CONSTRAINED VARIATIONAL PROBLEMS

LONG CHEN

ABSTRACT. These notes introduce constrained variational problems from several viewpoints: finite-dimensional Lagrange multipliers, integral and pointwise constraints, mixed formulations, and dual Hamiltonian formulations. The goal is to show how constraints can be handled either by eliminating variables or by introducing multipliers, and how these viewpoints lead to primal, mixed, and dual formulations.

CONTENTS

1. Constrained Optimization Problems 1
2. Integral Constraint 3
3. Pointwise Constraints 5
4. Mixed Formulation 9
4.1. Constraint formulation 9
4.2. Hamiltonian and Dual Formulation 11
4.3. An example 12
4.4. Duality 15
References 15

1. CONSTRAINED OPTIMIZATION PROBLEMS

As an introductory example, consider the two-dimensional constrained optimization problem

$$\begin{aligned} &\min_{x\in\mathbb{R}^2} f(x), \\ &\text{s.t. } g(x) = 0. \end{aligned} \tag{1}$$

A practical way to handle the constraint is to introduce the Lagrangian, with a multiplier λ enforcing the constraint:

$$L(x,\lambda) = f(x) + \lambda g(x).$$

We then look for its critical points:

$$\begin{aligned} \nabla_x L &= \nabla f(x) + \lambda \nabla g(x) = 0, \\ \partial_\lambda L &= g(x) = 0. \end{aligned}$$

We now explain why the Lagrange multiplier appears. For the unconstrained problem $\min_x f(x)$, the variable $x = (x_1, x_2)$ is free to move in $\mathbb{R}^2$. The constraint $g(x_1, x_2) = 0$ removes one degree of freedom, so the variables x_1 and x_2 are coupled and the admissible

Date: May 15, 2026.

points lie on a curve. Locally, when $\nabla g \neq 0$, the implicit function theorem allows us to parametrize this curve. More generally, we may write it as

$$x = x(t), \text{ so that } g(x(t)) = 0.$$

Thus the constrained problem becomes a one-dimensional unconstrained problem along the constraint curve.

Suppose $x_0 = x(0)$ is a local minimum on the constraint curve and $x'(0) \neq 0$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x(t))\Big|_{t=0} = \nabla f(x_0) \cdot x'(0) = 0.$$

Also, since $g(x(t)) = 0$ for all t near 0,

$$\frac{\mathrm{d}}{\mathrm{d}t} g(x(t))\Big|_{t=0} = \nabla g(x_0) \cdot x'(0) = 0.$$

Hence both $\nabla f(x_0)$ and $\nabla g(x_0)$ are orthogonal to the tangent vector $x'(0)$. Therefore they are parallel, and there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(x_0) + \lambda \nabla g(x_0) = 0.$$

See Fig. 1 for an illustration. When $\nabla g(x_0) \neq 0$, we can compute

$$\lambda = -\frac{(\nabla f(x_0), \nabla g(x_0))}{\|\nabla g(x_0)\|^2}.$$

If $x'(0) = 0$, then the chosen parametrization is degenerate at $t = 0$, and the above argument gives no information. In that case, one should reparametrize the constraint curve, or argue directly in terms of tangent vectors to the level set $g(x) = 0$.

If $\nabla g(x_0) = 0$, then x_0 is a singular point of the constraint set $g(x) = 0$. In this case, the level set need not be a smooth curve near x_0, and the standard Lagrange multiplier condition may fail. One then needs a more careful analysis of the local geometry of the constraint set, which is beyond the scope of these lecture notes.

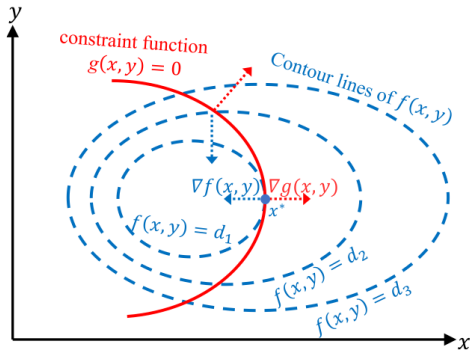


FIGURE 1. Minimization of $f(x, y)$ subject to the constraint $g(x, y) = 0$. At a constrained local minimum, the gradients of f and g are parallel, that is, $\nabla f + \lambda \nabla g = 0$.

This simple example shows the basic idea of a constrained problem. A constraint introduces a relation among otherwise free variables. One may use this relation to eliminate degrees of freedom, or introduce a new variable, called a Lagrange multiplier, to enforce the constraint.

2. Integral Constraint

Assume that $\Omega \subset \mathbb{R}^d$ is a bounded open domain. Let

$$L, G \in C^2\left(\bar{\Omega} \times \mathbb{R}^n \times \mathbb{R}^{n\times d}\right), \qquad u_D \in H^{1/2}\left(\partial\Omega, \mathbb{R}^n\right),$$

and define

$$\mathcal{M} = \left\{u \in H^1\left(\Omega, \mathbb{R}^n\right) \mid u|_{\partial\Omega} = u_D\right\}.$$

We follow [1, Lecture 7], but use a simpler parametrization argument.

Consider the constrained variational problem

$$\begin{aligned} &\min_{u\in\mathcal{M}} I(u), \qquad \text{s.t. } N(u) = 0, \\ &I(u) = \int_\Omega L(x, u, \nabla u)\,\mathrm{d}x, \\ &N(u) = \int_\Omega G(x, u, \nabla u)\,\mathrm{d}x. \end{aligned} \tag{2}$$

Let $u \in \mathcal{M} \cap N^{-1}(0)$ and let $\phi \in H^1_0(\Omega, \mathbb{R}^n)$. The variation $u + \varepsilon\phi$ may not satisfy the constraint. To fix this, we introduce a second direction and consider

$$u + \varepsilon\phi + \tau\psi,$$

where $\psi \in H^1_0(\Omega, \mathbb{R}^n)$ is fixed. With a slight abuse of notation, define

$$I(\varepsilon, \tau) = I(u + \varepsilon\phi + \tau\psi), \qquad N(\varepsilon, \tau) = N(u + \varepsilon\phi + \tau\psi).$$

We are now in the same situation as in the two-dimensional example of Section 1: the variables (ε, τ) are not free, because they must satisfy the constraint $N(\varepsilon, \tau) = 0$. Thus one variable must be eliminated.

Assume that we can find a parametrization $(\varepsilon(t), \tau(t))$ such that

$$N(\varepsilon(t), \tau(t)) = 0$$

for t near 0, with

$$(\varepsilon(0), \tau(0)) = (0, 0),$$

and that the minimum is attained at $t = 0$. Then

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} I(\varepsilon(t), \tau(t))\Big|_{t=0} = \delta I(u, \phi)\,\varepsilon'(0) + \delta I(u, \psi)\,\tau'(0) = 0, \\ \dfrac{\mathrm{d}}{\mathrm{d}t} N(\varepsilon(t), \tau(t))\Big|_{t=0} = \delta N(u, \phi)\,\varepsilon'(0) + \delta N(u, \psi)\,\tau'(0) = 0. \end{cases} \tag{3}$$

Assume $\delta N(u, \psi) \neq 0$. Then we can solve for $\tau'(0)$ from the second equation and substitute into the first to obtain

$$\left[\delta I(u, \phi) - \frac{\delta I(u, \psi)}{\delta N(u, \psi)}\,\delta N(u, \phi)\right]\varepsilon'(0) = 0.$$

If $\varepsilon'(0) \neq 0$, this implies

$$\delta I(u, \phi) + \lambda\,\delta N(u, \phi) = 0,$$

where

$$\lambda = -\frac{\delta I(u, \psi)}{\delta N(u, \psi)}.$$

Although this formula is derived using the auxiliary direction ψ, the multiplier λ is in fact independent of the choice of ψ.

It remains to justify the parametrization. We have

$$N(0,0) = 0, \qquad \partial_\tau N(0,0) = \delta N(u,\psi) \neq 0.$$

Therefore, by the implicit function theorem, for $|\varepsilon|$ small there exists a function $\tau = \tau(\varepsilon)$ such that

$$\tau(0) = 0, \qquad N(\varepsilon, \tau(\varepsilon)) = 0.$$

We may then choose $\varepsilon = t$, so that $\varepsilon'(0) = 1 \neq 0$.

Since ψ is fixed and ϕ is arbitrary, we obtain the following result.

Theorem 2.1. *Suppose $N^{-1}(0) \cap \mathcal{M} \neq \emptyset$. Let $u \in \mathcal{M}$ be a weak minimum of $I(u)$ under the constraint $N(u) = 0$, that is,*

$$I(u) = \min_{w \in \mathcal{M} \cap N^{-1}(0)} I(w).$$

If there exists $\psi \in H_0^1(\Omega, \mathbb{R}^n)$ such that $\delta N(u,\psi) \neq 0$, then there exists $\lambda \in \mathbb{R}$ satisfying

$$\delta I(u,\phi) + \lambda\, \delta N(u,\phi) = 0, \qquad \forall\, \phi \in H_0^1(\Omega, \mathbb{R}^n). \tag{4}$$

The first-order necessary condition (4) can also be derived by introducing the multiplier λ and the adjusted Lagrangian

$$\mathcal{L}(x,u,\nabla u,\lambda) = L(x,u,\nabla u) + \lambda\, G(x,u,\nabla u),$$

and then considering the saddle problem

$$\inf_{u \in \mathcal{M}} \sup_{\lambda \in \mathbb{R}} \int_\Omega \mathcal{L}(x,u,\nabla u,\lambda)\,\mathrm{d}x.$$

The corresponding first-order conditions are

$$\delta_u \int_\Omega \mathcal{L}(x,u,\nabla u,\lambda)\,\mathrm{d}x = 0,$$
$$\partial_\lambda \int_\Omega \mathcal{L}(x,u,\nabla u,\lambda)\,\mathrm{d}x = 0.$$

The variation ∂_λ gives back the constraint

$$\int_\Omega G(x,u,\nabla u)\,\mathrm{d}x = 0.$$

The first condition is the Euler–Lagrange equation for the adjusted Lagrangian:

$$-\nabla \cdot \big(L_p(x,u,\nabla u) + \lambda\, G_p(x,u,\nabla u)\big) + L_u(x,u,\nabla u) + \lambda\, G_u(x,u,\nabla u) = 0.$$

Equivalently,

$$-\nabla \cdot \mathcal{L}_p(x,u,\nabla u,\lambda) + \mathcal{L}_u(x,u,\nabla u,\lambda) = 0.$$

We now include the existence result from Evans [2, Chapter 8]. Consider the integral constraint

$$N(w) := \int_\Omega G(w)\,\mathrm{d}x = 0,$$

where $G : \mathbb{R} \to \mathbb{R}$ is a given smooth function, and define the admissible class

$$\mathcal{A} = \left\{ w \in H_0^1(\Omega) \mid N(w) = 0 \right\}.$$

Theorem 2.2 (Existence of constrained minimizer)**.** *Assume that L satisfies the coercivity inequality and is convex in the variable p. Assume that the admissible set $\mathcal{A}$ is nonempty and that*

$$|G'(z)| \leq C(|z|+1)$$

for some constant C. Then there exists $u \in \mathcal{A}$ such that

$$I(u) = \min_{w \in \mathcal{A}} I(w).$$

Proof. Choose a minimizing sequence $\{u_k\}_{k=1}^{\infty} \subset \mathcal{A}$ such that

$$I(u_k) \to \inf_{w \in \mathcal{A}} I(w).$$

By coercivity, $\{u_k\}$ is bounded in $H_0^1(\Omega)$. Hence, after passing to a subsequence,

$$u_{k_j} \rightharpoonup u \quad \text{weakly in } H_0^1(\Omega), \qquad u_{k_j} \to u \quad \text{strongly in } L^2(\Omega).$$

By weak lower semicontinuity,

$$I(u) \leq \liminf_{j \to \infty} I(u_{k_j}) = \inf_{w \in \mathcal{A}} I(w).$$

It remains to verify that $N(u) = 0$, so that $u \in \mathcal{A}$. Since $N(u_{k_j}) = 0$,

$$|N(u)| = |N(u) - N(u_{k_j})| \leq \int_\Omega |G(u) - G(u_{k_j})| \, \mathrm{d}x.$$

By the mean value theorem and the growth condition on G',

$$|G(u) - G(u_{k_j})| \leq C|u - u_{k_j}| \left(1 + |u| + |u_{k_j}|\right).$$

Therefore

$$\begin{aligned} |N(u)| &\leq C \int_\Omega |u - u_{k_j}| \left(1 + |u| + |u_{k_j}|\right) \, \mathrm{d}x \\ &\to 0 \qquad \text{as } j \to \infty. \end{aligned}$$

Hence $N(u) = 0$, and so $u \in \mathcal{A}$. □

3. Pointwise Constraints

We now consider the pointwise constraint

$$N(u(x)) = 0 \qquad \text{for a.e. } x \in \Omega \subset \mathbb{R}^d,$$

where $u : \Omega \to \mathbb{R}^n$ is a vector-valued function and $N : \mathbb{R}^n \to \mathbb{R}$ is smooth. We consider the constrained variational problem

$$\begin{aligned} &\min_{u \in \mathcal{M}} I(u), \qquad I(u) = \int_\Omega L(x, u, \nabla u) \, \mathrm{d}x, \\ &\text{s.t. } N(u(x)) = 0 \qquad \text{for a.e. } x \in \Omega, \end{aligned} \tag{5}$$

with the Dirichlet boundary condition

$$\mathcal{M} = \left\{ u \in H^1(\Omega, \mathbb{R}^n) \mid u|_{\partial\Omega} = u_D \right\}.$$

Here u is vector-valued and ∇u is matrix-valued:

$$L\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}, \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \begin{pmatrix} - - p_1 - - \\ \vdots \\ - - p_n - - \end{pmatrix} \right).$$

$$\mathbb{R}^d \qquad \mathbb{R}^n \qquad \mathbb{R}^{n \times d}$$

Again we follow [1, Lecture 7], but use a simpler presentation. Let $u \in \mathcal{M} \cap N^{-1}(0)$ and $\phi \in H_0^1(\Omega, \mathbb{R}^n)$. We first recall the unconstrained variation

$$u + \varepsilon\phi_1\vec{e}_1 = \begin{pmatrix} u_1 + \varepsilon\phi_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} I\left(u + \varepsilon\phi_1\vec{e}_1\right)\Big|_{\varepsilon=0} = \int_\Omega \left(L_{u_1}\phi_1 + L_{p_1}\cdot\nabla\phi_1\right)\,\mathrm{d}x = \int_\Omega \left(-\nabla\cdot L_{p_1} + L_{u_1}\right)\phi_1\,\mathrm{d}x.$$

Here we use integration by parts:

$$\int_\Omega L_{p_1}\cdot\nabla\phi_1\,\mathrm{d}x = -\int_\Omega (\nabla\cdot L_{p_1})\phi_1\,\mathrm{d}x.$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} I\left(u + \varepsilon\phi_1\vec{e}_1\right)\Big|_{\varepsilon=0} = 0 \qquad \forall\,\phi_1 \in H_0^1(\Omega)$$

implies the Euler–Lagrange system

$$-\nabla\cdot L_{p_i} + L_{u_i} = 0, \qquad i = 1, \cdots, n.$$

In compact form,

$$-\nabla\cdot L_p + L_u = 0,$$

where L_p is an $n \times d$ matrix field, and the divergence is taken row-wise. Thus $\nabla\cdot L_p$ is an $n \times 1$ vector field, which matches the dimension of L_u.

In the constrained case, the variation $u + \varepsilon\phi_1\vec{e}_1$ may not satisfy the constraint. To address this, we use a local parametrization of the constraint manifold. Formally, let

$$P : H^1(\Omega) \to H^1(\Omega)$$

be a local projection such that

$$N(P(v)) = 0, \qquad P(u) = u \quad \text{for } u \in \mathcal{M} \cap N^{-1}(0).$$

We then consider the first-order condition

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} I(P(u+\varepsilon\phi))\Big|_{\varepsilon=0} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\int_\Omega L\big(x, P(u+\varepsilon\phi), \nabla_x P(u+\varepsilon\phi)\big)\,\mathrm{d}x\Big|_{\varepsilon=0} = 0. \tag{6}$$

Here $P(u)$ is again a function on Ω. We write ∇_x for the derivative with respect to the spatial variable x, and ∇_u for the derivative with respect to the variable u.

Example 3.1. A basic example is the sphere

$$S^{n-1} = \{u \in \mathbb{R}^n : |u|^2 = 1\}, \qquad N(u) = |u|^2 - 1 = 0.$$

Near a point $u_0 \in S^{n-1}$ with $(u_0)_n \neq 0$, the implicit function theorem allows us to solve for the last component. Locally, the constraint manifold is parametrized by the first $n-1$ components, with local projection

$$P(v_1, \ldots, v_n) = \left(v_1, \ldots, v_{n-1}, \sqrt{1 - v_1^2 - \cdots - v_{n-1}^2}\right),$$

or the same formula with the negative sign. This projection is local: it depends on the chosen chart and breaks down across hemispheres. Here the constraint removes one degree of freedom by forcing u to lie on the sphere, and P restores the constraint by mapping back to the chosen local chart.

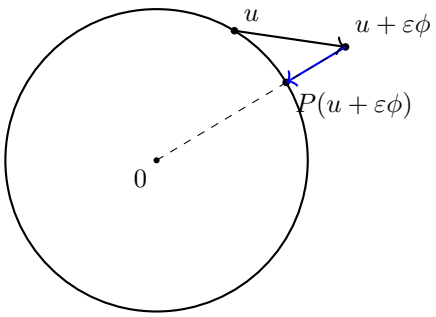


FIGURE 2. A local picture of the projection onto the constraint manifold. The point u lies on the sphere $N(u) = 0$. The perturbation $u + \varepsilon\phi$ leaves the sphere, and the projection P maps it back to the sphere.

It remains to determine the local projection and its derivative. The nondegeneracy condition is on the derivative of N with respect to u. Assume that

$$\nabla_u N(u(x)) \neq 0 \qquad \forall x \in \Omega.$$

Then for each $x_0 \in \Omega$, there exists a ball $B_r(x_0) \subset \Omega$ such that, after relabeling coordinates if needed,

$$N_{u_n}(u(x)) \neq 0, \qquad \forall\, x \in B_r(x_0).$$

By the implicit function theorem, there exists a C^2 function U such that

$$N\left(u_1 + \varepsilon\phi_1, u_2, \cdots, U\left(u_1 + \varepsilon\phi_1, \cdots, u_{n-1}\right)\right) = 0. \tag{7}$$

Thus, locally,

$$P((v_1, v_2, \ldots, v_n)) = (v_1, v_2, \ldots, v_{n-1}, U(v_1, \ldots, v_{n-1})).$$

That is, locally we use the constraint to eliminate one degree of freedom.

We then substitute

$$L\left(x, \begin{pmatrix} u_1 + \varepsilon\phi_1 \\ u_2 \\ \vdots \\ U\left(u_1 + \varepsilon\phi_1, \cdots, u_{n-1}\right) \end{pmatrix}, \begin{pmatrix} \nabla_x u_1 + \varepsilon\nabla_x\phi_1 \\ \nabla_x u_2 \\ \vdots \\ \nabla_x U\left(u_1 + \varepsilon\phi_1, \cdots\right) \end{pmatrix}\right).$$

The variation in (6) becomes

$$\begin{aligned} &L_{u_1}\phi_1 + L_{p_1}\cdot\nabla\phi_1 \\ &\qquad + L_{u_n}\partial_{u_1}U\,\phi_1 + L_{p_n}\cdot\partial_\varepsilon\nabla_x U\left(u_1 + \varepsilon\phi_1, \cdots\right). \end{aligned}$$

Since $\partial_\varepsilon\nabla_x = \nabla_x\partial_\varepsilon$, the last term is

$$L_{p_n}\cdot\nabla_x\left(\partial_{u_1}U\,\phi_1\right).$$

Applying integration by parts, we obtain

$$-\nabla\cdot L_{p_1} + L_{u_1} + \left(L_{u_n} - \nabla\cdot L_{p_n}\right)\partial_{u_1}U = 0. \tag{8}$$

To compute $\partial_{u_1}U$, differentiate the constraint $N(u) = 0$, that is, (7), with respect to u_1:

$$N_{u_1} + N_{u_n}\partial_{u_1}U = 0 \qquad \Longrightarrow \qquad \partial_{u_1}U = -\frac{N_{u_1}}{N_{u_n}}.$$

Substituting this into (8) and setting

$$\lambda = \frac{1}{N_{u_n}} \left(\nabla \cdot L_{p_n} - L_{u_n} \right),$$

we arrive formally at the following result.

Theorem 3.2. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain. Let*

$$L \in C^2 \left(\bar{\Omega} \times \mathbb{R}^n \times \mathbb{R}^{n \times d}, \mathbb{R} \right), \qquad N \in C^2 \left(\mathbb{R}^n, \mathbb{R} \right), \qquad u_D \in H^{1/2} \left(\partial\Omega, \mathbb{R}^n \right),$$

and

$$\mathcal{M} = \left\{ u \in H^1 \left(\Omega, \mathbb{R}^n \right) \mid u|_{\partial\Omega} = u_D \right\}.$$

Suppose $u \in \mathcal{M}$ is a local minimum under the pointwise constraint and is C^2 away from finitely many $(d-1)$-dimensional piecewise C^1 hypersurfaces. If

$$\nabla_u N \left(u(x) \right) \neq 0 \qquad \forall x \in \bar{\Omega},$$

then there exists a scalar multiplier $\lambda = \lambda(x)$ such that u satisfies the Euler–Lagrange equation of the adjusted Lagrangian $\mathcal{L} = L + \lambda N$:

$$-\nabla \cdot L_{p_i} + L_{u_i} + \lambda N_{u_i} = 0, \qquad 1 \leq i \leq n.$$

Proof. The previous argument was written for the variation in the first component, so it gives the equation for $i = 1$. The same argument applies to any variation in the components $u_1, \ldots, u_{n-1}$, and therefore yields

$$-\nabla \cdot L_{p_i} + L_{u_i} + \lambda N_{u_i} = 0, \qquad 1 \leq i \leq n-1.$$

For $i = n$, the equation follows directly from the definition

$$\lambda = \frac{1}{N_{u_n}} \left(\nabla \cdot L_{p_n} - L_{u_n} \right),$$

which is equivalent to

$$-\nabla \cdot L_{p_n} + L_{u_n} + \lambda N_{u_n} = 0.$$

Hence the full system

$$-\nabla \cdot L_{p_i} + L_{u_i} + \lambda N_{u_i} = 0, \qquad 1 \leq i \leq n,$$

holds.

The previous argument is local, since the implicit function theorem only gives a local parametrization of the constraint manifold near each point $x_0 \in \Omega$. Thus the Euler–Lagrange equation is first obtained only on a neighborhood of x_0. To conclude that it holds on the whole domain, we cover Ω by finitely many such neighborhoods $\{U_\alpha\}$ and choose a partition of unity $\{\eta_\alpha\}$ subordinate to this cover. For any test function $\phi \in C_c^\infty(\Omega, \mathbb{R}^n)$, we write

$$\phi = \sum_\alpha \eta_\alpha \phi.$$

Since each $\eta_\alpha \phi$ is supported in a single chart U_α, the local argument applies to each term. Summing over α, we obtain the weak form of the Euler–Lagrange system on all of Ω. Therefore the equation

$$-\nabla \cdot L_{p_i} + L_{u_i} + \lambda N_{u_i} = 0, \qquad 1 \leq i \leq n,$$

holds on the whole domain. □

Notice that for a scalar pointwise constraint $N(u(x)) = 0$, the multiplier $\lambda = \lambda(x)$ is a scalar function on Ω, not a vector field.

Example 3.3. A closely related example is the Stokes equation, where the constraint is

$$\nabla \cdot u = 0 \qquad \text{in } \Omega.$$

Here $u : \Omega \to \mathbb{R}^d$ is the velocity field, and the constraint means that the flow is incompressible. Consider the energy

$$I(u) = \int_\Omega \frac{1}{2}|\nabla u|^2 - (f, u)\,\mathrm{d}x$$

subject to the incompressibility constraint and suitable boundary conditions. Introducing a scalar Lagrange multiplier $p = p(x)$ for the constraint $\nabla \cdot u = 0$, we obtain the adjusted Lagrangian

$$\mathcal{L}(u, p) = \frac{1}{2}|\nabla u|^2 - (f, u) + p\,\nabla \cdot u.$$

The Euler–Lagrange equation is then

$$-\Delta u + \nabla p = f, \qquad \nabla \cdot u = 0.$$

In this example, the Lagrange multiplier p is exactly the pressure. Although the Stokes constraint depends on ∇u rather than only on u, it shows clearly how a pointwise constraint leads to an additional unknown with physical meaning.

4. Mixed Formulation

The weak form of the primal Euler–Lagrange equation is: find $u \in H^1(\Omega)$ with $u|_{\partial\Omega} = u_D$ such that

$$\int_\Omega L_p(x, u, \nabla u) \cdot \nabla v\,\mathrm{d}x + \int_\Omega L_u(x, u, \nabla u)\, v\,\mathrm{d}x = 0 \qquad \forall v \in H^1_0(\Omega). \tag{9}$$

Indeed, this is the first variation $\langle \delta I(u), v\rangle = 0$. We stop before integrating by parts. In this section, we derive mixed formulations using the dual variable and the Hamiltonian.

4.1. **Constraint formulation.** We treat the relation

$$p = \nabla u$$

as a constraint and introduce the modified Lagrangian

$$L(x, u, p) + \lambda \cdot (\nabla u - p),$$

where λ is a matrix-valued Lagrange multiplier of the same size as p. The corresponding functional is

$$J(u, p, \lambda) = \int_\Omega \left[L(x, u, p) + \lambda \cdot (\nabla u - p)\right]\,\mathrm{d}x.$$

Variation with respect to λ. For arbitrary $\delta\lambda$,

$$\delta_\lambda J = \int_\Omega \delta\lambda \cdot (\nabla u - p)\,\mathrm{d}x = 0.$$

Hence

$$\nabla u - p = 0 \qquad \text{in } \Omega.$$

Variation with respect to p. For arbitrary δp,

$$\delta_p J = \int_\Omega (L_p - \lambda) \cdot \delta p \, \mathrm{d}x = 0.$$

Therefore

$$\lambda = L_p.$$

Thus the Lagrange multiplier enforcing $p = \nabla u$ coincides with the dual variable in the Legendre transform. This illustrates a general principle: in many variational and optimization problems, dual variables arise naturally as Lagrange multipliers for constraints.

Variation with respect to u. Let $v = \delta u \in H_0^1(\Omega, \mathbb{R}^n)$. Then

$$\delta_u J = \int_\Omega \left(L_u \, v + \lambda \cdot \nabla v\right) \, \mathrm{d}x = 0.$$

Integrating by parts gives

$$\int_\Omega \lambda \cdot \nabla v \, \mathrm{d}x = -\int_\Omega (\nabla \cdot \lambda) \, v \, \mathrm{d}x,$$

where the divergence is taken row-wise. Hence

$$\int_\Omega (L_u - \nabla \cdot \lambda) \, v \, \mathrm{d}x = 0,$$

so

$$L_u - \nabla \cdot \lambda = 0 \qquad \text{in } \Omega.$$

Euler–Lagrange system. Combining the three variations gives

$$p = \nabla u, \tag{10a}$$

$$\lambda = L_p, \tag{10b}$$

$$L_u - \nabla \cdot \lambda = 0. \tag{10c}$$

If we substitute (10a) and (10b) into (10c), then we recover the standard Euler–Lagrange equation

$$L_u(x, u, \nabla u) - \nabla \cdot L_p(x, u, \nabla u) = 0.$$

Instead of eliminating p and λ pointwise, we may keep all three unknowns and work directly with a variational formulation of (10). One natural choice is

$$u \in H^1(\Omega; \mathbb{R}^n), \qquad p \in L^2(\Omega; \mathbb{R}^{n\times d}), \qquad \lambda \in L^2(\Omega; \mathbb{R}^{n\times d}),$$

with the Dirichlet condition imposed strongly:

$$u|_{\partial\Omega} = u_D.$$

The corresponding test spaces are

$$v \in H_0^1(\Omega; \mathbb{R}^n), \qquad q \in L^2(\Omega; \mathbb{R}^{n\times d}), \qquad \mu \in L^2(\Omega; \mathbb{R}^{n\times d}).$$

The mixed variational formulation is: find (u, p, λ) such that

$$\int_\Omega p \cdot q \, \mathrm{d}x = \int_\Omega \nabla u \cdot q \, \mathrm{d}x, \qquad \forall \, q \in L^2(\Omega; \mathbb{R}^{n\times d}), \tag{11a}$$

$$\int_\Omega \lambda \cdot \mu \, \mathrm{d}x = \int_\Omega L_p(x, u, p) \cdot \mu \, \mathrm{d}x, \qquad \forall \, \mu \in L^2(\Omega; \mathbb{R}^{n\times d}), \tag{11b}$$

$$\int_\Omega \lambda \cdot \nabla v \, \mathrm{d}x = -\int_\Omega L_u(x, u, p) \cdot v \, \mathrm{d}x, \qquad \forall \, v \in H_0^1(\Omega; \mathbb{R}^n). \tag{11c}$$

If we integrate the relation $p = \nabla u$ by parts, we obtain a different mixed formulation. In that case the natural spaces are

$$u \in L^2(\Omega;\mathbb{R}^n), \qquad p \in H(\mathrm{div};\Omega;\mathbb{R}^{n\times d}), \qquad \lambda \in H(\mathrm{div};\Omega;\mathbb{R}^{n\times d}),$$

and the boundary condition is imposed weakly. The corresponding test spaces are

$$v \in L^2(\Omega;\mathbb{R}^n), \qquad q \in H(\mathrm{div};\Omega;\mathbb{R}^{n\times d}), \qquad \mu \in H(\mathrm{div};\Omega;\mathbb{R}^{n\times d}).$$

The weak form becomes

$$\int_\Omega p\cdot q\,\mathrm{d}x + \int_\Omega u\,\nabla\cdot q\,\mathrm{d}x = \int_{\partial\Omega} u_D\,(q\cdot n)\,\mathrm{d}s, \qquad \forall\, q \in H(\mathrm{div};\Omega;\mathbb{R}^{n\times d}), \tag{12a}$$

$$\int_\Omega \lambda\cdot\mu\,\mathrm{d}x = \int_\Omega L_p(x,u,p)\cdot\mu\,\mathrm{d}x, \quad \forall\,\mu \in H(\mathrm{div};\Omega;\mathbb{R}^{n\times d}), \tag{12b}$$

$$\int_\Omega (\nabla\cdot\lambda)\cdot v\,\mathrm{d}x = \int_\Omega L_u(x,u,p)\cdot v\,\mathrm{d}x, \quad \forall\, v \in L^2(\Omega;\mathbb{R}^n). \tag{12c}$$

4.2. **Hamiltonian and Dual Formulation.** So far we have used the Lagrangian formulation. We now switch to the Hamiltonian formulation. Assume that L is convex with respect to p. The Hamiltonian, that is, the Legendre transform of L with respect to p, is

$$H(x,u,\sigma) = \sup_p \left((\sigma,p) - L(x,u,p)\right).$$

At the supremum, we have

$$\sigma = L_p, \qquad H(x,u,\sigma) = \sigma\cdot p - L(x,u,p).$$

Conversely,

$$L(x,u,p) = \sup_\sigma \left((\sigma,p) - H(x,u,\sigma)\right).$$

If we then substitute the pointwise relation $p = \nabla u$, the variational problem can be written formally as

$$\inf_u \int_\Omega L(x,u,\nabla u)\,\mathrm{d}x = \inf_u \sup_\sigma \int_\Omega \left[(\sigma,\nabla u) - H(x,u,\sigma)\right]\,\mathrm{d}x.$$

Taking the variation of u and σ, we obtain the mixed first-order system

$$\begin{aligned} \nabla u &= H_\sigma, \\ \nabla\cdot\sigma &= -H_u. \end{aligned}$$

Constitutive Relation: $\nabla u = H_\sigma$. Multiply by a test function $\tau \in H(\mathrm{div};\Omega)$ and integrate over Ω. Applying integration by parts to the left-hand side gives

$$\int_\Omega \tau\cdot\nabla u\,\mathrm{d}x = \int_{\partial\Omega} u(\tau\cdot n)\,\mathrm{d}s - \int_\Omega u(\nabla\cdot\tau)\,\mathrm{d}x.$$

In the mixed formulation, $u \in L^2(\Omega)$ does not have a classical trace on $\partial\Omega$, so the Dirichlet condition is imposed weakly by replacing the boundary value u with the prescribed data u_D. Thus we write

$$\int_\Omega H_\sigma\cdot\tau\,\mathrm{d}x - \int_\Omega u(\nabla\cdot\tau)\,\mathrm{d}x = \int_{\partial\Omega} u_D(\tau\cdot n)\,\mathrm{d}s.$$

Equilibrium Equation: $\nabla \cdot \sigma = -H_u$. Multiply by a test function $v \in L^2(\Omega)$ and integrate over Ω:

$$\int_\Omega (\nabla \cdot \sigma) v \,\mathrm{d}x = -\int_\Omega H_u v \,\mathrm{d}x.$$

The weak formulation is: find $u \in L^2(\Omega)$ and $\sigma \in H(\mathrm{div};\Omega)$ such that

$$\int_\Omega H_\sigma \cdot \tau \,\mathrm{d}x - \int_\Omega u\,(\nabla \cdot \tau)\,\mathrm{d}x = \int_{\partial\Omega} u_D(\tau \cdot n)\,\mathrm{d}s, \qquad \forall \tau \in H(\mathrm{div};\Omega), \tag{13a}$$

$$\int_\Omega (\nabla \cdot \sigma) v \;\mathrm{d}x = -\int_\Omega H_u v \;\mathrm{d}x, \qquad \forall v \in L^2(\Omega). \tag{13b}$$

These equations define the mixed formulation, while the Euler–Lagrange equation

$$L_u(x,u,\nabla u) - \nabla \cdot L_p(x,u,\nabla u) = 0$$

is called the primal formulation.

Compared with the mixed formulation of (u,p,λ), we see that the dual variable $\sigma = \lambda$ is exactly the Lagrange multiplier enforcing $p = \nabla u$. But these two mixed formulations are different: one is written in terms of the Lagrangian, and the other is written in terms of its Legendre dual Hamiltonian.

Remark 4.1. The system

$$\nabla u = H_\sigma, \qquad \nabla \cdot \sigma = -H_u$$

looks formally similar to Hamilton's equations

$$\dot q = H_p, \qquad \dot p = -H_q,$$

but the two should not be confused. Hamilton's equations are evolution equations in time on phase space, while the system above is a first-order PDE system in the spatial variable x, obtained from a constrained variational principle. It is therefore better viewed as a mixed Euler–Lagrange system or a primal-dual first-order system.

4.3. **An example.** Consider the quadratic Lagrangian

$$L(x,u,p) = \frac{1}{2}(a(x)p,p) - (f(x),u),$$

where $u : \Omega \to \mathbb{R}^n$, $p \in \mathbb{R}^{n\times d}$, $f : \Omega \to \mathbb{R}^n$, and $a(x)$ is a symmetric positive definite linear operator on $\mathbb{R}^{n\times d}$ for each $x \in \Omega$.

Primal formulation. Substituting $p = \nabla u$, the functional becomes

$$I(u) = \int_\Omega \left[\frac{1}{2}(a(x)\nabla u, \nabla u) - (f,u)\right] \mathrm{d}x.$$

Since

$$L_u = -f, \qquad L_p = a(x)p,$$

the Euler–Lagrange equation is

$$-\nabla \cdot (a(x)\nabla u) = f \qquad \text{in } \Omega.$$

Its weak form is: find $u \in H^1(\Omega)$ with $u|_{\partial\Omega} = u_D$ such that

$$\int_\Omega (a(x)\nabla u, \nabla v)\,\mathrm{d}x = \int_\Omega (f,v)\,\mathrm{d}x \qquad \forall v \in H^1_0(\Omega).$$

Mixed formulation in (u,p,λ). Introduce the constraint $p = \nabla u$ and define

$$J(u,p,\lambda) = \int_\Omega \left[\frac{1}{2}(a(x)p,p) - (f,u) + \lambda \cdot (\nabla u - p)\right] \mathrm{d}x.$$

The three variations give

$$p = \nabla u, \tag{14a}$$

$$\lambda = a(x)p, \tag{14b}$$

$$\nabla \cdot \lambda = -f. \tag{14c}$$

Substituting (14a) and (14b) into (14c) gives back

$$-\nabla \cdot (a(x)\nabla u) = f.$$

Strong mixed formulation in (u, p, λ). Find

$$u \in H^1(\Omega; \mathbb{R}^n), \qquad p \in L^2(\Omega; \mathbb{R}^{n\times d}), \qquad \lambda \in L^2(\Omega; \mathbb{R}^{n\times d}),$$

with $u|_{\partial\Omega} = u_D$, such that

$$\int_\Omega p \cdot q \,\mathrm{d}x = \int_\Omega \nabla u \cdot q \,\mathrm{d}x, \qquad \forall q \in L^2(\Omega; \mathbb{R}^{n\times d}), \tag{15a}$$

$$\int_\Omega \lambda \cdot \mu \,\mathrm{d}x = \int_\Omega a(x)p \cdot \mu \,\mathrm{d}x, \qquad \forall \mu \in L^2(\Omega; \mathbb{R}^{n\times d}), \tag{15b}$$

$$\int_\Omega \lambda \cdot \nabla v \,\mathrm{d}x = \int_\Omega (f, v) \,\mathrm{d}x, \qquad \forall v \in H^1_0(\Omega; \mathbb{R}^n). \tag{15c}$$

Weak mixed formulation in (u, p, λ). Find

$$u \in L^2(\Omega; \mathbb{R}^n), \qquad p \in H(\mathrm{div}; \Omega; \mathbb{R}^{n\times d}), \qquad \lambda \in H(\mathrm{div}; \Omega; \mathbb{R}^{n\times d}),$$

such that

$$\int_\Omega p \cdot q \,\mathrm{d}x + \int_\Omega u \,\nabla \cdot q \,\mathrm{d}x = \int_{\partial\Omega} u_D (q \cdot n) \,\mathrm{d}s, \qquad \forall q \in H(\mathrm{div}; \Omega; \mathbb{R}^{n\times d}), \tag{16a}$$

$$\int_\Omega \lambda \cdot \mu \,\mathrm{d}x = \int_\Omega a(x)p \cdot \mu \,\mathrm{d}x, \qquad \forall \mu \in H(\mathrm{div}; \Omega; \mathbb{R}^{n\times d}), \tag{16b}$$

$$\int_\Omega (\nabla \cdot \lambda) \cdot v \,\mathrm{d}x = -\int_\Omega (f, v) \,\mathrm{d}x, \qquad \forall v \in L^2(\Omega; \mathbb{R}^n). \tag{16c}$$

Dual formulation in (u, σ). The Hamiltonian is the Legendre transform of L with respect to p:

$$H(x, u, \sigma) = \sup_p \left[(\sigma, p) - \frac{1}{2}(a(x)p, p) + (f, u) \right].$$

The maximizer satisfies

$$\sigma = a(x)p, \qquad p = a(x)^{-1}\sigma.$$

Hence

$$H(x, u, \sigma) = \frac{1}{2}(a(x)^{-1}\sigma, \sigma) + (f, u).$$

The dual first-order system is therefore

$$\nabla u = H_\sigma = a(x)^{-1}\sigma, \tag{17a}$$

$$\nabla \cdot \sigma = -H_u = -f. \tag{17b}$$

Equivalently,

$$\sigma = a(x)\nabla u, \qquad -\nabla \cdot \sigma = f.$$

Its weak mixed form is: find

$$\sigma \in H(\mathrm{div}; \Omega; \mathbb{R}^{n\times d}), \qquad u \in L^2(\Omega; \mathbb{R}^n),$$

such that

(18a)
$$\int_\Omega \tau \cdot a(x)^{-1}\sigma \,\mathrm{d}x + \int_\Omega u \nabla\cdot \tau \,\mathrm{d}x = \int_{\partial\Omega} u_D(\tau\cdot n)\, \mathrm{d}s \qquad \forall \tau \in H(\mathrm{div};\Omega;\mathbb{R}^{n\times d}),$$

(18b)
$$-\int_\Omega (\nabla\cdot\sigma)\cdot v \,\mathrm{d}x = \int_\Omega (f,v)\,\mathrm{d}x \qquad \forall v \in L^2(\Omega;\mathbb{R}^n).$$

Example 4.2. A standard example is linear elasticity. Here $u:\Omega\to\mathbb{R}^d$ is the displacement, and $\sigma:\Omega\to\mathbb{S}$ is the stress tensor, where

$$\mathbb{S} := \{\tau \in \mathbb{R}^{d\times d} : \tau^{\intercal} = \tau\}$$

is the space of symmetric matrices. The symmetric gradient

$$\varepsilon(u) := \nabla^s u := \frac{\nabla u + \nabla u^\top}{2}$$

is the strain tensor.

Let $\mathbb{C}(x):\mathbb{S}\to\mathbb{S}$ be the elasticity tensor, assumed symmetric and positive definite on $\mathbb{S}$. The elastic energy is

$$I(u) = \int_\Omega \left[\frac{1}{2}\big(\mathbb{C}(x)\varepsilon(u),\varepsilon(u)\big) - (f,u)\right]\mathrm{d}x.$$

Primal formulation. The Euler–Lagrange equation is

$$-\nabla\cdot\sigma = f, \qquad \sigma = \mathbb{C}(x)\varepsilon(u).$$

Equivalently,

$$-\nabla\cdot\big(\mathbb{C}(x)\varepsilon(u)\big) = f.$$

This is called the displacement formulation, since u is the only unknown. Its weak form is: find $u\in H^1(\Omega;\mathbb{R}^d)$ with $u|_{\partial\Omega} = u_D$ such that

$$\int_\Omega \big(\mathbb{C}(x)\varepsilon(u),\varepsilon(v)\big)\,\mathrm{d}x = \int_\Omega (f,v)\,\mathrm{d}x \qquad \forall v\in H^1_0(\Omega;\mathbb{R}^d).$$

Dual mixed formulation. The complementary energy density is

$$A(\sigma) = \frac{1}{2}\big(\mathbb{C}(x)^{-1}\sigma,\sigma\big),$$

and the constitutive relation becomes

$$\mathbb{C}(x)^{-1}\sigma = \varepsilon(u).$$

The first-order system is

$$\mathbb{C}(x)^{-1}\sigma = \varepsilon(u), \qquad -\nabla\cdot\sigma = f.$$

This is called the stress-displacement mixed formulation, or simply the dual mixed formulation, since the stress σ is kept as an independent unknown.

Its weak form is: find

$$\sigma\in H(\mathrm{div},\mathbb{S};\Omega), \qquad u\in L^2(\Omega;\mathbb{R}^d),$$

such that

(19a)
$$\int_\Omega \mathbb{C}(x)^{-1}\sigma : \tau \,\mathrm{d}x - \int_\Omega u \cdot (\nabla \cdot \tau)\,\mathrm{d}x = \int_{\partial\Omega} u_D \cdot (\tau n)\ \mathrm{d}s \qquad \forall \tau \in H(\mathrm{div}, \mathbb{S}; \Omega),$$

(19b)
$$\int_\Omega (\nabla \cdot \sigma) \cdot v \,\mathrm{d}x = -\int_\Omega f \cdot v \,\mathrm{d}x \qquad \forall v \in L^2(\Omega; \mathbb{R}^d),$$

where

$$H(\mathrm{div}, \mathbb{S}; \Omega) := \{\tau \in L^2(\Omega; \mathbb{S}) : \nabla \cdot \tau \in L^2(\Omega; \mathbb{R}^d)\}.$$

In this formulation, σ is the stress tensor and u is the displacement. The primal formulation uses only u, while the dual mixed formulation keeps both σ and u.

4.4. **Duality.** When the Hamiltonian is separable, that is,

$$H(x, u, \sigma) = A(\sigma) + f(u),$$

the dual viewpoint becomes more transparent. We now keep σ and regard the equilibrium equation

$$\nabla\cdot\sigma + f'(u) = 0$$

as a constraint.

Introduce a scalar Lagrange multiplier λ and consider the functional

(20)
$$J(u, \sigma, \lambda) = \int_\Omega \Big(H(x, u, \sigma) + \lambda \left[\nabla\cdot\sigma + f'(u)\right] \Big)\,\mathrm{d}x.$$

Variation with respect to λ. For arbitrary variations of λ, we obtain

$$\nabla\cdot\sigma + f'(u) = 0.$$

Thus λ enforces the equilibrium equation.

Variation with respect to σ. Using (20) and integrating by parts, we obtain

$$\delta_\sigma J = \int_\Omega \big(H_\sigma - \nabla\lambda\big) \cdot \delta\sigma \,\mathrm{d}x + \int_{\partial\Omega} \lambda(\delta\sigma\cdot n)\ \mathrm{d}s.$$

If the boundary term vanishes, for example when $\lambda = 0$ or $\delta\sigma\cdot n = 0$ on $\partial\Omega$, then

$$H_\sigma = \nabla\lambda.$$

Comparing this with the constitutive relation

$$\nabla u = H_\sigma,$$

we get

$$\nabla u = \nabla\lambda, \qquad \Rightarrow \lambda = u + C,$$

where the constant C is determined by the boundary condition or by a normalization. Under homogeneous conditions, one may set $\lambda = u$.

This formal computation shows that the primal variable u can also be viewed as the multiplier enforcing the equilibrium equation in a dual formulation. In this sense, the primal and dual formulations are symmetric: in the primal mixed formulation, σ is the multiplier enforcing $p = \nabla u$, while in the dual formulation, u appears as the multiplier enforcing $\nabla\cdot\sigma = -H_u$.

REFERENCES

[1] K.-C. Chang. *Lecture Notes on Calculus of Variations*, volume 6. World Scientific, 2016.
[2] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, 1997.