Accelerated Gradient Methods for Geodesically Convex Optimization

Tractable Algorithms and Convergence Analysis

Jungbin Kim **Insoon Yang**

Department of Electrical and Computer Engineering
Seoul National University

ICML 2022

Convex optimization & Nesterov acceleration

For **convex optimization problems**

$$\min_{x \in \mathbb{R}^n} f(x),$$

Nesterov's accelerated gradient (NAG) is one of the fastest first-order method.

Table: Iteration complexities.[1]

	convex	μ-strongly convex
NAG	$O\left(\sqrt{\frac{L}{\epsilon}}\right)$	$O\left(\sqrt{\frac{L}{\mu}} \log \frac{L}{\epsilon}\right)$

[1]The required number of iterations to obtain an ϵ-approximate solution

Convex optimization on Riemannian manifolds

Geodesically convex (g-convex) optimization problem:

$$\min_{x \in N \subseteq M} f(x),$$

where

- M: Riemannian manifold (e.g., $\mathbb{R}^n$, sphere, hyperbolic space)
- N: geodesically convex set
- f: geodesically convex function.

Q) **Is there an algorithm that converges as fast as NAG?**

Acceleration on Riemannian manifolds?

Q) Is there an algorithm that converges as fast as NAG?

Some recent papers...

- (Liu et al., 2017): not computationally tractable.
- (Zhang & Sra, 2018): locally accelerated.
- (Ahn & Sra, 2020): globally, but eventually accelerated.
- (Alimisis et al., 2021): accelerated only in early stages.
- (Martínez-Rubio, 2022): fully accelerated, but they only consider the manifolds with constant sectional curvature.

Open problem: to achieve full acceleration in general

Our contributions

Q) Is there an algorithm that converges as fast as NAG?

- We propose Riemannian NAG (**RNAG**): new Riemannian optimization algorithm.
- Our algorithm **always achieves full acceleration.**

Standard assumptions

- The sectional curvature is bounded by $K_{\min}$ and $K_{\max}$.
- The diameter $\operatorname{diam}(N)$ of the domain is bounded above by D.

Note: these assumptions are common in the literature.

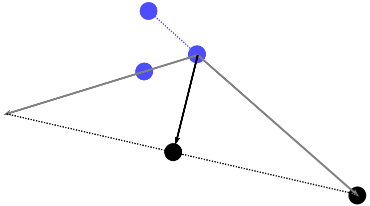
Proposed algorithm: RNAG

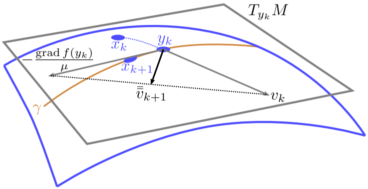
Original NAG

$$
\begin{aligned}
y_k &= x_k + \tau_k \bar{v}_k \\
x_{k+1} &= y_k + \left(-\alpha_k \operatorname{grad} f\left(y_k\right)\right) \\
v_k &= \bar{v}_k - \left(y_k - x_k\right) \\
\bar{\bar{v}}_{k+1} &= \beta_k v_k - \gamma_k \operatorname{grad} f\left(y_k\right) \\
\bar{v}_{k+1} &= \bar{\bar{v}}_{k+1} - \left(x_{k+1} - y_k\right)
\end{aligned}
$$

RNAG (ours)

$$
\begin{aligned}
y_k &= \exp_{x_k}\left(\tau_k \bar{v}_k\right) \\
x_{k+1} &= \exp_{y_k}\left(-\alpha_k \operatorname{grad} f\left(y_k\right)\right) \\
v_k &= \Gamma_{x_k}^{y_k}\left(\bar{v}_k - \log_{x_k}\left(y_k\right)\right) \\
\bar{\bar{v}}_{k+1} &= \beta_k v_k - \gamma_k \operatorname{grad} f\left(y_k\right) \\
\bar{v}_{k+1} &= \Gamma_{y_k}^{x_{k+1}}\left(\bar{\bar{v}}_{k+1} - \log_{y_k}\left(x_{k+1}\right)\right)
\end{aligned}
$$





From NAG to RNAG

Original NAG

$$
\begin{aligned}
y_k &= x_k + \tau_k \bar{v}_k \\
x_{k+1} &= y_k + \left(-\alpha_k \operatorname{grad} f\left(y_k\right)\right) \\
v_k &= \bar{v}_k - \left(y_k - x_k\right) \\
\bar{\bar{v}}_{k+1} &= \beta_k v_k - \gamma_k \operatorname{grad} f\left(y_k\right) \\
\bar{v}_{k+1} &= \bar{\bar{v}}_{k+1} - \left(x_{k+1} - y_k\right)
\end{aligned}
$$

RNAG (ours)

$$
\begin{aligned}
y_k &= \exp_{x_k}\left(\tau_k \bar{v}_k\right) \\
x_{k+1} &= \exp_{y_k}\left(-\alpha_k \operatorname{grad} f\left(y_k\right)\right) \\
v_k &= \Gamma_{x_k}^{y_k}\left(\bar{v}_k - \log_{x_k}\left(y_k\right)\right) \\
\bar{\bar{v}}_{k+1} &= \beta_k v_k - \gamma_k \operatorname{grad} f\left(y_k\right) \\
\bar{v}_{k+1} &= \Gamma_{y_k}^{x_{k+1}}\left(\bar{\bar{v}}_{k+1} - \log_{y_k}\left(x_{k+1}\right)\right)
\end{aligned}
$$

Modifications:

- Addition & Subtraction $\rightarrow$ Exponential map & Logarithm map
- Parallel transport

From NAG to RNAG

Original NAG	**RNAG (ours)**
$y_k = x_k + \tau_k \bar{v}_k$	$y_k = \exp_{x_k}\left(\tau_k \bar{v}_k\right)$
$x_{k+1} = y_k + \left(-\alpha_k \operatorname{grad} f\left(y_k\right)\right)$	$x_{k+1} = \exp_{y_k}\left(-\alpha_k \operatorname{grad} f\left(y_k\right)\right)$
$v_k = \bar{v}_k - \left(y_k - x_k\right)$	$v_k = \Gamma_{x_k}^{y_k}\left(\bar{v}_k - \log_{x_k}\left(y_k\right)\right)$
$\bar{\bar{v}}_{k+1} = \beta_k v_k - \gamma_k \operatorname{grad} f\left(y_k\right)$	$\bar{\bar{v}}_{k+1} = \beta_k v_k - \gamma_k \operatorname{grad} f\left(y_k\right)$
$\bar{v}_{k+1} = \bar{\bar{v}}_{k+1} - \left(x_{k+1} - y_k\right)$	$\bar{v}_{k+1} = \Gamma_{y_k}^{x_{k+1}}\left(\bar{\bar{v}}_{k+1} - \log_{y_k}\left(x_{k+1}\right)\right)$

Modifications:

- Addition & Subtraction $\rightarrow$ Exponential map & Logarithm map
- Parallel transport

Main results

Theorem (Convergence of RNAG, g-convex case)

When f is geodesically convex and L-smooth, RNAG (with some parameters) finds an ϵ-approximate solution in $O\left(\xi\sqrt{\frac{L}{\epsilon}}\right)$ iterations, where ξ is a constant depending on the bounds $K_{\min}$, $K_{\max}$, and D.

Table: Iteration complexities.

	convex	μ-strongly convex
NAG	$O\left(\sqrt{\frac{L}{\epsilon}}\right)$	$O\left(\sqrt{\frac{L}{\mu}}\log\frac{L}{\epsilon}\right)$
RNAG (ours)	$O\left(\xi\sqrt{\frac{L}{\epsilon}}\right)$	

Main results

Theorem (Convergence of RNAG, g-strongly convex case)

When f is geodesically μ-strongly convex and L-smooth, RNAG (with some parameters) finds an ϵ-approximate solution in $O\left(\xi\sqrt{\frac{L}{\mu}}\log\frac{L}{\epsilon}\right)$ iterations, where ξ is a constant depending on the bounds $K_{\min}$, $K_{\max}$, and D.

Table: Iteration complexities.

	convex	μ-strongly convex
NAG	$O\left(\sqrt{\frac{L}{\epsilon}}\right)$	$O\left(\sqrt{\frac{L}{\mu}}\log\frac{L}{\epsilon}\right)$
RNAG (ours)	$O\left(\xi\sqrt{\frac{L}{\epsilon}}\right)$	$O\left(\xi\sqrt{\frac{L}{\mu}}\log\frac{L}{\epsilon}\right)$

Ideas of proof

Q) Follows original proof.. What modificaitons do we need?

- Parameters: depending on $K_{\min}$, $K_{\max}$, D

$$
\begin{aligned}
y_k &= \exp_{x_k}\left(\boldsymbol{\tau_k}\bar{v}_k\right)\\
x_{k+1} &= \exp_{y_k}\left(-\boldsymbol{\alpha_k}\operatorname{grad} f\left(y_k\right)\right)\\
v_k &= \Gamma_{x_k}^{y_k}\left(\bar{v}_k - \log_{x_k}\left(y_k\right)\right)\\
\bar{\bar{v}}_{k+1} &= \boldsymbol{\beta_k} v_k - \boldsymbol{\gamma_k}\operatorname{grad} f\left(y_k\right)\\
\bar{v}_{k+1} &= \Gamma_{y_k}^{x_{k+1}}\left(\bar{\bar{v}}_{k+1} - \log_{y_k}\left(x_{k+1}\right)\right)
\end{aligned}
$$

- Potential function

$$
\phi_k = A_k\left(f\left(x_k\right) - f\left(x^*\right)\right) + B_k \underbrace{\left(\left\|\bar{v}_k - \log_{x_k}\left(x^*\right)\right\|_{x_k}^2 + \boldsymbol{(\xi - 1)\left\|\bar{v}_k\right\|_{x_k}^2}\right)}_{\text{Squared distances in } T_{x_k}M}
$$

- Novel *metric distortion* lemma

Continuous-time interpretation

Taking the step size $s \to 0$,

- RNAG-C converges to the following ODE:

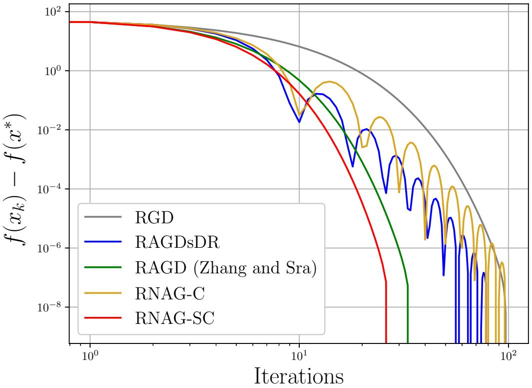
$$\nabla \dot{X} + \frac{1+2\xi}{t} \dot{X} + \operatorname{grad} f(X) = 0.$$

- RNAG-SC converges to the following ODE:

$$\nabla \dot{X} + \left(\frac{1}{\sqrt{\xi}} + \sqrt{\xi} \right) \sqrt{\mu} \dot{X} + \operatorname{grad} f(X) = 0.$$

- These ODEs correspond to those in (Alimisis et al., 2020) for modeling Riemannian acceleration.
- This analysis confirms the accelerated convergence of our algorithms through the lens of continuous-time flows.

Numerical experiment: Karcher mean of SPD matrices



- RAGDsDR: known accelerated method for g-convex functions.
- RAGD: known accelerated method for strongly g-convex functions.

(Note: full acceleration is not guaranteed for RAGDsDR and RAGD)

Concluding remark

Contributions

We proposed RNAG, **the first Riemannian optimization algorithm achieving full acceleration**.

Open questions

Effect of geometry (e.g., sectional curvature) on lower complexity bounds.

	convex	μ-strongly convex
NAG	$O\left(\sqrt{\frac{L}{\epsilon}}\right)$	$O\left(\sqrt{\frac{L}{\mu}}\log\frac{L}{\epsilon}\right)$
RNAG (ours)	$O\left(\xi\sqrt{\frac{L}{\epsilon}}\right)$	$O\left(\xi\sqrt{\frac{L}{\mu}}\log\frac{L}{\epsilon}\right)$

Acknowledgement: Supported in part by Samsung Electronics, National Research Foundation of Korea (MSIT 2020R1C1C1009766), Information and Communications Technology Planning and Evaluation (MSIT 2022-0-00124, 2022-0-00480).