On Dissipative Symplectic Integration with Applications to Gradient-Based Optimization

Guilherme França, Michael I. Jordan, and René Vidal

based on arXiv:2004.06840 [math.OC]

Problem setup

Suppose we have a **dissipative** Hamiltonian system:

$$\frac{dq^j}{dt} = \frac{\partial H}{\partial p_j}, \qquad \frac{dp_j}{dt} = -\frac{\partial H}{\partial q^j}, \qquad H = H(t, q, p),$$

where $q \in \mathcal{M}^n$ (smooth manifold) and $(q, p) \in T^*\mathcal{M}$ (cotangent bundle) $(j = 1, \ldots, n)$. Assume that its trajectories "efficiently" solve

$$\min_{q \in \mathcal{M}} f(q),$$

and that we understand the dynamics; i.e. stability, convergence rates, etc. A fundamental question is the following:

- **Which discretizations are able to preserve the stability and rates of convergence of such a continuous-time system?**
- The answer would give us a systematic way to derive efficient optimization algorithms ("acceleration") ...
- ... without the need for a discrete-time convergence analysis.

Problem setup

Conservative Hamiltonian systems are ubiquitous — $H(q, p)$ is independent of time. But the conservation of energy precludes convergence to a point; think about the harmonic oscillator.

This is the opposite behaviour of what we want in optimization. Thus, there is a clear need for "dissipation" — where $H(t, q, p)$ is explicit time-dependent — which leads us to another important question:

- **Can we map optimization algorithms into dissipative continuous-time dynamical systems that provide analytical insight into the behavior of the algorithm?**
- The answer would allow us to infer stability and convergence rates of such algorithms with a broader mathematical machinery than traditionally available.

Our approach

- Those two questions are somehow related. The ability to **preserve** convergence rates can be seen as some kind of **"invariance."**
- But "dissipation" is the complete opposite of "conservation," thus how can this even make sense?? A dissipative system presumably has no conservation law.
- Using **symplectic geometry**, we will show that a dissipative Hamiltonian system can be seen as a conservative Hamiltonian system in higher-dimensions (symplectification + gauge fixing).
- Together with **backward-error analysis** we can bring these ideas to discrete-time to obtain a framework (presymplectic integrators) where the stability and convergence rates of the continuous system are preserved (up to a small and controlled error).

Backward-error analysis

Consider an arbitrary dynamical system over a smooth manifold $\mathcal{M}$:

$$\dot{x}(t) = X(x(t)),$$

where X is the vector field and $\varphi_t = e^{tX}$ is its **flow map**.

A **numerical map** ϕ_h, of **order** $r \geq 1$, is an approximation ($h > 0$):

$$\|\phi_h(x) - \varphi_h(x)\| = \mathcal{O}(h^{r+1}) \qquad \text{for any } x \in \mathcal{M}.$$

Theorem

Every numerical method, ϕ_h, can be seen as the "exact flow" of a perturbed dynamical system:

$$\dot{x}(t) = \tilde{X}(x(t)), \qquad \tilde{X} = X + \Delta X_1 h + \Delta X_2 h^2 + \cdots .$$

These ideas have been developed since the late 90's in numerical analysis (Benettin, Giorgilli, Hairer, Reich, Lubich, . . .).

Backward-error analysis

The perturbed vector field $\tilde{X}$ has to be **truncated**. Denoting by $\varphi_{t,\tilde{X}} = e^{t\tilde{X}}$ the associated flow, one has:

Theorem (Benettin, Giorgilli, Hairer, Reich, ...)

There exists a family of (truncated) perturbed vector fields, $\|X(x) - \tilde{X}(x)\| = \mathcal{O}(h^r)$, *such that* $\|\phi_h(x) - \varphi_{h,\tilde{X}}(x)\| \le C h e^{-r} e^{-h_0/h}$.

- This tells us that the numerical flow is very close to the "perturbed flow" (exponentially small error).
- For typical numerical integrators this result is not very useful. One is rather interested in comparing ϕ_h to φ_h (not $\varphi_{h,\tilde{X}}$).
- However, this result becomes **extremely useful** if one can show that $\tilde{X}$ has the "same structure" as X. This is why **structure-preserving** methods are special; e.g., symplectic integrators.

Symplectic manifolds and conservative Hamiltonians

Definition

An even-dimensional smooth manifold $\mathcal{M}$ endowed with a closed nondegenerate 2-form ω is a **symplectic manifold**.[a]

[a] ω maps two vectors into a number and it is a totally *skew-symmetric* object, $\omega(X, Y) = -\omega(Y, X)$, thus it imposes a special geometry on $\mathcal{M}$.

As an analogy, in going from the real to complex numbers one introduces $i^2 = -1$. Here, in a matrix representation, one introduces $\omega^2 = -I$ over $\mathcal{M}$. Thus symplectic manifolds carry a sort of "complex structure."

- Symplectic geometry arises in several areas: classical mechanics, complex geometry, Lie groups and algebras, representation varieties, geometric quantization, and so on. They are worth studying in their own right and have a beautiful mathematical structure.

Symplectic manifolds and conservative Hamiltonians

The universality of symplectic manifolds and Hamiltonian systems follow from the following facts.

Theorem

The tangent bundle[a] $T^*\mathcal{M}$ *of any differentiable manifold* $\mathcal{M}$*, with coordinates* $q^1, \dots, q^n, p_1, \dots, p_n$*, is a symplectic manifold. The symplectic 2-form is given by* $\omega = \sum_j dp_j \wedge dq^j$.

[a] The tangent bunddle is just the collection of all cotangent spaces, i.e. the collection of all (tensor products of) dual vector spaces.

Theorem

A dynamical system with phase space $T^*\mathcal{M}$ *preserves the simplectic structure* ω *if and only if it is a conservative Hamiltonian system.*[a]

[a] One with a time-independent Hamiltonian $H = H(q, p)$.

Symplectic manifolds and conservative Hamiltonians

1. By "preserving" we mean that the Lie derivative of the Hamiltonian vector field obeys $\mathcal{L}_{X_H}\omega = 0$. This is the first fundamental property.
2. The second fundamental property is energy conservation: $\frac{dH}{dt} = 0$.

Definition

It is possible to construct a class of numerical integrators, ϕ_h, that exactly preserve ω: $\phi_h^* \circ \omega \circ \phi_h = \omega$. They are called **symplectic integrators**.

1. This implies that the perturbed dynamical system associated to ϕ_h obeys $\mathcal{L}_{\tilde{X}}\omega = 0$.[1] Thus, X_H and $\tilde{X}$ have the same structure!
2. The last theorem above implies that the perturbed system must be a Hamiltonian system, with a perturbed $\tilde{H}$, and for which $\frac{d\tilde{H}}{dt} = 0$.

[1] Recall that $\tilde{X}$ is the vector field of the perturbed system, associated to ϕ_h.

Why symplectic integrators are so successful?

We can now use the previous general backward-error analysis theorem:

Theorem (Benettin, Giorgilli)

Let ϕ_h be a symplectic integrator of order r. Assume H is Lipschitz. Then for large simulation times $t_\ell = h\ell = \mathcal{O}(h^r e^r e^{h_0/h})$, $\ell = 0, 1, \dots,$ we have

$$\underbrace{H \circ \phi_{t_\ell}}_{\text{discrete}} = \underbrace{H \circ \varphi_{t_\ell}}_{\text{continuous}} + \underbrace{\mathcal{O}(h^r)}_{\text{bounded error}}$$

1. A symplectic integrator preserves the symplectic form, ω, exactly;
2. It "almost" preserves the energy, H (up to a bounded error).

however . . . things break down in a dissipative setting!

There is one crucial assumption behind all of this: *the Hamiltonian is a constant of motion $H = const$.* Therefore, these arguments break down when H varies over time, i.e., in the absence of a conservation law.

Dissipative Hamiltonian systems

- Since $H(t, q, p)$ depends on time, the Hamiltonian is not conserved:

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} \neq 0.$$

- One can also show that the symplectic form is no longer preserved, $\mathcal{L}_{X_H}\omega \neq 0$. Thus, the phase space is no longer a symplectic manifold.
- **Actually, all we said before for conservative systems and symplectic integrators no longer holds with dissipation!**
- One can "naively" apply a symplectic integrator to a dissipative system, but there is no existing result that extends that "main theorem"—close preservation of H and long term stability—into a dissipative setting ...
- ... What is the geometry of the phase space? Does the numerical method reproduces the Hamiltonian? Does it has long time stability?

Symplectification

There is a generalization of symplectic manifolds:

Definition

A **presymplectic manifold** $\mathcal{M}$ has dimension $2n + \bar{n}$ ($\bar{n} \geq 0$), and a 2-form ω of rank $2n$ everywhere. (The presymplectic form ω is degenerate.)[a]

[a] In our case, $\bar{n} = 1$.

It is possible to construct a **conservative** Hamiltonian system, $\mathscr{H}$, on a higher-dimensional **symplectic manifold** $T^*\hat{\mathcal{M}}$, of dimension $2n + 2$. Let its coordinates be (q^μ, p_μ), for $\mu = 0, 1, \ldots, n$:

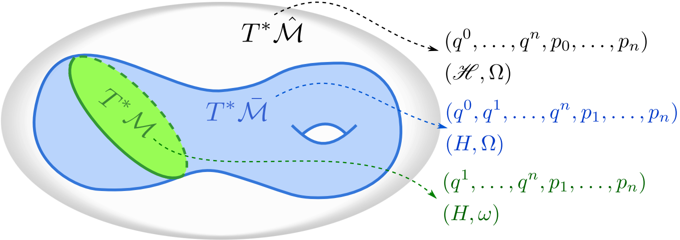
$$\frac{dq^\mu}{ds} = \frac{\partial \mathscr{H}}{\partial p_\mu}, \qquad \frac{dp_\mu}{ds} = -\frac{\partial \mathscr{H}}{\partial q^\mu}, \qquad \frac{d\mathscr{H}}{ds} = 0 \quad \text{(energy conservation)}.$$

Here s is a "new time parameter." Then it is possible to **embed** the original dissipative system into this symplectic manifold.

Symplectification

By removing the spurious degrees of freedom—gauge fixing—, i.e. setting $q^0 = t = s$ and $p_0 = H(s) \equiv H(q(s), p(s))$—this is a function of time!—, then the dissipative system lies on a hypersurface $\mathscr{H} = \text{const.}$ defined by:

$$\mathscr{H}(q^0, \ldots, q^n, p_0, \ldots, p_n) = p_0(s) + H(q^0, q^1, \ldots, q^n, p_1, \ldots, p_n).$$



Under this correspondence, the symplectic structure of the higher dimensional conservative system, Ω, recovers the "presymplectic structure" of the dissipative system, ω.

Presymplectic integrators

We define the following class of numerical methods:

Definition

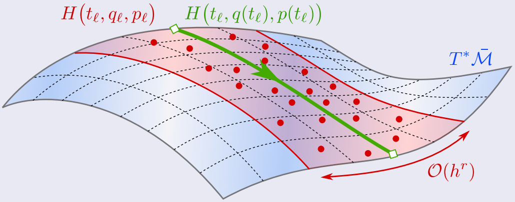
ϕ_h is a **presymplectic integrator** to the **dissipative** Hamiltonian system if it is a **reduction** of a symplectic integrator for its symplectification.

Theorem

Due to this correspondence, we can extend the range of standard theorems into a dissipative setting, where there is no conservation law. In particular, we can prove that the **decaying Hamiltonian is "preserved:"**

$$\underbrace{H \circ \phi_{t_\ell}}_{\text{numerical}} = \underbrace{H \circ \varphi_{t_\ell}}_{\text{continuous}} + \underbrace{\mathcal{O}(h^r)}_{\text{small error}}$$

$$\text{for } t_\ell \equiv h\ell = \mathcal{O}(h^r e^r e^{h_0/h})$$



Implications for optimization

We consider dissipative systems arising from the general class of Hamiltonians:

$$H \equiv e^{-\eta_1(t)} T(t,q,p) + e^{\eta_2(t)} f(q).$$

$\eta_1, \eta_2 \geq 0$ are increasing with t. In specific cases, we know how to obtain a continuous-time convergence rate: $f(q(t)) - f^\star \leq \mathcal{R}(t)$.

Corollary

A presymplectic integrator ϕ_h, of order $r \geq 1$, is a "rate-matching" discretization:

$$\underbrace{f(q_\ell) - f^\star}_{\text{discrete rate}} = \underbrace{f(q(t_\ell)) - f^\star}_{\text{continuous rate}} + \underbrace{\mathcal{O}\left(h^r e^{-\eta_2(t_\ell)}\right)}_{\text{tiny error}},$$

provided $e^{L_\phi t_\ell - \eta_1(t_\ell)} < \infty$ and for large $t_\ell \equiv h\ell = \mathcal{O}(h^r e^r e^{h_0/h})$.

Implications for optimization

- Under appropriate damping, presymplectic integrators can provide "rate-matching" discretizations.

- The error decreases with the order, $\sim h^r$, but is dominated by $\sim e^{-\eta_2(t)}$. Thus high-order integrators may not be necessary.

- If η_2 grows sufficiently fast, the error can be negligible; e.g. exponentially small.

- $\ell \sim h^{r-1} e^r e^{h_0/h}$ is astonishingly large; e.g., $h = 0.01$, $\ell \sim 10^{43}$.

- The strongest requirement is $e^{L_\phi t - \eta_1(t_\ell)} < \infty$, which "fixes" η_1. In particular, the "heavy ball damping", $\eta_1 = \gamma t$, or "Nesterov's damping", $\eta_1 = \gamma \log t$, can be seen as arising from this condition.

- Other choices may be possible, such as $\eta_1 = \gamma_1 \log t + \gamma_2 t^\delta$.

Example: the Bregman dynamics

The Bregman Hamiltonian provides a general approach to optimization (Wibisono, Wilson, MJ, PNAS 2016):

$$H = e^{\alpha+\gamma}\left\{ D_{h^\star}\left(\nabla h(q) + e^{-\gamma}p, \nabla h(q)\right) + e^{\beta} f(q) \right\},$$

where D_h is the Bregman divergence, obtained in terms of a convex function $h(x)$, and $h^\star$ is its convex dual. Under appropriate "scaling conditions" on α, β, γ, Hamilton's equations are equivalent to

$$\ddot{q} + \left(e^{\alpha} - \dot{\alpha}\right)\dot{q} + e^{2\alpha+\beta}\left[\nabla^2 h\left(q + e^{-\alpha}\dot{q}\right)\right]^{-1}\nabla f(q) = 0.$$

For a convex function f, one can show that this system has a convergence rate given by:

$$f(q(t)) - f^\star = \mathcal{O}\left(e^{-\beta(t)}\right).$$

Bregman dynamics: separable case

Choosing $h(x) = \frac{1}{2} x \cdot Mx$, the kinetic energy simplifies and we have

$$H = \tfrac{1}{2} e^{-\eta_1(t)} \, p \cdot M^{-1} p + e^{\eta_2(t)} f(q), \quad \eta_1 \equiv \gamma - \alpha, \quad \eta_2 \equiv \alpha + \beta + \gamma.$$

One can now apply any presymplectic integrator (many possible choices are available). For instance, one based on the popular leapfrog method yields

$$\begin{aligned}
t_{\ell+1/2} &= t_\ell + h/2, \\
q_{\ell+1/2} &= q_\ell + (h/2) e^{-\eta_1(t_{\ell+1/2})} M^{-1} p_\ell, \\
p_{\ell+1} &= p_\ell - h e^{\eta_2(t_{\ell+1/2})} \nabla f(q_{\ell+1/2}), \\
t_{\ell+1} &= t_{\ell+1/2} + h/2, \\
q_{\ell+1} &= q_{\ell+1/2} + (h/2) e^{-\eta_1(t_{\ell+1/2})} M^{-1} p_{\ell+1}.
\end{aligned}$$

One can now make several choices for M, α, β, and γ to obtain a specific optimization algorithm that will respect the continuous convergence rate.

Bregman dynamics: nonseparable case

It is possible to construct **explicit** methods even though the general Bregman Hamiltonian is **nonseparable**. This is done by duplicating the degrees of freedom:

$$\bar{H}(t, q, p, \bar{t}, \bar{q}, \bar{p}) \equiv H(t, q, \bar{p}) + H(\bar{t}, \bar{q}, p) + \frac{\xi}{2}\left(\|q-\bar{q}\|^2 + \|p-\bar{p}\|^2\right).$$

We thus propose the following numerical maps:

$$\phi_h^A \begin{pmatrix} t \\ q \\ p \\ \bar{t} \\ \bar{q} \\ \bar{p} \end{pmatrix} = \begin{pmatrix} t \\ q \\ p - h\nabla_q H(t, q, \bar{p}) \\ \bar{t} + h \\ \bar{q} + h\nabla_{\bar{p}} H(t, q, \bar{p}) \\ \bar{p} \end{pmatrix}, \quad \phi_h^B \begin{pmatrix} t \\ q \\ p \\ \bar{t} \\ \bar{q} \\ \bar{p} \end{pmatrix} = \begin{pmatrix} t + h \\ q + h\nabla_p H(\bar{t}, \bar{q}, p) \\ p \\ \bar{t} \\ \bar{q} \\ \bar{p} - h\nabla_{\bar{q}} H(\bar{t}, \bar{q}, p) \end{pmatrix},$$

$$\phi_h^C \begin{pmatrix} t \\ q \\ p \\ \bar{t} \\ \bar{q} \\ \bar{p} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2t \\ q + \bar{q} + \cos(2\xi h)(q - \bar{q}) + \sin(2\xi h)(p - \bar{p}) \\ p + \bar{p} - \sin(2\xi h)(q - \bar{q}) + \cos(2\xi h)(p - \bar{p}) \\ 2\bar{t} \\ q + \bar{q} - \cos(2\xi h)(q - \bar{q}) - \sin(2\xi h)(p - \bar{p}) \\ p + \bar{p} + \sin(2\xi h)(q - \bar{q}) - \cos(2\xi h)(p - \bar{p}) \end{pmatrix}.$$

A presymplectic integrator can then be constructed by composing these maps. For instance, with the Strang composition ($r = 2$):

$$\phi_{h/2}^A \circ \phi_{h/2}^B \circ \phi_h^C \circ \phi_{h/2}^B \circ \phi_{h/2}^A.$$

Conclusions

- We introduced "presymplectic integrators" which are suitable to simulating dissipative Hamiltonian systems.
- We showed how the important properties of symplectic integrators, which only apply for conservative systems, can be extended to dissipative system where there is no underlying conservation law.
- This has implications to optimization; e.g., it allowed us to show that presymplectic integrators may yield "rate-matching" optimization algorithms in quite general terms.
- No discrete-time convergence analysis was necessary. In this approach, it rather follows by construction.
- There is an entire class of algorithms that can be systematically constructed our framework, that will be guaranteed to preserve the stability and continuous-time rates of convergence.

Acknowledgements: work supported by ARO MURI W911NF-17-1-0304.