Conformal Symplectic and Relativistic Optimization

Guilherme França,[⋆,a,b] Jeremias Sulam,[b] Daniel P. Robinson,[b] and René Vidal[b]

[a] Computer Science Division, University of California, Berkeley
[b] Mathematical Institute for Data Science, Johns Hopkins University
[⋆] guifranca@berkeley.edu

Berkeley UNIVERSITY OF CALIFORNIA

JOHNS HOPKINS MATHEMATICAL INSTITUTE for DATA SCIENCE

Accelerated Methods

The two most popular accelerated optimization methods are the *heavy ball* method, also known as *classical momentum* (CM), given by

$$v_{k+1} = \mu v_k - \epsilon \nabla f(x_k), \qquad x_{k+1} = x_k + v_{k+1}, \tag{1}$$

as well as and *Nesterov's accelerated gradient* (NAG),

$$v_{k+1} = \mu v_k - \epsilon \nabla f(x_k + \mu v_k), \qquad x_{k+1} = x_k + v_{k+1}. \tag{2}$$

Both can bee seen as discretizations of a 2nd order ODE:

$$m\ddot{x}(t) + m\gamma \dot{x}(t) = -\nabla f(x(t)). \tag{3}$$

This system is special because it is a conformal Hamiltonian system.

Conformal Hamiltonian Systems

Given a Hamiltonian $H(x,p)$ the system is defined by a modified form of Hamilton's equations:

$$\dot{x} = \nabla_p H, \qquad \dot{p} = -\nabla_x H - \gamma p, \tag{4}$$

where $\gamma > 0$. Such systems are *dissipative*, i.e. $\frac{dH}{dt} \le 0$. For instance (3) is obtained with

$$H = \frac{\|p\|^2}{2m} + f(x). \tag{5}$$

Conformal Hamiltonian systems have a very special geometric structure, namely the phase space is a conformal symplectic manifold, endowed with a 2-form $\omega \equiv dx \wedge dp$ which contracts as

$$\omega(t) = e^{-\gamma t}\omega(0). \tag{6}$$

This is the most fundamental property of such systems.

Conformal Symplectic Integrators

Definition 1 *A* numerical integrator *is a map,* $(x_{k+1}, p_{k+1}) = \Phi_h(x_k, p_k)$, *where* $h > 0$ *is the step size and* $k = 1, 2, \dots$. *It is said to be of order* r *if* $\|\Phi_h(x,p) - \varphi_h(x,p)\| = O(h^{r+1})$ *where* φ *is the true flow.*

Definition 2 *A* conformal symplectic integrator *is a numerical map that preserves* (6) *exactly, i.e.* $\omega_{k+1} = e^{-\gamma h}\omega_k$.

This is interesting because:

- The most fundamental structure of the system is preserved.
- Numerical trajectories lie on the same "symplectic manifold."
- Stability of critical points are preserved.
- Continuous-time rates of convergence are also preserved (as recently shown [2] in much more generality than considered here).
- It is a principled way to discretize the system.

Two General Integratos

We construct to generic integrators. The first is order $r = 1$:

$$x_{k+1} = x_k + h\nabla_p H(x_k, p_{k+1}), \qquad p_{k+1} = p_k - h\nabla_x H(x_k, p_{k+1}). \tag{7}$$

This can be seen as a dissipative version of symplectic Euler. The second is order $r = 2$:

$$x_{k+1/2} = x_k + \tfrac{h}{2}\nabla_p H(x_{k+1/2}, e^{-\gamma h/2} p_k), \tag{8a}$$
$$p_{k+1/2} = e^{-\gamma h/2} p_k - \tfrac{h}{2}\left[\nabla_x H(x_{k+1/2}, e^{-\gamma h/2} p_k) + \nabla_x H(x_{k+1/2}, p_{k+1/2})\right], \tag{8b}$$
$$x_{k+1} = x_{k+1/2} + \tfrac{h}{2}\nabla_p H(x_{k+1/2}, p_{k+1/2}), \tag{8c}$$
$$p_{k+1} = e^{-\gamma h/2} p_{k+1/2}. \tag{8d}$$

This can be seen as a dissipative version of leapfrog.
Replacing the Hamiltonian (5) into (7) and (8) we can show the following.

Theorem 3 *Heavy ball is a 1st order conformal symplectic integrator.*

Theorem 4 *Nesterov's method is 1st order but not conformal symplectic. It contracts the symplectic form in a Hessian dependent manner:*

$$\omega_{k+1} = e^{-\gamma h}\left[I - \tfrac{h^2}{m}\nabla^2 f(x_k)\right]\omega_k + O(h^3). \tag{9}$$

Shadow Dynamical Systems

Which modified dynamical systems capture the behaviour of CM and NAG more closely?

Theorem 5 *CM is a 2nd order integrator to the following (perturbed) Hamiltonian system:*

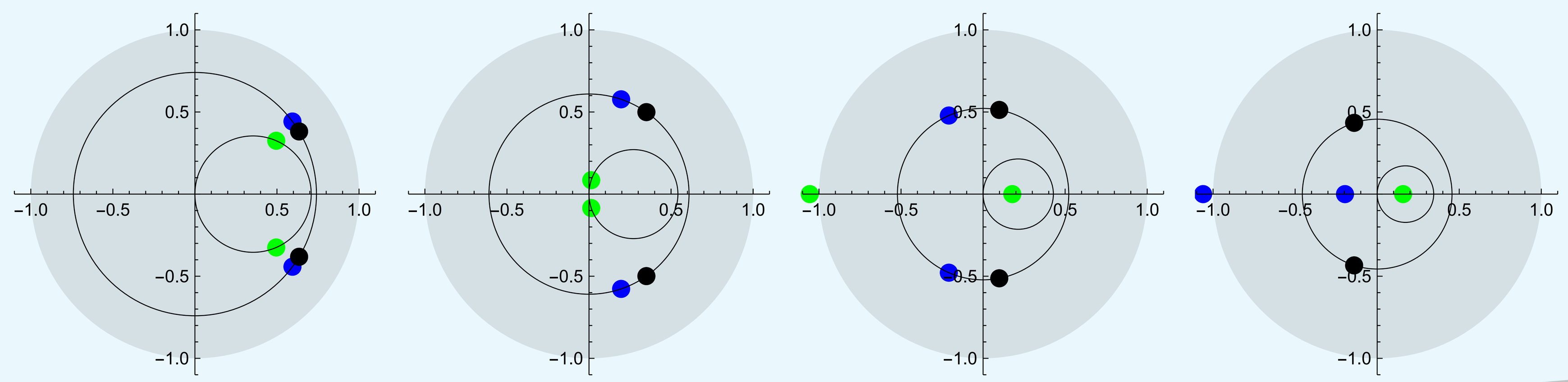
$$m\ddot{x} + m\gamma\dot{x} = -\left(I + \tfrac{h\gamma}{2}I - \tfrac{h^2\gamma^2}{4}I - \tfrac{h^2}{4m}\nabla^2 f(x)\right)\nabla f(x). \tag{10}$$

Theorem 6 *NAG is a 2nd order integrator to the following (non-Hamiltonian) system:*

$$m\ddot{x} + \left(m\gamma + h\nabla^2 f(x)\right)\dot{x} = -\left(I + \tfrac{h\gamma}{2}I - \tfrac{h^2\gamma^2}{4}I + \tfrac{h^2}{4m}\nabla^2 f(x)\right)\nabla f(x). \tag{11}$$

Trade-off Between Stability and Convergence Rate

A conformal symplectic method tends to be more stable (thus can use a larger step size), while the spurious dissipation of Nesterov introduce a slightly improved convergence rate. Below we have eigenvalues of the transition matrix; RGD (black), CM (blue), NAG (green).



Relativistic Optimization

Using $H = c\sqrt{\|p\|^2 + m^2c^2} + f(x)$ into (4) we obtain the dynamics of a relativistic particle under the influence of the potential $f(x)$ and dissipation:

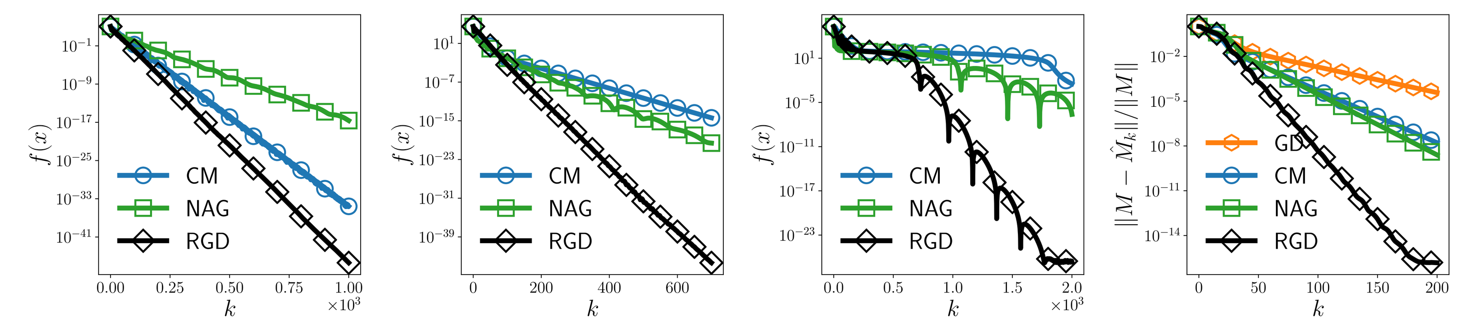
$$\dot{x} = \frac{cp}{\sqrt{\|p\|^2 + m^2c^2}}, \qquad \dot{p} = -\nabla f - \gamma p. \tag{12}$$

Using the general integrator (8), after some modifications, we obtain what we call *Relativistic Gradient Descent* (RGD):

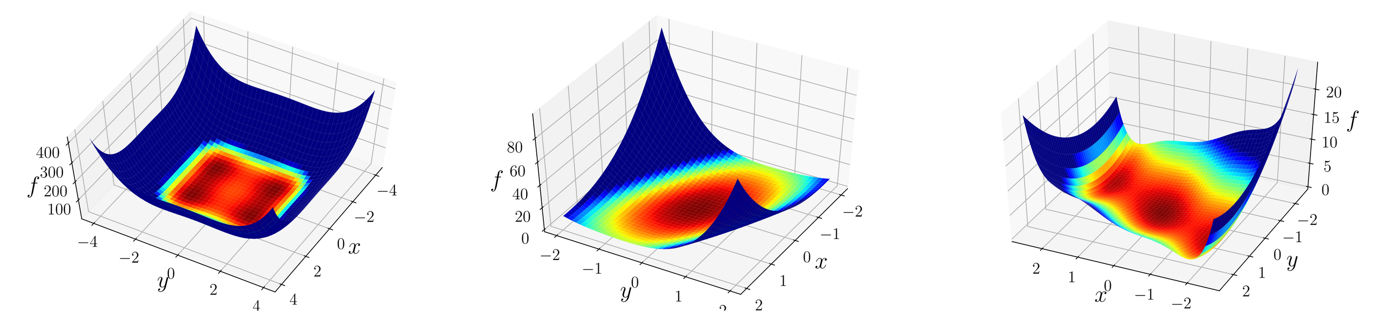
$$x_{k+1/2} = x_k + \sqrt{\mu}\left(\mu\delta\|v_k\|^2 + 1\right)^{-1/2} v_k, \tag{13a}$$
$$v_{k+1/2} = \sqrt{\mu}v_k - \epsilon\nabla f(x_{k+1/2}), \tag{13b}$$
$$x_{k+1} = \alpha x_{k+1/2} + (1-\alpha)x_k + \left(\delta\|v_{k+1/2}\|^2 + 1\right)^{-1/2} v_{k+1/2}, \tag{13c}$$
$$v_{k+1} = \sqrt{\mu}v_{k+1/2}. \tag{13d}$$

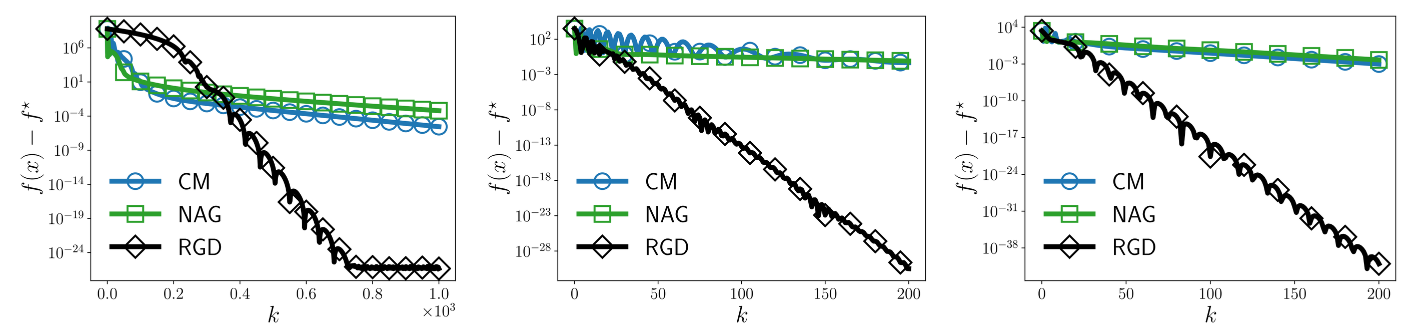
- RGD generalizes both CM ($\delta = 0$, $\alpha = 1$) and NAG ($\delta = 0$, $\alpha = 1$).
- RGD can be more stable since controls the kinetic energy.

Some Numerical Examples



correlated quadratic ($50D$) random quadratic ($500D$) Rosenbrock ($100D$) matrix completion ($100D$)





Qing Zakharov Camelback

References

[1] G. França, J. Sulam, D. P. Robinson, R. Vidal, *"Conformal symplectic and relativistic optimization,"* NeurIPS 2020, extended version at arXiv:1903.04100 [math.OC]

[2] G. França, M. I. Jordan, R. Vidal, *"On Dissipative Symplectic Integration with Applications to Gradient-Based Optimization,"* (2020) arXiv:2004.06840 [math.OC]

Acknowledgements: This work is supported by grants ARO MURI W911NF-17-1-0304, NSF 2031985, and NSF 1934931.

tZposter