Introduction to Hyperbolic Billiards

Heiko Gimperlein

A billiard is the dynamical system generated by the free motion of a point particle in a domain $Q \subset \mathbb{R}^2$ or $\mathbb{T}^2$ with piecewise smooth boundary $\partial Q = \Gamma_1 \cup \cdots \cup \Gamma_k$. When the particle hits a boundary point at which the normal vector is well-defined, it is elastically reflected so that the angle of incidence equals the angle of reflection. The motion corresponds to a Hamiltonian system with a constant potential in Q that becomes infinite at ∂Q, and we are going to assume that the reflections do not have an accumulation point. The particle trajectories then give rise to a Hamiltonian flow $\{\Phi_t\}_{t\in\mathbb{R}}$ on (a dense subset of) the phase space $\Omega = S^*Q/\sim$, where $\sim$ identifies the incoming and outgoing velocity vectors over ∂Q. Φ_t extends continuously to a multiple–valued map on Ω and preserves the Liouville measure μ. Its lack of differentiability at glancing points and corners is responsible for many of the particular difficulties of billiard systems, even though they correspond to a set of measure 0.

Question: Give an explicit example of a billiard with a cusp on the boundary, such that reflections accumulate in the tip in finite time.

The billiard dynamics is conveniently described in terms of wave fronts, i.e. families of particles parameterized by smooth curve segments with a continuous choice of a normal "velocity" vector field. Away from ∂Q, the curvature κ of the wave front evolves according to

$$\kappa_t = \frac{\kappa_0}{1+t\kappa_0}\ .$$

Upon reflection, κ jumps according to the mirror equation

$$\kappa_{t+0} = \kappa_{t-0} + \frac{2\mathcal{K}(q)}{\cos(\varphi)}$$

depending on the angle of incidence φ and the curvature of the boundary $\mathcal{K}(q)$ at the point of reflection. The two equations completely determine the evolution, though practically they are most useful for the hyperbolic billiards defined below.

If all trajectories are eventually reflected, the dynamics can be reduced to ∂Q: A particle hitting the boundary corresponds to a point in $\mathcal{M} = \bigcup_{i=1}^k \Gamma_i \times [-\frac{\pi}{2}, \frac{\pi}{2}]$, and the billiard map $\mathcal{F} : \mathcal{M} \to \mathcal{M}$ maps a reflection to the next one. As for the flow, it is first defined on regular trajectories and then extends by continuity to a multiple–valued map on all of $\mathcal{M}$. Conversely, we may recover Φ_t from $\mathcal{F}$ and the return–time $T : \mathcal{M} \to (0,\infty)$ of the particle to $\mathcal{M}$. When applied to the Liouville measure on Ω, this correspondence naturally yields the $\mathcal{F}$–invariant measure $\nu = \cos(\alpha)\ dr \wedge d\alpha$ on $\mathcal{M}$.

The billiard map is particularly useful to analyze the stability of the system. An important computation shows that if $(r_1, \alpha_1) = \mathcal{F}(r, \alpha)$ and the curvatures of the Γ_i are bounded, the derivative behaves like $\|D\mathcal{F}(r,\alpha)\| \sim \frac{1}{\cos(\alpha_1)}$, which is singular at glancing points. The singularities of $\mathcal{F}$ are hence contained in $\partial\mathcal{M}\cup\mathcal{F}^{-1}(\partial\mathcal{M})$, and the explicit formula for ν shows

$$\int_{\mathcal{M}} \log^+ \|D\mathcal{F}^{\pm 1}\| \; d\nu < \infty \; .$$

To discuss linearized stability, we may appeal to a general theorem by Oseledets:

Theorem: Let $\mathcal{M}$ be a Riemannian manifold (with corners), $\mathcal{N}\subset\mathcal{M}$ an open dense subset and $\mathcal{F}\in C^2(\mathcal{N},\mathcal{M})$ a diffeomorphism onto its image. Suppose that $\mathcal{F}$ preserves a probability measure ν and $\bigcap_{n\in\mathbb{Z}}\mathcal{F}^n(\mathcal{N})$ is of full measure. If $\int_{\mathcal{M}}\log^+\|D\mathcal{F}^{\pm 1}\|\; d\nu<\infty$, then for each x in a full–measure $D\mathcal{F}$–invariant subset $T_x\mathcal{M}$ admits a $D\mathcal{F}$ invariant decomposition $E_x^1\oplus\cdots\oplus E_x^{m(x)}$ such that

$$\lim_{n\to\pm\infty}\frac{1}{n}\log\|D\mathcal{F}^n v\| = \lambda_x^i \qquad (v\in E_x^i)\; .$$

If $\mathcal{F}$ is ergodic, the Lyapunov exponents λ_x^i are a.e. constant. Loosely speaking, $\|D\mathcal{F}^n|_{E_x^i}\|\sim e^{n\lambda_x^i}$, so that for $\lambda_x^i>0$ small perturbations in the direction E_x^i at time 0 are expected to grow exponentially in the future, making the long–term evolution "chaotic" and practically inaccessible.

If the Lyapunov exponents exist and are nonzero a.e., the map $\mathcal{F}$ is said to be hyperbolic. For two–dimensional billiards, the preservation of the Liouville measure assures $\lambda_x^1+\lambda_x^2=0$, and the billiard map is hyperbolic unless both exponents vanish.

Because of the above correspondence between the continuous and the discrete system, hyperbolicity turns out to be a property of the actual billiard flow on phase space: The tangent space of Ω splits into orthogonal subspaces $T^0\Omega$ and $T^\perp\Omega$ tangent resp. perpendicular to the flow, and the exponents associated to Φ_t on $T^\perp\Omega$ agree with those of $\mathcal{F}$. $T^0\Omega$ merely contributes an additional exponent 0.

A powerful method to determine the stable and unstable subspaces associated to the positive resp. negative Lyapunov exponents is the method of cone fields [1]. Here, a cone $\mathcal{C}_x$ is determined by a line $\mathcal{L}_x\subset T_x\mathcal{M}$ and an angle $\alpha_x\in(0,\frac{\pi}{2})$, such that $\mathcal{C}_x=\{v\in T_x\mathcal{M}:\angle(v,\mathcal{L}_x)\le\alpha_x\}$. Intuitively, application of $D\mathcal{F}^n$ to a cone should stretch out the unstable directions and "converge" to the unstable subspace E_x^u for $n\to\infty$. A convenient criterion goes back to Wojtkowski [4, 5]: If for almost all $x\in\mathcal{M}$ one finds cones $\mathcal{C}_x^u$ satisfying the strict invariance condition $D\mathcal{F}(\mathcal{C}_x^u)\subset \mathrm{int}\ \mathcal{C}^u_{\mathcal{F}(x)}\cup\{0\}$, then $\mathcal{F}$ is hyperbolic and $E_x^u=\bigcap_{n=0}^\infty D\mathcal{F}^n(\mathcal{C}^u_{\mathcal{F}^{-n}(x)})$. Replacing $\mathcal{F}$ by $\mathcal{F}^{-1}$ leads to the corresponding assertions for the stable subspace.

Once the infinitesimal case has been clarified, one would, of course, want to use it to investigate the long–time behavior of the flow. The appropriate notion is that of an unstable manifold, which is defined as a smooth curve $\mathcal{W}^u\subset\mathcal{M}$ such that $\mathcal{F}^{-n}$ is smooth on $\mathcal{W}^u$ for all $n\ge 1$ and $\lim_{n\to\infty}|\mathcal{F}^{-n}(\mathcal{W}^u)|=0$. Similarly, a stable manifold is an unstable manifold for $\mathcal{F}^{-1}$.

Dispersing billiards provide a prototypical example of a hyperbolic system. We consider the simple case where the Γ_i are nonintersecting, concave outward, smooth closed curves in $\mathbb{T}^2$ and all trajectories eventually get reflected. The two basic equations for the propagation of wave fronts assure that a dispersing wave front ($\kappa \geq 0$) will remain dispersive. A bit more work establishes hyperbolicity by applying Wojtkowski's criterion to the cone field

$$\mathcal{C}_x^u = \{(\text{linearized) wave fronts at } x \text{ with } \kappa_{-0} \geq 0\} \ .$$

To construct stable and unstable manifolds through a point $x \in \mathcal{M}$, which lies on a regular trajectory, we recall that the singularities of $\mathcal{F}$ are contained in $\partial\mathcal{M} \cup \mathcal{F}^{-1}(\partial\mathcal{M})$. Similarly, the singularities $\mathcal{S}_n$ of $\mathcal{F}^n$, $n \in \mathbb{Z}$, are given by a union of compact smooth curves in $\mathcal{M}$, and we let $Q_n(x)$ the connected component of $\mathcal{M} \setminus \mathcal{S}_n$ containing x. $\overline{Q_n}(x)$ is a curvilinear polygon that shrinks to a line for $n \to -\infty$, and we obtain the maximal unstable manifold containing x quite explicitly as $W^u = \bigcap_{n\geq 1} \overline{Q_{-n}}(x) \setminus \{\text{endpoints}\}$. Replacing $\mathcal{F}$ by $\mathcal{F}^{-1}$ yields a maximal stable manifold through x.

We conclude with some remarks on the local ergodicity of dispersive billiards:

Theorem *(Sinai* [3]*)***:** Any $x \in \mathcal{M}$ which lies on at most one smooth singularity curve has an open neighborhood that belongs to a single ergodic component of $\mathcal{F}$.

Like most proofs of ergodicity for hyperbolic systems, Sinai's proof relies on Hopf's method and informally proceeds in two steps:
1) A.e. stable or unstable manifold belongs to a single ergodic component of $\mathcal{F}$.
2) Generic $x, y \in \mathcal{M}$ are connected by a finite sequence of stable and unstable manifolds $\mathcal{W}_1, \ldots, \mathcal{W}_k$ such that $\mathcal{W}_i \cap \mathcal{W}_{i+1} \neq \emptyset$, a so–called Hopf chain.
While 2) is based on the construction of stable and unstable manifolds, the heuristic idea behind 1) is as follows: If $M_{1/2} \subset \mathcal{M}$ are distinct ergodic components, then $\mathcal{F}$ preserves the conditional ergodic measures $\mu_{1/2}$. The trajectory of μ_1–a.e. $x \in M_1$ is distributed according to μ_1, and similarly for $y \in M_2$. Therefore, given e.g. a stable manifold $\mathcal{W}^s$ intersecting M_1 and M_2, the trajectories of typical points $x \in \mathcal{W}^s \cap M_1$, $y \in \mathcal{W}^s \cap M_2$ are distributed according to μ_1 resp. μ_2. But lying on $\mathcal{W}^s$, the future trajectories of x and y converge to each other and hence must lead to identical distributions. Contradiction to $\mu_1 \neq \mu_2$.

References

[1] V. M. Alekseev, *Quasi random dynamical systems*, Mat. USSR Sbornik **7** (1969), 1–43.

[2] N. Chernov and R. Markarian, *Chaotic Billiards*, American Mathematical Society, Providence, RI, 2006.

[3] Ya. G. Sinai, *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russ. Math. Surv. **25** (1970), 137–189.

[4] M. Wojtkowski, *Invariant families of cones and Lyapunov exponents*, Ergod. Theory Dynam. Syst. **5** (1985), 145–161.

[5] M. Wojtkowski, *Principles for the design of billiards with nonvanishing Lyapunov exponents*, Comm. Math. Phys. **105** (1986), 391–414.