Differential Geometry and Calculus of Variations in AI and Physics

Howard Qian

August 16, 2025

1. Introduction

Differential geometry is the study of curves, surfaces, and higher-dimensional spaces using calculus and linear algebra. It explores properties including curvature, arclength, and area, and applications span from physics and computer graphics to robotics.

In this paper, we begin with an overview of fundamental ideas in differential geometry, including parametrized curves, arclength, and curvature [§2]. I will then narrow the focus to the Calculus of Variations [§3], a powerful framework for finding functions that optimize functionals. This naturally leads to the Euler–Lagrange equation, which lies at the heart of many physical laws and optimization problems. Building on this foundation, I will explore the applications of variational calculus, showing how variational principles appear in both physics and artificial intelligence. In physics, we will see applications in laws of motion and in optical laws such as Snell's Law. As for AI, I discuss topics such as smoothing splines for data modeling and variational autoencoders [§4].

2. Overview of Differential Geometry Concepts

Differential geometry is the mathematical study of curves, surface, and more generally, manifolds. It describes how geometric objects bend, twist, and stretch, and how those properties can be measured and quantified. Many principles in differential geometry are crucial for understanding physical phenomena and optimization problems, making it a precursor to variational calculus and its application to other fields. In this subsection, we will explore important concepts in differential geometry that construct the backbone for understanding variational calculus. Concepts are covered in [Rub] and [Car76].

To start, let's talk about parametrized curves.

2.1. Parametrized Curves

Definition 2.1. A parametrized curve is a continuous function $\gamma : (a, b) \to \mathbb{R}^n$ for some a, b with $-\infty \leq a < b \leq \infty$.

For instance, the unit circle with equation $x^2 + y^2 = 1$ in $\mathbb{R}^2$ can be expressed as the parametrization $\gamma : (-\infty, \infty) \to \mathbb{R}^2$ defined by $\gamma(t) = (\cos t, \sin t)$. Each portion of a parametrized curve is a smooth function, meaning that it is infinitely differentiable.

When taking the derivative of a parametrized curve, we differentiate it component wise. The derivative of a curve γ is denoted with symbol $\dot{\gamma}$. For instance:

$$\frac{d}{dt}\gamma(t) = \dot{\gamma}(t) = (\gamma_1'(t), \ldots, \gamma_n'(t)).$$

The derivative at time t, or $\dot{\gamma}(t)$, represents the tangent vector at time t. In terms of kinematics, which will be touched upon when discussing arclength, $\dot{\gamma}(t)$ denotes the velocity of the curve at time t, and further differentiating, $\ddot{\gamma}(t)$ denotes the acceleration.

Understanding parametrized curves is essential for variational calculus because in many optimization problems — such as finding the shortest path between two points — that path is represented as a parametrized curve to minimize a given functional.

2.2. Arclength

The arclength of the curve γ starting at point t_0 is the function $s(t)$ defined by

$$s(t) = \int_{t_0}^{t} \|\dot{\gamma}(t)\| \, du$$

where $\|\dot{\gamma}(t)\| = \sqrt{(\gamma_1'(t)^2 + \cdots + \gamma_n'(t))^2}$. By treating $\gamma(t)$ as the position on the curve at time t, and assuming that $t_0 = 0$, $s(t)$ is the distance traveled by time t. Naturally, $s'(t)$ is the rate of change of the distance traveled, or the speed.

In other words,

$$\text{speed} = s'(t) = \|\dot{\gamma}(t)\| \ .$$

Definition 2.2. A curve is unit-speed if its speed, $\|\dot{\gamma}(t)\|$ is always 1.

In the following sub-sections, we will assume that we are dealing with unit-speed curves.

2.3. Curvature

Precisely, curvature denotes how much we move away from the tangent line as we move minutely on the curve. For instance, say we have a unit-speed curve γ. When we move a short distance Δt away, from $\gamma(t)$ to $\gamma(t+\Delta t)$, we measure the projected position on the tangent line from our actual position at $\gamma(\Delta t)$. The vector of that displacement is $(\gamma(t+\Delta t)-\gamma(t))\cdot \mathbf{n}$ where $\mathbf{n}$ is the unit normal vector.

Definition 2.3. Let γ be a unit-speed curve. Its curvature $\kappa(t) = \|\ddot{\gamma}(t)\|$

In order to check this definition, we will test it against our proposed formula for the curvature of a circle - for a circle with radius R, its curvature is $\frac{1}{R}$.

A circle with center at (x_0, y_0) of radius R has unit-speed parametrization $\gamma(t) = (x_0 + R\cos\frac{t}{R}, y_0 + R\sin\frac{t}{R})$. Taking the second derivative, we get $\ddot{\gamma}(t) = (-\frac{1}{R}\cos\frac{t}{R}, -\frac{1}{R}\sin\frac{t}{R})$. Solving for curvature $\kappa(t)$:

$$\begin{aligned}
\kappa(t) &= \|\ddot{\gamma}(t)\| \\
&= \sqrt{\left(-\frac{1}{R}\cos\frac{t}{R}\right)^2 + \left(-\frac{1}{R}\sin\frac{t}{R}\right)^2} \\
&= \sqrt{\frac{1}{R^2}\cos^2\frac{t}{R} + \frac{1}{R^2}\sin^2\frac{t}{R}} \\
&= \sqrt{\frac{1}{R^2}\left(\cos^2\frac{t}{R} + \sin^2\frac{t}{R}\right)} \\
&= \sqrt{\frac{1}{R^2}} \\
&= \frac{1}{R}
\end{aligned}$$

For regular curves in $\mathbb{R}^3$,

$$\kappa = \frac{\|\ddot{\gamma} \times \dot{\gamma}\|}{\|\dot{\gamma}\|^3}$$

The notions of curvature and arclength explored form the basis of the next sections.

3. Calculus of Variations

The calculus of variations is the study of optimizing, or minimizing and maximizing, functionals.

Definition 3.1. A functional is a mapping from a set of functions to the real numbers. In other words, they are functions whose inputs are functions and whose outputs are real numbers. The domain of a functional is a function space, which can be though of as a vector space of functions.

3.1. The Euler-Lagrange Equation

Discussed in [Kha17] and [EJ23], processes for optimizing functionals are often subject to certain boundary conditions. A typical functional takes the form

$$f(x) = I = \int_{x_1}^{x_2} F(x, y(x), y'(x))dx$$

where u has endpoint or boundary conditions $y(x_1) = y_1$ and $y(x_2) = y_2$.

In addition, we introduce the function $\eta(x)$ where $\eta(x_1) = \eta(x_2) = 0$.

The function $\eta(x)$ is known as the variation function, and it represents a small perturbation to the original curve $y(x)$. The boundary conditions of $\eta(x)$ are set to ensure that we can explore nearby curves without violating the endpoint constraints.

Also, we define $\overline{y}(x) = y(x) + \epsilon\eta(x)$ that satisfies the same boundary conditions are y, where $y(x)$ is an extremal. Note that setting $\epsilon = 0$, $\overline{y}(x) = y(x)$.

With the equation for $\overline{y}$, we can write out:

$$\overline{y}'(x) = y'(x) + \epsilon\eta'(x)$$
$$\frac{\partial \overline{y}}{\partial \epsilon} = \eta(x)$$
$$\frac{\partial \overline{y}'}{\partial \epsilon} = \eta'(x)$$

We want to find the particular $\overline{y}(x)$ that makes $I(\epsilon)$ stationary. Since $y(x)$ is an extremal, $I(\epsilon)$ should be stationary at $\overline{y}(x) = y(x)$, and this occurs when $\epsilon = 0$. Thus,

$$\left.\frac{dI}{d\epsilon}\right|_{\epsilon=0} = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \int_{x_1}^{x_2} F(x, y(x), y'(x))dx = 0$$

Taking the partial derivative, we obtain:

$$\int_{x_1}^{x_2} \left.\frac{\partial}{\partial \epsilon} F(x, \overline{y}, \overline{y}')\right|_{\epsilon=0} dx = 0$$

Applying the chain rule, we get:

$$\frac{\partial}{\partial \epsilon} F(x, \overline{y}, \overline{y}') = \frac{\partial F}{\partial \overline{y}} \cdot \frac{\partial \overline{y}}{\partial \epsilon} + \frac{\partial F}{\partial \overline{y}'} \cdot \frac{\partial \overline{y}'}{\partial \epsilon}$$

Now, with the equations $\frac{\partial \overline{y}}{\partial \epsilon} = \eta(x)$ and $\frac{\partial \overline{y}'}{\partial \epsilon} = \eta'(x)$, we can simplify the integral. This yields

$$\int_{x_1}^{x_2} \left[\frac{\partial F}{\partial \overline{y}} \cdot \frac{\partial \overline{y}}{\partial \epsilon} + \frac{\partial F}{\partial \overline{y}'} \cdot \frac{\partial \overline{y}'}{\partial \epsilon} \right] \Bigg|_{\epsilon=0} dx = \int_{x_1}^{x_2} \left[\frac{\partial F}{\partial \overline{y}} \cdot \eta + \frac{\partial F}{\partial \overline{y}'} \cdot \eta' \right] \Bigg|_{\epsilon=0} dx = 0$$

We hope to leverage the boundary conditions of $\eta(x)$, and to do so, we need to factor out $\eta(x)$. To do this, we eliminate $\eta'(x)$ through integration by parts on the term $\int_{x_1}^{x_2} \left(\frac{\partial F}{\partial \overline{y}'} \cdot \eta' \right) dx$.

$$\begin{aligned}
&\int_{x_1}^{x_2} \left[\frac{\partial F}{\partial \overline{y}'} \cdot \eta' \right] dx \\
&\qquad = \frac{\partial F}{\partial \overline{y}'} \cdot \int_{x_1}^{x_2} \eta' dx - \int_{x_1}^{x_2} \left[\Big(\int \eta' dx \Big) \cdot \frac{d}{dx} \Big[\frac{\partial F}{\partial \overline{y}'} \Big] dx \right] \\
&\qquad = \frac{\partial F}{\partial \overline{y}'} \cdot (\eta(x_2) - \eta(x_1)) - \int_{x_1}^{x_2} \left[\eta \cdot \frac{d}{dx} \Big[\frac{\partial F}{\partial \overline{y}'} \Big] dx \right] \\
&\qquad = 0 - \int_{x_1}^{x_2} \left[\eta \cdot \frac{d}{dx} \Big[\frac{\partial F}{\partial \overline{y}'} \Big] dx \right]
\end{aligned}$$

Note that the first term is equal to 0 due to the boundary condition on η. Now plugging back into our original integral,

$$\begin{aligned}
&\int_{x_1}^{x_2} \left[\frac{\partial F}{\partial \overline{y}} \cdot \eta + \frac{\partial F}{\partial \overline{y}'} \cdot \eta' \right] \Bigg|_{\epsilon=0} dx \\
&\qquad = \int_{x_1}^{x_2} \left[\frac{\partial F}{\partial \overline{y}} \cdot \eta - \eta \cdot \frac{d}{dx} \Big(\frac{\partial F}{\partial \overline{y}'} \Big) \right] \Bigg|_{\epsilon=0} dx \\
&\qquad = \int_{x_1}^{x_2} \left[\eta \left(\frac{\partial F}{\partial \overline{y}} - \frac{d}{dx} \Big(\frac{\partial F}{\partial \overline{y}'} \Big) \right) \right] \Bigg|_{\epsilon=0} dx \\
&\qquad = 0
\end{aligned}$$

At $\epsilon = 0$, $\overline{y} = y + 0 \cdot \eta = y$, so

$$\int_{x_1}^{x_2} \left[\eta \left(\frac{\partial F}{\partial \overline{y}} - \frac{d}{dx} \Big(\frac{\partial F}{\partial \overline{y}'} \Big) \right) \right] \Bigg|_{\epsilon=0} dx = \int_{x_1}^{x_2} \left[\eta \left(\frac{\partial F}{\partial y} - \frac{d}{dx} \Big(\frac{\partial F}{\partial y'} \Big) \right) \right] dx = 0$$

Finally, to guarantee that the integral is equal to 0,

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\Big(\frac{\partial F}{\partial y'}\Big) = 0$$

The equation above is the **Euler-Lagrange equation**. If $y(x)$ is an extremal of the function I, then it must satisfy the Euler-Lagrange equation.

3.1.1. Using the Euler-Lagrange equation to find shortest path

Let $y(x)$ be a curve joining points (x_0, y_0) and (x_1, y_1). The length of the curve is

$$L(y) = \int_{x_0}^{x_1} \sqrt{1 + (y')^2} dx$$

Since the functional $F(x, y, y')$ doesn't depend on y, $\frac{\partial F}{\partial y} = 0$.
Thus,

$$0 - \frac{d}{dx}\left(\frac{\partial}{\partial y'}\sqrt{1 + (y')^2}\right) = 0$$

$$\implies \frac{d}{dx}\left(\frac{y'}{\sqrt{1 + (y')^2}}\right) = 0$$

Thus, $\frac{y'}{\sqrt{1+(y')^2}}$ is a constant. Letting $\frac{y'}{\sqrt{1+(y')^2}} = C$, we can solve for y':

$$y' = \frac{C}{\sqrt{1 - C^2}} = p$$

for some constant p.

Then,

$$y = px + b$$

where b is a constant. Therefore, using the Euler-Lagrange equation, the shortest path between two points on a plane is a straight line.

Another interesting appearance of the Euler-Lagrange equation is in the Brachistochrone curve, which is the curve that allows a particle to travel from one point to another in the shortest amount of time under the influence of gravity.

4. Applications of Variational Calculus

4.1. Physics

4.1.1. Newton's Second Law

A particle of mass m and speed $\dot{x}$ has kinetic energy $T = \frac{1}{2}m\dot{x}^2$ and potential energy $V(x)$. As a result, the Lagrangian L follows $L = T - V$. Using the Euler-Lagrange equation,

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{x}}\Big) = 0$$

With the equation $L = T - V = \frac{1}{2}m\dot{x}^2 - V(x)$, we are able to take the partial derivatives of L:

$$\frac{\partial L}{\partial x} = -\dot{V}(x) = -\frac{\partial V}{\partial x}$$
$$\frac{\partial L}{\partial \dot{x}} = m\dot{x}$$
$$\frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{x}}\Big) = m\ddot{x}$$

Thus,

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{x}}\Big) = -\frac{\partial V}{\partial x} - m\ddot{x} = 0$$

Since Force $= F = -\frac{dV}{dx}$,

$$-\frac{\partial V}{\partial x} - m\ddot{x} = F - m\ddot{x} = 0$$
$$F - m\ddot{x} = F - ma = 0 \to F = ma$$

4.1.2. Snell's Law

Another application of the Euler-Lagrange equation is on the derivation of Snell's Law. Snell's Law outlines how a light ray changes direction when it crosses two media with two different refractive indices. If medium 1 has index n_1 and medium 2 has index n_2, and the ray in medium 1 meets the interface at angle θ_1 and refracts into medium 2 at angle θ_2, Snell's Law shows that

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

Fermat's principle states that the physical path taken by light between two fixed points is the one that minimizes travel time, and thus, the path length of the ray is a functional of the light-ray curve.

For the ray $y(x)$ in a medium with index $n(y)$, meaning that the index is y-dependent, the length between two points x_1 and x_2 is

$$S(y) = \int_{x_1}^{x_2} n\Big(y(x)\Big)\sqrt{1 + y'(x)^2}dx$$

where $\sqrt{1 + y'(x)^2}$ is the arclength.

Applying the Euler-Lagrange equation,

$$0 = n'\Big(y(x)\Big)\sqrt{1 + y'(x)^2} - \frac{d}{dx}\frac{y'(x) \cdot n\Big(y(x)\Big)}{\sqrt{1 + y'(x)^2}}$$

Simplifying and equating, we are able to get

$$\frac{d}{dx}\left(\frac{n(y)}{\sqrt{1 + y'^2}}\right) = 0$$

Hence, $\frac{n(y)}{\sqrt{1+y'^2}}$ is a constant.

Finally, as $\cos\theta = \frac{1}{\sqrt{1+y'^2}}$,

$$\frac{n(y)}{\sqrt{1 + y'^2}} = n(y)\cos\theta = \text{constant}$$

Letting $\phi = \frac{\pi}{2} - \theta$,

$$n(y)\sin\phi = \text{constant}$$

[Cal] provides further details.

4.2. Deep Learning and AI

The principles of variational calculus, traditionally rooted in physics and geometry, have developed a growing role in the field of artificial intelligence and deep learning. Understandably, deep learning is an optimization problem; we seek to find a function that minimizes an overall "loss functional" over a wide space of parameters. This section will explore two key areas where variational calculus is applied: the use of smoothing splines for robust data modeling and signal extraction, and the development of variational autoencoders (VAEs).

4.2.1. Smoothing Splines

In many scientific and engineering contexts, the data we observe is far from perfect; data is often corrupted by noise from measurement errors, instrument limitations, or inherent randomness. In astronomy particularly, light curves - flux over time measurements - obtained from telescopes may contain irregularities and unwanted fluctuation due to atmospheric effects, stellar activity, or sensor noise. In machine learning, training data may be noisy due to imperfect labeling or real-world variability. If we fit such data, we will risk overfitting, capturing noise rather than the real underlying signal. We will apply the Euler-Lagrange equation to introduce a method of smoothing signal variability.

To start, we write out the functional as

$$J(y) = \int F(x, y, y', y'')dx = \int \left[(y(x) - y_{data}(x))^2 + \lambda(y''(x))^2\right]dx$$

where the first term $(y(x) - y_{data}(x))^2$ is the error term while the second term $\lambda(y''(x))^2$ represents the smoothness term. λ is thus the smoothing parameter.

We seek to find the function $y(x)$ to minimize the functional $J(y)$, and thus we apply the Euler-Lagrange equation up to the second derivative:

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) = 0$$

where in this case, $F = (y - y_{data})^2 + \lambda(y'')^2$

The individual partial derivatives can then be calculated as follows:

$$\frac{\partial F}{\partial y} = 2(y - y_{data})$$
$$\frac{\partial F}{\partial y'} = 0$$
$$\frac{\partial F}{\partial y''} = 2\lambda y''$$

Therefore, solving the Euler-Lagrange equation,

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) = 0$$
$$2(y - y_{data}) - \frac{d}{dx}0 + \frac{d^2}{dx^2}2\lambda y'' = 0$$
$$2(y - y_{data}) + 2\lambda y'''' = 0$$

$$\implies y - y_{data} + \lambda y'''' = 0$$

The end result, a fourth-order differential equation, is a spline. **SciPy**, A common python package for machine learning and AI, uses the principles behind this spline for its widely-used **UnivariateSpline** function.

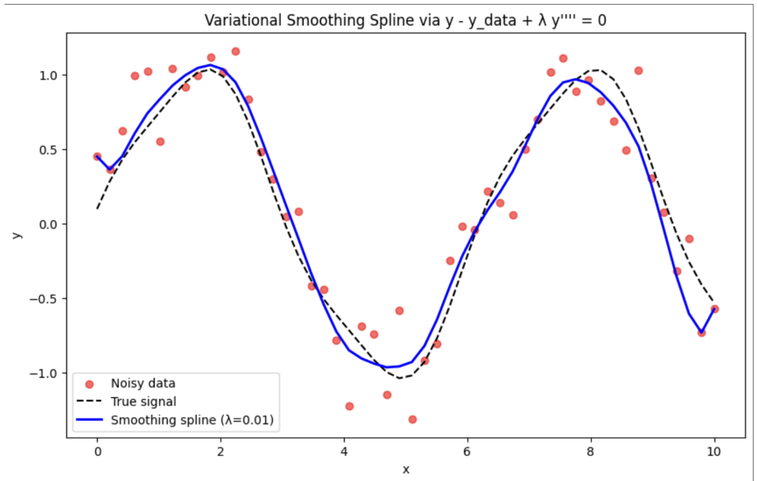

Figure 4.1: An example of the smoothing spline on noisy data. **Source:** Self-created using **numpy** and **matplotlib**.

4.2.2. Variational Autoencoders (VAEs)

Although this section is less connected to variational calculus, the intentions are similar. The variational inference aspect of VAEs seeks to minimize a certain difference metric between two distributions, which introduces the Evidence Lower Bound (ELBO) and is the typical functional VAEs optimize.

Autoencoders are neural network architectures designed to learn representations of data by compressing inputs into a bottleneck, or latent space, representation. Then, the latent presentation is used to reconstruct the original input. Autoencoders contain two main parts: the encoder, which is used to map the input to a lower-dimensional space, and the decoder, which reconstructs the original input. Autoencoders learn and train by reducing the loss function, which is taken by evaluating the difference between the reconstructed and original images. They serve several purposes, from dimensionality reduction and compression to feature learning. However, typical autoencoders are not generative models, meaning they are unable to model data distributions and cannot be used to generate new samples from the learned latent representations.

Introduced in [KW19], **variational autoencoders (VAEs)** improve upon the general autoencoder structure by introducing a probabilistic approach. Instead of the fixed bottleneck vector, VAEs encode mean and standard deviation vectors, and these inputs are parameters of a probability distribution. Typically, the distribution is Gaussian. These changes allow VAEs to generate new, realistic data. In addition, VAEs allow the sampling of any random point, producing a reasonable output. Applications of VAEs include image denoising and image compression.

References

[Cal] U. of California San Diego. *Calculus of Variations*. https://courses.physics.ucsd.edu/2010/Fall/physics110a/LECTURES/CH05.pdf.

[Car76] M. P. do Carmo. *Differential Geometry of Curves and Surfaces*. ISBN 978-0-13-212589-5. Rio de Janeiro, Brazil: Instituto de Matemática Pura e Aplicada (IMPA), 1976.

[EJ23] A. Edelman and S. G. Johnson. *Lecture 10: Calculus of Variations*. https://ocw.mit.edu/courses/18-s096-matrix-calculus-for-machine-learning-and-beyond-january-iap-2023/mit18_s096iap23_lec10.pdf. Lecture notes for 18.S096, Matrix Calculus for Machine Learning and Beyond, MIT OpenCourseWare. 2023.

[Kha17] F. of Khan. *Derivation of the Euler-Lagrange Equation*. https://www.youtube.com/watch?v=sFqp2lCEvwM&list=PLdgVBOaXkb9CD8igcUr9Fmn5WXLpE8ZE_&index=2. Derivation of the Euler-Lagrange Equation. 2017.

[KW19] D. P. Kingma and M. Welling. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, 12(4): 307–392, 2019.

[Rub] S. Rubinstein-Salzedo. *Differential Geometry*. Chapters 1 and 2.