A robust accelerated optimization algorithm for strongly convex functions

Saman Cyrus Bin Hu Bryan Van Scoy Laurent Lessard

University of Wisconsin–Madison

June 27, 2018

Iterative algorithms

Unconstrained optimization with f strongly convex.

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

$mI \preceq \nabla^2 f(x) \preceq LI$, and define $\kappa = \frac{L}{m}$ (condition number)

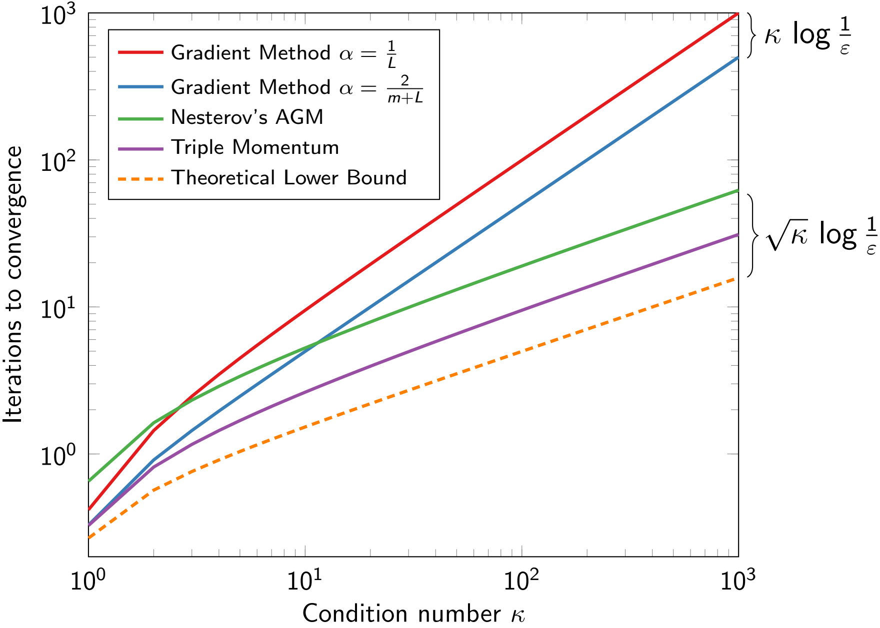
- **Gradient method:**

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

- **Nesterov's accelerated gradient method (AGM)**

$$\begin{aligned} y_k &= x_k + \beta(x_k - x_{k-1}) \\ x_{k+1} &= y_k - \alpha \nabla f(y_k) \end{aligned}$$

Iteration complexity



Noise robustness

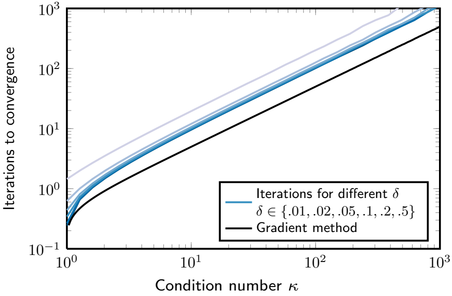
- ∇f is inexact/approximate.
- Many different noise models have been studied.
- We use relative deterministic noise:

$$\frac{\|\nabla f_{\text{noisy}} - \nabla f_{\text{exact}}\|}{\|\nabla f_{\text{exact}}\|} \leq \delta$$

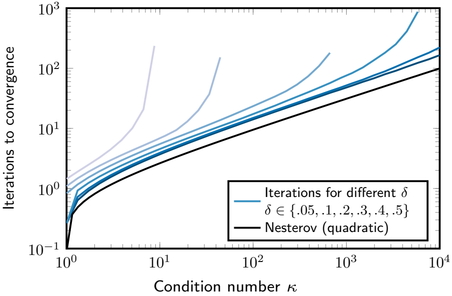
- Ex: round-off error.

Performance in the presence of noise

Gradient method: **slow and robust**



Nesterov's AGM: **fast and fragile**



Simulated by solving a small SDP
(Lessard et al., SIAM Journal on Opt., 2016)

Proposed algorithm: Robust momentum method

Iteration update:

$$\begin{aligned} x_{k+1} &= x_k + \beta(x_k - x_{k-1}) - \alpha \nabla f(y_k) \\ y_k &= x_k + \gamma(x_k - x_{k-1}) \end{aligned}$$

with parameters:

$$\alpha = \frac{\kappa(1-\rho)^2(1+\rho)}{L}, \quad \beta = \frac{\kappa\rho^3}{\kappa - 1}, \quad \gamma = \frac{\rho^3}{(\kappa-1)(1-\rho)^2(1+\rho)}$$

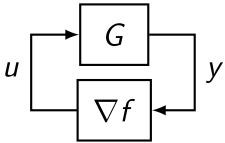
single tuning parameter ρ:

$$\underbrace{1 - \tfrac{1}{\sqrt{\kappa}}}_{\text{fast + fragile}} \quad \le \quad \rho \quad \le \quad \underbrace{1 - \tfrac{1}{\kappa}}_{\text{slow + robust}}$$

Control interpretation

Nesterov's AGM:

$$
\begin{aligned}
x_{k+1} &= y_k - \alpha \nabla f(y_k) \\
y_k &= x_k + \beta(x_k - x_{k-1})
\end{aligned}
$$



$$
G = \begin{cases}
\begin{bmatrix} x_{k+1} \\ x_k \end{bmatrix} = \begin{bmatrix} 1+\beta & -\beta \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix} + \begin{bmatrix} -\alpha \\ 0 \end{bmatrix} u_k \\
y_k = \begin{bmatrix} 1+\beta & -\beta \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix}
\end{cases}
$$

Frequency domain condition

Transfer function for the linear part:

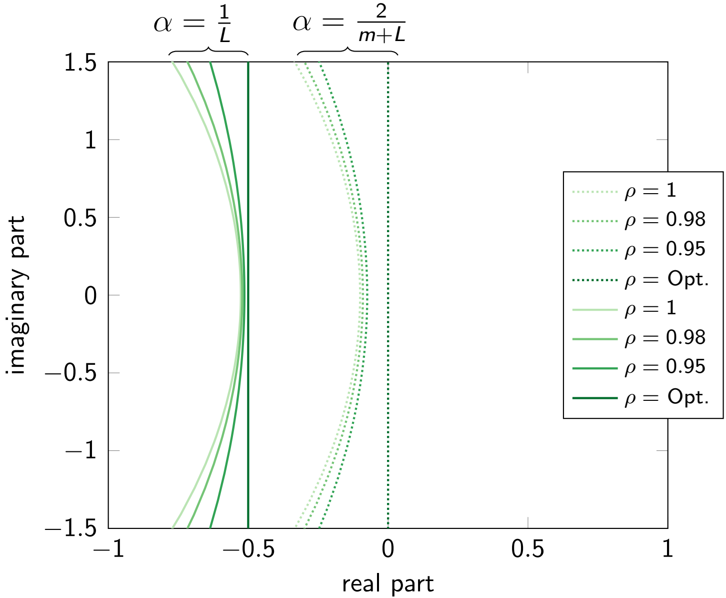
$$G(z) = -\alpha \frac{(1+\gamma)z - \gamma}{(z-1)(z-\beta)}$$

Sufficient condition for linear convergence: $x_k - x_\star = O(\rho^k)$

$$\mathrm{Re}\left[(\rho z^{-1} - 1)\,\frac{1 - LG(\rho z)}{1 - mG(\rho z)}\right] < 0 \qquad \text{for all } |z| = 1$$

- combination of Circle Criterion and Zames-Falb multipliers

Gradient method

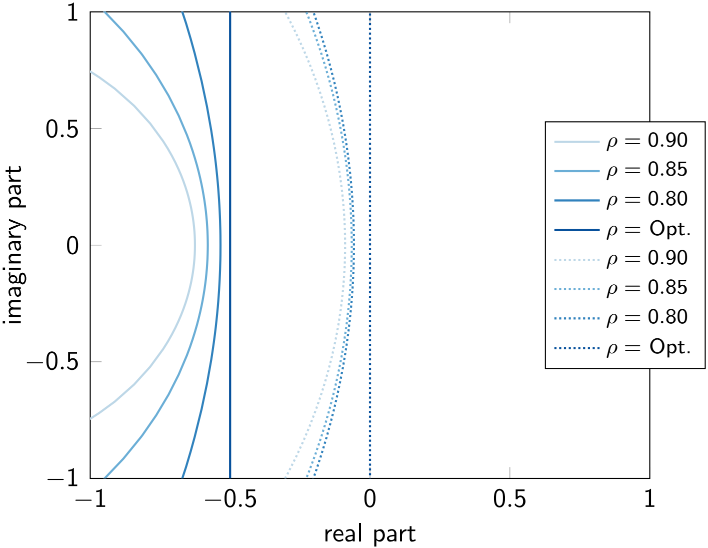


Idea: add a margin to improve robustness

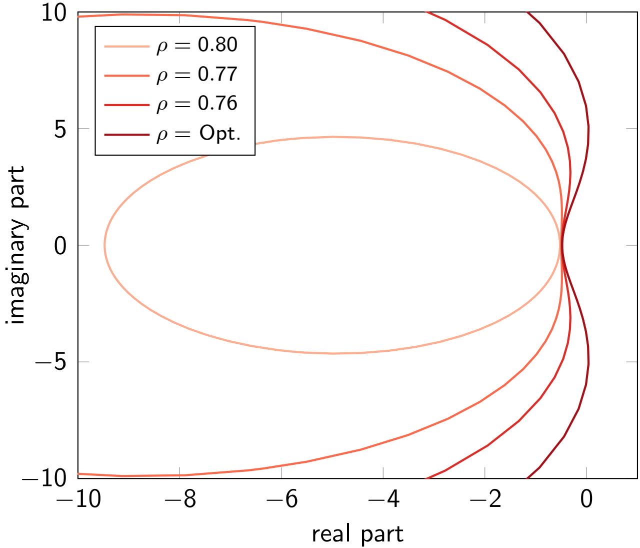
$$\mathrm{Re}\left[(\rho z^{-1} - 1)\,\frac{1 - LG(\rho z)}{1 - mG(\rho z)}\right] + \nu \leq 0 \qquad \text{for all } |z| = 1$$

- $\nu > 0$ is a robustness margin
- tune α, β, γ to obtain a vertical line at $-\nu$.

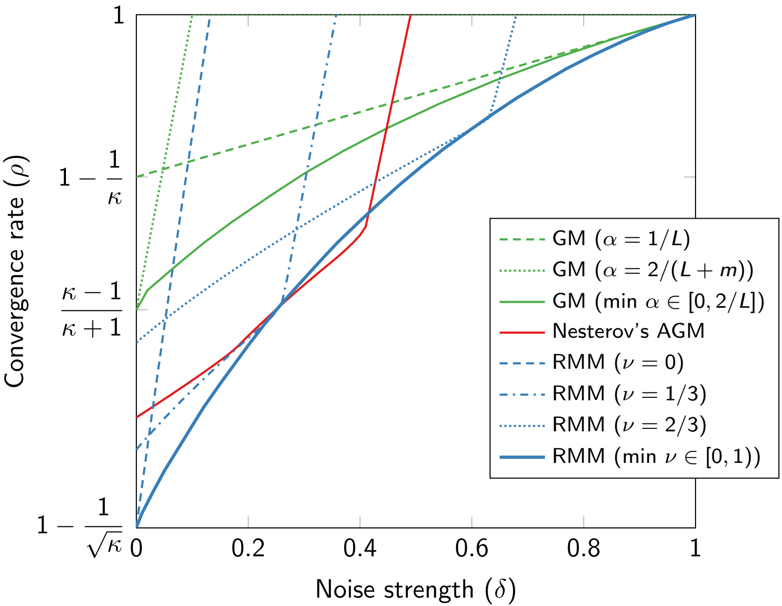
Robust Momentum Method



Nesterov's AGM — not a line!

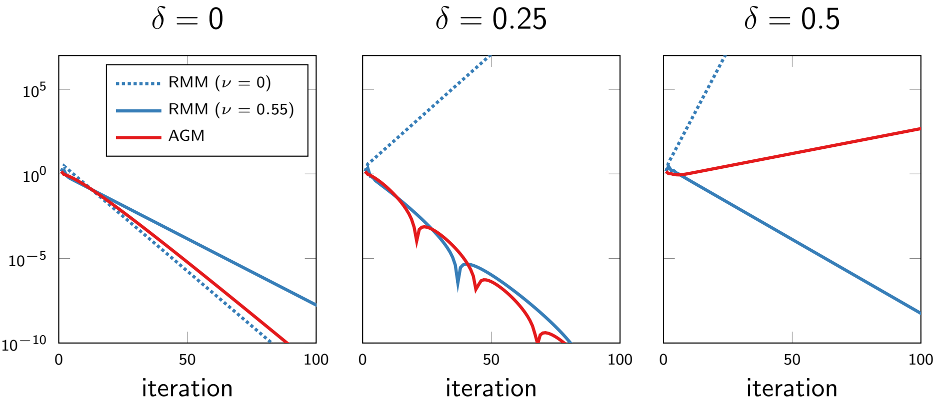


Trade-off: Performance vs noise



Simulation with relative gradient noise

$$f(x) = x_1^2 + 10x_2^2$$



Thank you