Lyapunov Functions for First-Order Methods

Tight Automated Convergence Guarantees

Adrien Taylor **Bryan Van Scoy** Laurent Lessard

35th International Conference on Machine Learning
Stockholm, Sweden
July 11, 2018

We consider *first-order* methods to solve the following problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbb{R}^d \end{array}$$

This work

Automatic analysis of optimization methods by solving a **small-sized semidefinite program**.
Convergence rate is **provably tight** when

- f is L-smooth and μ-strongly convex (denoted $f \in \mathcal{F}_{\mu,L}(\mathbb{R}^d)$), and
- the method is *iterative* with *fixed stepsizes*.

Combines:

- performance estimation problems (Drori & Teboulle, 2014)
- integral quadratic constraints (Lessard, Recht, Packard, 2016)

Uses smooth strongly convex interpolation (Taylor, Hendrickx, Glineur, 2017)

First-order iterative fixed-step method

$$y_k = \sum_{j=0}^{N} \gamma_j \, x_{k-j}$$

$$x_{k+1} = \sum_{j=0}^{N} \beta_j \, x_{k-j} - \alpha \, \nabla f(y_k)$$

degree N

stepsizes α, β_j, γ_j

initial conditions $x_j \in \mathbb{R}^d$ for $j = -N, \dots, 0$

Extensions:

- linesearch (or subspace search)
- restart (fixed/adaptive)

First-order iterative fixed-step method

$$y_k = \sum_{j=0}^{N} \gamma_j \, x_{k-j}$$

$$x_{k+1} = \sum_{j=0}^{N} \beta_j \, x_{k-j} - \alpha \, \nabla f(y_k)$$

Main result

$\mathrm{SDP}(\mu, L, \rho, \alpha, \beta_j, \gamma_j)$ is feasible

$$\Downarrow$$

the method converges linearly for all $f \in \mathcal{F}_{\mu,L}(\mathbb{R}^d)$ with rate at least ρ

First-order iterative fixed-step method

$$y_k = \sum_{j=0}^{N} \gamma_j \, x_{k-j}$$

$$x_{k+1} = \sum_{j=0}^{N} \beta_j \, x_{k-j} - \alpha \, \nabla f(y_k)$$

Main result

SDP$(\mu, L, \rho, \alpha, \beta_j, \gamma_j)$ is feasible

$$\Updownarrow$$

there exists a quadratic Lyapunov function

$$\Downarrow$$

the method converges linearly for all $f \in \mathcal{F}_{\mu,L}(\mathbb{R}^d)$ with rate at least ρ

Lyapunov function

Fundamental tool from control theory that can be used to verify stability of a dynamical system (Kalman & Bertram, 1960).

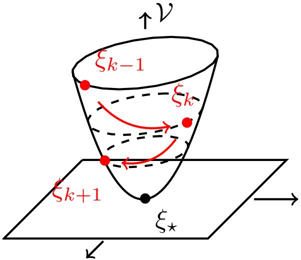
Notation	Control	Optimization
ξ	state	iterate, function, & gradient values
$\xi_\star$	fixed-point	optimal solution
$\mathcal{V}(\xi)$	energy in system	distance from optimality

nonnegative $\mathcal{V}(\xi) \geq 0$ for all ξ

zero at fixed-point $\mathcal{V}(\xi) = 0$ if and only if $\xi = \xi_\star$

radially unbounded $\mathcal{V}(\xi) \to \infty$ as $\|\xi\| \to \infty$

decreasing $\mathcal{V}(\xi_{k+1}) \leq \rho^2\, \mathcal{V}(\xi_k)$ for all k



Existence of a Lyapunov function provides a certificate of convergence.

Lyapunov function

Fundamental tool from control theory that can be used to verify stability of a dynamical system (Kalman & Bertram, 1960).

Notation	Control	Optimization
ξ	state	iterate, function, & gradient values
$\xi_\star$	fixed-point	optimal solution
$\mathcal{V}(\xi)$	energy in system	distance from optimality

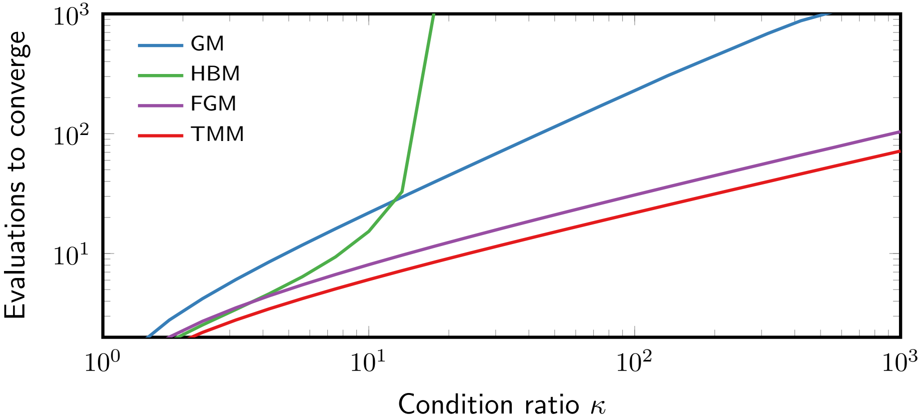
Suppose we choose the state as $\xi_k = \{(x_i, f(y_i), \nabla f(y_i)\}_{i=k-N}^{N}$ where x_k are the iterates. We then use the quadratic Lyapunov function candidate:

$$\mathcal{V}(\xi_k) = \begin{bmatrix} x_k - x_\star \\ \vdots \\ x_{k-N} - x_\star \\ \nabla f(y_k) \\ \vdots \\ \nabla f(y_{k-N}) \end{bmatrix}^\mathsf{T} (P \otimes I_d) \begin{bmatrix} x_k - x_\star \\ \vdots \\ x_{k-N} - x_\star \\ \nabla f(y_k) \\ \vdots \\ \nabla f(y_{k-N}) \end{bmatrix} + p^\mathsf{T} \begin{bmatrix} f(y_k) - f(x_\star) \\ \vdots \\ f(y_{k-N}) - f(x_\star) \end{bmatrix}$$

Numerical results

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

Method	α	β	γ
GM	$\frac{1}{L}$	0	0
HBM	$\frac{4}{(\sqrt{L}+\sqrt{\mu})^2}$	$\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^2$	0
FGM	$\frac{1}{L}$	$\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$	$\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$
TMM	$\frac{2\sqrt{L}-\sqrt{\mu}}{L\sqrt{L}}$	$\frac{(\sqrt{\kappa}-1)^2}{\kappa+\sqrt{\kappa}}$	$\frac{(\sqrt{\kappa}-1)^2}{2\kappa+\sqrt{\kappa}-1}$

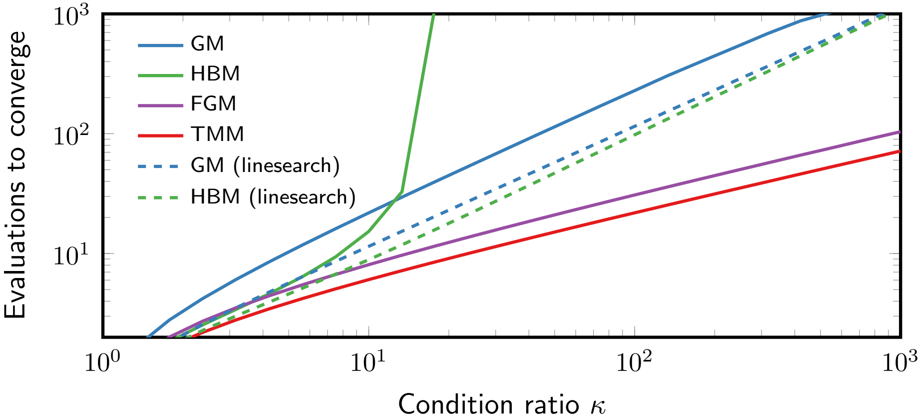


GM with linesearch

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

HBM with linesearch

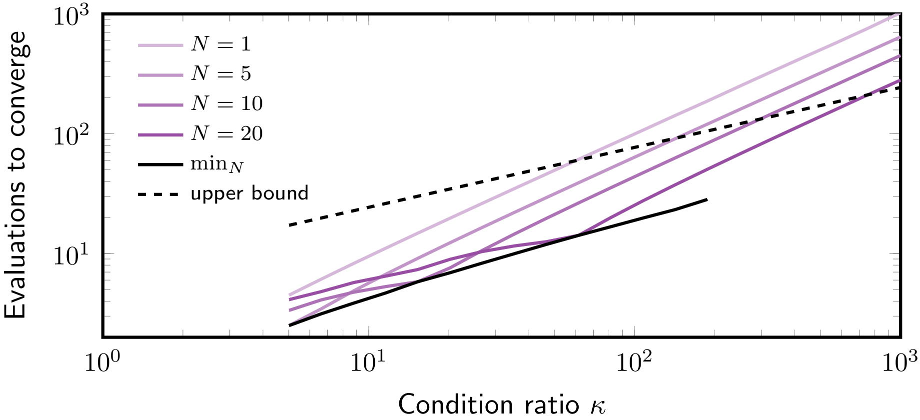
$$(\alpha, \beta) = \arg\min_{\alpha,\beta} f(x_k + \beta\,(x_k - x_{k-1}) - \alpha\, \nabla f(x_k))$$
$$x_{k+1} = x_k + \beta\,(x_k - x_{k-1}) - \alpha\, \nabla f(x_k)$$



FGM (Nesterov, 1983) with restarts every N iterations

$$x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$$
$$y_{k+1} = x_{k+1} + \beta_{k+1}(x_{k+1} - x_k)$$

Upper bound (O'Donoghue and Candès, 2015)
$$\rho(N_\star) \le \exp\left(\frac{-1}{e\sqrt{8\kappa}}\right)$$



Summary

Automatic analysis of optimization methods by solving a **small-sized semidefinite program**. Convergence rate is **provably tight** when

- f is L-smooth and μ-strongly convex, and
- the method is *iterative* with *fixed stepsizes*.

Same methodology may be used (possibly without tightness) to analyze:

- line/subspace search
- restart (fixed/adaptive)
- other function classes (besides smooth strongly convex)

Code

`https://github.com/QCGroup/quad-lyap-first-order`