Hamiltonian Descent Methods

University of Oxford

September 20, 2018

Joint work with Chris J. Maddison, Yee Whye Teh, Brendan O'Donoghue and Arnaud Doucet

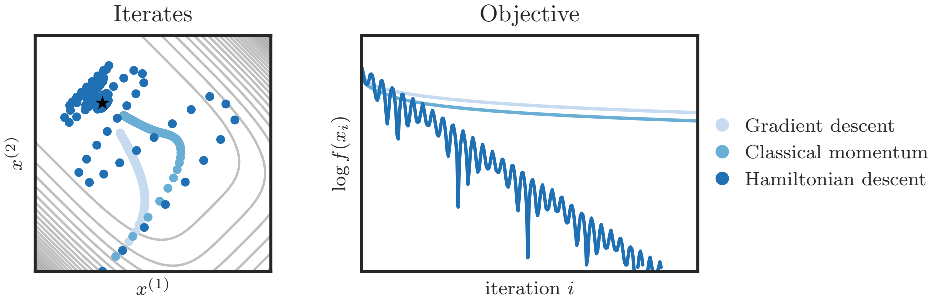


Figure: Optimizing $f(x) = [x^{(1)} + x^{(2)}]^4 + [x^{(1)}/2 - x^{(2)}/2]^4$ with three methods: gradient descent with fixed step size equal to $1/L_0$ where $L_0 = \lambda_{\max}(\nabla^2 f(x_0))$ is the maximum eigenvalue of the Hessian $\nabla^2 f$ at x_0; classical momentum, which is a particular case of our first explicit method with $k(p) = [(p^{(1)})^2 + (p^{(2)})^2]/2$ and fixed step size equal to $1/L_0$; and Hamiltonian descent, which is our first explicit method with $k(p) = (3/4)[(p^{(1)})^{4/3} + (p^{(2)})^{4/3}]$ and a fixed step size.

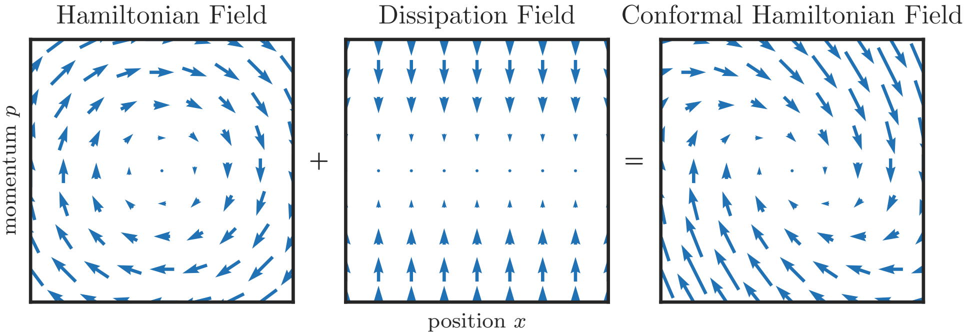


Figure: A visualization of a conformal Hamiltonian system.

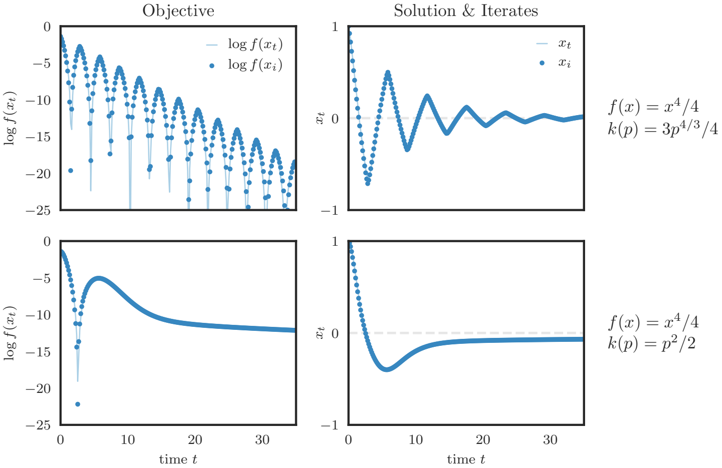


Figure: Importance of assumptions A. Solutions x_t and iterates x_i of our first explicit method on $f(x) = x^4/4$ with two different choices of k. Notice that $f_c^*(p) = 3p^{4/3}/4$ and thus $k(p) = p^2/2$ cannot be made to satisfy assumption A.4.

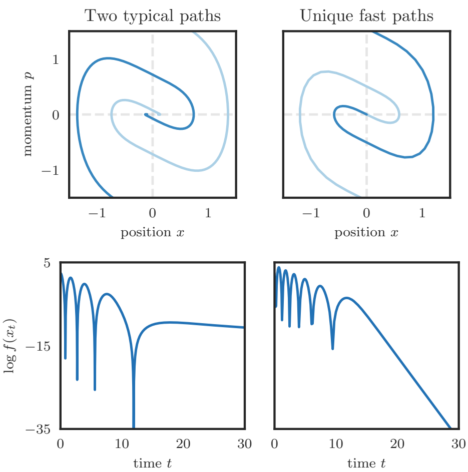
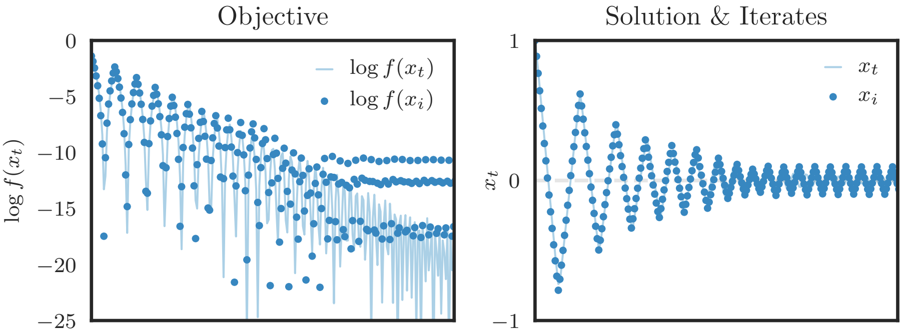


Figure: Solutions for $f(x) = x^4/4$ and $k(p) = x^2/2$. The right plots show a numerical approximation of $(x_t^{(\eta)}, p_t^{(\eta)})$ and $(-x_t^{(\eta)}, -p_t^{(\eta)})$. The left plots show a numerical approximation of $(x_t^{(\theta)}, p_t^{(\theta)})$ and $(-x_t^{(\theta)}, -p_t^{(\theta)})$ for $\theta = \eta + \delta \in \mathbb{R}$, which represent typical paths.



$$f(x) = x^4/4$$
$$k(p) = p^{8/7}7/8$$

Figure: Importance of discretization assumptions. Solutions x_t and iterates x_i of our first explicit method on $f(x) = x^4/4$. With an inappropriate choice of kinetic energy, $k(p) = p^{8/7}/(8/7)$, the continuous solution converges at a linear rate but the iterates do not.

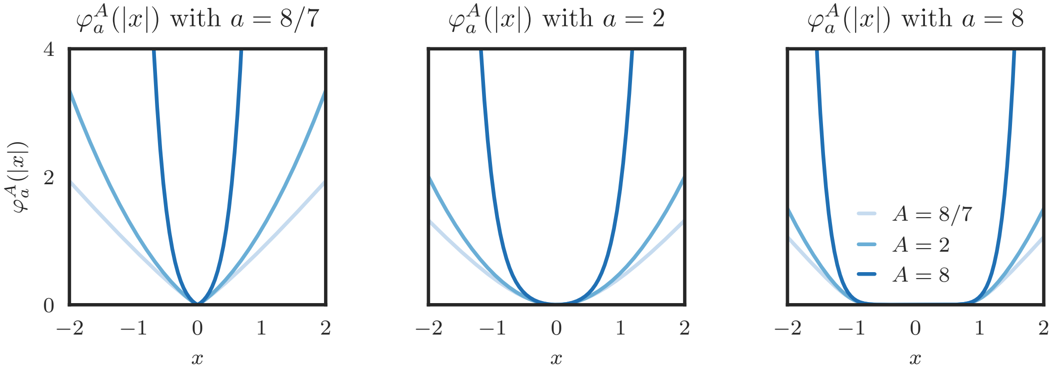


Figure: Power kinetic energies in one dimension.

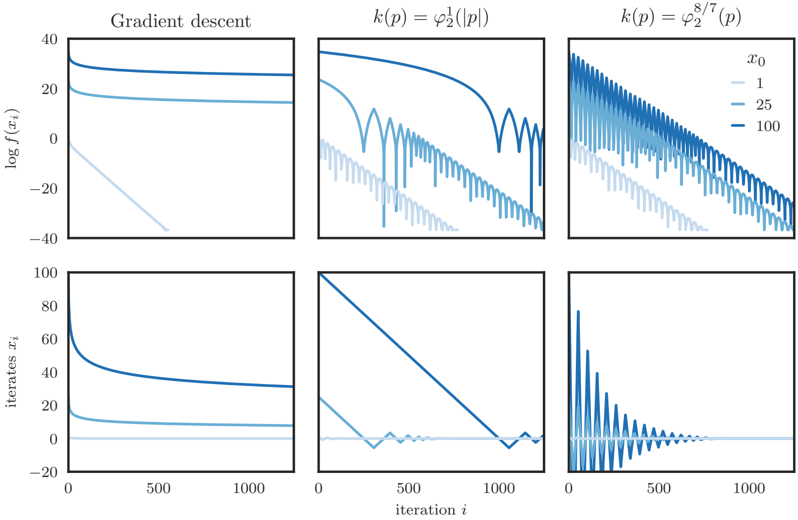


Figure: $f(x) = \varphi_2^8(x)$ with three different methods: gradient descent with the optimal fixed step size, Hamiltonian descent with relativistic kinetic energy, and Hamiltonian descent with the near dual kinetic energy.

Assumptions F.

F.1 $f : \mathbb{R}^d \to \mathbb{R}$ differentiable and convex with unique minimum $x_\star$.

F.2 $\|p\|_*$ is differentiable at $p \in \mathbb{R}^d \setminus \{0\}$ with dual norm $\|x\| = \sup\{\langle x, p\rangle : \|p\|_* = 1\}$.

F.3 $B = A/(A-1)$, and $b = a/(a-1)$.

F.4 There exist $\mu, L \in (0, \infty)$ such that for all $x \in \mathbb{R}^d$

$$\begin{aligned} f(x) - f(x_\star) &\geq \mu \varphi_b^B(\|x - x_\star\|) \\ \varphi_a^A(\|\nabla f(x)\|_*) &\leq L(f(x) - f(x_\star)). \end{aligned} \tag{1}$$

F.5 $b \geq 2$ and $B \geq 2$. $f : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable for all $x \in \mathbb{R}^d \setminus \{x_\star\}$ and there exists $L_f, D_f \in (0, \infty)$ such that for all $x \in \mathbb{R}^d \setminus \{x_\star\}$

$$\left(\varphi_{b/2}^{B/2}\right)^* \left(\frac{\lambda_{\max}^{\|\cdot\|}(\nabla^2 f(x))}{L_f}\right) \leq D_f(f(x) - f(x_\star)). \tag{2}$$

Assumptions G.

G.1 $f : \mathbb{R}^d \to \mathbb{R}$ differentiable and convex with unique minimum $x_\star$.

G.2 $\|p\|_*$ is differentiable at $p \in \mathbb{R}^d \setminus \{0\}$ with dual norm $\|x\| = \sup\{\langle x, p\rangle : \|p\|_* = 1\}$.

G.3 $B \in [2, \infty)$ and $A = B/(B-1)$.

G.4 There exist $\mu, L \in (0, \infty)$ such that for all $x \in \mathbb{R}^d$

$$\begin{aligned} f(x) - f(x_\star) &\geq \mu \varphi_2^B(\|x - x_\star\|) \\ \varphi_2^1(\|\nabla f(x)\|_*) &\leq L(f(x) - f(x_\star)). \end{aligned} \tag{3}$$

G.5 $B > 2$. Define

$$\psi(t) = \begin{cases} 0 & 0 \leq t < 1 \\ t - 3t^{\frac{1}{3}} + 2 & 1 \leq t \end{cases}. \tag{4}$$

$f : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable for all $x \in \mathbb{R}^d \setminus \{x_\star\}$ and there exists $L_f \in (0, \infty)$ such that for all $x \in \mathbb{R}^d \setminus \{x_\star\}$

$$\psi\left(\tfrac{B-1}{B-2}\varphi_1^{\frac{B-1}{B-2}}\left(\frac{\lambda_{\max}^{\|\cdot\|}(\nabla^2 f(x))}{L_f}\right)\right) \leq 3(f(x) - f(x_\star)). \tag{5}$$