}{2\pi i} \lim_{\epsilon \to 0^+} \int_{-\infty}^{\infty} d\omega \frac{\exp(-\beta\omega i)}{\omega - \epsilon i} F(\omega),$$

$$F(\omega) \triangleq \frac{\int \mathcal{D}x \exp\left(-\left(\lambda \int \dot{x}^2 + x^2(t_1) - 2x_1 x(t_1) - i\omega x(t_0)\right)\right)}{\int \mathcal{D}x \exp\left(-\left(\lambda \int \dot{x}^2 + x^2(t_1) - 2x_1 x(t_1)\right)\right)} =$$

$$\frac{\int \mathcal{D}x \exp\left(-\left(\int (\lambda \dot{x}^2 + \delta_{t_1} x^2 - 2x_1 \delta_{t_1} x - i\omega \delta_{t_0} x)\right)\right)}{\int \mathcal{D}x \exp\left(-\left(\int (\lambda \dot{x}^2 + \delta_{t_1} x^2 - 2x_1 \delta_{t_1} x)\right)\right)} \triangleq \frac{\int \mathcal{D}x \exp\left(-\left(\int L_2(x, \dot{x})\right)\right)}{\int \mathcal{D}x \exp\left(-\left(\int L_1(x, \dot{x})\right)\right)} \quad (10)$$

Where in Eq. 10 all terms were collected into the integral over t, in order to apply the method of computation summarized in Section 3.4. The delta functions are introduced to incorporate the discrete terms, $x^2(t_1) - 2x_1 x(t_1)$, into the integrand. This introduces two Lagrangians: one in the numerator, $L_2 = \lambda \dot{x}^2 + \delta_{t_1} x^2 - 2x_1 \delta_{t_1} x - i\omega \delta_{t_0} x$, and another in the denominator, $L_1 = \lambda \dot{x}^2 + \delta_{t_1} x^2 - 2x_1 \delta_{t_1} x$. Note that $L_2 = L_1 - i\omega \delta_{t_0} x$; therefore, the quadratic parts of L_2, L_1 are equal, hence the factor of the numerator and denominator path integrals which corresponds to the eigenvalues cancels out, since the eigenvalues depend only on the quadratic part (alas we will not be that fortunate with the path integrals of Sections 5 and 6, as there the quadratic parts are different).

So, all that is left is to compute the classical action of L_2, L_1. We start with L_1.

4.1 Calculating the Classical Action for L_1

Computation of the classical action for $L_1 = \lambda \dot{x}^2 + \delta_{t_1} x^2 - 2x_1 \delta_{t_1} x$ is straightforward. The Euler-Lagrange equation for the classical path x_{cl} is $\lambda \ddot{x} - \delta_{t_1} x + x_1 \delta_{t_1} = 0$. This means that at every point but t_1, x_{cl} is linear. Since it has to satisfy the natural boundary conditions, its values at $t = 0, 1$ are 0. Therefore, it is determined by three parameters: it equals at in the interval $[0, t_1]$ and $bt + c$ in the interval $[t_1, 1]$. The equations which need to be solved for these parameters are:

- Natural boundary condition at $t = 1$: $b + c = 0$.

- Continuity at t_1: $at_1 = bt_1 + c$.

- Behavior of the first derivative at t_1. To satisfy the Euler-Lagrange equation, the $\lambda \ddot{x}$ term has to contain a delta function at t_1. Since the first derivative from the left equals a and from the right it equals b, the second derivative is $\lambda(b - a)\delta_{t_1}$ (here we used the definition of the delta function as the derivative of the Heaviside step

10

function), so the equation is $\lambda(b-a) - at_1 + x_1 = 0$. Incidentally, this explains why x_{cl} must be continuous at t_1; if not, the first derivative will be a delta function, and the second derivative the derivative of a delta function, which will not be canceled out by $\delta_{t_1} x + x_1 \delta_{t_1}{}^2$.

Next, the equations are solved and substituted into the action $\lambda \int \dot{x}^2 + x^2(t_1) - 2x_1 x(t_1)$, yielding $A_1(x_{cl}) = \frac{x_1{}^2(1-t_1)t_1}{-\lambda+t_1{}^2-t_1}$.

4.2   Calculating the Classical Action for L_2 and $\Pr(t_0, \beta)$

This proceeds very much as for L_1, but here we have the extra term $-i\omega\delta_{t_0} x$. Clearly the classical path is a linear spline with knots at t_1, t_0. Thus there are six parameters to solve for (two for each linear segment), and six equations: two for the natural boundary conditions, two for continuity at t_1 and t_0, and two for canceling out the delta function terms. As before, the equations are solved and substituted into the classical action; the resulting expression is a quadratic polynomial in ω, denote it $A_2(x_{cl}) \triangleq a_2\omega^2 + a_1\omega i + a_0$. a_2, a_1, a_0 are real numbers which depend on λ, t_0, t_1, x_1; also, since $L_2 = L_1 - i\omega\delta_{t_0} x$, $a_0 = A_1(x_{cl})$. From Eq. 10 we have

$$\Pr(t_0, \beta) = \frac{1}{2\pi i} \lim_{\epsilon \to 0^+} \int_{-\infty}^{\infty} d\omega \frac{\exp(-\beta\omega i)}{\omega - \epsilon i} \frac{\exp(-A_2(x_{cl}))}{\exp(-A_1(x_{cl}))} =$$

$$\frac{1}{2\pi i} \lim_{\epsilon \to 0^+} \int_{-\infty}^{\infty} d\omega \frac{\exp(-\beta\omega i)}{\omega - \epsilon i} \exp\left(-(a_2\omega^2 + a_1\omega i)\right) \qquad (11)$$

which is real[3] and can be computed in closed form involving the error function[4].

It is immediate to see that the cumulative distribution corresponds to a normal distribution, hence the point value $x(t_0)$ is a normal random variable, and the pointwise standard deviation of this normal distribution can be interpreted as the width of confidence intervals around the interpolant (this is in agreement with the results in [26, 7]). For an example, see Fig. 1.

---

[2]This also follows from x_{cl} being a member of the Sobolev space $W^{1,2}$ of functions with a square integrable first derivative [1].

[3]Summing the integrand for ω and $-\omega$ yields a pure imaginary number, and since the integration is from $-\infty$ to ∞ the entire integral is pure imaginary; after division by i we're left with a real number.

[4]Generally, $\frac{1}{2\pi i} \lim_{\epsilon \to 0^+} \int d\omega \frac{\exp(-(a\omega^2 + b\omega i))}{\omega - \epsilon i} = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{b}{2\sqrt{a}}\right)\right]$.
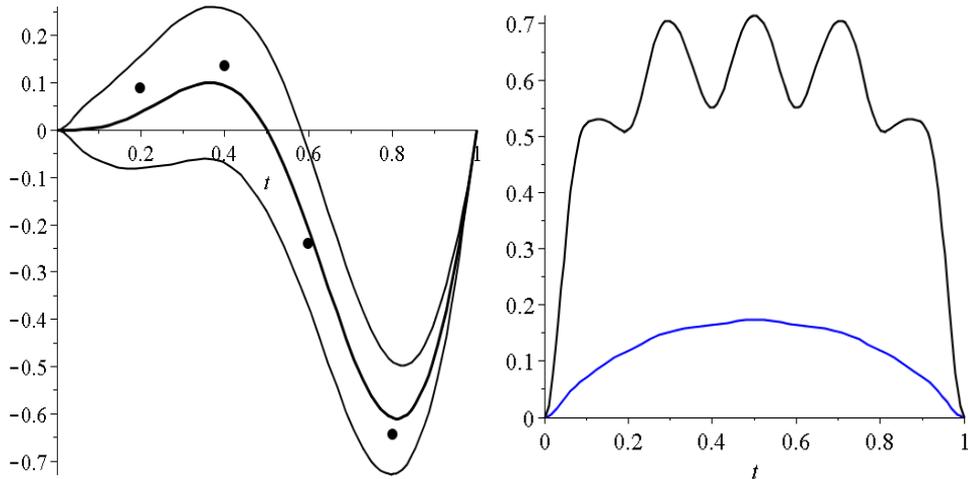
Figure 1: The uncertainty of the value at a point. Left: an example of a restored function is plotted (thicker line, center), as well as the upper and lower confidence intervals, of width equal to one standard deviation [26, 7]. The t_i values are $0.2, 0.4, 0.6, 0.8, \lambda = 1$, and the sample points are marked by dark solid circles. Right: same sample points, with the point uncertainty plotted for $\lambda = 1$ (blue) and $\lambda = 0.1$ (black). A smaller λ means a smaller penalty on "rough" (oscillating) functions, hence there is more freedom in the space of possible solutions, which means that the uncertainty is higher. As λ further decreases, the uncertainty at every point but the sample points will tend to infinity; see also [7].

5 Calculating the Path Integral for the Energy

Next we turn to the more involved problem of computing the energy path integral (i.e. the case in which the template y equals 0). This will be a natural preparation for the general case, since the quadratic parts of L_2 are equal for both cases. Again, assume for the while a single sample point, t_1, with the measured value denoted x_1. Following as in Section 4, we find:

$$\Pr(E, \beta) = \frac{1}{2\pi i} \lim_{\epsilon \to 0^+} \int_{-\infty}^{\infty} d\omega \frac{\exp(-\beta \omega i)}{\omega - \epsilon i} F(\omega),$$

$$F(\omega) \triangleq \frac{\int \mathcal{D}x \exp\left(-\left(\lambda \int \dot{x}^2 + x^2(t_1) - 2x_1 x(t_1) - i \int \omega x^2\right)\right)}{\int \mathcal{D}x \exp\left(-\left(\lambda \int \dot{x}^2 + x^2(t_1) - 2x_1 x(t_1)\right)\right)} =$$

$$\frac{\int \mathcal{D}x \exp\left(-\left(\int (\lambda \dot{x}^2 + \delta_{t_1} x^2 - 2x_1 \delta_{t_1} x - i\omega x^2)\right)\right)}{\int \mathcal{D}x \exp\left(-\left(\int (\lambda \dot{x}^2 + \delta_{t_1} x^2 - 2x_1 \delta_{t_1} x)\right)\right)} \triangleq \frac{\int \mathcal{D}x \exp\left(-\left(\int L_2(x, \dot{x})\right)\right)}{\int \mathcal{D}x \exp\left(-\left(\int L_1(x, \dot{x})\right)\right)} \quad (12)$$

Again we have two Lagrangians: the one in the numerator, $L_2 = \lambda \dot{x}^2 + \delta_{t_1} x^2 - 2x_1 \delta_{t_1} x - i\omega x^2$, and the one in the denominator is equal to L_1 of Section 4 (the denominator is still the normalizing factor). Note that $L_2 = L_1 - i\omega x^2$. It remains to compute the path

12

integral for the numerator, but also the L_1 and L_2 eigenvalues (which no longer cancel out, since now the quadratic components in L_1 and L_2 are different). The overall result will still be a one-dimensional integral over ω, alas a more complicated one.

5.1 Calculating the Eigenvalues Corresponding to L_1, L_2

Next, we calculate the eigenvalues of the linear operator T which satisfies $\int xT(x) = \int(\lambda \dot{x}^2 + \delta_{t_1} x^2)$. It is readily seen that $T(x) = -\lambda \ddot{x} + \delta_{t_1} x$ (this follows immediately by applying integration by parts and from the natural boundary conditions). For α to be an eigenvalue, there must be a solution to the differential equation $-\lambda \ddot{x} + \delta_{t_1} x = \alpha x$. Since T is obviously positive definite, only positive α's should be sought. As before, in every point but t_1 we must have $-\lambda \ddot{x} = \alpha x$, so the solution is a "trigonometric spline", $f_{\textbf{left}}(t) \triangleq a \cos\left(\sqrt{\frac{\alpha}{\lambda}}t\right) + b \sin\left(\sqrt{\frac{\alpha}{\lambda}}t\right)$ in the interval $[0, t_1]$ and $f_{\textbf{right}} \triangleq c \cos\left(\sqrt{\frac{\alpha}{\lambda}}t\right) + d \sin\left(\sqrt{\frac{\alpha}{\lambda}}t\right)$ in the interval $[t_1, 1]$. The resulting equations are $f_{\textbf{left}}(0) = f_{\textbf{right}}(1) = 0$ (boundary conditions), $f_{\textbf{left}}(t_1) = f_{\textbf{right}}(t_1)$ (continuity), and $-\lambda(\dot{f}_{\textbf{right}}(t_1) - \dot{f}_{\textbf{left}}(t_1)) + f_{\textbf{right}}(t_1)) = 0$ (the condition for the delta function terms to cancel out). These four equations define a linear system, denoted S, which has a non-trivial solution iff $\det(S)$, viewed as a function of α, is equal to 0; therefore the eigenvalues are the solutions of the equation

$$\det(S) = \lambda\gamma \sin(\gamma) - \cos(\gamma) + \cos(\gamma)(\cos(\gamma t_1))^2 + \sin(\gamma t_1)\cos(\gamma t_1)\sin(\gamma) = 0, \quad \gamma \triangleq \sqrt{\frac{\alpha}{\lambda}}$$

This is a transcendental equation, which is solved numerically. For large values of γ (or α), the term $\gamma \sin(\gamma)$ dominates, and therefore the solutions quickly approach $\gamma = k\pi$ or $\alpha = \lambda k^2 \pi^2$ for natural k. This is not surprising, since had it not been for the $\delta_{t_1} x$ term in the differential equation, the eigenvalues would have exactly equaled $\lambda k^2 \pi^2$; but this perturbation term is small and does not change the asymptotic behavior of the eigenvalues. As λ increases, the term $\lambda \ddot{x}$ in the differential equation becomes more dominant, and the eigenvalues converge to $\lambda k^2 \pi^2$ faster (see Fig. 2). Starting with $\lambda k^2 \pi^2$ as an initial guess, it is very easy to solve for the eigenvalues. Fig. 2 depicts the location of the k-th eigenvalues vs. $\lambda k^2 \pi^2$.

The discussion above concerns the L_1 eigenvalues. Fortunately, since $L_2 = L_1 - i\omega x^2$, the L_2 eigenvalues are simply the L_1 ones, minus $i\omega$. Since we have to divide the L_2 integral by the L_1 integral, the two products of the square roots of the eigenvalues yield a factor equal to $\prod_{k=1}^{\infty}\left(1 - \frac{i\omega}{\alpha_k}\right)^{-\frac{1}{2}}$. For relatively small values of ω, the terms of the product approach 1 very quickly and the infinite product can be truncated. It is easy to estimate
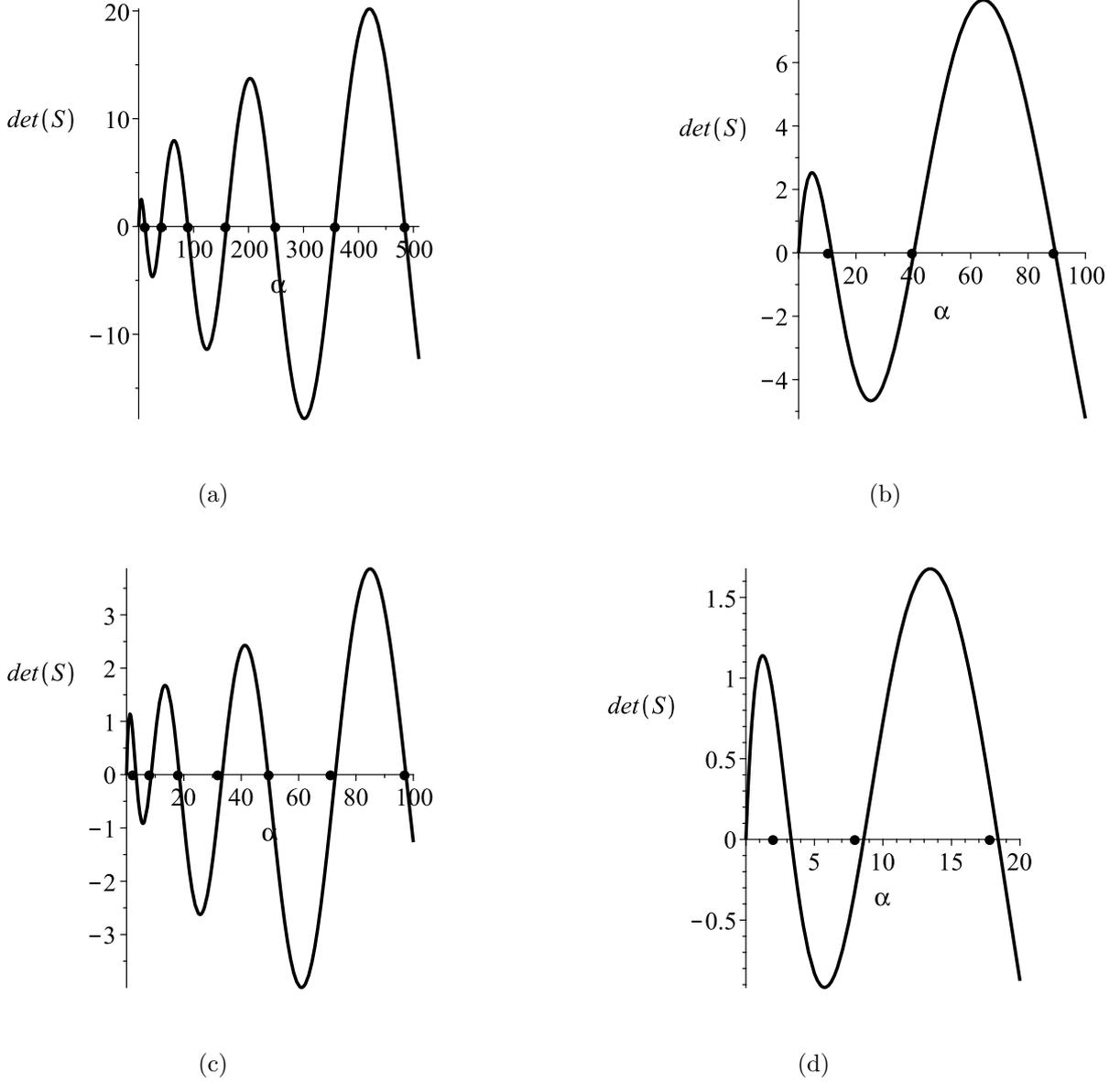
Figure 2: Plots of $\det(S)$. Roots correspond to eigenvalues of the function space operator $T(x) = -\lambda\ddot{x} + \delta_{t_1}x$. Dark dots on the horizontal axis depict the locations of $\lambda k^2\pi^2$. (a) $t_1 = 0.4, \lambda = 1$. (b) Same as (a), smaller range of α. (c) and (d) are the same as (a) and (b), with $\lambda = 0.2$. In both cases, the eigenvalues converge to $\lambda k^2\pi^2$, with convergence faster when λ is larger.

the residual when truncating at $k = N$: for small h, $1 - h \approx \exp(-h)$, so

$$\prod_{k=N}^{\infty}\left(1 - \frac{i\omega}{\alpha_k}\right)^{-\frac{1}{2}} \approx \prod_{k=N}^{\infty}\exp\left(\frac{\omega i}{2\alpha_k}\right) \approx \prod_{k=N}^{\infty}\exp\left(\frac{\omega i}{2\lambda k^2\pi^2}\right) \approx \exp\left(\frac{\omega i}{2\lambda\pi^2 N}\right) \qquad (13)$$

So the infinite product can be truncated at a sufficiently large N (e.g. for which $\lambda\pi^2 N^2 > 10\omega$), with the truncated part replaced by the right-hand side of Eq. 13.

5.2 L_2 Classical Action for the Energy Path Integral

This proceeds almost exactly as in Section 4.2 – with the difference that the quadratic term of the Lagrangian now contains an additional term, $-i\omega x^2$. So, the Euler-Lagrange equation is $\lambda\ddot{x} - \delta_{t_1}x + x_1\delta_{t_1} + i\omega x = 0$. Hence, for all points but t_1, the solution x_{cl} is no longer linear, but satisfies $\lambda\ddot{x} + i\omega x = 0$. The solution of this differential equation is a "complex trigonometric spline", with the part in $[0, t_1]$ equal to $a\exp(st)[\cos(st) - i\sin(st)] + b\exp(-st)[\cos(st) + i\sin(st)]$, for $s \triangleq \sqrt{\frac{\omega}{2\lambda}}$, and similarly in the $[t_1, 1]$ interval. The equations are therefore similar in principle to those of the L_1 case: natural boundary conditions (two equations), a continuity at t_1 condition (one equation), and a condition for the δ_{t_1} factors to cancel out (one equation). These equations are still linear in a, b, c, d, but the solution, and the classical action, are more complicated than for L_1 and, for the sake of brevity, will not be reproduced here.

5.3 Computing $\Pr(E, \beta)$

Putting it all together, denote the classical action of the L_1 Lagrangian by $A_1(\lambda, t_1, x_1)$ (note that it does not depend on ω), and same for L_2 and $A_2(\lambda, t_1, x_1, \omega)$. Denote the eigenvalues by $\alpha_k, k = 1 \ldots \infty$. Then $\Pr(E, \beta)$ equals

$$\exp(A_1(\lambda, t_1, x_1))\frac{1}{2\pi i} \lim_{\epsilon \to 0^+} \int_{-\infty}^{\infty} d\omega \frac{\exp(-\beta\omega i)}{\omega - \epsilon i} \exp(-A_2(\lambda, t_1, x_1, \omega)) \prod_{k=1}^{\infty} \left(1 - \frac{i\omega}{\alpha_k}\right)^{-\frac{1}{2}} \quad (14)$$

Fig. 3 depicts a plot of the integrand for $\lambda = 3, t_1 = 0.4, x_1 = 1, \beta = 0.05$ (the result of the integration is real for the same reason as for Eq. 11, see footnote 3).

We have evaluated this one-dimensional integral numerically. The term $\prod_{k=1}^{\infty} \left(1 - \frac{i\omega}{\alpha_k}\right)^{-\frac{1}{2}}$ is handled as explained in Section 5.1. As ω increases, the factor $\exp(-\beta\omega i)$ oscillates very rapidly, so the value of the integral is very small for regions with large ω.

The numerical computation of the integral can be further facilitated by noting that for small ω (which constitute the "critical region" of integration, in which the integrand is relatively large), the term containing the eigenvalues is very close to 1, hence only a very small number of eigenvalues can be used to a good approximation. Moreover, it is possible to explicitly calculate $\lim_{\omega \to 0} \frac{A_1(\lambda, t_1, x_1) - A_2(\lambda, t_1, x_1, \omega)}{\omega} \triangleq l_1 + l_2 i$, which means that in the critical region, the term $\frac{\exp(A_1(\lambda, t_1, x_1))}{\exp(A_2(\lambda, t_1, x_1\omega))}$ is very well approximated by $\exp((l_1 + l_2 i)\omega)$. Using the theory of exponential integrals, $\lim_{\epsilon \to 0^+} \int_{-\omega_0}^{\omega_0} d\omega \frac{\exp((l_1 + l_2 i)\omega)}{\omega - \epsilon i}$ can be calculated in closed form. Hence, we can divide the integration domain into two parts, for a small
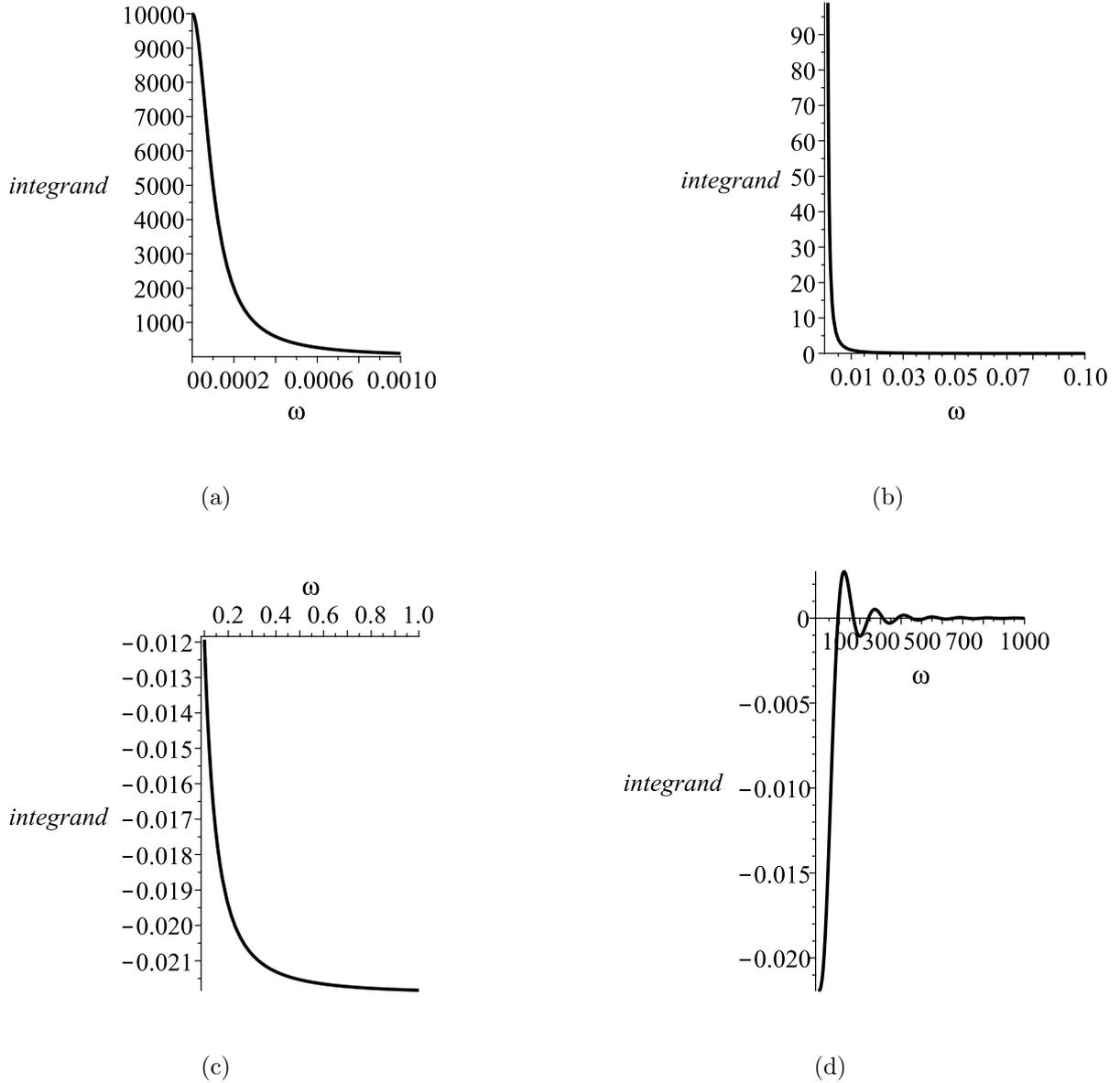
15

Figure 3: The integrand of Eq. 14, in various ranges: in (a) from 0 to 10^{-3}, (b) from 10^{-3} to 10^{-1}, (c) from 10^{-1} to 10^1, (b) from 10^1 to 10^3. $\lambda = 3, t_1 = 0.4, x_1 = 1, \beta = 0.05$.

ω_0: $[-\omega_0, \omega_0]$ and $[-\infty, -\omega_0] \cup [\omega_0, \infty]$. In the first domain, we can use a very accurate approximation; in the second, the integrand behaves nicely and the integral can be very efficiently computed using numerical methods. Also, since in the second domain ω is bounded away from 0, we can discard the ϵ in the integrand (and, of course, the limit over ϵ).

5.3.1 More than one Sample Point

For more than one sample point, the computation proceeds in very much the same fashion. Assume e.g. two sample points, t_1, t_2 with sample values x_1, x_2. The Lagrangian

L_1 equals $\lambda \dot{x}^2 + \delta_{t_1} x^2 - 2x_1 \delta_{t_1} x + \delta_{t_2} x^2 - 2x_2 \delta_{t_2} x$. The Euler-Lagrange equation is $\lambda \ddot{x} - \delta_{t_1} x + x_1 \delta_{t_1} - \delta_{t_2} x + x_2 \delta_{t_2} = 0$. Again, at any point but t_1, t_2 the solution is linear, so the overall solution is a piecewise linear spline with knots at t_1, t_2. There are now six parameters to recover – two at every linear segment – and six equations: two for the natural boundary conditions, two for continuity (at t_1 and t_2), and two for canceling out the coefficients of the delta functions at t_1, t_2. The calculation of the classical path for L_2 and the computation of the eigenvalue proceed similarly. Fig 4. contains some plots of $\Pr(E, \beta)$.
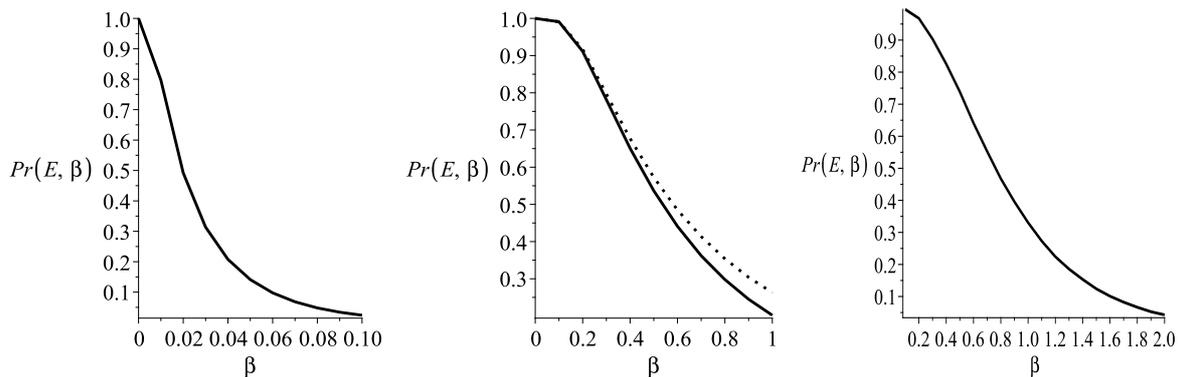

Figure 4: Some plots of $\Pr(E, \beta)$. Left: $\lambda = 3, x_1 = 1, t_1 = 1/2$. Center: $\lambda = 0.1, x_1 = 1, t_1 = 1/2$ (solid line), and $\lambda = 0.1, x_1 = 0.1, t_1 = 1/2$ (dotted line). $\Pr(E, \beta)$ increases when λ decreases, since a smaller λ means that the penalty on roughness is smaller, hence rough functions are more probable, allowing more freedom as t moves away from the interval's boundary and sampling points, so the energy is typically larger (see also Fig. 1 for the pointwise uncertainty with different λ's). $\Pr(E, \beta)$ also increases when the sample point is moved from $t = \frac{1}{2}$ to $t = 0.1$, since here, too, the functions have more freedom in the larger part of the interval which is far from the sampling locations. Loosely speaking, "nailing" the function at $\{0, \frac{1}{2}, 1\}$ is more restrictive than "nailing" it in $\{0, 0.1, 1\}$. Right: example with more sample points: $t_1 = 0.2, t_2 = 0.4, t_3 = 0.6, t_4 = 0.8, x_1 = 1.1, x_2 = 2.0, x_3 = -1.3, x_4 = 0.5, \lambda = 0.1$.

6   The Distribution of $d(x, y)$

Almost everything required to solve the problem posed in the beginning of this paper – computing the distribution of the distance from the signal to a fixed template – was accomplished in the previous sections. All we need now is to compute the quotient

of the path integrals in Eq. 12, with $i \int \omega x^2$ replaced by $i \int \omega (x - y)^2$, where y is the fixed template. So, the denominator (with L_1) does not change at all, and neither does the quadratic term of the numerator (since y is a constant function). The only change is in the differential equation defining the classical path, which now becomes $\lambda \ddot{x} - \delta_{t_1} x + x_1 \delta_{t_1} + i \omega (x - y) = 0$. The solution is the same as before, except for a non-homogeneous part accounting for y. This requires the solution of $\lambda \ddot{x} + i \omega x = i \omega y$, a simple differential equation whose solution is given by $C_1 \sin(\sqrt{\alpha} t) + C_2 \cos(\sqrt{\alpha} t) + \sqrt{\alpha} \left(\int \cos(\sqrt{\alpha} t) y(t) \right) \sin(\sqrt{\alpha} t) - \left(\int \sin(\sqrt{\alpha} t) y(t) \right) \cos(\sqrt{\alpha} t))$ for $\alpha = \omega i$, and C_1, C_2 constants determined by the boundary conditions. . For many important y's, this solution assumes a very simple form, e.g. for $y = \sin(k\pi t)$ it equals $\frac{\omega i \sin(k\pi t)}{\omega i - \lambda k^2 \pi^2}$. These functions, for $k = 1 \ldots \infty$, form an orthogonal base for the functions with natural boundary conditions considered here; further, most functions can be very well approximated with only a small number of the sine functions. Thus one can either use the sine functions approximation or solve directly using the solution provided above. All that remains is the straightforward computation of the L_2 classical action.

6.1 Compatibility of Model with Reality

To test the compatibility of the theoretic model with real signals (see also [14] for a comparison of a similar prior with the distribution of real images), we took many samples of such signals (specifically, rows of images sampled randomly from a large database of background images[5]). Each row was normalized to satisfy the boundary conditions and the x-coordinate scaled to the unit interval, and also normalized to unit length. The t_i, x_i used were as in Fig. 3, and the rows chosen were those satisfying them up to an error of 0.1. A sample of template functions y was defined by $a_1 \sin(\pi t) + a_2 \sin(2\pi t) + a_3 \sin(3\pi t)$, where a_i were randomly chosen in the interval $[-1, 1]$. Values of β were randomly chosen in the interval $[0, 2]$. For every choice of y and β, $\Pr(y, \beta)$ was computed analytically, and its empirical analogue, defined as the percentage of the database signals x for which $\int (x - y)^2 \geq \beta$, was also computed. The analytic vs. empirical probabilities are depicted in the scatter plot in Fig. 5, showing a good fit between the two probabilities.
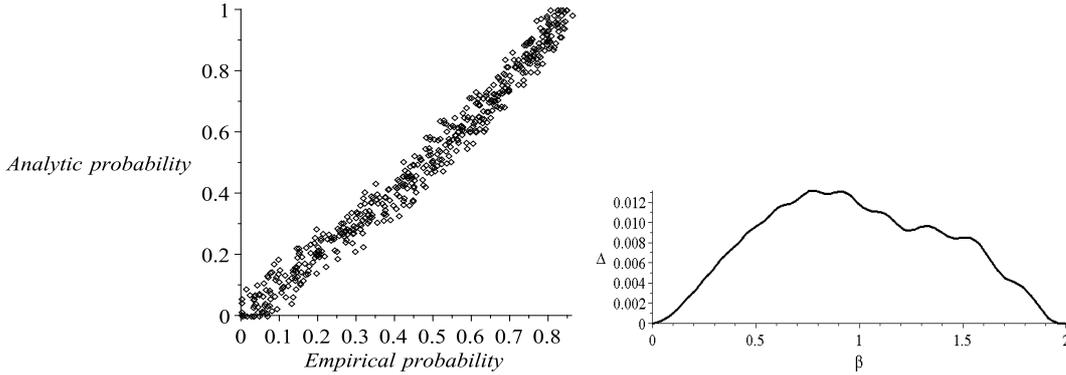
---

[5] http://cbcl.mit.edu/cbcl/software-datasets/FaceData2.html

Figure 5: Left: scatter plot of analytic probability (the probability that $\int(x-y)^2 < \beta$) vs. the empirical probability (percentage of sampled functions x for which $\int(x-y)^2 < \beta$) scatter plot. Right: plot of absolute difference (Δ) between analytic and empirical as a function of beta.

7   The General Case: Arbitrarily many Sample Points and Higher Dimensions

In this section we study a general method for computing the probability defined in Eq. 4, for any number of points and in any dimension (number of variables). Here we compute not the c.d.f of $\int(x-y)^2$, but that p.d.f (from which the c.d.f can be obtained by integration). The main difference between the approach in this section and the one in Sections 4-6 is the generality of the formulation and the application of the Green's functions of the relevant differential operators, which easily extends to more than one variable. For functions in one variable, and a modest number of points, the method described in Sections 4-6 is simpler and more direct.

To simplify the presentation, we assume that $\lambda = 1$ and $y = 0$ (i.e. we compute the p.d.f of the energy). From the calculations in Section 6 it follows that the case of a general y differs from that of $y = 0$ only in the linear term of the Lagrangian, hence the derivations in both cases are similar.

We start with the 1D case and then point out the modifications required for higher dimensions. Especially, we will prove that in dimensions higher than 1 the expectation of the energy diverges to infinity, and this has to be remedied either by introducing a "cutoff" or using a different prior.

Proceeding as before, the p.d.f for a fixed value β equals

$$\frac{\int \mathcal{D}x \exp\left(-\left(\int \dot{x}^2 + \sum(x^2(t_i) - 2\sum x_i x(t_i))\right)\right) \delta(\int x^2 - \beta)}{\int \mathcal{D}x \exp\left(-\left(\int \dot{x}^2 + \sum(x^2(t_i) - 2\sum x_i\ x(t_i))\right)\right)} \quad (15)$$

after denoting $V(t) = \sum \delta_{t_i}$ and $J(t) = \sum x_i \delta_{t_i}$, and using the well-known expansion for

19

the delta function $\delta(t) = \int d\omega \exp(\omega t i)$, Eq. 15 becomes

$$\int d\omega \exp(-\beta \omega i) \frac{\int \mathcal{D}x \exp\left(-x\left(-\nabla^2 + V - \omega i\right) x^T - xJ\right)}{\int \mathcal{D}x \exp\left(-x\left(-\nabla^2 + V\right) x^T\right)} \tag{16}$$

Where we have used the Laplacian notation as a hint for the general treatment in higher dimensions. Denoting $-\nabla^2 + V - \omega i \triangleq G_\omega$, Eq. 16 (sans the integral over ω) assumes the compact form

$$\frac{\int \mathcal{D}x \exp\left(-xG_\omega x^T - xJ\right)}{\int \mathcal{D}x \exp\left(-xG_0 x^T - xJ\right)} \tag{17}$$

which equals

$$\sqrt{\frac{|G_0|}{|G_\omega|}} \exp\left(J(G_\omega^{-1} - G_0^{-1})J^T\right) \tag{18}$$

since $J(t) = \sum x_i \delta_{t_i}$, it follows that $JG_\omega^{-1}J^T = \sum_{i,j} x_i x_j G^{(\omega)}(t_i, t_j)$, where $G^{(\omega)}$ is the Green's function of G_ω; we discuss its computation in Appendix II.

Next, we compute the ratio of the functional determinants, $\frac{|G_0|}{|G_\omega|}$. It equals

$$\frac{|G_0|}{|G_\omega|} = \frac{|-\nabla^2 + V|}{|-\nabla^2 + V - \omega i|} = \frac{|-\nabla^2|}{|-\nabla^2 - \omega i|} \frac{|I + (\nabla^2)^{-1}V|}{|I + (\nabla^2 - \omega i)^{-1}V|} \tag{19}$$

Due to the presence of V, which is a sum of delta functions at the sample points, $\frac{|I+(\nabla^2)^{-1}V|}{|I+(\nabla^2-\omega i)^{-1}V|}$ can be computed over the finite-dimensional subspace spanned by these delta functions (see for example [8]). The computation of $\frac{|-\nabla^2|}{|-\nabla^2-\omega i|}$ cannot be reduced to a finite-dimensional subspace; further, it is the root of the divergence problem in higher dimensions, to be discussed in Section 7.1. For our problem, the natural boundary conditions make it easy to define a basis consisting of eigenfunctions of minus the Laplacian – $\sin(k\pi t), k = 1 \ldots \infty$, with the k-th eigenvalue equal to $\pi^2 k^2$. Therefore, the ratio of the functional determinants, $\frac{|-\nabla^2|}{|-\nabla^2-\omega i|}$, is equal to

$$\frac{\prod\limits_{k=1}^{\infty} \pi^2 k^2}{\prod\limits_{k=1}^{\infty} (\pi^2 k^2 - \omega i)} = \left[\prod_{k=1}^{\infty}\left(1 - \frac{\omega i}{\pi^2 k^2}\right)\right]^{-1} = \frac{\sqrt{\omega i}}{\sin(\sqrt{\omega i})} \tag{20}$$

(where the last equality follows from the well-known infinite product expansion for the sine function). However, for the two-dimensional case, the eigenfunctions of the Laplacian are $\sin(k\pi t)\sin(l\pi s), k, l = 1 \ldots \infty$ with eigenvalues $\pi^2(k^2 + l^2), k, l = 1 \ldots \infty$. Therefore, the ratio of determinants equals $\left[\prod\limits_{k,l=1}^{\infty}\left(1 - \frac{\omega i}{\pi^2(k^2+l^2)}\right)\right]^{-1} = 0$ (since $\sum\limits_{k,l=1}^{\infty} \frac{1}{k^2+l^2}$ diverges). Next we look more closely at why the p.d.f in the two-dimensional case is zero, and suggest two methods to overcome this problem.

7.1 Divergence in Dimensions Larger than 1

To understand why the computation of the p.d.f of $\int x^2$ becomes a problem in dimensions two and higher, let us start with a simple exercise in one dimension: compute the expectation of $\int x^2$, given only the boundary conditions (no measurements; see also Section 3.1). This expectation is equal to

$$\frac{\int \mathcal{D}x(\int x^2)\Pr(x)}{\int \mathcal{D}x\Pr(x)} = \frac{\int \mathcal{D}x(\int x^2)\exp\left(-\int \dot{x}^2\right)}{\int \mathcal{D}x\exp\left(-\int \dot{x}^2\right)} \tag{21}$$

The expression in Eq. 21 is most easily computed in the Fourier basis (the boundary conditions imply that $sin(k\pi t), k = 1 \ldots \infty$, is an orthogonal basis for the inner products on the unit interval defined by $\int xy$ and $\int \dot{x}\dot{y}$). Denoting $x(t) = \sum u_k \sin(k\pi t)$, the expression in Eq. 21 becomes

$$\frac{\int du_1 du_2 \ldots (\sum u_k^2)\exp\left(-\sum \frac{\pi^2 k^2 u_k^2}{2}\right)}{\int du_1 du_2 \ldots \exp\left(-\sum \frac{\pi^2 k^2 u_k^2}{2}\right)} \tag{22}$$

which, after some straightforward manipulations and using the identity $\frac{\int du_k u_k^2 \exp\left(-a^2 u_k^2\right)}{\int du_k \exp\left(-a^2 u_k^2\right)} = \frac{1}{2a^2}$, turns out to equal (up to some multiplicative constant) $\sum_k \frac{1}{k^2}$, which is known to converge.

In two-dimensions, however, the situation is different. Denoting the variables t, s, an orthonormal basis is provide by $\sin(k\pi t)\sin(l\pi s), k, l = 1 \ldots \infty$, and a calculation very similar to the one-dimensional case yields that the expectation of the energy is $\sum_{k,l} \frac{1}{k^2+l^2}$, which diverges to infinity – not surprisingly, this is exactly the reason for which, as demonstrated before, the p.d.f is identically zero. Intuitively, the first-order smoothness term is strong enough to enforce convergence in one dimension, but not in two and above. The fact that the expectation in these dimensions diverges explains why the p.d.f at any finite value is equal to 0.

7.1.1 Resolving the Divergence

The simplest solution to the divergence in dimensions greater than one is to apply a frequency cutoff, that is, not to allow elements $\sin(k\pi t)\sin(l\pi s)$ for which $\sqrt{k^2 + l^2} > \Lambda$, where Λ is the cutoff. In practical problems this typically makes sense, as the sampled signal is assumed to be band-limited. Another option is to use the second-order smoothness term, in which the $\int \int (x_t^2 + x_s^2)$ roughness measure in the prior is replaced by $\int \int (x_{tt}^2 + 2x_{ts}^2 + x_{ss}^2)$. A calculation quite similar to the one for the first-order smoothness term yields that the expectation is equal to $\sum_{k,l} \frac{1}{(k^2+l^2)^2}$, which converges to a finite

sum. Save for this, the treatment of the two-dimensional case follows exactly as in the one-dimensional one; in Appendix II the Green's function in two dimensions is described.

Similar considerations to those here prove that the second-order smoothness term guarantees convergence up to three dimensions. For functions with more than variables, higher-order terms are required.

8 Finite-Dimensional Approximation

Computing a Gaussian integral with the term $\exp(-xAx^t)$ is much easier when the matrix A is diagonal. For the integrals that need to be computed in this work, A is replaced by an operator L, which consists of a derivative operator ($-\nabla^2$ when using the first order smoothness term and ∇^4 for the second order smoothness term), plus a sum of discrete terms represented by delta functions. The derivative component, which is the dominant one, can be easily diagonalized by going to the trigonometric basis. Using this basis presents another advantage: the eigenvalues of L increase very rapidly as a function of the frequency k of the basis elements (they equal k^2 for $-\nabla$ and k^4 for ∇^4, for eigenfunctions with frequency k). While, due to the delta functions, the trigonometric functions are no longer eigenfunctions of the operator L, they still make its matrix very strongly diagonally dominant, in a sense to be now made precise. This means that very good approximations can be obtained using a finite-dimensional subspace of modest dimension; we now proceed to define these subspaces and bound the error between the p.d.f computed on them to that computed over the entire space. This section is rather technical, so we have tried to make the presentation short; some small details are therefore omitted from the proofs.

While the following analysis holds for any characteristic of the p.d.f we seek to compute, it is especially convenient to carry out for the moment generating function, $M(s)$[6]. We now describe how to approximate $M(s)$ for the energy p.d.f. Assuming a second-order smoothness term:

$$M(s) = \frac{\int \mathcal{D}x \, \exp\left(-\left(\lambda \int \ddot{x}^2 - s \int x^2 + \sum_{k=1}^{N} x^2(t_k) - 2 \sum_{k=1}^{N} x_k x(t_k)\right)\right)}{\int \mathcal{D}x \, \exp\left(-\left(\lambda \int \ddot{x}^2 + \sum_{k=1}^{N} x^2(t_k) - 2 \sum_{k=1}^{N} x_k x(t_k)\right)\right)} \tag{23}$$

Replacing the entire function space with the finite-dimensional subspace

---

[6]For a random variable X, $M(s)$ is defined as the expectation of $\exp(sX)$. The derivatives of M at $s = 0$ are equal to the moments of X.

$S_n = \mathbf{Span}\{\sin(\pi t) \ldots \sin(n\pi t)\}$. $M(s)$ then assumes the form

$$\frac{\int\limits_{\mathcal{R}^n} dx \exp\left(-x(A_n - (s/2)I)x^t + (b_n, x)\right)}{\int\limits_{\mathcal{R}^n} dx \exp\left(-xA_nx^t + (b_n, x)\right)} =$$

$$\sqrt{\frac{|A_n|}{|A_n - (s/2)I|}} \exp\left(\frac{b_n((A_n - (s/2)I)^{-1} - A_n^{-1})b_n^t}{4}\right) \quad (24)$$

where

$$(A_n)_{u,v} = \begin{cases} \frac{\lambda\pi^4 u^4}{2} + \sum\limits_{k=1}^{N} \sin^2(\pi u t_k) & u = v \\ \sum\limits_{k=1}^{N} \sin(\pi u t_k)\sin(\pi v t_k) & u \neq v \end{cases}$$

$$b_n(u) = 2\sum\limits_{k=1}^{N} x_k \sin(\pi u t_k)$$

Since the constant $\frac{\pi^4}{2}$ frequently appears in the forthcoming analysis, we will hereafter denote it by c.

Next we proceed to analyze some simple properties of the matrix A_n, which determines the rate at which the integral over S_n approximates the integral over the entire space. Since it depends only on the locations of the sample points, we treat two cases – uniformly distributed and random sample points.

8.1 Uniformly distributed sample points

Often, the locations of the sample points are uniformly distributed, $t_k = \frac{k}{N+1}$. In this case A_n becomes

$$(A_n)_{u,v} = \begin{cases} \frac{\lambda\pi^4 u^4}{2} + \frac{N+1}{2} & u = v,\ u \neq 0 \bmod (N+1) \\ \frac{\lambda\pi^4 u^4}{2} & u = v, u = 0 \bmod (N+1) \\ \frac{N+1}{2} & u \neq v, |u - v| = 2(N+1)m,\ m \in \mathbb{N} \\ -\frac{N+1}{2} & u \neq v, u + v = 2(N+1)m,\ m \in \mathbb{N} \end{cases}$$

Two additional computational shortcuts are:

- In the uniformly distributed case the eigenvalues of A_n can be computed offline, hence computing the determinants and other expressions in $M(s)$ can be done very quickly, since only the vector b_n is data-dependent.

- In the uniformly distributed case, if $n < N$ (that is, many sample points) then A_n is *diagonal*, making the computation very fast.

8.2 Non-uniformly distributed sample points

The diagonal of A_n is similar to the one for the uniformly distributed case, with $\frac{N+1}{2}$ replaced by an expression whose average is very close to $\frac{N}{2}$, and the off-diagonal elements have an absolute value with an average of about $\frac{\sqrt{N}}{2}$. To see this, recall that the off-diagonal elements equal $\sum_{k=1}^{N} \sin(\pi u t_k) \sin(\pi v t_k)$. If the t_k are random, so is $\sin(\pi u t_k) \sin(\pi v t_k)$, which makes it a random variable with zero expectation and variance equal to $\int_{0}^{1} dt \sin^2(\pi u t) \sin(\pi v t) = \frac{1}{4}$. So in this case too, A_n is very strongly diagonally dominant.

8.3 Convergence with increasing n

The question is at what rate do the expressions

$$D_n \triangleq \frac{|A_n|}{|A_n - (s/2)I|}$$

$$E_n \triangleq \exp\left(b_n((A_n - (s/2)I)^{-1} - A_n^{-1})b_n^t\right)$$

which determine $M(s)$ (Eq. 24) converge to a limit as $n \to \infty$. This depends on the rate of increase of A_n's eigenvalues and on the size of b_n's components. Next we analyze the behavior of the determinant $|A_n|$ in the more difficult case of non-uniform sample points.

8.4 Approximating the Determinant: Non-Uniform Sample Points

Since A_n is strongly diagonally dominant, it is reasonable that for $n > m$, $|A_m| \prod_{k=m+1}^{n} A_n(k,k)$ is a good approximation to A_n. We now evaluate the quality of this approximation and offer a heuristic improvement which performs very well in practice. Let us start with the case $n = m + 1$. To keep notations simple, $A(k,l)$ will denote the (k,l) element of all matrices (there are no grounds for confusion as all matrices are identical on their common range). We have $|A_{m+1,m+1}| = A(m+1,m+1)|A_{m,m}| + \sum_{k=1}^{m} A(m+1,k)M(m+1,k)$, where $M(m+1,k)$ is the k'th minor. The following simple estimate for $M(m+1,k)$ turns out to be very accurate: since it contains all the diagonal elements of $A_{m,m}$ except $A(k,k)$, and since the absolute value of the element which appears there instead of $A(k,k)$ is roughly equal to $\frac{\sqrt{N}}{2}$ (Section 8.2), $M(m+1,k)$ can be approximated by $\frac{\sqrt{N}}{2A(m,m)}|A_m|$. Summing over the minors and dividing by $|A_{m,m}|$ yields that $\frac{|A_{m+1,m+1}|}{|A_{m,m}|}$ is roughly equal to $A_{m,m}$ multiplied by an error factor of $1 + \epsilon(m)$, where $\epsilon(m)$ is a random variable with zero expectation and standard deviations equal to $\frac{\sqrt{N}}{4\lambda cm^4}\sqrt{\sum_{k=1}^{m} \frac{1}{c^2\lambda^2 k^8}}$. The error in estimating $|A_n|$ from $|A_m|$ is bounded by summing up $\epsilon(k)$ for $m < k \leq n$ (this follows since the $\epsilon(k)$ factors are very small and $\prod[1 + \epsilon(k)]$ can be approximated by $1 + \sum \epsilon(k)$).

8.5 Summary of Results

Proceeding in a similar fashion as in Section 8.4, bounds on the quality of approximation over a finite-dimensional subspace can be derived for the exponential term E_n (Section 8.3) and for the two-dimensional case. Summarizing, we obtain that when using the second-order smoothness term, then for one variable the value of $M(s)$ on an appropriately chosen subspace of dimension n differs from the exact value by a factor bounded by $O\left(\frac{\sqrt{N}}{c\lambda n^3}\right)$, and for two variables the factor is $O\left(\frac{\sqrt{N}}{c\lambda n^2}\right)$ (the difference in the power of n is due to the fact that in the two-dimensional case, $O(n^2)$ basis functions are required to cover the frequencies lower than n).

Note that as λ decreases, more basis functions are required to achieve a fixed accuracy. This is because a smaller λ means that higher frequencies (which increase the signal's roughness and thus decrease its prior probability) are penalized to a smaller extent than for a larger λ. Also, note that the number of required basis functions increases with the number of sample points N, but only at a rate of $\sqrt{N}$.

In Fig. 6, the quality of the finite-dimensional approximation is demonstrated for calculating the first four moments of the p.d.f, using the value and derivatives of the moment generating function at zero.
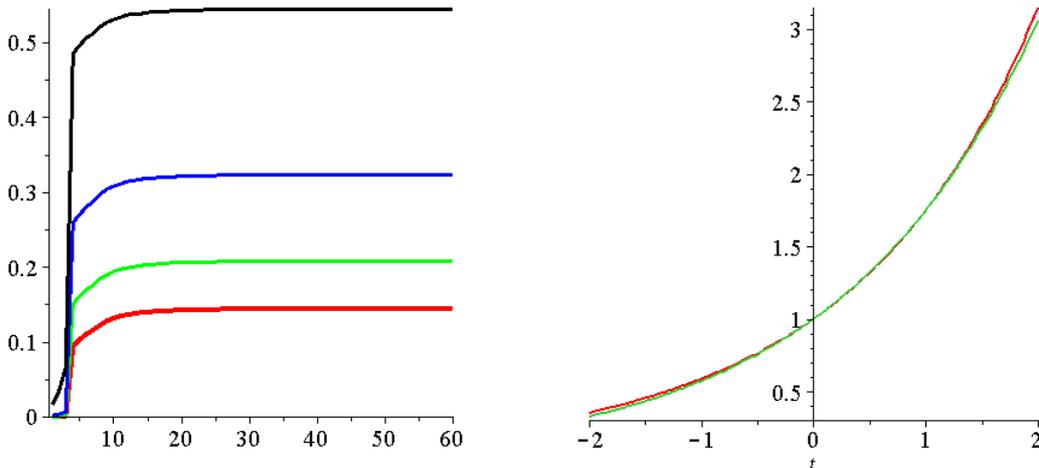

Figure 6: Left: example of the quality of finite-dimensional approximation. Horizontal axis is the dimension, vertical the value of the first four moments (black for first, blue for second, green for third, red for fourth). The moments were computed by differentiating $M(s)$ at $s = 0$ by finite differences. Data consisted of 50 sample points from $x(t) = \sin(4\pi t)$ with noise uniformly sampled in $[-0.05, 0.05]$, $\lambda = 0.0001$. Right: moment generating function for the same data (red), super-imposed with $\exp(M(1)t)$ (green).

9   Conclusions and Future Work

A rigorous method for computing the distribution of a signal's distance from a template, given discrete noisy samples, was provided. Important sub-cases which are also solved are the distribution of the pointwise values and the energy. The presented method can handle any linear or quadratic functional on signals. Future work will address other functionals and probability measures, and questions such as where to sample in order to optimally reduce the functional's uncertainty.

10   Acknowledgment

References

[1] R. A. Adams. *Sobolev Spaces*. Academic Press, 1975.

[2] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

[3] Tony F. Chan and Chiu-Kwong Wong. Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7(3):370–375, 1998.

[4] R.P. Feynman and A.R. Hibbs. *Quantum Mechanics and Path Integrals*. McGraw-Hill, 1965.

[5] N.P. Galatsanos and A.K. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Trans. Image Processing*, 1(3):322–336, July 1992.

[6] S. Geman and D.Geman. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6:721–741, June 1984.

[7] D. Keren and M. Werman. Probabilistic analysis of regularization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(10):982–995, 1993.

[8] D. Keren and M. Werman. A full bayesian approach to curve and surface reconstruction. *Journal of Mathematical Imaging and Vision*, 11(1):27–43, 1999.

[9] F.M. Larkin. Gaussian measure in Hilbert space and applications in numerical analysis. *Rocky Mountain Journal of Mathematics*, 2:379–421, 1972.

[10] A.B. Lee, K.S. Pedersen, and D. Mumford. The nonlinear statistics of high-contrast patches in natural images. *International Journal of Computer Vision*, 54(1-3):83–103, August 2003.

[11] D. Lee. Approximation of linear operators on a wiener space. *Rocky Mountain Journal of Mathematics*, 16:641–659, 1986.

[12] M. Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20(1-2):99–120, 2004.

[13] M. Osadchy and D. Keren. A rejection-based method for event detection in video. *IEEE Trans. Circuits Syst. Video Techn.*, 14(4):534–541, 2004.

[14] M. Osadchy and D. Keren. Incorporating the boltzmann prior in object detection using svm. In *CVPR (2)*, pages 2095–2101, 2006.

[15] M. Osadchy, D. Keren, and C. Gotsman. Antifaces: A novel, fast method for image detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(7):747–761, 2001.

[16] S. Ramani, D. Van De Ville, T. Blu, and M. Unser. Nonideal sampling and regularization theory. *IEEE Transactions on Signal Processing*, 56(3):1055–1070, 2008.

[17] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Advanced Lectures on Machine Learning*, pages 63–71, 2003.

[18] Turner R.E. and Sahani M. Modeling natural sounds with modulation cascade processes. In *NIPS*, 2007.

[19] Turner R.E. and Sahani M. Statistical inference for single- and multi-band probabilistic amplitude demodulation. In *ICASSP*, 2010.

[20] L. S. Schulman. *Techniques and Applications of Path Integration*. John Wiley & Sons, 1981.

[21] R. Shankar. *Principles of Quantum Mechanics*. Springer, 1994.

[22] J. Stalker. *Complex Analysis: Fundamentals of the Classical Theory of Functions*. Birkhauser, 1998.

[23] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-Level Vision*. Kluwer, 1989.

[24] D. Terzopoulos. Multi-level surface reconstruction. In A. Rosenfeld, editor, *Multiresolution Image Processing and Analysis*. Springer-Verlag, 1984.

[25] A.N Tikhonov and V.Y Arsenin. *Solution of Ill-Posed Problems*. Winston and Sons, 1977.

[26] G. Wahba. Bayesian 'confidence intervals' for the cross-validated smoothing spline. *J. Roy. Stat. Soc. B.*, 45(1):133–150, 1983.

[27] Y. Weiss and W.T. Freeman. What makes a good model of natural images? In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

11   Author Information

Daniel Keren received the PhD degree from the Hebrew University in 1991, and then spent three years as a postdoctoral researcher at Brown University. Since then, he has been with the Computer Science Department at the University of Haifa. His main research interests include computer vision, regularization, and monitoring of large-scale distributed systems

Michael Werman Ph.D. 1986, The Hebrew University, currently professor of computer science at The Hebrew University. His research interest is mainly in designing computer algorithms and mathematical tools for analyzing, understanding and synthesizing pictures

Joshua Feinberg is a theoretical physicist, mainly working on non-perturbative quantum field theory and its applications. He's with the physics department of Haifa University at Oranim since 1997. Before that held post-doc positions at the Kavli Inst. of Theoretical Physics in UC Santa Barbara and at the theory group in the physics department in UT Austin.

Appendix I: the Gelfand-Yaglom Solution and its Application to the Path Integrals in this Paper

In a famous paper [2, 1], Gelfand and Yaglom expressed the solution of the path integral

$$\int \mathcal{D}x \exp\left(-\int \left(\lambda \dot{x}^2 + b(t)x^2\right)\right)$$

as

$$\sqrt{\frac{1}{2\pi f(1)}} \tag{1}$$

where f is the solution of the differential equation $\ddot{f} - 2b(t)f = 0$, with boundary conditions $f(0) = 0, \dot{f}(0) = 1$.

The Gelfand-Yaglom solution thus allows to compute the determinant of the operator in the exponent, or the infinite product of the eigenvalues. In our case, solving their differential equation will allow to compute the value of

$$\prod_{k=1}^{\infty} \left(1 - \frac{i\omega}{\alpha_k}\right) \tag{2}$$

There is, however, a delicate problem in applying this solution method to our path integrals. This problem is due to the square root term in the solution (Eq. 25). In typical quantum mechanics and statistical mechanics applications, the Lagrangian is either real or purely imaginary, and all the eigenvalues (and hence their product) are positive; hence, there is no ambiguity in computing the square root. Our path integrals, on the other hand, contain both real and imaginary parts in the exponent, hence the infinite

product in Eq. 26 has a non-zero phase. It is well-known that it is impossible to define a single square root branch for which

$$\sqrt{z_1 z_2 \ldots z_n} = \sqrt{z_1}\sqrt{z_2}\ldots\sqrt{z_n}$$

for every set of complex numbers $z_1, z_2 \ldots z_n$. Therefore, it is not possible to directly evaluate the square root of the solution of the differential equation. For example, if the determinant's phase is $\frac{\pi}{3}$, we cannot tell whether the phase of its square root is $\frac{\pi}{6}$ or $\frac{7\pi}{6}$. To overcome this problem, one can, for example, compute the total phase of the solution, i.e $\sum_{k=1}^{\infty} \arctan\left(\frac{\omega}{\alpha_k}\right)$, and divide it by 2. Clearly this total phase equals 0 for $\omega = 0$ (there is no imaginary part), and it has to be tracked from there. This method of solution uses the fact that the total phase is a continuous function of ω. The process can be made simpler by using the asymptotic properties of the eigenvalues α_k, as discussed in Section 5.1.

Appendix II: the Green's Functions Used in this Paper

The derivations in Section 7 make use of the Green's function of an operator which is the sum of a derivative term and a finite number of delta functions. Next we offer a brief overview of the Green's function for the derivative in one and two dimensions, and describe a general method for calculating the Green's function of the sum of a linear operator L and a finite number of delta functions, given the Green's function of L.

0.1 Green's Function of a Linear Operator

Given an operator L on a function space F, where as before the functions are denoted $x(t)$, the *Green's function* of L is a function $G(t_1, t_2)$ which satisfies $LG(t_1, t_2) = \delta(t_1 - t_2)$. It is perhaps easier to think about G as a parametric family of functions, indexed by t_1 (that is, $G(t_1, t_2) \triangleq G_{t_1}(t_2)$), which satisfies $LG_{t_1} = \delta_{t_1}$. For example, in the function space dealt with in this paper (functions $x(t)$ on the unit interval, with boundary conditions $x(0) = x(1) = 0$), the Green's function of the operator $-\nabla^2$, which is used in Section 7, is the piecewise linear function

$$G(t_1, t_2) = \begin{cases} (1 - t_1)t_2 & \text{if } t_2 \geq t_1 \\ (1 - t_2)t_1 & \text{if } t_1 \geq t_2 \end{cases}$$

it can easily be checked that, for a fixed t_1, G_{t_1} satisfies the boundary condition and that minus its second derivative is indeed a delta function at t_1, hence it satisfies the requirement for the Green's functions of minus the second derivative.

0.2 Green's Function for Higher-Order Smoothness Terms and Dimensions Higher than 1

It is easy to verify that if a linear operator L admits a set of eigenfunctions Φ_k with eigenvalues λ_k, which also satisfy the so-called *completeness relation* $\sum_k \Phi_k(t_1)\Phi_k(t_2) = \delta(t_1 - t_2)$, then L's Green's function equals $\sum_k \frac{\Phi_k(t_1)\Phi_k(t_2)}{\lambda_k}$. Hence if we're using a second order smoothness term in one dimension (yielding $L(x) = \frac{d^4x}{dt^4}$) then $G(t_1, t_2) = \sum_k \frac{\sin(kt_1)\sin(kt_2)}{\pi^4 k^4}$. Similarly, in two dimensions, the second-order smoothness term yields $L = \nabla^4$, or $L(x) =$

$\frac{\partial^4 x}{\partial^4 t} + 2\frac{\partial^4 x}{\partial^2 t \partial^2 s} + \frac{\partial^4 x}{\partial^4 s}$, and $G((t_1, s_1), (t_2, s_2)) = \sum_{k,l} \frac{\sin(k\pi t_1)\sin(l\pi s_1)\sin(k\pi t_2)\sin(l\pi s_2)}{(k^2+l^2)^2}$.

0.3   Adding Delta Functions

In Section 7 the general solution of computing the p.d.f (in any dimension) necessitated the computation of the Green's function of an operator which is the sum of $-\nabla^2$ or ∇^4 and delta functions at the sample points. Generally, assume G is the Green's function of an operator L, and we need to compute the Green's function of L plus delta functions. For simplicity, assume only two deltas are present (the general case is solvable in exactly the same manner). Denote $\mathcal{L} = L + \delta_{t_1} + \delta_{t_2}$, and define $\mathcal{G}_t = G_t + \alpha_1 G_{t_1} + \alpha_2 G_{t_2}$. We will now show that for a correct choice of α_1 and α_2, $\mathcal{G}$ is a Green's function for $\mathcal{L}$. We must have $\mathcal{L}\mathcal{G}_t = \delta_t$. Applying the definition of $\mathcal{L}$ and $\mathcal{G}_t$:

$\mathcal{L}\mathcal{G}_t = (L + \delta_{t_1} + \delta_{t_2})(G_t + \alpha_1 G_{t_1} + \alpha_2 G_{t_2}) =$

$LG_t + \alpha_1 LG_{t_1} + \alpha_2 LG_{t_2} + G_t(t_1)\delta_{t_1} + \alpha_1 G_{t_1}(t_1)\delta_{t_1} + \alpha_2 G_{t_2}(t_1)\delta_{t_1} +$

$G_t(t_2)\delta_{t_2} + \alpha_1 G_{t_1}(t_2)\delta_{t_2} + \alpha_2 G_{t_2}(t_2)\delta_{t_2} =$

$\delta_t + (\alpha_1 + G_t(t_1) + \alpha_1 G_{t_1}(t_1) + \alpha_2 G_{t_2}(t_1))\delta_{t_1} + (\alpha_2 + G_t(t_2) + \alpha_1 G_{t_1}(t_2) + \alpha_2 G_{t_2}(t_2))\delta_{t_2}$

all that remains now is to choose α_1, α_2 which solve the set of equations

$\alpha_1 + G_t(t_1) + \alpha_1 G_{t_1}(t_1) + \alpha_2 G_{t_2}(t_1) = 0, \quad \alpha_2 + G_t(t_2) + \alpha_1 G_{t_1}(t_2) + \alpha_2 G_{t_2}(t_2) = 0$

this method works for any dimension and choice of smoothness term.

References

[1] I.M. Gel'and and A.M. Yaglom. Integration in functional spaces and its applications in quantum physics. *Uspekhi Mat. Nauk*, 11(1-67):77–114,

1956.

[2] I.M. Gel'and and A.M. Yaglom. Integration in functional spaces and its applications in quantum physics. *Journal of Mathematical Physics*, 1(1):48–69, 1960.