Math 408

L-functions and Sphere Packing

Steven Miller: sjm1@williams.edu

https://web.williams.edu/Mathematics/sjmiller/public_html/408Fa20/

Lecture 36: December 11, 2020

Calculus of Variations

New upper bounds on sphere packings I

By Henry Cohn and Noam Elkies*

Abstract

We develop an analogue for sphere packing of the linear programming bounds for error-correcting codes, and use it to prove upper bounds for the density of sphere packings, which are the best bounds known at least for dimensions 4 through 36. We conjecture that our approach can be used to solve the sphere packing problem in dimensions 8 and 24.

Contents

https://annals.math.princeton.edu/wp-content/uploads/annals-v157-n2-p09.pdf

We do not know how to use Theorem 3.1 to match the best density bound known in high dimensions, that of Kabatiansky and Levenshtein [KL]. However, it provides a new proof of the second-best bound known, due to Levenshtein [Lev]:

$$\Delta \leq \frac{j_{n/2}^{n}}{(n/2)!^2 4^n},$$

where j_t is the smallest positive zero of the Bessel function J_t. (For more information about the asymptotics of this bound and how it compares with other bounds, see page 19 of [CS], but note that equation (42) is missing the exponent in $j_{n/2}^{n}$.) We will show how to use a calculus of variations argument to find functions that prove that bound. This approach is analogous to that used by Levenshtein. Yudin [Y] has also given a proof of Levenshtein's bound that seems reminiscent of our general approach, but not identical.

https://politics.theonion.com/cia-realizes-its-been-using-black-highlighters-all-thes-1819568147

Optimization in one dimension:

function $f(x)$ is: continuous, differentiable

Take derivative, find critical points: x st $f'(x) = 0$

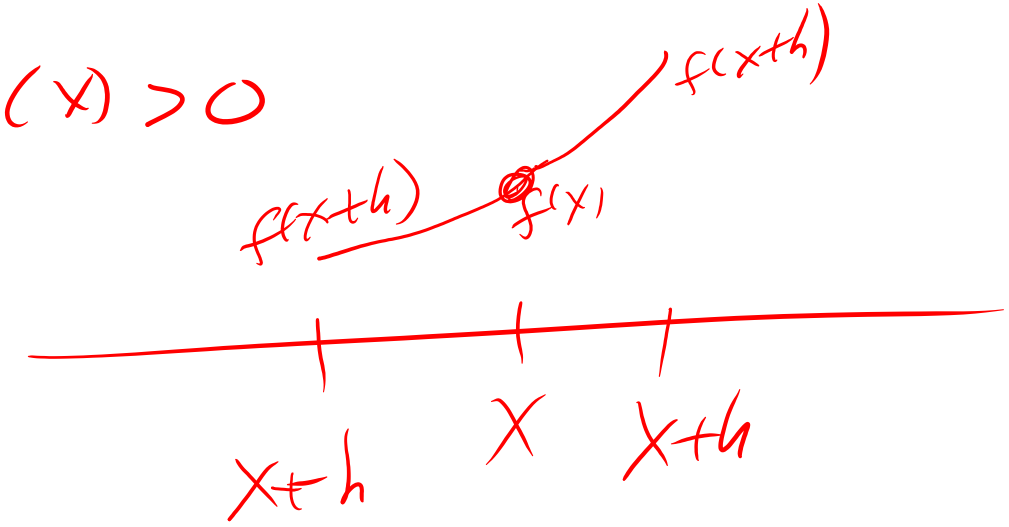
ENDPOINTS: at most 2 in one-dimension

Much is the algebra in solving $\underbrace{f'(x) = 0}_{\text{calculus}}$

Optimization in one dimension:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

if $f'(x) > 0$

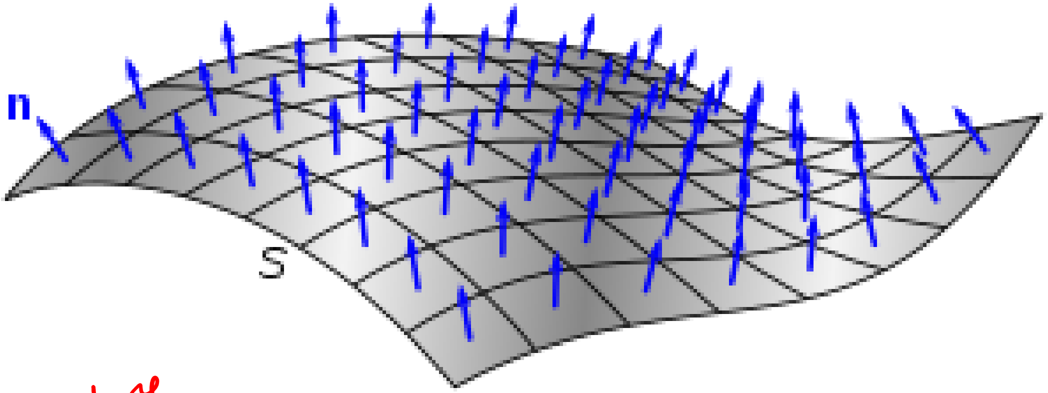


no max or min

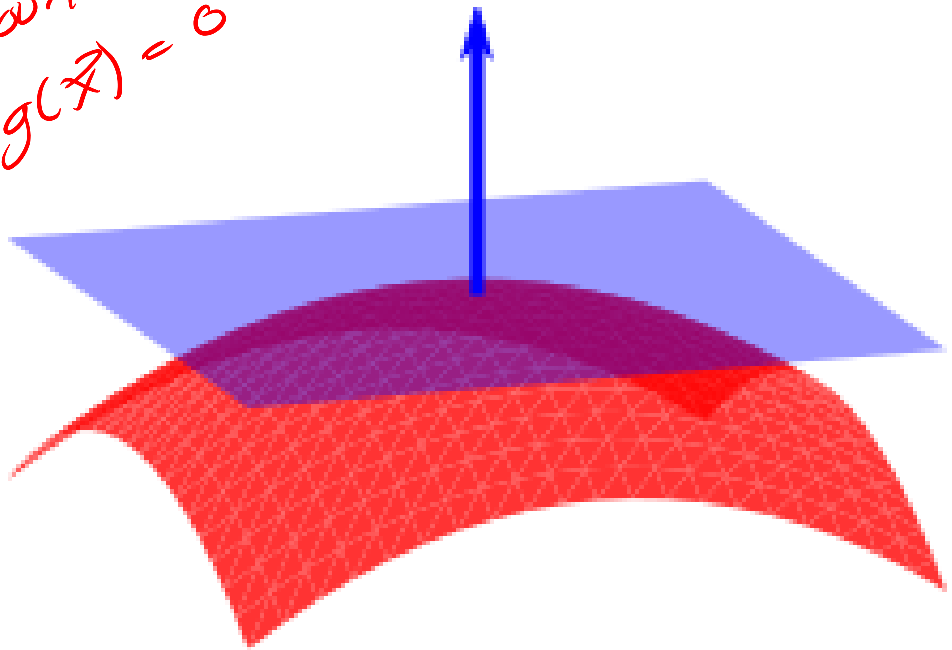
similar result if $f'(x) < 0$

So $f'(x) = 0$ is a necessary but not sufficient condition for a max/min.

Optimization in several dimensions:



Boundary
$g(\vec{x}) = 0$



https://en.wikipedia.org/wiki/Normal_(geometry)

Interior: $\nabla f(\vec{x}) = \vec{0}$

$\left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right) = \nabla f(\vec{x})$

Boundary: $C: [0,1] \to$ Surface

$g(t) = f(c(t))$ one-dim

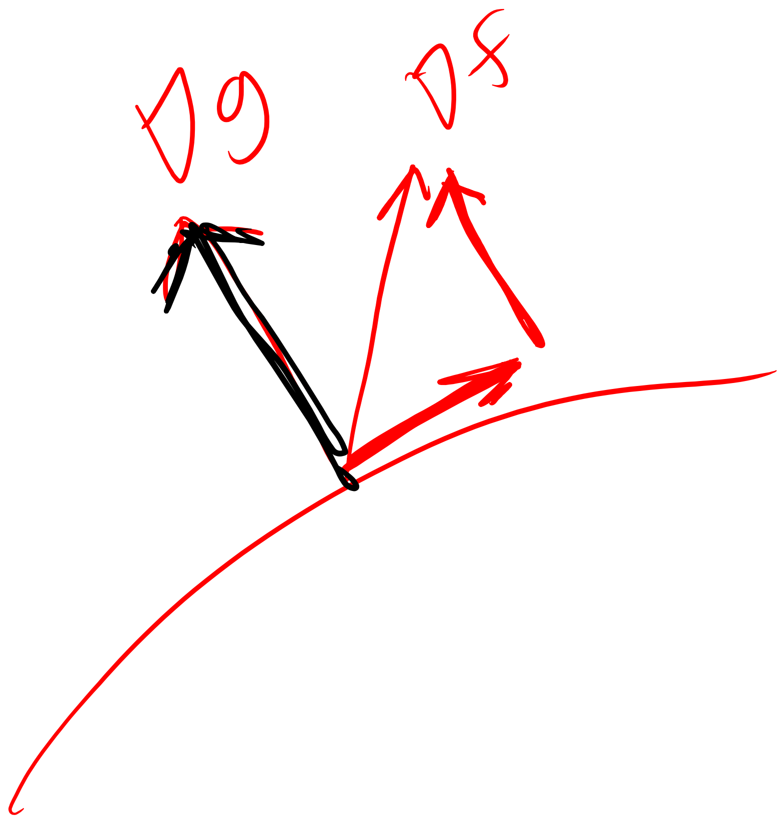
$g'(x) = (\nabla f)(c(t)) \cdot c'(t)$

Directional Derivative

Lagrange Multipliers: $\nabla f = \lambda \nabla g$ and $g(\vec{x}) = 0$

n+1 vars: $x_1, \dots, x_n, \lambda$

n+1 eqs: $\partial f / \partial x_i = \lambda\, \partial g / \partial x_i$ and $g(\vec{x}) = \vec{0}$



$\nabla f \neq \lambda \nabla g$

can't be a min/max

need ∇f and ∇g to be parallel

Advanced: Think about say $g_1(\vec{x}) = \cdots = g_k(\vec{x}) = 0$

Differentiability

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

- What is differentiability in several variables?
- Is it enough for the partials $\partial f/\partial x_i$ to exist?

Differentiability

- What is differentiability in several variables?

- Is it enough for the partials $\partial f/\partial x_i$ to exist?

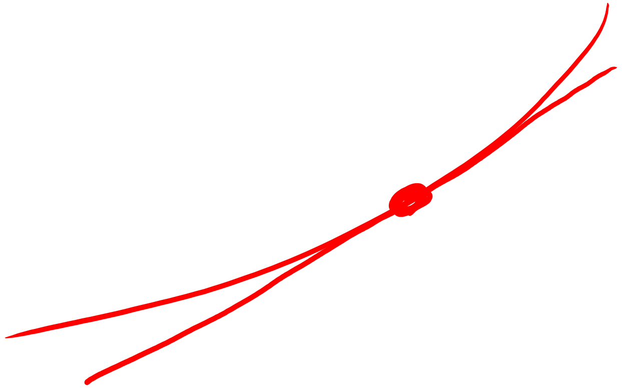
Consider $f(x,y) = (xy)^{1/3}$.

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} = 0$$

Note the partials are zero but if we consider $y = x^2$ then along this path the derivative is not zero (chain rule…). So partials are not enough.

Derivative in one dimension:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$



$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

← value → ← tangent line approximation →

$$\lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x-a)}{x-a} = 0$$

$y = f(a) + f'(a)(x-a)$

Derivative in several dimensions:

$$\lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x-a)}{x-a} = 0$$

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} \frac{||f(\mathbf{x}) - f(\boldsymbol{a}) - \nabla f(\boldsymbol{a}) \bullet (\boldsymbol{x}-\boldsymbol{a})||}{||\boldsymbol{x}-\boldsymbol{a}||} = 0$$

What are the derivs of

$f(x) = x^2$

$f'(x) = 2x$

$f(x+h) = (x+h)^2$

Binomial Theorem

$(x+h)^n = x^n + nx^{n-1}h +$

terms with $h^2, h^3, \ldots$

$f(x) = x^{3/2}$

$f'(x) = \frac{3}{2} x^{\frac{1}{2}}$

Can't use binomial thm!

$g(x) = f(x)^2 = x^3$

$g'(x) = 2f(x)f'(x) = 3x^2$

$f'(x) = \frac{3x^2}{2f(x)} \ldots$

$$\frac{(x+h)^{3/2} - x^{3/2}}{h} * \frac{(x+h)^{\frac{3}{2}} + x^{\frac{3}{2}}}{(x+h)^{\frac{3}{2}} + x^{3/2}}$$

top $(x+h)^3 - x^3$: Binomial thm

lim bottom is $2x^{3/2}h$

$f(x) = x^{\sqrt{2}}$

$f'(x) = \sqrt{2}\, x^{\sqrt{2}-1}$

$f(x) = x^{\sqrt{2}} = e^{\sqrt{2}\log x}$

Chain Rule

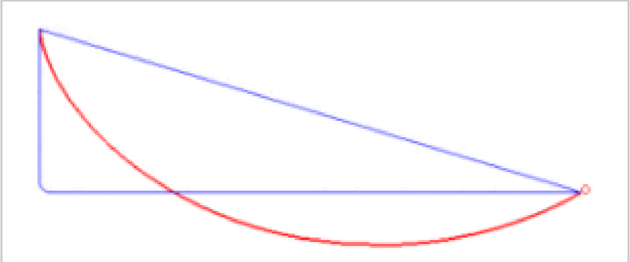
Rule:

$$\boxed{(x^n)' = nx^{n-1}}$$

Brachistochrone curve

From Wikipedia, the free encyclopedia

In mathematics and physics, a **brachistochrone curve** (from Ancient Greek βράχιστος χρόνος *(brákhistos khrónos)* 'shortest time'),[1] or curve of fastest descent, is the one lying on the plane between a point *A* and a lower point *B*, where *B* is not directly below *A*, on which a bead slides frictionlessly under the influence of a uniform gravitational field to a given end point in the shortest time. The problem was posed by Johann Bernoulli in 1696.



The curve of fastest descent is not a straight or polygonal line (blue) but a cycloid (red).

The brachistochrone curve is the same shape as the tautochrone curve; both are cycloids. However, the portion of the cycloid used for each of the two varies. More specifically, the brachistochrone can use up to a complete rotation of the cycloid (at the limit when A and B are at the same level), but always starts at a cusp. In contrast, the tautochrone problem can only use up to the first half rotation, and always ends at the horizontal.[2] The problem can be solved using tools from the calculus of variations and optimal control.[3]

The curve is independent of both the mass of the test body and the local strength of gravity. Only a parameter is chosen so that the curve fits the starting point *A* and the ending point *B*.[4] If the body is given an initial velocity at *A*, or if friction is taken into account, then the curve that minimizes time will differ from the tautochrone curve.

https://en.wikipedia.org/wiki/Brachistochrone_curve

Johann and his brother Jakob Bernoulli derived the same solution, but Johann's derivation was incorrect, and he tried to pass off Jakob's solution as his own.[7] Johann published the solution in the journal in May of the following year, and noted that the solution is the same curve as Huygens's tautochrone curve. After deriving the differential equation for the curve by the method given below, he went on to show that it does yield a cycloid.[8][9] However, his proof is marred by his use of a single constant instead of the three constants, v_m, $2g$ and D, below.

Bernoulli allowed six months for the solutions but none were received during this period. At the request of Leibniz, the time was publicly extended for a year and a half.[10] At 4 p.m. on 29 January 1697 when he arrived home from the Royal Mint, Isaac Newton found the challenge in a letter from Johann Bernoulli.[11] Newton stayed up all night to solve it and mailed the solution anonymously by the next post. Upon reading the solution, Bernoulli immediately recognized its author, exclaiming that he "recognizes a lion from his claw mark". This story gives some idea of Newton's power, since Johann Bernoulli took two weeks to solve it.[4][12] Newton also wrote, "I do not love to be dunned [pestered] and teased by foreigners about mathematical things...", and Newton had already solved Newton's minimal resistance problem, which is considered the first of the kind in calculus of variations.

https://en.wikipedia.org/wiki/Brachistochrone_curve

Theorem 1 (Fundamental Lemma of the Calculus of Variations). *Let $f : [0,1] \to \mathbb{R}^n$ be a continuous function which obeys*

$$\int_0^1 \langle f(t), h(t) \rangle \, dt = 0$$

for all C^2 functions $h : [0,1] \to \mathbb{R}^n$ with $h(0) = h(1) = 0$. Then $f \equiv 0$.

https://www.maths.ed.ac.uk/~jmf/Teaching/Lectures/CoV.pdf

$$\int_0^1 f(t) \cdot h(t) \, dt = 0$$

$h(0) = h(1) = 0$

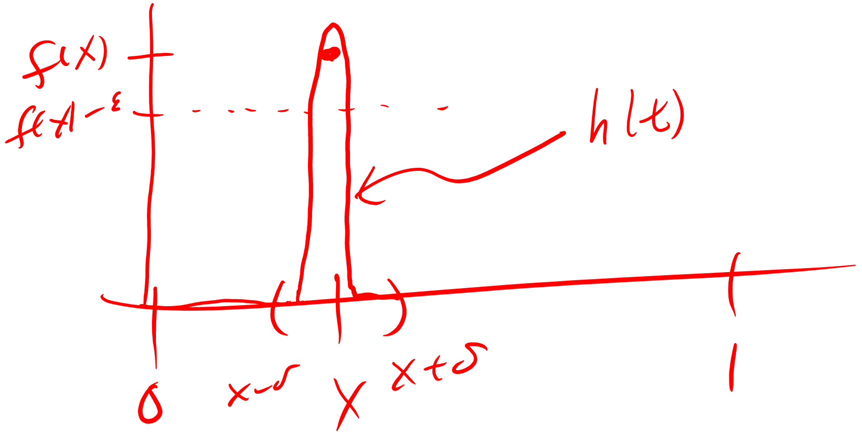
h twice cont diff

Theorem 1 (Fundamental Lemma of the Calculus of Variations). *Let $f : [0,1] \to \mathbb{R}^n$ be a continuous function which obeys*

$$\int_0^1 \langle f(t), h(t) \rangle \, dt = 0$$

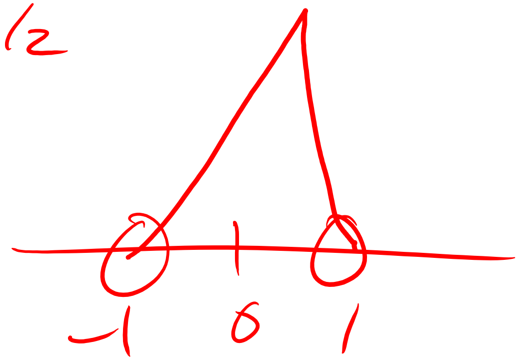
for all C^2 functions $h : [0,1] \to \mathbb{R}^n$ with $h(0) = h(1) = 0$. Then $f \equiv 0$.

Proof: Assume $\exists\, x$ st $f(x) \neq 0$ (w/log $f(x) > 0$) Doing $n=1$ but the more generally

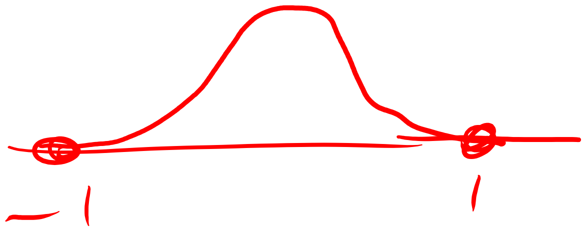


$h(t) = 0$ if $|x-t| > \delta/2$

$h(t) \geq 0$ spike



Test function $h(x) = \begin{cases} e^{\frac{1}{x^2-1}} & \text{for } |x|<1 \\ 0 & \text{for } |x| \geq 1 \end{cases}$



5. The Euler–Lagrange equation

Let $\mathcal{C}_{P,Q}$ be the space of C^2 curves $x : [0,1] \to \mathbb{R}^n$ with $x(0) = P$ and $x(1) = Q$. Let $L : \mathbb{R}^{2n+1} \to \mathbb{R}$ be a sufficiently differentiable function (typically smooth in applications) and let us consider the functional $S : \mathcal{C}_{P,Q} \to \mathbb{R}$ defined by

$$S[x] = \int_0^1 L(x(t), \dot{x}(t), t)\, dt \ .$$

The function L is called the **lagrangian** and the functional S is called the **action**. Extremising S will yield a differential equation for x. Recall that a path x is a critical point for the action if, for all endpoint-fixed variations ε, we have

$$\frac{d}{ds} S[x + s\varepsilon]\bigg|_{s=0} = 0 \ .$$

Differentiating under the integral sign, we find

$$S[x] = \int_0^1 L(x(t), \dot{x}(t), t)\, dt\ .$$

Differentiating under the integral sign, we find

$$\begin{aligned}
0 &= \int_0^1 \frac{d}{ds} L(x + s\varepsilon, \dot{x} + s\dot{\varepsilon}, t)\Big|_{s=0} dt \\
&= \int_0^1 \left(\sum_{i=1}^n \frac{\partial L}{\partial x^i}\varepsilon^i + \sum_{i=1}^n \frac{\partial L}{\partial \dot{x}^i}\dot{\varepsilon}^i \right) dt \\
&= \int_0^1 \sum_{i=1}^n \left(\frac{\partial L}{\partial x^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} \right) \varepsilon^i dt\ ,
\end{aligned}$$

where we have integrated by parts and used that $\varepsilon(0) = \varepsilon(1) = 0$. Using the Fundamental Lemma, this is equivalent to

$$\boxed{\frac{\partial L}{\partial x^i} = \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i}}\ , \tag{5}$$

for all $i = 1, 2, \ldots, n$. This is the Euler–Lagrange equation.

Conservative force [edit]

A particle of mass *m* moves under the influence of a conservative force derived from the gradient ∇ of a scalar potential,

$$\mathbf{F} = -\nabla V(\mathbf{r}) \,.$$

If there are more particles, in accordance with the above results, the total kinetic energy is a sum over all the particle kinetic energies, and the potential is a function of all the coordinates.

Cartesian coordinates [edit]

The Lagrangian of the particle can be written

$$L(x, y, z, \dot{x}, \dot{y}, \dot{z}) = \frac{1}{2} m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - V(x, y, z) \,.$$

The equations of motion for the particle are found by applying the Euler–Lagrange equation, for the *x* coordinate

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}}\right) = \frac{\partial L}{\partial x} \,,$$

with derivatives

$$\frac{\partial L}{\partial x} = -\frac{\partial V}{\partial x} \,, \quad \frac{\partial L}{\partial \dot{x}} = m\dot{x} \,, \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}}\right) = m\ddot{x} \,,$$

hence

$$m\ddot{x} = -\frac{\partial V}{\partial x} \,,$$

and similarly for the *y* and *z* coordinates. Collecting the equations in vector form we find

$$m\ddot{\mathbf{r}} = -\nabla V$$

which is Newton's second law of motion for a particle subject to a conservative force.

I think you have some problems, because you use an incorrect notation. Let me rewrite your original problem:

$$\begin{aligned} \text{minimize} \quad & J(y) = \int_{x_0}^{x_1} F(x, y(x), y'(x))\,\mathrm{d}x \\ \text{subject to} \quad & G(x, y(x), y'(x)) = 0 \quad \text{for all } x \in [x_0, x_1]. \end{aligned}$$

Here, $F : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $G : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^n$. Do you see the differences? J only depends on the function y, whereas the integrand F and the constraint G depend on real numbers.

Now (if a constraint qualification is satisfied), you get a multiplier $\lambda : [x_0, x_1] \to \mathbb{R}^n$ (compare with section 6.2 in your link: you get a multplier for each constraint, that is, for each x), such that the derivative of the Lagrangian

$$J(y) + \int_{t_0}^{t_1} G(x, y(x), y'(x))\,\lambda(x)\,\mathrm{d}x$$

with respect to y is zero (that is, the derivative of your lagrangian w.r.t. the optimization variable). Now, you can continue like for the derivation of the euler-lagrange equation.

https://math.stackexchange.com/questions/279518/constrained-variational-problems-intuition

16. Position, momentum, and the Uncertainty Principle

Suppose f is a function in $L^2(\mathbf{R})$ with norm 1. Then $|f(x)|^2$, which has integral 1, can be thought of as a probability density, so that the probability of that a point x lies in an interval $I \subset \mathbf{R}$ is

$$\text{(16.1)} \qquad \operatorname{Prob}\{x \in I\} = \int_I |f(x)|^2 \, dx.$$

Then the expected value of the position of the point is

$$\text{(16.2)} \qquad E = \int_{\mathbf{R}} x|f(x)|^2 \, dx.$$

The variance (square of the standard deviation) is a measure of how spread out the probability distribution is; it is

$$\text{(16.3)} \qquad V = \int_{\mathbf{R}} (x-E)^2 |f(x)|^2 \, dx;$$

It is small only if most of the mass of $|f|^2$ is concentrated near the mean E. The behavior of the Fourier transform under scaling (see Problem 15.5) suggests that if f has small variance then $\hat{f}$ may be expected to have large variance. Note that we need a factor of $1/2\pi$ to make the Fourier transform have norm 1, so the mean and variance for $\hat{f}$ are

$$\text{(16.4)} \qquad \hat{E} = \frac{1}{2\pi} \int_{\mathbf{R}} \xi |\hat{f}(\xi)|^2 d\xi$$

$$\text{(16.5)} \qquad \hat{V} = \frac{1}{2\pi} \int_{\mathbf{R}} (\xi - \hat{E})^2 |\hat{f}(\xi)|^2 d\xi.$$

The remark above about the relation between the variances V and $\hat{V}$ can be given quantitative form.

Proposition. *If f is an element of $L^2(\mathbf{R})$ such that $||f|| = 1$, then the product of the variances of f and of $\hat{f}$, $V \cdot \hat{V}$, is at least $\frac{1}{4}$.*

Sketch of proof. We use the results of some of the problems above. Let Q and P be the linear transformations on functions

$$Qf(x) = xf(x); \qquad Pf(x) = \frac{1}{i}\frac{df(x)}{dx}.$$

Then the Fourier transform of Pf is $\xi \hat{f}(\xi)$, so

$$\text{(16.6)} \qquad V = ||(Q-E)f||^2; \qquad \hat{V} = ||(P-\hat{E})\hat{f}||^2.$$

Denoting the identity operator by I, note that

$$\text{(16.7)} \qquad PQ - QP = -iI; \qquad (Qf,g) = (f,Qg); \qquad (Pf,g) = (f,Pg).$$

It follows from (16.7) and the Cauchy-Schwartz inequality that

$$\text{(16.8)} \qquad \begin{aligned} 1 = ||f||^2 = (f,f) &= i(PQf - QPf, f) = i[(Qf,Pf) - (Pf,Qf)] \\ &= 2\operatorname{Im}(Pf,Qf) \le 2||Qf|| \cdot ||Pf||. \end{aligned}$$

Now it is also true that

$$(P - \hat{E}I)(Q - EI) - (Q - EI)(P - \hat{E}I) = -iI$$

so we may repeat the calculation (16.8) with $Q - EI$ in place of Q and $P - \hat{E}I$ in place of P to obtain the desired inequality.

The simplest case in quantum mechanics consists of a single particle in one dimension. Its *wave function* is an element $\psi \in L^2(\mathbf{R})$ having norm 1. Any physical measurement is characterized by a linear transformation T defined on some subspace of $L^2(\mathbf{R})$ which has the property $(Tf,f) = (f,Tf)$ for all f in its domain. The theory is probabilistic: if the wave function of the particle at a given moment is ψ then the mean and variance of the measurement of the quantity associated to T are

$$\text{(16.9)} \qquad E_T = (T\psi, \psi); \qquad V_T = ||(T - E_T I)^2 \psi||^2.$$

In the usual representation of the wave function, the *position operator* is the operator Q above and the *momentum operator* is hP, where P is the operator above and $h > 0$ is Planck's constant. Thus the inequality proved above gives the quantitative form of the relationship between uncertainty in measurement of position and uncertainty in measurement of velocity known as the *Heisenberg Uncertainty Principle:*

$$\text{(16.10)} \qquad \sqrt{V_Q} \cdot \sqrt{V_{hP}} \ge \frac{h}{2}.$$