Week 11. Convex Hamiltonians and the Legendre Transform

Tran: §2.4 (Theorem 2.16, pp. 61–62); §2.6 (Theorems 2.32, 2.34 and Corollary 2.36, pp. 72–75)

1. Overview

In the previous weeks, we introduced deterministic optimal control, the Dynamic Programming Principle, Hamilton–Jacobi–Bellman equations, viscosity solutions, and comparison principles. In this lecture, we impose an additional structural assumption: convexity of the Hamiltonian in the momentum variable.

Convexity allows us to pass from the Hamiltonian viewpoint to the Lagrangian viewpoint through the Legendre transform. It also gives additional stability and characterization results for viscosity solutions. These ideas prepare the representation formulas of Week 12.

2. Learning Objectives

By the end of this lecture, students should be able to:

1. state the convexity and superlinearity assumptions used in Tran §2.4;
2. define the Legendre transform and prove the dual inequality;
3. compute the Legendre transform of quadratic, power-type, and exponential Hamiltonians;
4. explain the equality condition in Fenchel duality;
5. state the Barron–Jensen theorem and explain the role of convexity;
6. use generalized differentials to characterize Lipschitz viscosity solutions;
7. distinguish general inf-stability of supersolutions from convex inf-stability of solutions.

3. Standard Assumptions: Convexity and Superlinearity

Throughout the first part of this lecture, assume that $H(x,p)$ satisfies Tran (2.11):

- $H \in C^1(\mathbb{R}^n \times \mathbb{R}^n)$;
- $H \in \mathrm{BUC}(\mathbb{R}^n \times B(0,R))$ for every $R > 0$;
- for each fixed $x \in \mathbb{R}^n$, the map $p \mapsto H(x,p)$ is convex;
- H is superlinear in p, uniformly in x:

$$\lim_{|p|\to\infty} \inf_{x\in\mathbb{R}^n} \frac{H(x,p)}{|p|} = +\infty.$$

Superlinearity is stronger than coercivity. It says that $H(x,p)$ eventually dominates every linear function of p, uniformly in x.

Convexity in p is natural in deterministic optimal control. Under the course convention,

$$H(x,p) = \sup_{v\in V}\{-b(x,v)\cdot p - f(x,v)\}.$$

For each fixed x, this is the supremum of affine functions of p, and is therefore convex. Superlinearity is a separate assumption: the bounded drifts used earlier give a Hamiltonian that is globally Lipschitz in p, so it cannot grow superlinearly.

4. Legendre Transform

In the Cauchy equation $u_t + H(x,Du) = 0$, the momentum p stands for the spatial slope Du of the unknown value function. If we propose a path, we know its position and velocity, but we do not yet know Du. To compare paths, we want a running cost that we can compute from the motion itself.

For a fixed position x and a prescribed velocity v, define the **Lagrangian** by taking the Legendre transform in p:

$$L(x,v) = \sup_{p\in\mathbb{R}^n}\{p\cdot v - H(x,p)\}.$$

Why take this supremum? To see what it measures, suppose for a moment that u and the path γ are smooth. At time s, write $x=\gamma(s)$, $v=\dot\gamma(s)$, and $p=Du(\gamma(s),s)$. By the chain rule and the PDE,

$$\frac{d}{ds}u(\gamma(s),s) = u_t(\gamma(s),s) + p\cdot v = p\cdot v - H(x,p) \le L(x,v).$$

Thus $L(x,v)$ is the smallest upper bound on this rate that works for **every** slope p. It gives a **cost per unit time**: moving from x at velocity v for a short time h costs approximately $hL(x,v)$. For instance, when $H(p)=|p|^2/2$, Example 11.1 gives $L(v)=|v|^2/2$, so doubling the speed quadruples the cost per unit time.

Writing $u_0 = u(\cdot,0)$, we integrate along a path ending at x at time t to obtain

$$u(x,t) \le u_0(\gamma(0)) + \int_0^t L(\gamma(s),\dot\gamma(s))\,ds.$$

The right-hand side is the cost of **one path**: its initial cost plus the running cost accumulated along it. Week 12 shows that, under the representation theorem's assumptions, the infimum of these costs over all paths ending at x equals $u(x,t)$, even when u is not differentiable.

Here v is the variable paired with p, not necessarily the original control. With our earlier convention $H=\sup\{-b\cdot p - f\}$, it is $-b$ that is paired with p. This is why the discounted control formula for a path starting at x uses $L(\gamma,-\dot\gamma)$.

5. Basic Properties of the Legendre Transform

Under Tran (2.11), Theorem 2.16 gives:

1. $L(x,v)$ is finite for every (x,v);
2. for each fixed x, $v\mapsto L(x,v)$ is convex;
3. L is superlinear in v, uniformly in x;
4. the biconjugate recovers the Hamiltonian:

$$L^*(x,p) = H^{**}(x,p) = H(x,p).$$

Thus passing from H to L loses no information.

To see why L is finite, fix x and v. Superlinearity gives

$$p\cdot v - H(x,p) \to -\infty \qquad \text{as } |p|\to\infty.$$

Continuity in p then gives a maximum on a sufficiently large closed ball. For fixed x, $L(x,\cdot)$ is convex because it is a supremum of affine functions of v. To see superlinearity, fix $A>0$. Taking $p=Av/|v|$ for $v\neq 0$ gives

$$L(x,v) \ge A|v| - M_A, \qquad M_A := \sup_{x\in\mathbb{R}^n,\,|p|\le A}|H(x,p)| < \infty.$$

Divide by $|v|$, let $|v|\to\infty$, and then take A arbitrarily large. The bound is independent of x. We prove the biconjugacy claim in §6.

Tran notes that the C^1 assumption is used to simplify the proof. The biconjugacy conclusion remains valid for proper lower semicontinuous convex Hamiltonians without differentiability.

6. Fenchel Duality Inequality

For every $x, p, v \in \mathbb{R}^n$,

$$H(x,p) + L(x,v) \geq p \cdot v.$$

Indeed, by definition,

$$L(x,v) = \sup_{q \in \mathbb{R}^n} \{q \cdot v - H(x,q)\}.$$

Taking $q = p$ gives

$$L(x,v) \geq p \cdot v - H(x,p),$$

which is equivalent to the dual inequality.

Equality means that p attains the supremum defining $L(x,v)$, or equivalently that

$$H(x,q) \geq H(x,p) + v \cdot (q-p) \qquad \text{for every } q.$$

In other words, v belongs to the convex subdifferential $\partial_p H(x,p)$, which for a convex function agrees with $D_p^- H(x,p)$ from Week 7. When H is differentiable in p, it reduces to $v = D_p H(x,p)$.

This also proves biconjugacy under (2.11): the dual inequality gives $L^*(x,p) = \sup_v \{p \cdot v - L(x,v)\} \leq H(x,p)$. Choosing $v = D_p H(x,p)$ gives equality, hence $L^*(x,p) = H(x,p)$.

For the quadratic pair $H(p) = p^2/2$ and $L(v) = v^2/2$, fixing v produces the affine lower support $p \mapsto pv - L(v)$ of H. Equality occurs exactly at the tangency point $p = v$, which is the smooth equality condition $v = D_p H(p)$.

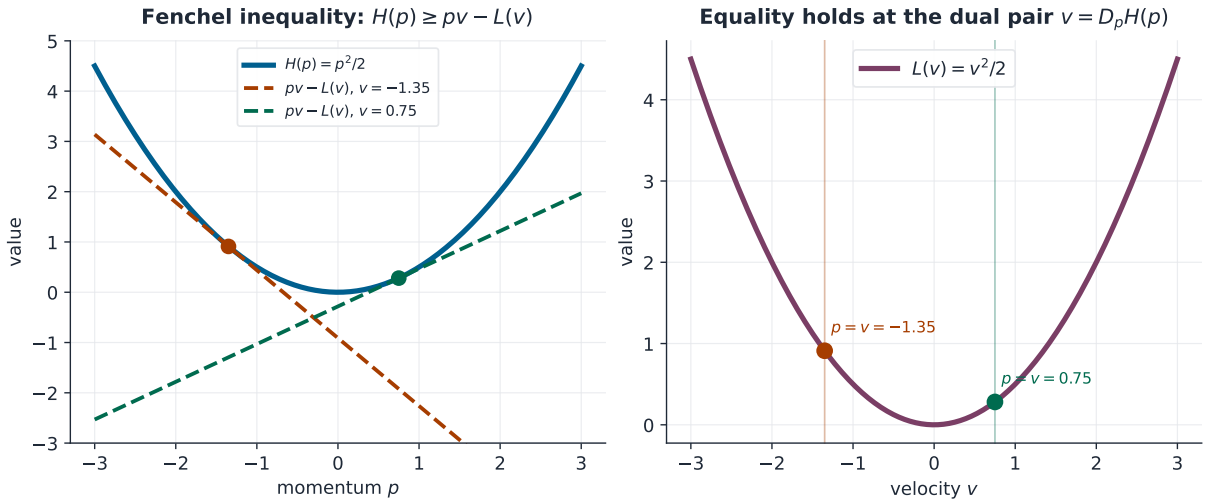


Figure 11.1. Fenchel duality for the quadratic Hamiltonian. Each velocity v determines an affine function $pv - L(v)$ below $H(p)$; the contact point identifies the dual momentum-velocity pair.

7. Standard Examples

To compare the three Lagrangians, consider the same one-dimensional trip in each example: start at 0 and arrive at 1 at time 1. One path moves at speed 1 throughout. The other waits for half the time, then moves at speed 2. Both have initial cost $u_0(0)$, so we need only compare their accumulated running costs. For the two examples with a potential, take $V = 0$ for this comparison.

Example 11.1: Quadratic Hamiltonian

If

$$H(x,p) = \frac{1}{2}|p|^2 + V(x),$$

then completing the square gives

$$p \cdot v - \tfrac{1}{2}|p|^2 - V(x) = \tfrac{1}{2}|v|^2 - V(x) - \tfrac{1}{2}|p - v|^2.$$

The maximum occurs at $p = v$, so $L(x, v) = \frac{1}{2}|v|^2 - V(x)$.

For the steady path, the cost is $L(1) = 1/2$ over one unit of time. The delayed path costs $\frac{1}{2}L(0)+\frac{1}{2}L(2) = 1$. Moving twice as fast for half as long doubles the total running cost.

More generally, let $V = 0$. To travel from y to x in time $t > 0$, any absolutely continuous path between them satisfies

$$\int_0^t \tfrac{1}{2}|\dot\gamma(s)|^2\, ds \;\geq\; \frac{1}{2t}\left|\int_0^t \dot\gamma(s)\, ds\right|^2 = \frac{|x-y|^2}{2t}.$$

The straight path with constant velocity $(x-y)/t$ attains equality in Cauchy–Schwarz. Thus the least cost of moving from y to x is $|x-y|^2/(2t)$. If the initial cost is $u_0(y)$, the remaining choice is the starting point: minimize $u_0(y) + |x-y|^2/(2t)$ over y.

Example 11.2: Power-Type Hamiltonian

If

$$H(x, p) = \frac{|p|^m}{m} + V(x), \qquad m > 1,$$

and m' is the conjugate exponent,

$$\frac{1}{m} + \frac{1}{m'} = 1,$$

then, for $v \neq 0$, Cauchy–Schwarz shows that it suffices to take p in the direction of v. Writing $r = |p|$, maximize $r|v| - r^m/m$: the maximizing radius satisfies $r^{m-1} = |v|$. Substituting gives

$$L(x, v) = \frac{|v|^{m'}}{m'} - V(x), \qquad m' = \frac{m}{m-1}.$$

The same formula holds at $v = 0$, where $p = 0$ is optimal.

For the same two paths with $V = 0$, the cost of moving steadily is $1/m'$. The delayed path costs

$$\tfrac{1}{2}L(0) + \tfrac{1}{2}L(2) = \frac{2^{m'-1}}{m'} > \frac{1}{m'}.$$

The factor $2^{m'-1}$ shows how the penalty for delaying depends on the exponent. When V varies with position, each path also contributes $-\int_0^1 V(\gamma(s))\, ds$. One can still compare their costs from their positions and velocities alone, but the speed comparison by itself no longer decides which costs less.

8. Example 11.3: Exponential Growth

Let

$$H(p) = e^{|p|} - 1.$$

We compute

$$L(v) = \sup_{p \in \mathbb{R}^n} \{p \cdot v - (e^{|p|} - 1)\}.$$

For $v \neq 0$ and fixed $r = |p|$, the largest possible $p \cdot v$ is $r|v|$, attained in the direction of v. For $v = 0$, the maximum occurs at $p = 0$. Thus the problem reduces to

$$\sup_{r\geq 0}\{|v|r - e^r + 1\}.$$

If $|v| \leq 1$, the maximum is attained at $r = 0$, so $L(v) = 0$. If $|v| > 1$, the critical point satisfies

$$e^r = |v|, \qquad r = \log|v|.$$

Therefore

$$L(v) = \begin{cases} 0, & |v| \leq 1, \\ |v|\log|v| - |v| + 1, & |v| > 1. \end{cases}$$

On the steady path the speed is 1, so its accumulated running cost is zero. Waiting also costs zero, but traveling at speed 2 for the remaining half of the trip costs

$$\tfrac{1}{2}L(2) = \tfrac{1}{2}(2\log 2 - 1) = \log 2 - \tfrac{1}{2} > 0.$$

Speeds above 1 are allowed; they simply carry a positive cost. Here we can tell from the two speed profiles alone that the steady path is cheaper, without solving for u or Du.

$H(p) = e^{|p|} - 1$ is not differentiable at $p = 0$, but the supremum defining L still applies, as noted after Tran Theorem 2.16.

9. Convex Viscosity Theory: Standing Assumption (2.23)

For the remainder of the lecture, consider

$$\lambda u + H(x, Du) = 0 \quad \text{in } \mathbb{R}^n, \qquad \lambda \geq 0.$$

Tran assumption (2.23) is:

- $H \in \mathrm{BUC}(\mathbb{R}^n \times B(0,R))$ for every $R > 0$;
- $p \mapsto H(x,p)$ is convex for every x;
- H is coercive in p, uniformly in x:

$$\lim_{|p|\to\infty} \inf_{x\in\mathbb{R}^n} H(x,p) = +\infty.$$

Superlinearity is not required for the results in this section.

10. Barron–Jensen Theorem

Theorem 2.32. Assume $\lambda \geq 0$ and that H satisfies (2.23). For $u \in \mathrm{Lip}(\mathbb{R}^n)$, the following statements are equivalent:

1. u is a viscosity subsolution of $\lambda u + H(x, Du) = 0$;
2. u is an almost-everywhere subsolution:

$$\lambda u(x) + H(x, Du(x)) \leq 0 \qquad \text{for a.e. } x \in \mathbb{R}^n.$$

Proof

Suppose first that u is a viscosity subsolution. At a differentiability point, $Du(x) \in D^+u(x)$, so there is an upper test φ with $D\varphi(x) = Du(x)$ (Week 7). The subsolution test gives (2) at this point. By Rademacher's theorem, a Lipschitz function is differentiable almost everywhere, so (2) follows.

Conversely, assume (2). We want a smooth approximation of u for which the subsolution inequality holds at every point, up to an error that tends to zero. Take a smooth, nonnegative weight η_ε supported in $B(0,\varepsilon)$ with integral one, and set

$$u^\varepsilon(x) = \int_{\mathbb{R}^n} u(y)\eta_\varepsilon(x-y)\,dy, \qquad Du^\varepsilon(x) = \int_{\mathbb{R}^n} Du(y)\eta_\varepsilon(x-y)\,dy.$$

The first formula takes a weighted average of the values of u within distance ε of x. Since u is Lipschitz, the gradient of this average is the same weighted average of its almost-everywhere gradients, as in the second formula.

Fix $R > \mathrm{Lip}(u)$, so $|Du(y)| < R$ almost everywhere. By the uniform continuity in (2.23), there is a **modulus of continuity** ω_R: a nondecreasing function that bounds how much H can change when the position changes but the momentum stays fixed. Thus

$$|H(x,p) - H(y,p)| \le \omega_R(|x-y|) \quad (|p| < R), \qquad \omega_R(r) \longrightarrow 0 \quad \text{as } r \downarrow 0.$$

For example, if $H(\cdot,p)$ is Lipschitz with the same constant C_R for all $|p| < R$, we may take $\omega_R(r) = C_R r$. With R fixed, we now estimate

$$\begin{aligned} \lambda u^\varepsilon(x) + H(x, Du^\varepsilon(x)) &\le \int [\lambda u(y) + H(x, Du(y))]\,\eta_\varepsilon(x-y)\,dy \\ &\le \int [\lambda u(y) + H(y, Du(y))]\,\eta_\varepsilon(x-y)\,dy + \omega_R(\varepsilon) \\ &\le \omega_R(\varepsilon). \end{aligned}$$

In the first line, keep x fixed: by convexity in p, Jensen's inequality bounds H of the average gradient by the average of H. In the second line, keep $Du(y)$ fixed: replacing x by y costs at most $\omega_R(\varepsilon)$ because only $|x-y| \le \varepsilon$ contributes. The weights integrate to one, so averaging does not increase this error. In the last line, (2) makes the integrand nonpositive almost everywhere.

Thus the smooth functions u^ε satisfy the subsolution inequality at every x with error at most $\omega_R(\varepsilon)$. As $\varepsilon \downarrow 0$, the local averages converge to u locally uniformly and, since R was fixed, the error tends to zero. Stability of viscosity subsolutions gives (1). □

11. Characterization by Generalized Differentials

Theorem 2.34. Assume $\lambda \ge 0$ and that H satisfies (2.23). A function $u \in \mathrm{Lip}(\mathbb{R}^n)$ is a viscosity solution of $\lambda u + H(x, Du) = 0$ if and only if

$$\lambda u(x) + H(x,p) = 0 \qquad \text{for every } x \in \mathbb{R}^n \text{ and every } p \in D^- u(x).$$

Recall that $D^- u(x)$ consists of the gradients of lower test functions at x. The usual supersolution test gives only ≥ 0 on this set. Convexity gives the opposite inequality there as well. This characterizes solutions; it does not by itself select an optimal feedback at a corner.

Proof

Suppose u is a solution and $p \in D^- u(x)$. Choose a lower test ψ with $\psi(x) = u(x)$ and $D\psi(x) = p$. The supersolution test gives $\lambda u(x) + H(x,p) \ge 0$. To prove the opposite inequality, we cannot apply the subsolution test directly: ψ touches from below.

Instead, use the smooth approximations from Theorem 2.32. Since u is also a subsolution, they satisfy the estimate in §10 at **every** point. To find a point where their gradient agrees with that of ψ, first replace $\psi(y)$ by $\psi(y) - |y-x|^2$. This leaves its value and gradient at x unchanged, but now $u(y) - \psi(y) \ge |y-x|^2$ on a small closed ball about x.

Let x_ε minimize $u^\varepsilon - \psi$ on this ball. We have $x_\varepsilon \to x$: outside any smaller ball about x, the quadratic gap is bounded away from zero, whereas $u^\varepsilon - \psi$ converges uniformly to $u - \psi$ and its value at x tends to zero. Hence the minimum is interior for small ε, and the derivative there vanishes: $Du^\varepsilon(x_\varepsilon) = D\psi(x_\varepsilon)$. The estimate from §10 yields

$$\lambda u^\varepsilon(x_\varepsilon) + H(x_\varepsilon, D\psi(x_\varepsilon)) \le \omega_R(\varepsilon).$$

Since $u^\varepsilon(x_\varepsilon) \to u(x)$ and $D\psi(x_\varepsilon) \to p$, continuity of H gives $\lambda u(x) + H(x,p) \le 0$. Together with the supersolution inequality, this proves equality.

Conversely, assume equality holds for every $p \in D^-u(x)$ at every x. This gives the supersolution test directly. At a differentiability point the first-order expansion gives $Du(x) \in D^-u(x)$, so $\lambda u(x) + H(x, Du(x)) = 0$ there. Such points have full measure by Rademacher's theorem. In particular, (2) of Theorem 2.32 holds, and that theorem gives the subsolution property. □

12. Inf-Stability in the Convex Case

For any continuous Hamiltonian, the infimum of viscosity supersolutions is again a viscosity supersolution, under the usual finiteness and continuity assumptions. This statement does not require convexity.

Convexity gives the stronger conclusion in Tran Corollary 2.36:

- if u_1 and u_2 are Lipschitz solutions of $\lambda u + H(x, Du) = 0$, then $\min\{u_1, u_2\}$ is also a solution;
- if $\{u_i\}_{i\in I}$ is a family of Lipschitz solutions, then

$$u(x) = \inf_{i\in I} u_i(x)$$

  is also a solution, provided u is finite and continuous.

Proof

Let $u = \min\{u_1, u_2\}$. It is Lipschitz. At a strict inequality, u agrees locally with the smaller branch. At a tie, a lower test for u touches both branches from below. Conversely, any slope supporting both branches to first order also supports their minimum. Hence at a tie,

$$D^-u(x) = D^-u_1(x) \cap D^-u_2(x).$$

For every $p \in D^-u(x)$, the equality in Theorem 2.34 therefore holds on a branch whose value is $u(x)$. Applying the theorem to u proves the assertion for two solutions and, by induction, for finite minima.

For a general family the infimum need not be attained. On the closed ball $\overline{B_k(0)}$, the open sets $\{u_i < u+1/k\}$ cover the ball because $u = \inf_i u_i$ is finite and continuous. A finite subcover gives a finite minimum m_k with $u \le m_k < u + 1/k$ on that ball. Thus $m_k \to u$ locally uniformly. Each m_k is a solution, so stability gives the result. □

13. Failure for a Second-Order Equation

Tran Exercise 31 considers

$$|u'|^3 - u'' - 1 = 0 \qquad \text{in } \mathbb{R}.$$

The functions

$$u_1(x) = x, \qquad u_2(x) = -x$$

are classical solutions. Their minimum is

$$u_3(x) = \min\{u_1(x), u_2(x)\} = -|x|.$$

Away from 0, u_3 solves the equation classically. At 0 no differentiable lower test exists: its slope would have to be at least 1 from the left and at most -1 from the right. Thus u_3 is a viscosity supersolution everywhere. The subsolution test, however, fails at 0. To see this, choose $K > \frac{1}{2}$ and set

$$\varphi(x) = -Kx^2.$$

Since $Kx^2 \le |x|$ for $|x| \le 1/K$, φ touches u_3 from above at the origin. However,

$$|\varphi'(0)|^3 - \varphi''(0) - 1 = 2K - 1 > 0,$$

so the subsolution inequality fails.

This example does not show a failure caused by a nonconvex first-order Hamiltonian. It shows that the first-order convex inf-stability theorem does not extend directly to this second-order equation.

14. Summary

Convexity gives Hamilton–Jacobi equations a dual variational structure. Under convexity and superlinearity, the Legendre transform produces a finite, convex, superlinear Lagrangian and the biconjugate recovers the Hamiltonian. The fundamental link is

$$H(x,p) + L(x,v) \geq p \cdot v.$$

Convexity also has decisive consequences for viscosity theory. The Barron–Jensen theorem identifies viscosity and almost-everywhere subsolution inequalities for Lipschitz functions. Theorem 2.34 characterizes solutions through the subdifferential, and Corollary 2.36 gives solution-level inf-stability.

15. Exercises

Exercise 11.1. Prove the Fenchel inequality

$$H(x,p) + L(x,v) \geq p \cdot v$$

directly from the definition of L, and characterize equality using the convex subdifferential.

Exercise 11.2. Let

$$H(x,p) = \frac{|p|^m}{m} + V(x), \qquad m > 1.$$

Compute its Legendre transform and verify explicitly that the conjugate exponents satisfy $1/m + 1/m' = 1$.

Exercise 11.3. Complete the calculation of the Legendre transform of $H(p) = e^{|p|} - 1$. Check continuity and convexity of L at $|v| = 1$.

Exercise 11.4. Using Theorem 2.34, prove the two-solution case of Tran Corollary 2.36:

$$u = \min\{u_1, u_2\}.$$

Identify $D^- u(x)$ at points where $u_1(x) = u_2(x)$.

Exercise 11.5. For

$$u_3(x) = -|x|, \qquad \varphi(x) = -Kx^2,$$

show that φ touches u_3 from above at 0 for every $K > 0$. Then choose any $K > 1/2$ and use this test to show that u_3 is not a viscosity subsolution of

$$|u'|^3 - u'' - 1 = 0.$$

Exercise 11.6. Explain why the infimum of supersolutions being a supersolution does not imply Corollary 2.36. Which additional viscosity inequality must be recovered, and where is convexity used?

16. References

- H. V. Tran, *Hamilton–Jacobi Equations: Theory and Applications*, AMS Graduate Studies in Mathematics 213, 2021.
- E. N. Barron and R. Jensen, "Semicontinuous viscosity solutions for Hamilton–Jacobi equations with convex Hamiltonians," *Communications in Partial Differential Equations* 15 (1990), 1713–1742.