Lecture 6: Lagrange multipliers: Equilibrium, minimization, and duality

A system which has the chance to exchange energy with its exterior will tend to a state with minimal energy, and its motion on the way to rest also is usually extremal in some particular sense. The Lagrange equations of motion give stationarity – if not necessarily minimality - to the "action integral"

$$I(t) = \int_0^t \left(T(\mathbf{u}, \dot{\mathbf{u}}) - W(\mathbf{u})\right) dt$$

and equilibrium has $T = 0$ so equilibrium $\mathbf{u}^*$ must make W stationary,

$$\mathbf{u}^* = \arg\min_{\mathbf{u}} W(\mathbf{u})$$

as can be demonstrated by variational calculus. The minimization is over all functions $\mathbf{u}(t)$. How can we formulate similar equations when the minimization is constrained to a smaller set of $\mathbf{u}$?

We consider first m linearly independent linear equality constraints,

$\mathbf{u}^T\mathbf{b}_i = c_i$, $i = 1,\ldots,m$, and will prove that:

If $\mathbf{u}^*$ is a stationary point, then the gradient vector

$$\mathbf{g}(\mathbf{u}^*) = \nabla W(\mathbf{u}^*) = \left(\frac{\partial W}{\partial u_1}(\mathbf{u}^*), \frac{\partial W}{\partial u_2}(\mathbf{u}^*), \ldots, \frac{\partial W}{\partial u_n}(\mathbf{u}^*)\right)^T$$

is a linear combination of the $\mathbf{b}_i$, $\mathbf{g} = \sum_{k=1}^{m} c_k \mathbf{b}_k$

Sketch:

First, we prove by e.g. QR-factorization (Exercise: do it!) that:

If rank($\mathbf{B}$) = m and for every $\mathbf{x}$ such that $\mathbf{x}^T\mathbf{b}_i = 0$, $i = 1,\ldots,m$ also $\mathbf{x}^T\mathbf{g} = 0$, then

$$\mathbf{g} = \sum_{k=1}^{m} c_k \mathbf{b}_k \ .$$

Suppose $\mathbf{u}^*$ is a constrained minimum. Then $W(\mathbf{u}^*) < W(\mathbf{u})$ for all $\mathbf{u}$ in a neighborhood of $\mathbf{u}^*$ which satisfy $(\mathbf{u}-\mathbf{u}^*)^T\mathbf{b}_i = 0$, and it follows that $(\mathbf{u}-\mathbf{u}^*)^T\mathbf{g} = 0$. But then, by the above, there are numbers λ_k such that $\mathbf{g} = \sum_{k=1}^{m} \lambda_k \mathbf{b}_k = \mathbf{B}\boldsymbol{\lambda}$.

This condition is expressible as $\nabla L = 0, L = W - \boldsymbol{\lambda}^T(\mathbf{B}^T\mathbf{u} - \mathbf{c})$ and the λ's are called *Lagrange* multipliers.

For non-linear constraints

$h_i(\mathbf{u}) = 0$, $i = 1,\ldots,m$

the story is the analogue: A necessary condition for a constrained extremum at $(\mathbf{u}^*, \boldsymbol{\lambda}^*)$ is

$$\nabla_{\mathbf{u}}\left[W(\mathbf{u}^*) + \sum \lambda_k^* h_k(\mathbf{u}^*)\right] = 0, \nabla_{\boldsymbol{\lambda}}\left[W(\mathbf{u}^*) + \sum \lambda_k^* h_k(\mathbf{u}^*)\right] = \mathbf{h}(\mathbf{u}^*) = 0$$

Example: The pendulum in Cartesian coordinates

(will be back in Dr. Hanke's lectures on Differential-Algebraic systems)

Let us describe the position of the mass point in Cartesian coordinates (x,z), origin at the joint:

$W = mgz$; $x^2+z^2=l^2$;

$L = mgz + \lambda(x^2+z^2-l^2)$

Let (x,z,λ) be the optimum. Then

$$\frac{\partial L}{\partial x} = 2x\lambda = 0;$$

$$\frac{\partial L}{\partial z} = mg + 2z\lambda = 0;$$

$$\frac{\partial L}{\partial \lambda} = x^2 + z^2 - l^2 = 0$$

and the only solution is $x = 0$, $z = +\!-l$, $\lambda = +\!-mg/2l$, $W_{\min} = -mgl$.

Observe: The value of λ is the "marginal cost" of the constraint, $\frac{\partial W_{\min}}{\partial l^2}$;

Exercise: Formulate the constraint as $\sqrt{x^2+z^2} = l$ and solve the problem again. What is the value of λ in this problem? Then write the equations for force balance, and eliminate the unknown force in the rod – its value at equilibrium is …?

For mechanical systems, the Lagrange multipliers associated with kinematic constraints are exactly the forces necessary to make the solution satisfy the constraints.

This formulation is completely symmetric in the *primal* variables (x,z) and the Lagrange multiplier(s) λ, also called *dual* variables. Solve from the necessary conditions for the primals as functions of the λ:

$x^* = 0$ (it just happens not to depend on λ)

$z^* = -mg/2\lambda$

so we define the dual function ϕ as

$$\phi(\lambda) = \min_{x,z} L(x,z,\lambda) = L(\mathbf{u}^*(\lambda),\lambda) = \ldots = -\frac{m^2g^2}{4\lambda} - l^2\lambda\,.$$

where the effects of the constraints are built in.

Consider positive λ:

Negative z, energy minimum, $\phi(\lambda)$ has a maximum at $\lambda = mg/(2l)$ but no finite minimum.

for negative λ the reverse: positive z, energy maximum, $\phi(\lambda)$ has a minimum.

Duality allows us to exchange an equality constrained minimization for the primal problem for an unconstrained maximization of the dual function. That this is generally so is expressed by the

Local Duality Theorem (Luenberger, p314)

Suppose the problem

$$\min_{\mathbf{u}} W(\mathbf{u}) \text{ subject to } \mathbf{h}(\mathbf{u}) = 0$$

has a local solution at $\mathbf{u}^*$ with value r^* and Lagrange multipliers $\boldsymbol{\lambda}^*$.

Suppose also that (technical conditions)

$\mathbf{u}^*$ is a regular point of the constraints,

the Hessian of the Lagrangian $L(\mathbf{u}^*(\boldsymbol{\lambda}),\boldsymbol{\lambda})$ is positive definite.

Then the dual *unconstrained* problem

$$\max_{\boldsymbol{\lambda}} \phi(\lambda) = \max_{\boldsymbol{\lambda}} L(\mathbf{u}^*(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda}} \min_{\mathbf{u}} L(u, \lambda)$$

has a local solution at $\boldsymbol{\lambda}^*$ with value r^* and $\mathbf{u}^*$ as the point corresponding to $\boldsymbol{\lambda}^*$ in the definition of ϕ.

The technical conditions are important. They explain why in our problem it does not make sense to try $\min_{\mathbf{u}} \max_{\boldsymbol{\lambda}}$. If we reverse roles of $\mathbf{u}$ and $\boldsymbol{\lambda}$ there are more constraints than variables so there is no regular point; remember, the gradients of the constraint functions must be linearly independent at a regular point. Note that the case of more *equality* constraints than variables is decidedly uncommon and requires dependency between the constraints for there to be sensible solutions.

Another idea is to *penalize* deviation from the algebraic relation. For the mechanical system, make the rigid rod elastic but very stiff,

$$W = mgz + \frac{1}{2}K(r-l)^2, T = \frac{1}{2}m(\dot{x}^2 + \dot{z}^2), r = \sqrt{x^2 + z^2}$$

$$m\ddot{x} = -K(r-l)\frac{x}{r}$$

$$m\ddot{z} = -mg - K(r-l)\frac{z}{r}$$

(we used $\frac{\partial r}{\partial x} = \frac{x}{r}, etc.$ to derive the equations of motion.)

Equilibrium says $x = 0$ (or $r - l = 0$) and $z = +\!-l - mg/K$: errors $O(1/K)$. Note that we used a mechanically correct elastic energy as penalization, but could have used some other function.

The consequences for the equilibrium are slight, but not so for the dynamics: The eigenfrequencies become

$\sqrt{\frac{g}{l}}$, for the "tangential" motion, and $\sqrt{\frac{g}{l} + \frac{K}{m}}$, for the "radial"

Large K gives good equilibrium accuracy, but also a "*stiff*" system with a time-scale much smaller than the oscillations of the rigid system.

Exercise: Check the expressions for the eigenfrequencies, and in particular how x and z-motions decouple, by computing the 4 x 4 Jacobian of the ODE system and its eigenvalues and –vectors.