# A Brief Introduction to Calculus of Variations

Marc Niethammer

## 1 Background

The derivative for "regular" functions on $\mathbb{R}$ is defined as

$$\dot{f}(x) := \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon} = \frac{df}{dx}.$$

The directional derivative, in the direction $u$, for a function defined on $\mathbb{R}^n$ is

$$f_{\boldsymbol{u}}(\boldsymbol{x}) = \lim_{\epsilon \to 0} \frac{f(\boldsymbol{x}+\epsilon\boldsymbol{u}) - f(\boldsymbol{x})}{\epsilon} = \frac{\partial f}{\partial \epsilon}(\boldsymbol{x}+\epsilon\boldsymbol{u})|_{\epsilon=0}.$$

To find a stationary point (i.e., a local maximum or minimum) the derivative $\dot{f}(x)$ needs to vanish. In higher dimensions, the gradient $\nabla f$ needs to be identical to zero. The latter may also be expressed as $f_{\boldsymbol{u}} = 0 \ \forall \boldsymbol{u}$[1]. In a sense these simple derivatives are the most basic form of the calculus of variations. Usually, however, one talks about calculus of variations in the context of determining functions that minimize functionals. In other words, the argument (for which a given quantity, the functional) is extremal is no longer a point in space, but a function out of an appropriate class of functions. Given a functional $J : F \mapsto \mathbb{R}$ (where $F$ is the desired space of functions) the Gâteaux variation is (compare this to the directional derivative)

$$\delta J(y; v) := \lim_{\epsilon \to 0} \frac{J(y+\epsilon v) - J(y)}{\epsilon} = \frac{\partial}{\partial \epsilon} J(y+\epsilon v)|_{\epsilon=0}, \ y \in F, \epsilon \in \mathbb{R},$$

where $v$ is a testfunction that needs to be chosen consistent with the problem's boundary conditions. (More on this later, for now, just think as an example of a function with two fixed end points, at these fixed endpoints the test function $v$ cannot vary, the solution is known, it needs to be fixed to 0). The testfunction $v$ plays the role of the $\boldsymbol{u}$ in the previously introduced directional derivative. $\delta J$ needs to vanish for all $v$ to have an extremum. Figure 1 illustrates the finite dimensional derivatives and the concept of the Gâteaux variation.

Before moving on to an example let's review the differentiation rule for integrals. Given

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

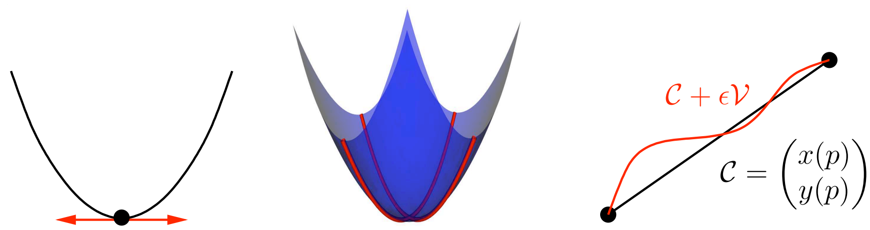
the derivative with respect to $\epsilon$ is

$$\frac{dI}{d\epsilon} = f(x_2, \epsilon)\frac{dx_2}{d\epsilon} - f(x_1, \epsilon)\frac{dx_1}{d\epsilon} + \int_{x_1(\epsilon)}^{x_2(\epsilon)} \frac{\partial f}{\partial \epsilon} \ dx.$$

Together with the integration by parts rule

$$\int (\dot{gf}) \ dx = gf(+c) = \int g\dot{f} \ dx + \int \dot{g}f \ dx,$$

[1] This will be connected to calculus of variations later.




(a) $f(x) = x^2$ (b) $f(x,y) = x^2 + y^2$ (c) $E(\mathcal{C}) = \int_{p=0}^{1} \|\mathcal{C}_p\| \; dp = L$

Variational calculus is "differential calculus for functions."

---

Analogous definitions (infinitesimal perturbations):

(a) $\frac{df(x)}{dx} = \lim_{\epsilon\to 0} \frac{f(x+\epsilon)-f(x)}{\epsilon} = 0$

(b) $f_{\boldsymbol{u}}(\boldsymbol{x}) = \lim_{\epsilon\to 0} \frac{f(\boldsymbol{x}+\epsilon\boldsymbol{u})-f(\boldsymbol{x})}{\epsilon} = \frac{\partial f}{\partial \epsilon}(\boldsymbol{x}+\epsilon\boldsymbol{u})|_{\epsilon=0} = 0 \;\forall\; \boldsymbol{u}$

(c) $\delta E(\mathcal{C};\mathcal{V}) = \lim_{\epsilon\to 0} \frac{E(\mathcal{C}+\epsilon\mathcal{V})-E(\mathcal{C})}{\epsilon} = \frac{\partial}{\partial \epsilon} E(\mathcal{C}+\epsilon\mathcal{V})|_{\epsilon=0} = 0 \;\forall\; \mathcal{V}$

Figure 1: Illustration of the principle of calculus of variations as "differential calculus for functions."

or with definite integration bounds

$$\int_a^b \dot{g} f \; dx = [gf]_a^b - \int_a^b g\dot{f} \; dx.$$

this is pretty much all one needs to know in practice to do calculus of variations with functionals having functions defined on $\mathbb{R}$ as their argument. Here is an unpractical example: What is the Gâteaux variation of the functional

$$F(y) = \int_a^b f(x, y, \dot{y}) dx?$$

Sticking strictly to the previous definition (probably always the safest bet, instead of using ready-made solutions) yields

$$\delta F(y;v) = \frac{\partial}{\partial \epsilon} F(y+\epsilon v)|_{\epsilon=0} = \frac{\partial}{\partial \epsilon} \int_a^b f(x, y+\epsilon v, \dot{y}+\epsilon \dot{v}) \; dx|_{\epsilon=0} = \int_a^b \frac{\partial f}{\partial y} v + \frac{\partial f}{\partial \dot{y}} \dot{v} \; dx.$$

For an extremal point, $\delta F(y;v) = 0 \;\forall v$. To check what the condition on $y$ needs to be, integration of part brings relief, since

$$\delta F(y;v) = \int_a^b \frac{\partial f}{\partial y} v - \frac{d}{dx}\left(\frac{\partial f}{\partial \dot{y}}\right) v \; dx + \left[\frac{\partial f}{\partial \dot{y}} v\right]_a^b = \int_a^b \left(\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial \dot{y}}\right)\right) v \; dx.$$

The boundary term dropped out since we assume the values of $f$ fixed at the boundary and thus $v(a) = v(b) = 0$. Alternatively, the boundary term yields the natural boundary conditions, i.e., $\frac{\partial f}{\partial \dot{y}} = 0$ at $x = a$ and/or $x = b$. Finally, the last result shows that for the Gâteaux variation to vanish for all $v$, the condition (known as the Euler-Lagrange equation)

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial \dot{y}}\right) = 0. \tag{1}$$

needs to hold. (This derivation easily extends to cases with an arbitrary order of derivatives as the arguments of the functional, leading to more complex Euler-Lagrange equations.) If the solution to the Euler-Lagrange equation is known, the problem is solved. Unfortunately, many times a closed form solution is not known. One way to cope with this problem is to use a gradient descent technique. Incidentally, the left hand side of the Euler-Lagrange equation can be regarded as an infinite-dimensional gradient in the $L_2$ sense, i.e.,

$$\nabla F = \frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial \dot{y}}\right),$$

yielding the gradient descent scheme

$$F_t = -\nabla F.$$

Why is this? We can write the Gâteaux variation in inner product form (which we have done all along, since the quantities of interest are real functions, but so far never explicitly spelled out). Then

$$\delta F(y; v) = \int gv \; dx = < g, v > .$$

The Cauchy Schwarz inequality is

$$| < g, v > |^2 \;\leq\; < g, g > \cdot < v, v > .$$

The function $g$ is given. If we assume without loss of generality that

$$< v, v > = c^2 < g, g >, \; c = const. \in \mathbb{R},$$

then

$$| < g, v > |^2 \leq c^2 < g, g >^2 .$$

Consequently, $v = cg$ maximizes the inner product. Thus, $g$ can be regarded as an infinite-dimensional gradient, $g = \nabla F$. Moving in its opposite direction amounts to the direction decreasing the energy $F$ maximally fast, as desired.

## 2 Example: Shortest distance between two points

Here is a more practical example. Assume the path connecting two points $(a, y(a))$ and $(b, y(b))$ can be expressed as a function. Then the curve is given by

$$\mathcal{C}(x, y(x)) = \begin{pmatrix} x \\ y(x) \end{pmatrix}$$

and its derivative with respect to $x$ is

$$\dot{\mathcal{C}}(x, y(x)) = \begin{pmatrix} 1 \\ \dot{y} \end{pmatrix}.$$

The length of the curve may be written (this is the functional to be minimized) as

$$L = \int_a^b \|\dot{\mathcal{C}}\| \; dx = \int_a^b \sqrt{1 + \dot{y}^2} \; dx. \qquad (2)$$

With $f(x, y, \dot{y}) = \sqrt{1 + \dot{y}^2}$ it follows

$$\frac{\partial f}{\partial y} = 0, \quad \frac{\partial f}{\partial \dot{y}} = \frac{\dot{y}}{\sqrt{1 + \dot{y}^2}}.$$

Thus the Euler-Lagrange equation (plugging into Equation 1) becomes

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

This implies

$$\frac{\dot{y}}{\sqrt{1+\dot{y}^2}} = const. \implies \dot{y} = const,$$

which proves that the shortest distance between two points in terms of the length defined in Equation 2 is given by the length of the straight line connecting the two.

# 3 Example: Shortest distance between two points, the parametric version

Assuming that the curve connecting two points in a plane (as in the previous section) can be expressed as a function is an oversimplification of the problem. If a point in the plane is given by the coordinates $(\alpha, \beta)$ and both these coordinates are independently described as a function over a parametrization $p \in [0, 1]$, then

$$\mathcal{C} = \begin{pmatrix} \alpha(p) \\ \beta(p) \end{pmatrix}, \quad \mathcal{C}_p = \begin{pmatrix} \alpha_p \\ \beta_p \end{pmatrix}.$$

The length functional then becomes

$$L = \int_{p=0}^{1} \|\mathcal{C}_p\| \ dp.$$

The Gâteaux variation is

$$\delta L(\mathcal{C}; V) = \frac{\partial}{\partial \epsilon} \int_{p=0}^{1} \|\mathcal{C}_p + \epsilon V_p\| \ dp|_{\epsilon=0}.$$

Going through the usual motions yields

$$\begin{aligned}
\delta L(\mathcal{C}; V) &= \int_{p=0}^{1} \frac{1}{\|\mathcal{C}_p\|}\mathcal{C}_p \cdot V_p \ dp = \int_{p=0}^{1} \underbrace{\mathcal{T}}_{tangent} \cdot V_p \ dp \\
&= \int_{p=0}^{1} -\frac{1}{\|\mathcal{C}_p\|}\frac{\partial}{\partial p}(\mathcal{T}) \cdot V \underbrace{\|\mathcal{C}_p\| dp}_{ds} + [\mathcal{T} \cdot V]_{p=0}^{1} \\
&= \int_{0}^{l} -\frac{\partial}{\partial s}(\mathcal{T}) \cdot V ds,
\end{aligned}$$

where $s$ denotes arclength. From this follows that $\mathcal{T} = const$. Thus the curve connecting the two points is a straight line.

# 4 Example: Multidimensional domain, Laplace's equation and the heat equation

Before delving into the multidimensional problem is it useful to review Green's theorem (here in 2D) which will lead to the analog to integration by parts in multiple dimensions. Green's theorem in two dimensions states

$$\iint_{\Omega} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) \ dx \ dy = \iint_{\Omega} (P dy - Q dx).$$

Substituting $P = \eta G$ and $Q = \eta F$ results in the two-dimensional analogue to integration by parts

$$\iint_\Omega (G \quad F)\,\nabla\eta\; dx\; dy = -\iint_\Omega \eta\left(\frac{\partial G}{\partial x} + \frac{\partial F}{\partial y}\right)\; dx\; dy + \int_{\partial\Omega} \eta(Gdy - Fdx), \tag{3}$$

where $\Omega$ is the domain of integration and $\partial\Omega$ is its boundary. Given the functional

$$L(\nabla I) = \int_\Omega \frac{1}{2}\|\nabla I\|^2\; d\Omega,$$

the Gâteaux variation is

$$\delta L(I;V) = \frac{\partial}{\partial\epsilon}\int_\Omega \frac{1}{2}\|\nabla I + \epsilon\nabla V\|^2\; d\Omega|_{\epsilon=0} = \int_\Omega \nabla I\cdot\nabla V. \tag{4}$$

With $\eta = V$, $G = I_x$ and $F = I_y$ and the help of Equation 3, Equation 4 becomes

$$\delta L(I;V) = -\int_\Omega (I_{xx} + I_{yy})V\; d\Omega + \int_{\partial\Omega} V(I_x\; dy - I_y\; dx) = -\int_\Omega \Delta IV\; d\Omega + \int_{\partial\Omega} V\frac{\partial I}{\partial n}\; dS, \tag{5}$$

where $dS$ denotes the surface element, $n$ the surface normal, and $\Delta I$ the Laplacian of $I$. The boundary term drops out for appropriate Dirichlet and Neumann conditions and the Euler-Lagrange equation becomes

$$\Delta I = 0,$$

which is Laplace's equation. Interpreting the Euler-Lagrange equation as an infinite-dimensional gradient leads to

$$I_t = \Delta I,$$

which is the heat equation.

# 5 Example:Optical flow

Optical flow is the simplest approach to perform image registration, i.e., to map two images to each other. Section 5.1 discusses the classic Horn and Schunck optimal flow formulation. Section 5.2 discusses an optical flow formulation without using the optical flow constraint. Both approaches are based on the assumption of image intensity constancy (for corresponding points), i.e.,

$$I(\boldsymbol{x}, t) = const. \tag{6}$$

## 5.1 Horn and Schunck

Taking the total derivative of Equation 6 yields the optical flow constraint equation

$$\frac{dI}{dt}(\boldsymbol{x}, t) = I_t + \nabla_{\boldsymbol{x}} I\frac{\partial\boldsymbol{x}}{\partial t} = I_t + \nabla I\cdot\boldsymbol{v} = 0, \tag{7}$$

with $\boldsymbol{v} = (v^1, v^2)^T$ the velocity vector. The Horn and Schunck energy (in two dimensions) is

$$L(\boldsymbol{v}) = \frac{1}{2}\int_\Omega (I_t + \nabla I\cdot\boldsymbol{v})^2\; d\Omega + \alpha\frac{1}{2}\int_\Omega \|\nabla v^1\|^2 + \|\nabla v^2\|^2\; d\Omega,$$

where the second term is used to regularize the resulting vector field to cope with the aperture problem; we know its corresponding variation from the Laplace equation example above. Define

$$L_1(\boldsymbol{v}) = \frac{1}{2}\int_\Omega (I_t + \nabla I\cdot\boldsymbol{v})^2\; d\Omega.$$

Its Gâteaux variation is

$$\delta L_1(\boldsymbol{v};\boldsymbol{V}) = \frac{\partial}{\partial\epsilon}\int_\Omega \frac{1}{2}\left(I_t + \nabla I\cdot(\boldsymbol{v}+\epsilon\boldsymbol{V})\right)^2 \ d\Omega|_{\epsilon=0} = \int_\Omega (I_t + \nabla I\cdot\boldsymbol{v})\nabla I\cdot\boldsymbol{V} \ d\Omega.$$

Combining the result for $\delta L_1(\boldsymbol{v};\boldsymbol{V})$ with the Laplace equation result (5) for vanishing boundary conditions yields the Euler-Lagrange equation

$$(I_t + \nabla I\cdot\boldsymbol{v})\nabla I - \alpha\begin{pmatrix}\Delta v^1\\ \Delta v^2\end{pmatrix} = \mathbf{0}.$$

The corresponding gradient descent scheme is thus

$$\begin{aligned} v_t^1 &= -(I_t + \nabla I\cdot\boldsymbol{v})I_x + \alpha\Delta v^1,\\ v_t^2 &= -(I_t + \nabla I\cdot\boldsymbol{v})I_y + \alpha\Delta v^2. \end{aligned}$$

### 5.2 "Nonlinear" optical flow

Instead of using the optical flow constraint for the minimization, the intensity mapping can be formulated explicitly, without the need to resort to a linear approximation (by taking the derivative). The energy then becomes

$$L(\boldsymbol{v}) = \frac{1}{2}\int_\Omega \left(I_1(\boldsymbol{x}-\boldsymbol{v}) - I_2(\boldsymbol{x})\right)^2 \ d\Omega + \alpha\frac{1}{2}\int_\Omega \|\nabla v^1\|^2 + \|\nabla v^2\|^2 \ d\Omega$$

where $I_1$ and $I_2$ denote the images at time points $t_1$ and $t_2$ respectively. The Gâteaux variation of the second term is again already known, the one of the first is

$$\delta L_1(\boldsymbol{v};\boldsymbol{V}) = \frac{\partial}{\partial\epsilon}\int_\Omega \left(I_1(\boldsymbol{x}-\boldsymbol{v}-\epsilon\boldsymbol{V}) - I_2(\boldsymbol{x})\right)^2 \ d\Omega|_{\epsilon=0} = \int_\Omega (I_1(\boldsymbol{x}-\boldsymbol{v}) - I_2(\boldsymbol{x}))\nabla I_1(\boldsymbol{x}-\boldsymbol{v})\cdot(-\boldsymbol{V}) \ d\Omega.$$

Thus the gradient descent based on the Euler-Lagrange equation becomes

$$\begin{aligned} v_t^1 &= (I_1(\boldsymbol{x}-\boldsymbol{v}) - I_2(\boldsymbol{x}))I_x(\boldsymbol{x}-\boldsymbol{v}) + \alpha\Delta v^1,\\ v_t^1 &= (I_1(\boldsymbol{x}-\boldsymbol{v}) - I_2(\boldsymbol{x}))I_y(\boldsymbol{x}-\boldsymbol{v}) + \alpha\Delta v^2. \end{aligned}$$

These equations are structurally very similar to the ones obtained for the Horn and Schunck optical flow. However, they make use of the explicit coordinate mapping.

## 6 A tiny little bit of differential geometry for planar curves

Given the planar parametrized curve

$$\mathcal{C} = \begin{pmatrix}x(p)\\ y(p)\end{pmatrix},$$

the tangent to it is

$$\mathcal{T} = \frac{\mathcal{C}_p}{\|\mathcal{C}_p\|} = \mathcal{C}_s,$$

where

$$\frac{\partial}{\partial s} = \frac{1}{\|\mathcal{C}_p\|}\frac{\partial}{\partial p}.$$

Since $\|\mathcal{T}\|^2 = 1$, it follows that

$$\frac{\partial}{\partial s}\|\mathcal{T}\|^2 = \mathcal{T}_s\cdot\mathcal{T} + \mathcal{T}\cdot\mathcal{T}_s = 2\mathcal{T}_s\cdot\mathcal{T} = 0.$$

Thus, $\mathcal{T}_s$ is orthogonal to $\mathcal{T}$ and

$$\mathcal{T}_s = \mathcal{C}_{ss} = \kappa \mathcal{N},$$

where $\mathcal{N}$ is the unit inward normal to the curve and

$$\kappa = \mathcal{C}_{ss} \cdot \mathcal{N}$$

is the signed Euclidean curvature. From before, the Gâteaux variation for $L = \int \|\mathcal{C}_p\| \, dp$ is

$$\delta L(\mathcal{C}; V) = \int_{\mathcal{C}} -\frac{\partial}{\partial s}(\mathcal{T}) \cdot V \, ds + \underbrace{[\mathcal{T} \cdot V]_{p=0}^{1}}_{=0 \ (closed\ curve)} .$$

With the new found differential geometry knowledge this reduces to

$$\delta L(\mathcal{C}; V) = \int_{\mathcal{C}} -\kappa \mathcal{N} \cdot V \, ds.$$

Thus the gradient descent for a length minimizing flow is

$$\mathcal{C}_t = \kappa \mathcal{N},$$

the curvature flow.

# 7 Example: Deriving the geodesic active contour

With the knowledge from Section 6 the evolution equation for the geodesic active contour can be derived. The functional is

$$L(\mathcal{C}, \mathcal{C}_p) = \int_0^1 g(\mathcal{C}) \|\mathcal{C}_p\| \, dp = \int_0^l g(\mathcal{C}) \, ds,$$

where $g > 0$. It corresponds to a weighted length. Using the developed machinery for calculus of variations the Gâteaux variation is

$$\begin{aligned}
\delta L(\mathcal{C}; V) &= \frac{\partial}{\partial \epsilon} \int_0^1 g(\mathcal{C} + \epsilon V) \|\mathcal{C}_p + \epsilon V_p\| \, dp|_{\epsilon=0} \\
&= \int_0^1 \frac{\partial}{\partial \epsilon} g(\mathcal{C} + \epsilon V)|_{\epsilon=0} \|\mathcal{C}_p\| + g(\mathcal{C}) \frac{\partial}{\partial \epsilon} \|\mathcal{C}_p + \epsilon V_p\||_{\epsilon=0} \, dp \\
&= \int_0^1 \nabla g(\mathcal{C}) \cdot V \|\mathcal{C}_p\| + g(C) \mathcal{T} \cdot V_p \, dp \\
&= \int_0^1 \nabla g(\mathcal{C}) \cdot V \|\mathcal{C}_p\| - \frac{\partial}{\partial p} \left( g(\mathcal{C}) \mathcal{T} \right) \cdot V \, dp + \underbrace{[g(\mathcal{C}) \mathcal{T} \cdot V]_0^1}_{=0 \ (closed\ curve)} \\
&= \int_0^1 \nabla g(\mathcal{C}) \cdot V \|\mathcal{C}_p\| - \frac{\partial}{\partial s} \left( g(\mathcal{C}) \mathcal{T} \right) \cdot V \|\mathcal{C}_p\| \, dp \\
&= \int_0^l \left( \nabla g(\mathcal{C}) - (\nabla g(\mathcal{C}) \cdot \mathcal{T}) \mathcal{T} - g(\mathcal{C}) \kappa \mathcal{N} \right) \cdot V \, ds \\
&= \int_0^l \left( (\nabla g(\mathcal{C}) \cdot \mathcal{N}) \mathcal{N} - g(\mathcal{C}) \kappa \mathcal{N} \right) \cdot V \, ds.
\end{aligned}$$

The corresponding gradient descent flow is the evolution equation of the geodesic active contour (geometric curve evolution):

$$\mathcal{C}_t = g(\mathcal{C}) \kappa \mathcal{N} - (\nabla g(\mathcal{C}) \cdot \mathcal{N}) \mathcal{N}.$$

## 8 Where to go from here ...

For many problems in computer vision and image analysis calculus of variations boils down to

1) Being able to take derivatives and becoming a master in the chain rule.
2) Being able to perform integration by parts, also in higher dimensions (Green's theorem).
3) Choosing the right numerical implementation to solve the resulting partial differential equations (finite differences, finite elements, boundary elements, spectral approaches, etc.).

See [2, 1] for more details, for example on how to integrate constraints through Lagrangian multipliers and how to use variational calculus for dynamic problems in the context of optimal control. In particular, Lanczos' classic book [1] is well worth the read and motivates everything from a mechanics perspective.

## References

[1] C. Lanczos. *The Variational Principles of Mechanics*. Courier Dover Publications, 1986.

[2] J. L. Troutman. *Variational Calculus and Optimal Control*. Springer, 1995.