

Direct Runge-Kutta Discretization Achieves Acceleration

Jingzhao Zhang , Aryan Mokhtari, Suvrit Sra, Ali Jadbabaie

NeurIPS 2018

Acceleration in first order convex optimization

Optimize smooth convex function: $\min_{x \in \mathbb{R}^d} f(x)$

Acceleration in first order convex optimization

Optimize smooth convex function: $\min_{x \in \mathbb{R}^d} f(x)$

Gradient Descent:

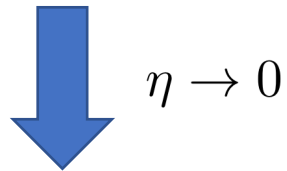
$$x_{k+1} = x_k - \eta \nabla f(x_k)$$

Acceleration in first order convex optimization

Optimize smooth convex function: $\min_{x \in \mathbb{R}^d} f(x)$

Gradient Descent:

$$x_{k+1} = x_k - \eta \nabla f(x_k)$$



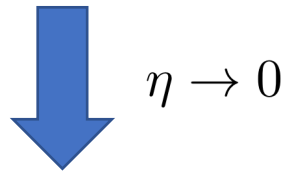
$$\dot{x} = -\nabla f(x)$$

Acceleration in first order convex optimization

Optimize smooth convex function: $\min_{x \in \mathbb{R}^d} f(x)$

Gradient Descent:

$$x_{k+1} = x_k - \eta \nabla f(x_k)$$



$$\dot{x} = -\nabla f(x)$$

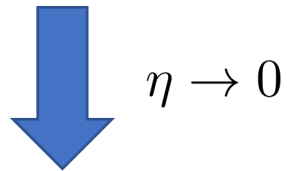
$$f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t}\right)$$

Acceleration in first order convex optimization

Optimize smooth convex function: $\min_{x \in \mathbb{R}^d} f(x)$

Gradient Descent:

$$x_{k+1} = x_k - \eta \nabla f(x_k)$$



$$\dot{x} = -\nabla f(x)$$

$$f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t}\right)$$

Accelerated Gradient Descent [Nesterov 1983]:

$$x_{k+1} = y_k - \eta \nabla f(y_k)$$

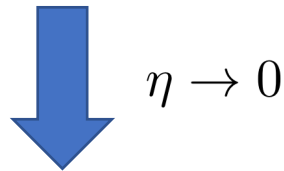
$$y_{k+1} = x_{k+1} + \beta(x_{k+1} - x_k)$$

Acceleration in first order convex optimization

Optimize smooth convex function: $\min_{x \in \mathbb{R}^d} f(x)$

Gradient Descent:

$$x_{k+1} = x_k - \eta \nabla f(x_k)$$



$$\dot{x} = -\nabla f(x)$$

$$f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t}\right)$$

Accelerated Gradient Descent [Nesterov 1983]:

$$x_{k+1} = y_k - \eta \nabla f(y_k)$$

$$y_{k+1} = x_{k+1} + \beta(x_{k+1} - x_k)$$



$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

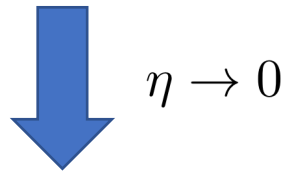
[SBC 2015] Su, Weijie, Stephen Boyd, and Emmanuel Candes. "A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights." *Advances in Neural Information Processing Systems*. 2014.

Acceleration in first order convex optimization

Optimize smooth convex function: $\min_{x \in \mathbb{R}^d} f(x)$

Gradient Descent:

$$x_{k+1} = x_k - \eta \nabla f(x_k)$$



$$\dot{x} = -\nabla f(x)$$

$$f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t}\right)$$

Accelerated Gradient Descent [Nesterov 1983]:

$$x_{k+1} = y_k - \eta \nabla f(y_k)$$

$$y_{k+1} = x_{k+1} + \beta(x_{k+1} - x_k)$$



$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0$$

$$f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t^2}\right)$$

[SBC 2015] Su, Weijie, Stephen Boyd, and Emmanuel Candes. "A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights." *Advances in Neural Information Processing Systems*. 2014.

Convergence in continuous time

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0 \quad f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t^2}\right)$$

Convergence in continuous time

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0 \quad f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t^2}\right)$$

$t \rightarrow t^{p/2}$
Arbitrary acceleration
by change of variable



[WWJ 2016]

$$\ddot{x} + \frac{2p+1}{t}\dot{x} + Cp^2t^{p-2}\nabla f(x) = 0 \quad f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t^p}\right)$$

[WWJ 2016] Wibisono, A., Wilson, A. C., & Jordan, M. I. (2016). A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47), E7351-E7358.

Convergence in continuous time

$$\ddot{x} + \frac{3}{t}\dot{x} + \nabla f(x) = 0 \quad f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t^2}\right)$$

$t \rightarrow t^{p/2}$
Arbitrary acceleration
by change of variable



[WWJ 2016]

$$\ddot{x} + \frac{2p+1}{t}\dot{x} + Cp^2 \boxed{t^{p-2}} \nabla f(x) = 0 \quad f(x(t)) - f(x^*) = \mathcal{O}\left(\frac{1}{t^p}\right)$$

However, smooth convex optimization algorithms
cannot achieve faster rate than: $\mathcal{O}\left(\frac{1}{t^2}\right)$

[WWJ 2016] Wibisono, A., Wilson, A. C., & Jordan, M. I. (2016). A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47), E7351-E7358.

Question: How to relate the convergence rate in continuous time ODE to the convergence rate of a discrete optimization algorithm?

Question: How to relate the **convergence rate** in continuous time **ODE** to the convergence rate of **a discrete optimization algorithm**?

Our approach: Discretize the ODE with **known Runge-Kutta integrators** (e.g. Euler, midpoint, RK44) and provide **theoretical guarantees** for convergence rates.

Main theorem:

For a p -flat, $(s+2)$ -differentiable convex function, if we discretize the ODE with order- s Runge-Kutta integrator, we have

$$f(x(t)) - f(x^*) = \mathcal{O}\left(t^{-\frac{ps}{s+1}}\right)$$

Main theorem:

For a p -flat, $(s+2)$ -differentiable convex function, if we discretize the ODE with order- s Runge-Kutta integrator, we have

$$f(x(t)) - f(x^*) = \mathcal{O}(t^{-\frac{ps}{s+1}})$$

p -flat:

$p = 2$: Gradient is Lipschitz continuous.

$p = 4$: $\|x\|_4^4$

$p = N$: $\log(e^{-x})$

Main theorem:

For a p -flat, $(s+2)$ -differentiable convex function, if we discretize the ODE with order- s Runge-Kutta integrator, we have

$$f(x(t)) - f(x^*) = \mathcal{O}(t^{-\frac{ps}{s+1}})$$

p -flat:

$p = 2$: Gradient is Lipschitz continuous.

$p = 4$: $\|x\|_4^4$

$p = N$: $\log(e^{-x})$

Order- s : Discretization error scales as $\mathcal{O}(h^{s+1})$, h is the step size .

Main theorem:

For a p -flat, $(s+2)$ -differentiable convex function, if we discretize the ODE with order- s Runge-Kutta integrator, we have

$$f(x(t)) - f(x^*) = \mathcal{O}(t^{-\frac{ps}{s+1}})$$

p -flat:

$p = 2$: Gradient is Lipschitz continuous.

$p = 4$: $\|x\|_4^4$

$p = N$: $\log(e^{-x})$

Objective	Integrator	Rate
L-smooth ($p=2$)	RK44 ($s=4$)	$\mathcal{O}(t^{-8/5})$
$\ x\ _4^4$ ($p=4$)	Midpoint($s=2$)	$\mathcal{O}(t^{-8/3})$

Order- s : Discretization error scales as $\mathcal{O}(h^{s+1})$; h is the step size.

Our poster session:

Thu Dec 6th 05:00 -- 07:00 PM

Room 210 & 230 AB

Poster Number: 9