LINEAR PROGRAMMING AND NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS

DIOGO A. GOMES

Abstract. In this paper we consider infinite dimensional linear programming problems and study its connections with nonlinear partial differential equations.

Contents

1. Model Problems

This survey paper is dedicated to the study of infinite dimensional linear programming problems and its connections with nonlinear partial differential equations. In this first section we present three model problems: Mather's deterministic and stochastic problems, and the Monge-Kantorowich mass transport problem.

1.1. **Mather's problem.**

The trajectories $\mathbf{x}(\cdot)$ of a system in classical mechanic are determined through a variational principle which asserts that they are critical points of the action:

$$\int_0^T L(\mathbf{x},\dot{\mathbf{x}})dt,$$

in which L is the Lagrangian, whcih is the difference between kinetic and potential energy. A case of interest is the one in which $L(x, v)$:

$\mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$ ($\mathbb{T}^n$ is the n dimensional torus), smooth, strictly convex in v, and $L \geq 0$.

An important issue is to understand the limit as $T \to \infty$. To study this problem, is useful to consider certain measures which encode the asymptotic behaviour of minimizing curves. In fact, if $\mathbf{x} : \mathbb{R} \to \mathbb{T}^n$ is globally Lipschitz, one can associate to $\mathbf{x}$ a probability measure μ by

$$\int_{\mathbb{T}^n \times \mathbb{R}^n} \phi(x, v) d\mu = \lim_{T \to \infty} \frac{1}{T} \int_0^T \phi(\mathbf{x}, \dot{\mathbf{x}}) dt,$$

in which the limit is taken through an appropriate subsequence.

Since for all $\varphi(x) : \mathbb{T}^n \to \mathbb{R}$

$$\int_0^T \dot{\mathbf{x}} D_x \varphi(\mathbf{x}) dt = O(1)$$

this implies

$$\int_{\mathbb{T}^n \times \mathbb{R}^n} v D_x \varphi(x) d\mu = 0.$$

It is therefore natural to consider the **Mather's problem**

$$\min \int_{\mathbb{T}^n \times \mathbb{R}^n} L d\mu,$$

in which the minimum is taken over all positive probability measures μ in $\mathbb{T}^n \times \mathbb{R}^n$ which satisfy

$$\int_{\mathbb{T}^n \times \mathbb{R}^n} v D_x \varphi(x) d\mu = 0,$$

for all $\varphi(x) : \mathbb{T}^n \to \mathbb{R}$. Both the function to be minimized and the constraints are linear in μ.

1.2. **Stochastic Mather's problem.** Instead of deterministic trajectories, we may consider stochastic processes

$$dx = \nu dt + \sigma dW,$$

in which ν is a bounded progressively measurable process, $\sigma > 0$ and W is a $n-$dimensional Brownian motion. Then we want to minimize the expected value of the action

$$E \int_0^T L(x, \nu) dt.$$

As before, we associate to each of these stochastic processes a probability measure, μ, by taking weak limits:

$$\int_{\mathbb{T}^n\times\mathbb{R}^n} \phi(x,v)d\mu = \lim_{T\to\infty}\frac{1}{T}\int_0^T \phi(x,\nu)dt,$$

in which the limit is taken through an appropriate subsequence.

The Dynkin's formula applied to $\varphi(x(t))$:

$$E(\varphi(x(T))-\varphi(x)) = E\int_0^T \nu D_x\varphi(x(t)) + \frac{\sigma^2}{2}\Delta\varphi(x(t))dt$$

yields

$$\int_{\mathbb{T}^n\times\mathbb{R}^n} vD_x\varphi(x) + \frac{\sigma^2}{2}\Delta\varphi(x)d\mu = 0,$$

for all $\varphi(x):\mathbb{T}^n\to\mathbb{R}$.

The **stochastic Mather's problem**[Gom02] consists in minimizing

$$\min\int_{\mathbb{T}^n\times\mathbb{R}^n} Ld\mu,$$

among all positive probability measures μ in $\mathbb{T}^n\times\mathbb{R}^n$ which satisfy

$$\int_{\mathbb{T}^n\times\mathbb{R}^n}\left[vD_x\varphi(x) + \frac{\sigma^2}{2}\Delta\varphi(x)\right]d\mu = 0,$$

for all $\varphi(x):\mathbb{T}^n\to\mathbb{R}$.

1.3. **Monge-Kantorowich problem.** The Monge-Kantorowich optimal mass transport problem (consult the survey paper [Eva99] for more references) consists in moving a certain amount of mass with the least possible cost. More precisely, one is given two measures μ^+ and μ^- in $\mathbb{R}^n$ which satisfy the mass balance condition

$$\int_{\mathbb{R}^n} d\mu^+ = \int_{\mathbb{R}^n} d\mu^-.$$

The problem consists in finding a map $s:\mathbb{R}^n\to\mathbb{R}^n$ which transports μ^+ into μ^-, that is,

$$\int_{\mathbb{R}^n}\varphi(s(x))d\mu^+ = \int_{\mathbb{R}^n}\varphi(y)d\mu^-,$$

for all $\varphi\in C_c^\infty(\mathbb{R}^n)$, in other words $s^\# \mu^+ = \mu^-$, and minimizes the cost

$$\int_{\mathbb{R}^n}|x-s(x)|^2d\mu^+(x).$$

Given a map s for which $s^{\#}\mu^+ = \mu^-$, one can define a measure in $\mathbb{R}^{2n}$ by
$$\int_{\mathbb{R}^{2n}} \phi(x,y)d\mu = \int_{\mathbb{R}^n} \phi(x,s(x))d\mu^+.$$
Furthermore, the marginals $\mu|_x = \mu^+$ and $\mu|_y = \mu^-$.

It is therefore natural to consider the relaxed **Monge Kantorowich problem**
$$\min \int_{\mathbb{R}^{2n}} |x-y|^2 d\mu,$$
among all measures which satisfy $\mu|_x = \mu^+$ and $\mu|_y = \mu^-$, that is,
$$\int_{\mathbb{R}^{2n}} \varphi(x)d\mu = \int_{\mathbb{R}^n} \varphi(x)d\mu^+,$$
and
$$\int_{\mathbb{R}^{2n}} \varphi(y)d\mu = \int_{\mathbb{R}^n} \varphi(y)d\mu^-.$$
The strategy consists in establishing a solution μ to the relaxed problem. Then to prove that the measure μ is supported on a graph $(x, s(x))$, and therefore the relaxed problem turns out to be equivalent to the original one.

2. Mather's problem

In Mather's problem (both deterministic or stochastic) the constraint
$$\int_{\mathbb{T}^n\times\mathbb{R}^n} vD_x\varphi(x) + \frac{\sigma^2}{2}\Delta\varphi(x)d\mu = 0,$$
$(\sigma \geq 0)$ is linear in v. Therefore, if we replace a minimizing measure $\mu(x,v)$ by a new measure $\tilde{\mu}$ given by
$$\int_{\mathbb{T}^n\times\mathbb{R}^n} \phi(x,v)d\tilde{\mu}(x,y) = \int_{\mathbb{T}^n} \phi(x,\bar{v}(x))\theta(x)dx,$$
in which
$$\bar{v}(x) = \int_{\mathbb{R}^n} v\mu(x,v)dv$$
and
$$\theta(x) = \int_{\mathbb{R}^n} \mu(x,v)dv,$$
we also obtain
$$\int_{\mathbb{T}^n\times\mathbb{R}^n} vD_x\varphi(x) + \frac{\sigma^2}{2}\Delta\varphi(x)d\tilde{\mu} = 0.$$

Furthermore, the convexity of L in v yields

$$\int L d\tilde{\mu} \leq \int L d\mu.$$

If L is strictly convex, the inequality is strict unless $v = \bar{v}(x)$, μ-almost everywhere.

Therefore we have:

Theorem 1. *The minimizing measure μ in the Mather's problem (stochastic or deterministic) is supported on a graph*

$$(x, v) = (x, \bar{v}(x)).$$

Additionally, the projection on the x-coordinates of this measure satisfies

$$-\nabla \cdot (\bar{v}(x)\theta(x)) + \frac{\sigma^2}{2}\Delta\theta = 0,$$

in the weak sense.

Suppose now that $L = \frac{|v|^2}{2} - U(x)$. Then, by using (formally) Lagrange multipliers, Mather's problem is equivalent to the unconstrained minimization

$$\min_{\theta, v(x)} \int \left(\frac{|v|^2}{2} - U(x) + vD_x\varphi + \frac{\sigma^2}{2}\Delta\varphi + \overline{H} \right) \theta dx,$$

in which $(\varphi, \overline{H})$ are the Lagrange multipliers for the constraints.

The Euler-Lagrange equation, obtained by making variations $v \to v + \epsilon w$ and $\theta \to \theta + \epsilon\eta$, yields

$$v = -D_x\varphi(x),$$

and

$$\frac{|v|^2}{2} - U(x) + vD_x\varphi + \frac{\sigma^2}{2}\Delta\varphi + \overline{H} = 0.$$

This implies

$$-\frac{\sigma^2}{2}\Delta\varphi + H(D_x\varphi, x) = \overline{H}, \tag{1}$$

with

$$H(p, x) = \frac{|p|^2}{2} + U(x).$$

As an applications of these results we prove an estimate for the second derivatives of the solution of the Hamilton-Jacobi equation. To that effect differentiate (1) twice with respect to the direction x

$$-\frac{\sigma^2}{2}\Delta(\varphi_{xx}) + D_x\varphi D_x(\varphi_{xx}) + |D_x\varphi_x|^2 + U_{xx} = 0.$$

since $v = -D_x\varphi$ we have

$$\int -\frac{\sigma^2}{2}\Delta(\varphi)_{xx} + D_x\varphi D_x(\varphi_{xx})d\mu = 0,$$

and so

$$\int |D^2\varphi|^2 d\mu \leq C.$$

Some of the previous results can be understood rigorously in terms of duality. In fact, since each of the problems considered is an infinite dimensional linear programming problem it admits a dual problem. In the case of the Aubry-Mather problem, the dual turns out to be equivalent to

$$\inf_\phi \sup_x -\frac{\sigma^2}{2}\Delta\phi + H(D_x\phi, x).$$

The duality theory asserts that the value of this infimum is

$$-\int_{\mathbb{T}^n\times\mathbb{R}^n} L d\mu,$$

for any minimizing measure μ, which also agrees with the value $\overline{H}$, which is the unique numeber for which the equation

$$-\frac{\sigma^2}{2}\Delta u + H(D_x u, x) = \overline{H}$$

has a periodic solution. This last statement can be checked directly. In fact, if φ is the solution to (1) then

$$\inf_\phi \sup_x -\frac{\sigma^2}{2}\Delta\phi + H(D_x\phi, x) \leq \sup_x -\frac{\sigma^2}{2}\Delta\varphi + H(D_x\varphi, x) = \overline{H}.$$

Additionally, for any ϕ periodic, $\varphi - \phi$ has a minimum at some point x_0. At this point $D_x\varphi = D_x\phi$, and $\Delta\varphi \geq \Delta\phi$, so

$$\begin{aligned}\sup_x -\frac{\sigma^2}{2}\Delta\phi + H(D_x\phi, x) &\geq -\frac{\sigma^2}{2}\Delta\phi(x_0) + H(D_x\phi, x_0)\\ &\geq -\frac{\sigma^2}{2}\Delta\varphi(x_0) + H(D_x\varphi, x_0) = \overline{H}.\end{aligned}$$

3. Monge-Kantorowich problem

To obtain the Euler-Lagrange for the Monge-Kantorowich problem we make the additional assumption that μ^+ and μ^- have densities ρ^+ and ρ^-. Suppose that $s(x)$ is an optimal transport map and μ is an optimal measure in $\mathbb{R}^{2n}$ associated to $s(x)$. Let w be a divergence free vector field in $\mathbb{R}^n$, and φ_τ be the flow associated with the ordinary differential equation:

$$\frac{d}{d\tau}z = \frac{w(z)}{\rho^+(z)}.$$

Since w is divergence free

$$\nabla \cdot \left(\rho^+ \frac{d}{d\tau}\varphi_\tau\right) = 0,$$

which imples that $\varphi_\tau^\# \mu^+ = \mu^+$. Let the measure μ_τ be given by

$$\int \phi(x,y) d\mu_\tau = \int \phi(\varphi_\tau(x), y) d\mu.$$

Since $\mu_0 = \mu$, and μ is optimal, we have

$$\frac{d}{d\tau}\int |x-y|^2 d\mu_\tau \bigg|_{\tau=0} = 0,$$

that is,

$$2\int (\varphi_\tau(x) - y) \cdot \frac{d}{d\tau}\varphi_\tau(x) d\mu \bigg|_{\tau=0} = 0.$$

This implies

$$\int (x - s(x)) w(x) = 0.$$

Since this holds for all divergence free vector fields, the function $x - s(x)$ is a gradient. Thus

$$s(x) = D_x \Psi(x), \tag{2}$$

for some $\Psi(x)$. The condition $s^\# \mu^+ = \mu^-$, which, by the change of coordinates formula, is equivalent to

$$\rho^+(x) = \rho^-(s(x)) \det Ds(x).$$

Using (2), the last equation can be rewritten as the **Monge-Ampére equation**

$$\rho^+(x) = \rho^-(D\Psi(x)) \det D^2\Psi(x).$$

4. Otto's calculus and nonlinear diffusions

In this section we follow [Vil] and study the application of Monge-Kantorowich problems to the theory of nonlinear diffusions. These ideas have the origin in the paper [JKO98].

The value of Monge-Kantorowich problem measures the distance between the two measures μ^+ and μ^-. In fact, one can define a distance between two probability measures, the **Wasserstein distance**, by

$$d_W^2(\mu^+,\mu^-) = \inf \int |x-y|^2 d\mu,$$

in which the infimum is taken among all measures μ which satisfy $\mu|_x = \mu^+$ and $\mu|_y = \mu^-$.

Suppose μ^+ and μ^- have densities ρ_0 and ρ_1, respectively. In the Brenier-Benamou formulation [BB00] one writes the Wasserstein distance as

$$d_W^2(\mu^+,\mu^-) = \inf \int_0^1 \int_{\mathbb{R}^n} \rho |v|^2 dx dt,$$

in which the infimum is taken over vector fields v and time dependent densities ρ which satisfy

$$\frac{d}{dt}\rho_t + \nabla \cdot (\rho_t v) = 0,$$

$\rho_{t=0} = \rho_0$, and $\rho_{t=1} = \rho_1$. The equality follows from two observations. First, given an optimal vector field v one can construct a map by solving the equation

$$\dot{x} = v(x),$$

and considering its flow φ_t. One has $\rho_t = \varphi_t^{\#}\rho_0$ and so $\varphi_1^{\#}\mu^+ = \mu^-$. Furthermore

$$\int_0^1 \int \rho_t |v|^2 = \int_0^1 \int \rho_0 |v(\varphi_t(x))|^2 \geq \int \rho_0 \left| \int_0^1 v(\varphi_t(x)) \right|^2 \geq$$
$$\geq d_W^2(\mu^+,\mu^-),$$

by Jensen's inequality.

To prove the other inequality one has to interpolate between ρ_0 and ρ_1. To do so, consider the optimal transport map $s(x)$. and let

$$s_t(x) = (1-t)x + t s(x),$$

$$v_t(y) = s(s_t^{-1}(y)) - s_t^{-1}(y),$$

and $\rho_t = s_t^{\#}\rho_0$. One can check that

$$\frac{d\rho_t}{dt} + \nabla\cdot(v_t\rho_t) = 0,$$

and

$$\int \rho_t |v_t|^2 = \int \rho_0 |x - s(x)|^2 = d_W^2(\mu^+, \mu^-).$$

This formulation suggests than the Wasserstein distance may be written as the infimum of an integral of $\left\|\frac{d\rho_t}{dt}\right\|_\rho^2$ with an appropriate norm:

$$d_W^2(\rho_0, \rho_1) = \inf \int_0^1 \left\|\frac{d\rho_t}{dt}\right\|_\rho^2 dt,$$

in which the infimum is taken over all time-dependente densities ρ_t with $\rho_{t=0} = \rho_0$, and $\rho_{t=1} = \rho_1$. The norm

$$\left\|\frac{d\rho_t}{dt}\right\|_\rho^2 = \inf_v \int \rho |v|^2 dt. \tag{3}$$

in which v is such that

$$\frac{d\rho_t}{dt} + \nabla\cdot(v_t\rho_t) = 0. \tag{4}$$

The Euler-Lagrange equation for (3), obtained by making the variation $v \to v + \epsilon\frac{w}{\rho}$, with w a divergence free vector field, is:

$$\int vw = 0,$$

which implies that v is a gradient:

$$v = \nabla\varphi.$$

Thus (4) reads then

$$\nabla\cdot(\rho\nabla\varphi) = -\frac{d\rho_t}{dt},$$

which is an elliptic equation for φ.

The norm $\|\cdot\|_\rho$ comes from Riemannian structure on the tangent space of probability measures

$$\langle\xi, \eta\rangle = \int \rho\nabla\phi\cdot\nabla\psi,$$

with

$$\nabla\cdot(\rho\nabla\phi) = -\xi \qquad \nabla\cdot(\rho\nabla\psi) = -\eta.$$

Given a Riemannian structure, the gradient of a functional $F(\rho)$ is defined as:

$$\langle \operatorname{grad} F(\rho), \eta\rangle = DF\eta = \int \frac{\delta F}{\delta \rho}\eta,$$

in which $\frac{\delta F}{\delta \rho}$ is the gradient of the functional with respect to the L^2 structure. Then

$$\langle \operatorname{grad} F(\rho), \eta\rangle = \int \rho \nabla\phi \cdot \nabla \psi,$$

with

$$\nabla \cdot (\rho\nabla\phi) = -\operatorname{grad} F(\rho).$$

Integrating by parts

$$\langle \operatorname{grad} F(\rho), \eta\rangle = \int \operatorname{grad} F(\rho)\psi = \int \frac{\delta F}{\delta \rho}\eta.$$

Similarly,

$$\nabla \cdot (\rho\nabla\psi) = \eta,$$

and so

$$\langle \operatorname{grad} F(\rho), \eta\rangle = \int \operatorname{grad} F(\rho)\psi = -\int \frac{\delta F}{\delta \rho}\nabla \cdot (\rho\nabla\psi),$$

since ψ is arbitrary this yields

$$\operatorname{grad} F(\rho) = -\nabla \cdot (\rho\nabla\frac{\delta F}{\delta \rho}).$$

For instance, if

$$F(\rho) = \int \rho^\gamma$$

with $\gamma > 1$ one has

$$\operatorname{grad} F(\rho) = -C\Delta\rho^\gamma,$$

and so, the Porous media equation

$$\frac{d}{dt}\rho = C\Delta\rho^\gamma,$$

is a gradient flow with respect to the metric induced by the Wasserstein metric.

References

[BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[Eva99] Lawrence C. Evans. Partial differential equations and Monge-Kantorovich mass transfer. In *Current developments in mathematics, 1997 (Cambridge, MA)*, pages 65–126. Int. Press, Boston, MA, 1999.

[Gom02] Diogo Aguiar Gomes. A stochastic analogue of Aubry-Mather theory. *Nonlinearity*, 15(3):581–603, 2002.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17 (electronic), 1998.

[Vil] C. Villani. Optimal transportation, dissipative pdes and functional inequalities.