mathematical methods - week 7

Method of steepest descent

Georgia Tech PHYS-6124

Homework HW #7 due Thursday, October 8, 2020

== show all your work for maximum credit,
== put labels, title, legends on any graphs
== acknowledge study group member, if collective effort
== if you are LaTeXing, here is the source code

Exercise 7.1 *In high dimensions any two vectors are (nearly) orthogonal* 16 points

Bonus points
Exercise 7.2 *Airy function for large arguments* 10 points

Total of 16 points = 100 % score. Extra points accumulate, can help you later if you miss a few problems.

edited October 10, 2020

Week 7 syllabus   September 29, 2020

Arfken, Weber & Harris [1] Chapter 12 *Further Topics in Analysis*; (click here) Chapter 13 *Gamma function* (click here). saddle point method; Gamma, Airy function estimates; beta function is also often encountered.

- *Asymptotic evaluation of integrals: perturbation expansions; Laplace, saddle, steepest descent leading term.*
  - AWH 12.6 *Asymptotic series*
  - AWH 12.7 *Method of steepest descents*
  - Grigoriev lecture notes
- *Steepest descent I: Gamma function, Sterling formula.*
- *Steepest descent II, for physicists: Zero-dimensional field theory - perturbation expansion is an asymptotic series.*
  - Sect. 7.1 *Saddle-point expansions are asymptotic*
- *Steepest descent III, for data scientists: How tall is my graduate student?*
  - Exercise 7.1 *In high dimensions any two vectors are (nearly) orthogonal*

Optional reading

- AWH 11.6 *Singularities*; Branch-cut integrals
- *If they only got the phase in the Fresnel integral right, QM would look different*
- *Got problems? Do them like a journalist*
- *I heard it through the grapevine: how to pick a tolerable adviser?*
- *You think you are stressed? Try finishing your thesis*

7.1 Saddle-point expansions are asymptotic

The first trial ground for testing our hunches about field theory is the *zero-dimensional field theory*, the field theory of a lattice consisting of one point, in case of the "ϕ^4 theory" given a humble 1-dimensional integral

$$Z[J] = \int \frac{d\phi}{\sqrt{2\pi}} e^{-\phi^2/2 - g\phi^4/4 + \phi J} \,. \tag{7.1}$$

The idea of the saddle-point expansions is to keep the Gaussian part $\phi^2/2$ ("free field", with a quadratic H_0 "Hamiltonian") as is, and expand the rest (H_I "interacting Hamiltonian") as a power series, and then evaluate the perturbative corrections using the moments formula

$$\int \frac{d\phi}{\sqrt{2\pi}} \phi^n e^{-\phi^2/2} = \left(\frac{d}{dJ}\right)^n e^{J^2/2}\Big|_{J=0} = (n-1)!! \quad \text{if } n \text{ even, } 0 \text{ otherwise} \,.$$

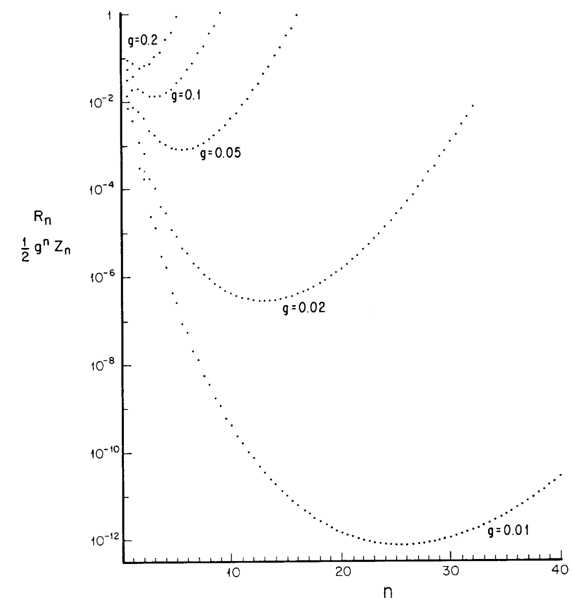

Figure 7.1: Plot of the saddle-point estimate of Z_n vs. the exact result (7.2) for $g = 0.1$, $g = 0.02$, $g = 0.01$.

In this zero-dimensional theory the n-point correlation is a number exploding combinatorially, as $(n-1)!!$. And here our troubles start.

To be concrete, let us work out the *exact* zero-dimensional ϕ^4 field theory in the saddle-point expansion to all orders:

$$\begin{aligned} Z[0] &= \sum_n Z_n g^n \,, \\ Z_n &= \frac{(-1)^n}{n! 4^n} \int \frac{d\phi}{\sqrt{2\pi}} \phi^{4n} e^{-\phi^2/2} = \frac{(-1)^n}{16^n n!} \frac{(4n)!}{(2n)!} \,. \end{aligned} \tag{7.2}$$

The Stirling formula $n! = \sqrt{2\pi}\, n^{n+1/2} e^{-n}$ yields for large n

$$g^n Z_n \approx \frac{1}{\sqrt{n\pi}} \left(\frac{4g}{e} n \right)^n \,. \tag{7.3}$$

As the coefficients of the parameter g^n are blowing up combinatorially, no matter how small g might be, the perturbation expansion is not convergent! Why? Consider again (7.1). We have tacitly assumed that $g > 0$, but for $g < 0$, the potential is unbounded for large ϕ, and the integrand explodes. Hence the partition function in not analytic at the $g = 0$ point.

Is the whole enterprise hopeless? As we shall now show, even though divergent, the perturbation series is an *asymptotic* expansion, and an asymptotic expansion can be extremely good [6]. Consider the residual error after inclusion of the first n perturbative corrections:

$$
\begin{aligned}
R_n &= \left| Z(g) - \sum_{m=0}^{n} g^m Z_m \right| \\
&= \int \frac{d\phi}{\sqrt{2\pi}} e^{-\phi^2/2} \left| e^{-g\phi^4/4} - \sum_{m=0}^{n} \frac{1}{m!} \left(-\frac{g}{4}\right)^m \phi^{4m} \right| \\
&\leq \int \frac{d\phi}{\sqrt{2\pi}} e^{-\phi^2/2} \frac{1}{(n+1)!} \left(\frac{g\phi^4}{4}\right)^{n+1} = g^{n+1} \left|Z_{n+1}\right| . \qquad (7.4)
\end{aligned}
$$

The inequality follows from the convexity of exponentials, a generalization of the inequality $e^x \geq 1 + x$. The error decreases as long as $g^n |Z_n|$ decreases. From (7.3) the minimum is reached at $4g\, n_{min} \approx 1$, with the minimum error

$$
g^n Z_n|_{min} \approx \sqrt{\frac{4g}{\pi}}\, e^{-1/4g}. \qquad (7.5)
$$

As illustrated by the figure 7.1, a perturbative expansion can be, for all practical purposes, very accurate. In Quantum ElectroDynamics, or QED, this argument had led Dyson to suggest that the QED perturbation expansions are good to $n_{min} \approx 1/\alpha \approx 137$ terms. Due to the complicated relativistic, spinorial and gauge invariance structure of perturbative QED, there is not a shred of evidence that this is so. The very best calculations performed so far stop at $n \leq 5$.

Predrag I find Córdova, Heidenreich, Popolitov and Shakirov [4] *Orbifolds and exact solutions of strongly-coupled matrix models* very surprising. The introduction is worth reading. They compute analytically the matrix model (QFT in zero dimensions) partition function for trace potential

$$
S[X] = \mathrm{tr}\,(X^r)\,, \qquad \text{integer } r \geq 2\,. \qquad (7.6)
$$

Their "non-perturbative ambiguity" in the case of the $N = 1$ cubic matrix model seem to amount to the Stokes phenomenon, i.e., choice of integration contour for the Airy function.

Unlike the weak coupling expansions, the strong coupling expansion of

$$
Z = \frac{1}{2\pi} \int dx e^{-\frac{1}{2g^2}x^2 - x^4}\,, \qquad (7.7)
$$

is convergent, not an asymptotic series.

There is a negative dimensions type duality $N \to -N$, their eq. (3.27). The loop equations, their eq. (2.10), are also interesting - they seem to essentially be the Dyson-Schwinger equations and Ward identities in my book's [5] formulation of QFT.

7.2 Notes on life in extreme dimensions

You can safely ignore this section, it's "math methods," as much as Predrag's musings about current research...

Exercise 7.1 is something that anyone interested in computational neuroscience [9] and/or machine learning already knows. It is also something that many a contemporary physicist *should* know; a daily problem for all of us, from astrophysics to fluid physics to biologically inspired physics is how to visualize large, extremely large data sets.

Possibly helpful references:

Distribution of dot products between two random unit vectors. They denote $Z = \langle X, Y \rangle = \sum X_i Y_i$. Define

$$f_{Z_i}(z_i) = \int_{-\infty}^{\infty} f_{X_i,Y_i}(x, \frac{z_i}{x}) \frac{1}{|x|} dx$$

then since $Z = \sum Z_i$,

$$f_Z(z) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{Z_1,\ldots,Z_D}(z_1, \ldots, z_d)\, \delta(z - \sum z_i)\, dz_1 \ldots dz_d\,.$$

There is a Georgia Tech paper on this [12]. See also cosine similarity and Mathworld. There is even a python tutorial. `scikit-learn` is supposed to be 'The de facto Machine Learning package for Python'.

Remark 7.1. High-dimensional flows and their visualizations. Dynamicist's vision of turbulence was formulated by Eberhard Hopf in his seminal 1948 paper [11]. Computational neuroscience grapples with closely related visualization and modeling issues [7, 8]. Much about high-dimensional state spaces is counterintuitive. The literature on why the expectation value of the angle between any two high-dimensional vectors picked at random is 90^o is mostly about spikey spheres: see the draft of the Hopcroft and Kannan [3] book and Ravi Kannan's course; lecture notes by Hermann Flaschka on *Some geometry in high-dimensional spaces*; Wegman and Solka [13] visualizations of high-dimensional data; Spruill paper [12]; a lively mathoverflow.org thread on "Intuitive crutches for higher dimensional thinking."

The 'good' coordinates, introduced in ref. [10] are akin in spirit to the low-dimensional projections of the POD modeling [2], in that both methods aim to capture key features and dynamics of the system in just a few dimensions. But the ref. [10] method is very different from POD in a key way: we construct basis sets from *exact solutions of the fully-resolved dynamics* rather than from the empirical eigenfunctions of the POD. Exact solutions and their linear stability modes (a) characterize the spatially-extended states precisely, as opposed to the truncated expansions of the POD, (b) allow for different basis sets and projections for different purposes and different regions of state space, (c) these low-dimensional projections are not meant to suggest low-dimensional ODE models; they are only visualizations, every point in these projections is still a point the full state space, and (d) the method is not limited to Fourier mode bases.

References

[1] G. B. Arfken, H. J. Weber, and F. E. Harris, *Mathematical Methods for Physicists: A Comprehensive Guide*, 7th ed. (Academic, New York, 2013).

[2] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, "The dynamics of coherent structures in the wall region of turbulent boundary layer", J. Fluid Mech. **192**, 115–173 (1988).

[3] A. Blum, J. Hopcroft, and R. Kannan, *Foundations of Data Science* (Cambridge Univ. Press, Cambridge UK, 2020).

[4] C. Córdova, B. Heidenreich, A. Popolitov, and S. Shakirov, "Orbifolds and exact solutions of strongly-coupled matrix models", Commun Math Phys **361**, 1235–1274 (2018).

[5] P. Cvitanović, *Field Theory*, Notes prepared by E. Gyldenkerne (Nordita, Copenhagen, 1983).

[6] R. B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic Press, London, 1973).

[7] A. Farshchian, J. A. Gallego, J. P. Cohen, Y. Bengio, L. E. Miller, and S. A. Solla, Adversarial domain adaptation for stable brain-machine interfaces, in International Conference on Learning Representations (2019), pp. 1–14.

[8] J. A. Gallego, M. G. Perich, R. H. Chowdhury, S. A. Solla, and L. E. Miller, "Long-term stability of cortical population dynamics underlying consistent behavior", Nature Neuroscience **23**, 260–270 (2020).

[9] J. A. Gallego, M. G. Perich, S. N. Naufel, C. Ethier, S. A. Solla, and L. E. Miller, "Cortical population activity within a preserved neural manifold underlies multiple motor behaviors", Nat. Commun. **9**, 4233 (2018).

[10] J. F. Gibson, J. Halcrow, and P. Cvitanović, "Visualizing the geometry of state-space in plane Couette flow", J. Fluid Mech. **611**, 107–130 (2008).

[11] E. Hopf, "A mathematical example displaying features of turbulence", Commun. Pure Appl. Math. **1**, 303–322 (1948).

[12] M. C. Spruill, "Asymptotic distribution of coordinates on high dimensional spheres", Elect. Comm. in Probab. **12**, 234–247 (2007).

[13] E. J. Wegman and J. L. Solka, "On some mathematics for visualizing high dimensional data", Sankhya: Indian J. Statistics, Ser. A **64**, 429–452 (2002).

Exercises

7.1. **In high dimensions any two vectors are (nearly) orthogonal.** Among humble plumbers laboring with extremely high-dimensional ODE discretizations of fluid and other PDEs, there is an inclination to visualize the ∞-dimensional state space flow by projecting it onto a basis constructed from a few random coordinates, let's say the 2nd Fourier mode along the spatial x direction against the 4th Chebyshev mode along the y direction. It's easy, as these are typically the computational degrees of freedom. As we will now show, it's easy but not smart, with vectors representing the dynamical states of interest being almost orthogonal to any such random basis.

Suppose your state space $\mathcal{M}$ is a real 10 247-dimensional vector space, and you pick from it two vectors $x_1, x_2 \in \mathcal{M}$ at random. What is the angle between them likely to be?

In the literature you might run into this question, formulated as the 'cosine similarity'

$$\cos(\theta_{12}) = \frac{{x_1}^\top \cdot x_2}{|\,x_1\,|\,|\,x_2\,|}\,. \tag{7.8}$$

Two vectors with the same orientation have a cosine similarity of 1, two vectors at 90^o have a similarity of 0, and two vectors diametrically opposed have a similarity of -1. By asking for 'angle between two vectors' we have implicitly assumed that there exist is a dot product

$$ {x_1}^\top \cdot x_2 = |\,x_1\,|\,|\,x_2\,|\cos(\theta_{12})\,, $$

so let's make these vectors unit vectors, $|\,x_j\,| = 1$. When you think about it, you would be hard put to say what 'uniform probability' would mean for a vector $x \in \mathcal{M} = \mathbb{R}^{10\,247}$, but for a unit vector it is obvious: probability that x direction lies within a solid angle $d\Omega$ is $d\Omega/(\text{unit hyper-sphere surface})$.

So what is the surface of the unit sphere (or, the total solid angle) in d dimensions? One way to compute it is to evaluate the Gaussian integral

$$ I_d = \int_{-\infty}^{\infty} dx_1 \cdots dx_d \, e^{-\frac{1}{2}\left(x_1^2+\cdots+x_d^2\right)} \tag{7.9}$$

in cartesian and polar coordinates. Show that

(a) In cartesian coordinates $I_d = (2\pi)^{d/2}$.

(b) Show, by examining the form of the integrand in the polar coordinates, that for an arbitrary, even complex dimension $d \in \mathbb{C}$

$$ S_{d-1} = 2\pi^{d/2}/\Gamma(d/2)\,. \tag{7.10}$$

In QFT, or Quantum Field Theory, integrals over 4-momenta are brought to polar form and evaluated as functions of a complex dimension parameter d. This procedure is called the 'dimensional regularization'.

(c) Recast the integrals in polar coordinate form. You know how to compute this integral in 2 and 3 dimensions. Show by induction that the surface S_{d-1} of unit d-ball, or the total solid angle in even and odd dimensions is given by

$$ S_{2k} = \frac{2(2\pi)^k}{(2k-1)!!}\,, \qquad S_{2k+1} = \frac{2\pi^{k+1}}{k!}\,. \tag{7.11}$$

However irritating to Data Scientists (these are just the Gamma function (7.10) written out as factorials), the distinction between even and odd dimensions is not silly - in Cartan's classification of all compact Lie groups, special orhtogonal groups $\mathrm{SO}(2k)$ and $\mathrm{SO}(2k{+}1)$ belong to two distinct infinite families of special orthogonal symmetry groups, with implications for physics in 2, 3 and 4 dimensions. For example, by the hairy ball theorem, there can be no non-vanishing continuous tangent vector field on even-dimensional d-spheres; you cannot smoothly comb hair on a 3-dimensional ball.

(d) Check your formula for $d = 2$ (1-sphere, or the circle) and $d = 3$ (2-sphere, or the sphere).

(e) What limit does S_d does tend to for large d? (Hint: it's not what you think. Try Sterling's formula).

So now that we know the volume of a sphere, what is a the most likely angle between two vectors x_1, x_2 picked at random? We can rotate coordinates so that x_1 is aligned with the 'z-axis' of the hypersphere. An angle θ then defines a meridian around the 'z-axis'.

(f) Show that probability $P(\theta)d\theta$ of finding two vectors at angle θ is given by the area of the meridional strip of width $d\theta$, and derive the formula for it:

$$P(\theta) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(d/2)}{\Gamma((d-1)/2)} .$$

(One can write analytic expression for this in terms of beta functions, but it is unnecessary for the problem at hand).

(g) Show that for large d the probability $P(\theta)$ tends to a normal distribution with mean $\theta = \pi/2$ and variance $1/d$.

So, in d-dimensional vector space the two random vectors are nearly orthogonal, within accuracy of $\theta = \pi/2 \pm 1/d$.

Null distribution: For data which can be negative as well as positive, the null distribution for cosine similarity is the distribution of the dot product of two independent random unit vectors. This distribution has a mean of zero and a variance of $1/d$ (where d is the number of dimensions), and although the distribution is bounded between -1 and +1, as d grows large the distribution is increasingly well-approximated by the normal distribution.

In high dimensions any two vectors are (nearly) orthogonal - If I am 2 meters tall, how tall does a graduate student look to me, if grad students are randomly distributed in a million directions?

If you are a humble plumber simulating turbulence, and trying to visualize its state space and the notion of a vector space is some abstract hocus-pocus to you, try thinking this way. Your 2nd Fourier mode basis vector is something that wiggles twice along your computation domain. Your turbulent state is very wiggly. The product of the two functions integrated over the computational domain will average to zero, with a small leftover. We have just estimated that with dumb choices of coordinate bases this leftover will be of order of $1/10\,247$, which is embarrassingly small for displaying a phenomenon of order ≈ 1.

Several intelligent choices of coordinates for state space projections are described in ChaosBook section 2.4, the web tutorial ChaosBook.org/tutorials, and Gibson *et al.* [10].

Sara A. Solla and P. Cvitanović

7.2. **Airy function for large arguments.** Important contributions as stationary phase points may arise from extremal points where the first non-zero term in a Taylor expansion of the phase is of third or higher order. Such situations occur, for example, at bifurcation points or in diffraction effects, such as waves near sharp corners, waves creeping around obstacles, etc.. In such calculations, one meets Airy functions integrals of the form

$$Ai(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dy\, e^{i(xy - \frac{y^3}{3})} . \tag{7.12}$$

Calculate the Airy function $Ai(x)$ using the stationary phase approximation. What happens when considering the limit $x \to 0$? Estimate for which value of x the stationary phase approximation breaks down.