TMA 4180 Optimeringsteori
Variational Calculus and Classical Mechanics

H. E. Krogstad

Revision 2010

This informal note gives a summary of variational calculus for functions with norms and a very small introduction to the variational formulations of the equations in classical mechanics. Some background in linear analysis may be helpful.

1 Local Extrema

In order to define a *local extremum* it is first of all necessary to define what we mean by *local*. For vectors this is expressed by the distance between the vectors, say $|x-y|$. Recall that function $f(x)$ defined on a domain $\mathcal{D}$ has a local minimum at x_0 if there is a $\delta > 0$ such that $f(x_0) \leq f(x)$ for all x such that $\{x \in \mathcal{D}\ ; |x - x_0| < \delta\}$.

The distance measure for functions are the *function norms*. You should consult a linear analysis textbook if this is completely new to you. We have often been considering functions $y \in C[a,b]$ or $C^1[a,b]$, and the norms here are the so-called maximum norms

$$
\begin{aligned}
\|y\|_{C[0,1]} &= \max_{x\in[a,b]} |y(x)|, \\
\|y\|_{C^1[0,1]} &= \max_{x\in[a,b]} |y(x)| + \max_{x\in[a,b]} \left|y'(x)\right|.
\end{aligned} \tag{1}
$$

For the last case there are other choices as well.

In the literature (and below!) you will often meet the short notation $\|y\|_\infty$ for $\max_{x\in[a,b]} |y(x)|$.

Let us define a *δ-ball centred at y_0* as

$$
\mathcal{B}(y_0, \delta) = \{y\ ;\ \|y - y_0\| < \delta\}. \tag{2}
$$

Consider a functional $J(y)$ defined for $y \in \mathcal{D} \subset C^1[a,b]$, say. The functional J has a *local minimum* at y_0 if there is a $\delta > 0$ such that

$$
J(y_0) \leq J(y) \text{ for all } y \in \mathcal{B}(y_0, \delta) \cap \mathcal{D}. \tag{3}
$$

Local maxima are defined similarly.

2 Differentiable Functionals and the Fréchet Derivative

A collection $\mathcal{D}$ of functions is called a *linear space* if for all $y_1, y_2 \in \mathcal{D}$, also $ay_1 + by_2 \in \mathcal{D}$ for all $a, b \in \mathbb{R}$. A functional L is *linear* if it is defined on a linear space and

$$
L(ay_1 + by_2) = aL(y_1) + bL(y_2), \text{ for all } y_1, y_2 \in \mathcal{D},\ a, b \in \mathbb{R}.
$$

Most of the linear functionals we encounter are also continuous. In general, a functional J is *continuous* at z if

$$\left|J(y) - J(z)\right| \underset{\|y-z\|\to 0}{\longrightarrow} 0. \tag{4}$$

For a linear functional, continuity (for all $y \in \mathcal{D}$) turns out to be equivalent to the existence of an inequality

$$\left|L(y)\right| \leq \|L\| \cdot \|y\|, \tag{5}$$

where the number $\|L\|$ is defined as $\|L\| = \sup_{\|y\|\leq 1} \left|L(y)\right|$. Try to prove that $\|L\|$ defined in this way is also the smallest number that can be used in (5). The notation suggests what happens to be the case, namely that $\|L\|$ is a norm on the linear space of linear functionals.

In most cases we have met, the Gâteaux derivative may be written as a linear and continuous functional, $\delta J(y_0; v) = L_{y_0}(v)$.

Example: Let

$$J(y) = \frac{1}{2}\int_a^b y^2(x)\,dx,\ \ y \in C[a,b],\ \ a, b \text{ finite}. \tag{6}$$

Then

$$\delta J(y_0; v) = \int_a^b y_0(x)\,v(x)\,dx. \tag{7}$$

But the integral is linear, so that $L_{y_0}(v)$, defined for all $v \in C[a,b]$ by

$$L_{y_0}(v) = \int_a^b y_0(x)\,v(x)\,dx, \tag{8}$$

will be linear as well. Moreover,

$$\left|L_{y_0}(v)\right| \leq \int_a^b \left|y_0(x)\,v(x)\right| dx \leq \int_a^b \left|y_0(x)\right| dx \ \cdot \max_{x\in[a,b]} \left|v(x)\right| = \int_a^b \left|y_0(x)\right| dx \cdot \|v\|_\infty. \tag{9}$$

Thus, it is obvious that $\|L_{y_0}\| \leq \int_a^b \left|y_0(x)\right| dx < \infty$, since y_0 is a continuous and hence bounded function. In this case it is actually possible to prove that $\|L_{y_0}\| = \int_a^b \left|y_0(x)\right| dx$.

Definition: *The functional J is differentiable at y if it is possible to write*

$$J(y+v) = J(y) + L_y(v) + o(\|v\|), \tag{10}$$

where L_y is linear and continuous.

Recall that the notation $o(\|v\|)$ means

$$\frac{o(\|v\|)}{\|v\|} \underset{\|v\|\to 0}{\longrightarrow} 0. \tag{11}$$

Of course, in this case also $\delta J(y; v) = L_y(v)$ (Prove it yourself!). The linear functional L_y is called the *Fréchet derivative* of J at y. Note that we require that

$$\frac{J(y+v) - J(y) - L_y(v)}{\|v\|} \underset{\|v\|\to 0}{\longrightarrow} 0 \tag{12}$$

for the Fréchet derivative to exist, whereas the Gâteaux derivative does not even need norms.

The Gâteaux derivative is analogous to the directional derivative, and the Fréchet derivative is analogous to the gradient.

Comment: For functions the existence of all directional derivatives at a point does not imply that the function is differentiable in the point: Take a look at the function defined on $\mathbb{R}^2$ in polar coordinates as $f(\mathbf{x}) = r\cos(3\theta)$. All directional derivatives exist at the origin (check it!), but it is impossible to construct a tangent plane there. The situation is similar for the Gâteaux and Fréchet derivatives.

3 Functionals of Vector Functions

Let us consider a vector of functions,

$$\mathbf{y}(x) = \begin{pmatrix} y_1(x) \\ \vdots \\ y_n(x) \end{pmatrix} \tag{13}$$

for $y_i \in C[a,b]$, say. Such a collection of vector functions will be a linear space with a norm, defined, e.g. as $\|\mathbf{y}\| = \max_j \|y_j\|_\infty$, or $\|\mathbf{y}\| = \sum_{j=1}^n \max_j \|y_j\|_\infty$.

The definition of functionals on vector functions does not cause any particular difficulties, and the Gâteaux derivative of functionals acting on vector functions is defined exactly as before.

Consider then a differentiable functional $J(\mathbf{y})$, where $\delta J(\mathbf{y};\mathbf{v}) = L_{\mathbf{y}}(\mathbf{v})$. Because of the linearity of $L_{\mathbf{y}}$, we may split the action of $L_{\mathbf{y}}$ on $\mathbf{v}$ into a sum,

$$L_{\mathbf{y}}(\mathbf{v}) = \sum_{j=1}^n L_{\mathbf{y}}\left(\begin{bmatrix} 0 \\ \vdots \\ v_j \\ \vdots \\ 0 \end{bmatrix}\right) = \sum_{j=1}^n L_j(v_j). \tag{14}$$

Similar to regular functions, it is tempting to write

$$\nabla J(\mathbf{y}) = (L_1, L_2, \cdots, L_n), \tag{15}$$

and hence,

$$\delta J(\mathbf{y};\mathbf{v}) = L_{\mathbf{y}}(\mathbf{v}) = \nabla J(\mathbf{y})\,\mathbf{v}. \tag{16}$$

Let us consider the *Standard functional* in this setting,

$$F(\mathbf{y}) = \int_a^b f\left(x, \mathbf{y}(x), \mathbf{y}'(x)\right) dx, \tag{17}$$

where f now is a function of $1+2n$ arguments. We assume that f is as smooth as necessary, at least differentiable with continuous derivatives in the last $2n$ arguments.

By means of partial derivatives and partial integration, we obtain, exactly as before,

$$\begin{aligned}\delta F(\mathbf{y};\mathbf{v}) &= \int_a^b \left.\frac{df\left(x,\mathbf{y}(x)+\varepsilon\mathbf{v}(x),\mathbf{y}'(x)+\varepsilon\mathbf{v}'(x)\right)}{d\varepsilon}\right|_{\varepsilon=0} dx \\ &= \int_a^b \left\{\sum_{j=1}^n \frac{\partial f}{\partial y_j} v_j + \sum_{j=1}^n \frac{\partial f}{\partial y_j'} v_j'\right\} dx \qquad (18) \\ &= \left[\sum_{j=1}^n \frac{\partial f}{\partial y_j'}\left(x,\mathbf{y},\mathbf{y}'\right) v_j(x)\right]_a^b + \int_a^b \left\{\sum_{j=1}^n \left(\frac{\partial f}{\partial y_j} - \frac{d}{dx}\frac{\partial f}{\partial y_j'}\right) v_j\right\} dx \qquad (19)\end{aligned}$$

In order for $\delta F(\mathbf{y};\mathbf{v})$ to be 0, it is sufficient to find a solution to the n *coupled* Euler equations,

$$\frac{\partial f}{\partial y_j}\left(x,\mathbf{y},\mathbf{y}'\right) - \frac{d}{dx}\frac{\partial f}{\partial y_j'}\left(x,\mathbf{y},\mathbf{y}'\right) = 0\ ,\ j = 1,\cdots,n, \qquad (20)$$

and the corresponding boundary conditions,

$$\left[\sum_{j=1}^n \frac{\partial f}{\partial y_j'}\left(x,\mathbf{y},\mathbf{y}'\right) v_j(x)\right]_a^b = 0. \qquad (21)$$

Also in this case, if the endpoints for all $\mathbf{y}(x)$ are fixed, then $\mathbf{v}(a) = \mathbf{v}(b) = 0$, and the boundary term vanishes. In other situations, we need to impose natural boundary conditions.

In the following it is convenient to use the short-hand notation

$$\frac{\partial f}{\partial \mathbf{y}}\left(x,\mathbf{y},\mathbf{y}'\right) - \frac{d}{dx}\frac{\partial f}{\partial \mathbf{y}'}\left(x,\mathbf{y},\mathbf{y}'\right) = 0 \qquad (22)$$

for all n equations in (20).

4 Variational Principles in Mechanics

Classical mechanics is a very old and important application of variational calculus, but the field is far from "old-fashioned"! Deriving the equations of motion for high-speed robots, aircraft or satellites would be impossible without variational formulations, and Troutman (Chapter 8) contains much more material than we are able to cover here. The most famous textbook on the topic is probably H. Goldstein: *Classical Mechanics*, Addison Westley.

You should be warned that what is covered below is just a tiny fraction of the field, but it may perhaps convince you about the strength of the method.

Consider the well-known Newton's Law for the motion of a point mass under the influence of a time independent external force,

$$\frac{d}{dt}(m\dot{\mathbf{x}}) = \mathbf{F}(\mathbf{x})\,,\ (\dot{\mathbf{x}} = d\mathbf{x}/dt). \qquad (23)$$

Could this equation be the Euler equation of some standard functional, say

$$\frac{d}{dt}\frac{\partial L(\mathbf{x},\dot{\mathbf{x}})}{\partial \dot{\mathbf{x}}} - \frac{\partial L(\mathbf{x},\dot{\mathbf{x}})}{\partial \mathbf{x}} = 0? \qquad (24)$$

From the left hand term of Eqn. 23, we should have that

$$\frac{\partial L}{\partial \dot{\mathbf{x}}} = m\dot{\mathbf{x}}, \tag{25}$$

or,

$$L(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2} m |\dot{\mathbf{x}}|^2 - V(\mathbf{x}) \tag{26}$$

for some function V of $\mathbf{x}$. This would satisfy the first part. In the same way, the second part will be satisfied if

$$\mathbf{F}(\mathbf{x}) = \frac{\partial V}{\partial \mathbf{x}} = \left(\frac{\partial V}{\partial x_1}, \frac{\partial V}{\partial x_2}, \frac{\partial V}{\partial x_3} \right) = \nabla V(\mathbf{x}). \tag{27}$$

This may begin to look familiar, since $T = \frac{1}{2} m |\dot{\mathbf{x}}|^2$ is equal to the *kinetic energy* of the mass particle, $V(\mathbf{x})$ is its *potential energy,* and the force derived from V as $\mathbf{F} = \nabla V$. Thus,

$$L(\mathbf{x}, \dot{\mathbf{x}}) = T(\dot{\mathbf{x}}) - V(\mathbf{x}) = \frac{1}{2} m |\dot{\mathbf{x}}|^2 - V(\mathbf{x}). \tag{28}$$

The function L is called the *Lagrange function* or the *Lagrangian in mechanics*, and the corresponding functional,

$$\mathcal{L}(\mathbf{x}) = \int_a^b L(\mathbf{x}(t), \dot{\mathbf{x}}(t))\, dt \tag{29}$$

is called the *action integral*.

It turns out that this situation is quite general. The Euler equations of the action integral with $L = T - V$ are the equations of motion for the mechanical system. The solution to the Euler equations, along with the appropriate boundary conditions, defines a *stationary point* ($\delta L(\mathbf{x}; \mathbf{v}) = 0$) for the action integral. However, $\mathcal{L}$ is in general *not* convex, which means that we can not guarantee this is a minimum.

Constraints on the motion are a problem in the formulation above. The motion is often restricted to move on a surface, say $h(\mathbf{x}) = 0$. We met this problem before, and recall that constraints as this one define a *manifold* of feasible points,

$$\mathcal{S} = \{\mathbf{x}\ ;\ h_i(\mathbf{x}) = 0,\ i = 1, \cdots, M\}. \tag{30}$$

The points on $\mathcal{S}$ may often be parametrized in terms of a set of new parameters $\mathbf{q}$, varying freely, so that $\mathbf{x} = \mathbf{G}(\mathbf{q})$. This is for example true locally around every point where the LICQ holds. The observation made Hamilton discover a way to remove the constraints from the problem, but still keeping the same form of the action integral. Since $\mathbf{x}(t) = \mathbf{G}(\mathbf{q}(t))$, we also have $\dot{\mathbf{x}} = \frac{\partial \mathbf{G}}{\partial \mathbf{q}} \dot{\mathbf{q}}$, and we may rewrite the action integral as

$$\tilde{\mathcal{L}}(\mathbf{q}) \triangleq \mathcal{L}(\mathbf{G}(\mathbf{q})) = \int_a^b L\left(\mathbf{G}(\mathbf{q}(t)), \frac{\partial \mathbf{G}}{\partial \mathbf{q}}(\mathbf{q}(t))\, \dot{\mathbf{q}}(t) \right) dt = \int_a^b \tilde{L}(\mathbf{q}(t), \dot{\mathbf{q}}(t))\, dt, \tag{31}$$

where now $\mathbf{q}$ moves freely. This is a standard functional in $\mathbf{q}$, and we may go on and solve the Euler equations (dropping ˜)

$$\frac{d}{dt} L_{\dot{\mathbf{q}}_i}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) - L_{\dot{\mathbf{q}}}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) = 0 \tag{32}$$

along with the boundary conditions. The new variables are called *generalized coordinates*.

The recipe is thus as follows:

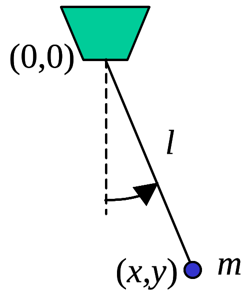

Figure 1: Simple pendulum.

1. Identify the generalized coordinates, $\mathbf{q}$, which may be varied freely, and at the same time define how the system moves.
2. Express the kinetic (T) and potential (V) energies in terms of $\mathbf{q}$ and $\dot{\mathbf{q}}$.
3. Form $L = T - V$ and solve the Euler equations,

$$\frac{d}{dt} L_{\dot{\mathbf{q}}} \left(\mathbf{q}\left(t\right), \dot{\mathbf{q}}\left(t\right)\right) - L_{\mathbf{q}} \left(\mathbf{q}\left(t\right), \dot{\mathbf{q}}\left(t\right)\right) = 0. \tag{33}$$

   Identify and fit the appropriate boundary conditions.
4. Obtain $\mathbf{x}\left(t\right)$ from $\mathbf{x}\left(t\right) = \mathbf{G}\left(\mathbf{q}\left(t\right)\right)$.

4.1 Some Examples

4.1.1 Equilibrium

At an equilibrium point, there is no motion and $T = 0$. The Euler equations become

$$\frac{\partial L}{\partial \mathbf{q}_i} = \frac{\partial V}{\partial \mathbf{q}_i} = 0 \tag{34}$$

As expected, the equilibrium points are stationary points of the potential energy.

4.1.2 The Pendulum

Consider a standard mathematical pendulum consisting of a stiff, weightless arm of length l and having mass m, as shown in Fig. 1.

The position of the pendulum is completely defined by θ, which is then the natural choice for a generalized coordinate. Moreover, $x = l\sin\theta$, $y = -l\cos\theta$. It is easy to express the kinetic and potential energies in terms of θ:

$$T = \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2\right) = \frac{ml^2}{2}\left(\dot{\theta}^2 \cos^2\theta + \dot{\theta}^2 \sin^2\theta\right) = \frac{ml^2}{2}\dot{\theta}^2, \tag{35}$$

$$V = mg\left(l + y\right) = mgl\left(1 - \cos\theta\right). \tag{36}$$

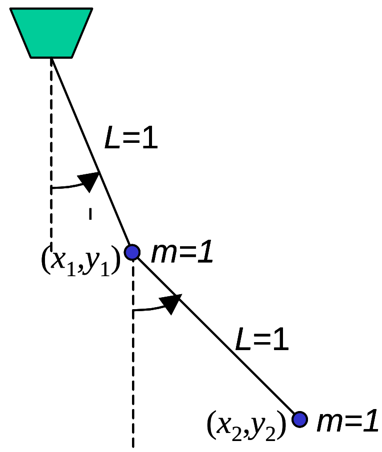

Figure 2: A double pendulum.

The Lagrangian is then

$$L = T - V = \frac{ml^2}{2}\dot{\theta}^2 - mgl\left(1 - \cos\theta\right), \tag{37}$$

and the Euler equation follows immediately:

$$\frac{d}{dt}L_{\dot{\theta}} - L_{\theta} = \frac{d}{dt}\left(ml^2\dot{\theta}\right) + mgl\sin\theta = 0, \tag{38}$$

that is,

$$\ddot{\theta} + \frac{g}{l}\sin\theta = 0. \tag{39}$$

You could consider some simple generalizations, like a hanging pendulum which is free to swing in two directions, a pendulum where the arm is a spring etc. (see also the Web-references below).

4.1.3 The Double Pendulum

Figure 2 shows a double pendulum, for simplicity shown in dimensionless form with equal masses (=1) and stiff arms (=1). We set the acceleration of gravity, g, equal to 1.

The obvious generalized coordinates are θ_1 and θ_2 shown in the figure, and using the suspension point as the origin, the positions of the masses are given by

$$\begin{aligned} x_1 &= \sin\theta_1,\ y_1 = -\cos\theta_1, \\ x_2 &= \sin\theta_1 + \sin\theta_2,\ y_2 = -\cos\theta_1 - \cos\theta_2. \end{aligned} \tag{40}$$

Show first that

$$v_1^2 = \dot{\theta}_1^2, \tag{41}$$

$$v_2^2 = \dot{\theta}_1^2 + \dot{\theta}_2^2 + 2\dot{\theta}_1\dot{\theta}_2\cos\left(\theta_1 - \theta_2\right), \tag{42}$$

and then,

$$T = \dot{\theta}_1^2 + \frac{\dot{\theta}_2^2}{2} + \dot{\theta}_1\dot{\theta}_2\cos\left(\theta_1 - \theta_2\right), \tag{43}$$

$$V = -\cos\theta_1 - \left(\cos\theta_1 + \cos\theta_2\right). \tag{44}$$

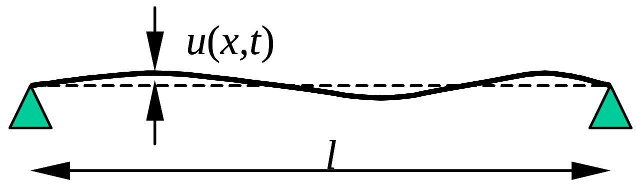

Figure 3: An elastic string.

The expression for Lagrangian is therefore

$$L = T - V = \dot{\theta}_1^2 + \frac{\dot{\theta}_2^2}{2} + \dot{\theta}_1\dot{\theta}_2 \cos(\theta_1 - \theta_2) + 2\cos\theta_1 + \cos\theta_2, \tag{45}$$

and the equations of motion follows:

$$\begin{aligned}\frac{d}{dt}L_{\dot{\theta}_1} - L_{\theta_1} &= \frac{d}{dt}\left[2\dot{\theta}_1 + \dot{\theta}_2\cos(\theta_1 - \theta_2)\right] - \dot{\theta}_1\dot{\theta}_2\sin(\theta_1 - \theta_2) + 2\sin\theta_1 \\ &= 2\ddot{\theta}_1 + \ddot{\theta}_2\cos(\theta_1 - \theta_2) + \dot{\theta}_2\left(\dot{\theta}_1 - \dot{\theta}_2\right)\sin(\theta_1 - \theta_2) - \dot{\theta}_1\dot{\theta}_2\sin(\theta_1 - \theta_2) + 2\sin\theta_1 \\ &= 2\ddot{\theta}_1 + \ddot{\theta}_2\cos(\theta_1 - \theta_2) + \dot{\theta}_2^2\sin(\theta_1 - \theta_2) + 2\sin\theta_1 = 0,\end{aligned} \tag{46}$$

and

$$\frac{d}{dt}L_{\dot{\theta}_2} - L_{\theta_2} = 2\ddot{\theta}_2 + \ddot{\theta}_1\cos(\theta_1 - \theta_2) - \dot{\theta}_1^2\sin(\theta_1 - \theta_2) + \sin\theta_2 = 0, \tag{47}$$

The equations may be restated as a system of four first order equations, and are said to show chaotic motion. More information about the double pendulum and Java applets showing the motion may be found on the *Internet*. Check out

- `scienceworld.wolfram.com/physics/DoublePendulum.html.`
- `www.myphysicslab.com/dbl_pendulum.html` (many interesting cases and animations!)
- `www.maths.tcd.ie/~plynch/SwingingSpring/doublependulum.html` (nice applet!)

The equations for the double pendulum are terrible to derive without variational calculus!

4.1.4 The Elastic String

In the final example we consider a tight elastic string supported at fixed ends, see Fig. 3. When the string moves, there is kinetic energy coming from the mass of the string (say ρ kilos per meter), and potential energy coming from the stretching.

We consider small excursions and assume that all points on the string move in one plane and orthogonally to the x-axis. This means that it is possible to write the kinetic energy at any instant of time as

$$T = \frac{1}{2}\int_0^l \rho u_t(x,t)^2\, dx. \tag{48}$$

For the potential energy we let the equilibrium (string without motion) define $V=0$. The length of the string at any instant of time is

$$H=\int_0^l \sqrt{1+u_x(x,t)^2}dx, \tag{49}$$

and assuming linear elasticity and small deflections (or rather, $|u_x(x,t)| \ll 1$),

$$V=\mu(H-l)=\mu\left(\int_0^l\left(\sqrt{1+u_x(x,t)^2}-1\right)dx\right)\approx\frac{\mu}{2}\int_0^l u_x(x,t)^2\,dx, \tag{50}$$

for some constant of elasticity, μ.

The action integral is therefore

$$\mathcal{L}(u)=\int_0^T\int_0^l\left(\frac{\rho}{2}u_t(x,t)^2-\frac{\mu}{2}u_x(x,t)^2\right)dxdt, \tag{51}$$

and the Gâteaux derivative,

$$\delta\mathcal{L}(u;v)=\int_0^T\int_0^l(\rho u_t v_t-\mu u_x v_x)\,dxdt=\int_0^T\int_0^l(U_1v_t+U_2v_x)\,dxdt, \tag{52}$$

where $\mathbf{U}=(U_1,U_2)=(\rho u_t,-\mu u_x)$. The trick we used for the soap-film equation also works here:

$$\begin{aligned}\delta\mathcal{L}(u;v)&=\int_0^T\int_0^l(U_1v_t+U_2v_x)\,dxdt\\&=\int_0^T\int_0^l\left\{\frac{\partial}{\partial t}(U_1v)+\frac{\partial}{\partial x}(U_2v)-U_{1t}v-U_{2x}v\right\}dxdt\\&=\int_0^T\int_0^l\{\nabla\cdot(\mathbf{U}v)-(U_{1t}+U_{2x})\,v\}\,dxdt\\&=\oint_{\partial\mathcal{R}}\mathbf{n}\cdot(\mathbf{U}v)\,ds-\int_0^T\int_0^l(U_{1t}+U_{2x})\,vdxdt\end{aligned} \tag{53}$$

The integral around the boundary of $\mathcal{R}=[0,T]\times[0,l]$ vanishes on the x-boundaries since $v=0$, and on the t-boundaries it will also vanish if we choose appropriate boundary conditions. What is of interest to us is the Euler Equation,

$$U_{1t}+U_{2x}=0, \tag{54}$$

which, as expected, is equal to the wave equation,

$$\rho u_{tt}-\mu u_{xx}=0. \tag{55}$$

You could try to look at some of the approximations we introduced above and derive more accurate equations. The classic book *The Theory of Sound* by Lord Rayleigh contains the derivation of the fourth order equation for a stiff string (which, by a closer look, turns out to be more relevant for an oscillating iron bar in a prison cell window).