# Synthesis of First-Order Convex Solvers

Govind Chari

University of Washington

Nov 2023

# Review Paper

- Authored by Dennis Gramlich, Christian Ebenbauer, Carsten W. Scherer
- Systems & Control Letters, 2022[1]
- Lyapunov-based synthesis of gradient-based algorithms for optimization and saddle-point problems

[1] Gramlich, Dennis et al. "Synthesis of accelerated gradient algorithms for optimization and saddle point problems using Lyapunov functions and LMIs." Systems & Control Letters, 2022.

# Blackbox Optimization

## Optimization Problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

## Gradient-based Algorithm

$$z_{k+1} = Az_k + B\nabla f(Cz_k) \quad \text{where } x_k = Cz_k$$

Gradient Descent:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{c|c} I_d & -\eta I_d \\ \hline I_d & 0_d \end{array}\right]$$

Nesterov's Accelerated Gradient:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{cc|c} (1+\beta)I_d & -\beta I_d & -\alpha I_d \\ I_d & 0_d & 0_d \\ \hline (1+\beta)I_d & -\beta I_d & 0_d \end{array}\right]$$

# Why Synthesis?

Synthesis provides a systematic way of generating optimal optimization algorithms

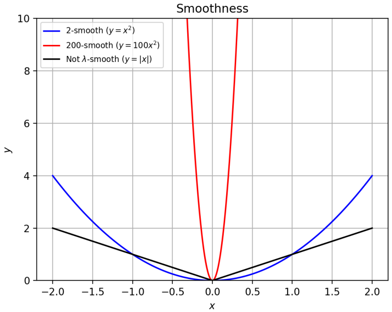
# Function Properties

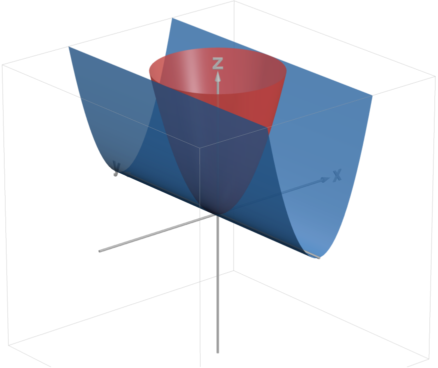
## $\lambda$-smoothness

$$\|\nabla f(x) - \nabla f(y)\| \leq \lambda \|x - y\|$$

## $\mu$-strong convexity

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|_2^2$$




$$\kappa = \frac{\lambda}{\mu}$$

# Lyapunov Analysis

Nonlinear Dynamical system:

$$x_{k+1} = g(x_k)$$
$$x^* = g(x^*)$$

Lyapunov Function Assumptions:

$$\alpha \|x - x^*\|_2^2 \leq V(x) \leq \beta \|x - x^*\|_2^2 \quad \forall x \in \mathbb{R}^n$$
$$V(x_{k+1}) - \rho^2 V(x_k) \leq 0 \quad \forall x \in \mathbb{R}^n$$

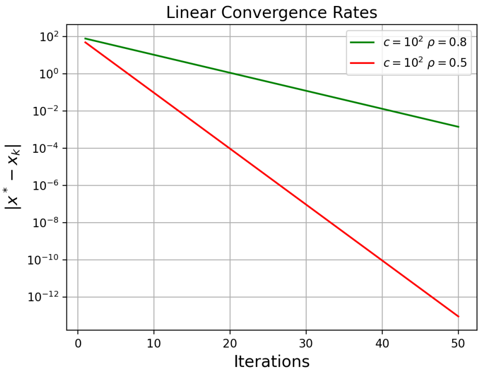
Global Exponential Stability:

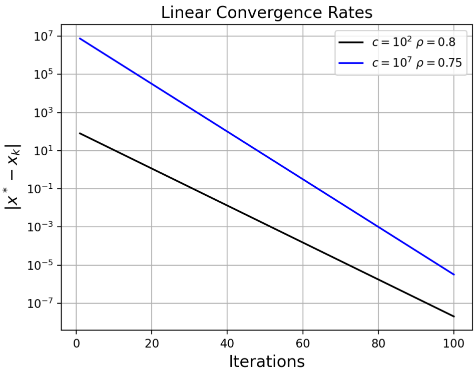
$$\|x^* - x_k\| \leq \sqrt{\frac{\beta}{\alpha}} \rho^k \|x^* - x_0\| \quad \forall x_0 \in \mathbb{R}^n$$

# Linear Convergence

We can upper bound the convergence of algorithms with linear convergence as follows

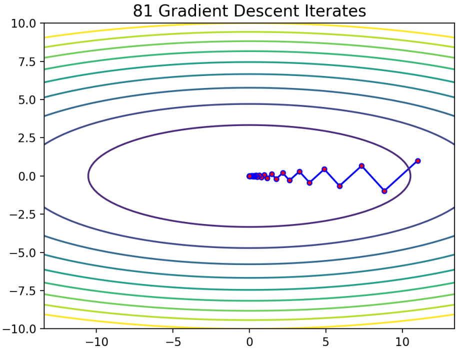
$$|x^* - x_k| \leq c\rho^k$$



# Drawbacks of Gradient Descent

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



# Accelerated Algorithms

Polyak's Heavy Ball:

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

Nesterov's Accelerated Gradient:

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

Triple Momentum:

$$\xi_{k+1} = (1+\beta)\xi_k - \beta \xi_{k-1} - \alpha \nabla f(y_k)$$
$$y_k = (1+\gamma)\xi_k - \gamma \xi_{k-1}$$
$$x_k = (1+\delta)\xi_k - \delta \xi_{k-1}$$

# Accelerated Algorithms: Rates $f(x_k) - f^*$

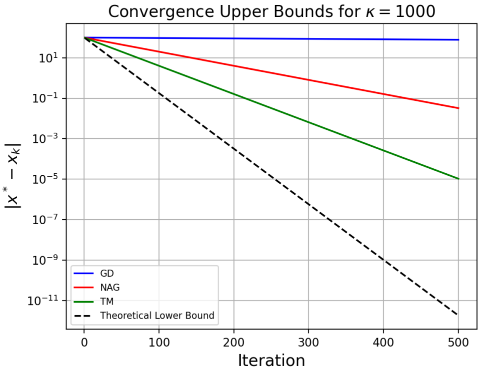
- Smooth $\rightarrow$ Sublinear Convergence
- Smooth $+$ Strongly Convex $\rightarrow$ Linear Convergence

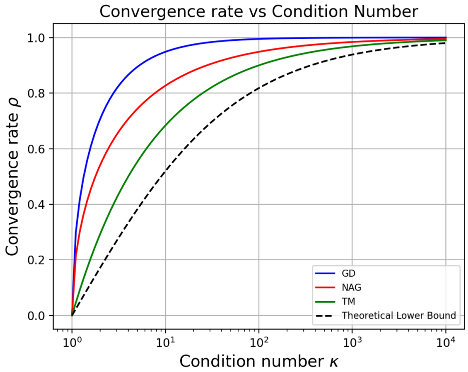
For smooth and strongly convex functions, the algorithms have linear convergence, and their rate to the *optimal objective*, $\rho$ is shown in the table.

In the table, $k$ is the iteration counter, and $\kappa$ is the condition number.

| Algorithm | Smooth | Smooth and Strongly Convex |
|---|---|---|
| Gradient Descent | $\mathcal{O}(1/k)$ | $1 - 1/\kappa$ |
| NAG | $\mathcal{O}(1/k^2)$ | $1 - 1/\sqrt{\kappa}$ |
| TM | - | $(1 - 1/\sqrt{\kappa})^2$ |

# Accelerated Algorithms

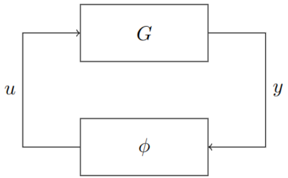


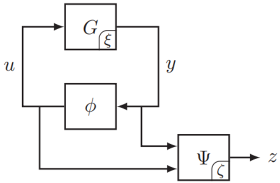



# Algorithm Analysis with IQCs[1]

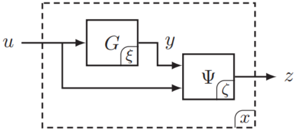
$$G : \begin{cases} \xi_{k+1} = A\xi_k + Bu_k \\ y_k = C\xi_k \end{cases}$$

$$u_k = \phi(y_k)$$



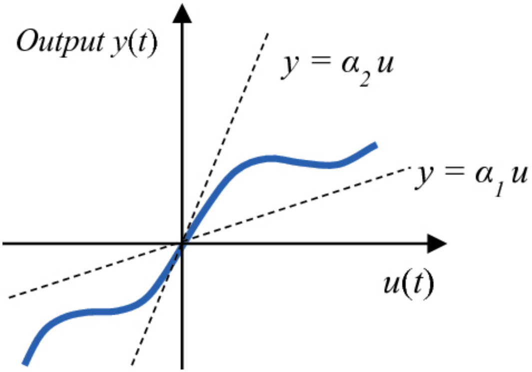

(a) The auxiliary system $\Psi$ produces $z$, which is a filtered version of the signals $y$ and $u$.

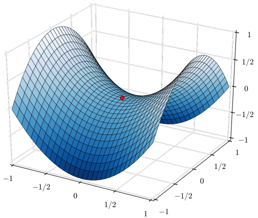



(b) The nonlinearity $\phi$ is replaced by a constraint on $z$, so we may remove $\phi$ entirely.

[1] Lessard, Laurant et al. "Analysis and design of optimization algorithms via integral quadratic constraints." SIAM Journal on Optimization, 2016.

# Sector Bounded Nonlinearity






Function Sector Conditions

$$\frac{1}{2}\|y-x\|_M^2 \leq f(y) - f(x) - \nabla f(x)^\top (y-x) \leq \frac{1}{2}\|y-x\|_L^2$$

Write Lyapunov Function

$$V(x) = \begin{bmatrix} x \\ \nabla f(Cx) \end{bmatrix}^\top \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} x \\ \nabla f(Cx) \end{bmatrix} + f(Cx) - f(0) - \frac{1}{2}\nabla f(Cx)^\top L \nabla f(Cx)$$

# Lyapunov Approach for Synthesis

Upper-bound Lyapunov Decrement

$$V_f(x^+) - \rho^2 V_f(x) \leq$$

$$\begin{pmatrix} x \\ w \\ \hline x^+ \\ w^+ \end{pmatrix}^\top \left(\begin{array}{cc|cc} -\rho^2 \boldsymbol{P}_{11} & -\rho^2 \boldsymbol{P}_{12} & 0 & 0 \\ -\rho^2 \boldsymbol{P}_{21} & -\rho^2 \boldsymbol{P}_{22} & 0 & 0 \\ \hline 0 & 0 & \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ 0 & 0 & \boldsymbol{P}_{21} & \boldsymbol{P}_{22} \end{array}\right) \begin{pmatrix} x \\ w \\ \hline x^+ \\ w^+ \end{pmatrix}$$

$$+ \begin{pmatrix} x \\ w \\ \hline x^+ \\ w^+ \end{pmatrix}^\top \left(\begin{array}{cc|cc} 0 & 0 & 0 & -\frac{\lambda}{2}\boldsymbol{C}^\top \\ 0 & 0 & 0 & \frac{\lambda}{2}\widetilde{\boldsymbol{L}}^\dagger \\ \hline 0 & 0 & 0 & \frac{1}{2}\boldsymbol{C}^\top \\ -\frac{\lambda}{2}\boldsymbol{C} & \frac{\lambda}{2}\widetilde{\boldsymbol{L}}^\dagger & \frac{1}{2}\boldsymbol{C} & -\widetilde{\boldsymbol{L}}^\dagger \end{array}\right) \begin{pmatrix} x \\ w \\ \hline x^+ \\ w^+ \end{pmatrix},$$

*where* $w = \nabla f(\boldsymbol{C}x)$, $w^+ = \nabla f(\boldsymbol{C}x^+)$ *and* $x^+ = \boldsymbol{A}x + \boldsymbol{B}w$.

Find Sufficient LMI for Lyapunov Decrement

$$
\begin{pmatrix} \boldsymbol{I}_n & 0 & 0 \\ 0 & \boldsymbol{I}_d & 0 \\ \hline \widetilde{\boldsymbol{A}} & \boldsymbol{B} & 0 \\ 0 & 0 & \boldsymbol{I}_d \end{pmatrix}^\top
\left(\begin{array}{cc|cc} -\rho^2 \boldsymbol{P}_{11} & -\rho^2 \boldsymbol{P}_{12} & 0 & 0 \\ -\rho^2 \boldsymbol{P}_{21} & -\rho^2 \boldsymbol{P}_{22} & 0 & 0 \\ \hline 0 & 0 & \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ 0 & 0 & \boldsymbol{P}_{21} & \boldsymbol{P}_{22} \end{array}\right) (\star)
$$

$$
+ \begin{pmatrix} \boldsymbol{I}_n & 0 & 0 \\ 0 & \boldsymbol{I}_d & 0 \\ \hline \widetilde{\boldsymbol{A}} & \boldsymbol{B} & 0 \\ 0 & 0 & \boldsymbol{I}_d \end{pmatrix}^\top
\left(\begin{array}{cc|cc} 0 & 0 & 0 & -\frac{\lambda}{2}\boldsymbol{C}^\top \\ 0 & -r\boldsymbol{\Pi} & 0 & \frac{\lambda}{2}\widetilde{\boldsymbol{L}}^\dagger \\ \hline 0 & 0 & 0 & \frac{1}{2}\boldsymbol{C}^\top \\ -\frac{\lambda}{2}\boldsymbol{C} & \frac{\lambda}{2}\widetilde{\boldsymbol{L}}^\dagger & \frac{1}{2}\boldsymbol{C} & -\widetilde{\boldsymbol{L}}^\dagger - r\boldsymbol{\Pi} \end{array}\right) (\star)
$$

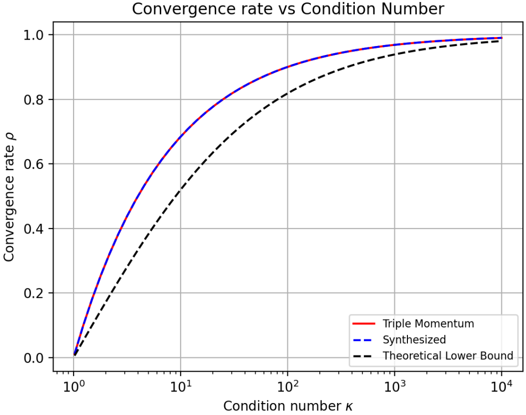
$$
\prec 0
$$

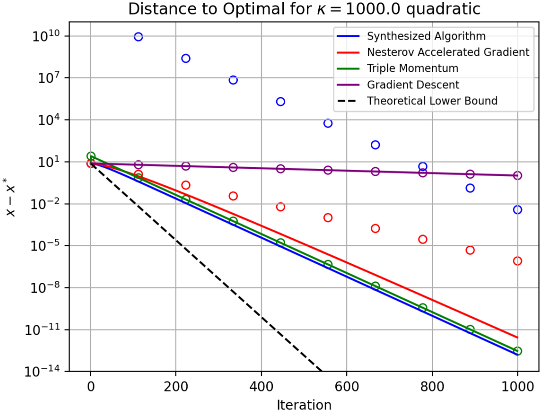
Construct synthesis LMI

$$\left(\begin{array}{cc|c|cc} -\rho^2 \boldsymbol{P}_{11} & -\rho^2 \boldsymbol{P}_{12} & * & * & * \\ -\rho^2 \boldsymbol{P}_{21} & -\rho^2 \boldsymbol{P}_{22} - r\boldsymbol{\Pi} & * & * & * \\ \hline \frac{1}{2}\boldsymbol{J}_2\hat{\boldsymbol{A}} - \frac{\lambda}{2}\boldsymbol{C} & \frac{1}{2}\boldsymbol{J}_2\hat{\boldsymbol{B}} + \frac{\lambda}{2}\widetilde{\boldsymbol{L}}^{\dagger} & -\widetilde{\boldsymbol{L}}^{\dagger} - r\boldsymbol{\Pi} & * & * \\ \hline \hat{\boldsymbol{A}} & \hat{\boldsymbol{B}} & \boldsymbol{P}_{12} & -\boldsymbol{P}_{11} & -\boldsymbol{P}_{12} \\ \boldsymbol{J}_3\hat{\boldsymbol{A}} & \boldsymbol{J}_3\hat{\boldsymbol{B}} & \boldsymbol{P}_{22} & -\boldsymbol{P}_{21} & -\boldsymbol{P}_{22} \end{array}\right) \prec 0$$

# Numerical Results

$$\|x^* - x_k\| \leq c\rho^k$$



# Related Work

- Using IQCs to derive Triple Momentum[1]
- Extension to non-strongly convex functions[2]
- Synthesizing gradient-based algorithms that are robust to additive noise[3]

[1] Van Scoy, Bryan et al. "The fastest known globally convergent first-order method for minimizing strongly convex functions." IEEE Control System Letters, 2018.

[2] Fazlyab, Mahyar et al. "Analysis of optimization algorithms via integral quadratic constraints: nonstrongly convex problems." SIAM Journal of Optimization, 2018.

[3] Van Scoy, Bryan et al. "The Speed-Robustness Trade-Off for First-Order Methods with Additive Gradient Noise." arXiv, 2021.

# Future Work

- Use IQCs to get tighter bounds on synthesized algorithms
- Synthesize preconditioners using IQCs
- Use dissipativity find tighter bounds on convergence of first-order conic optimization algorithms[1]

[1] Yu, Yue et al. "Proportional-integral projected gradient method for conic optimization." Automatica, 2022.