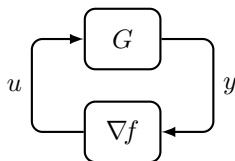


# Lyapunov-based approach to the analysis of iterative optimization algorithms

**Bryan Van Scoy**  
Miami University

**Laurent Lessard**  
Northeastern University

# Algorithm analysis



Represents a gradient-based iterative optimization algorithm. e.g.,

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

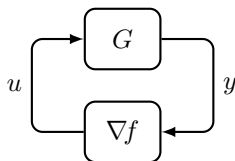
has the following equations:

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

$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix}$$

$$u_k = \nabla f(y_k)$$

# Algorithm analysis



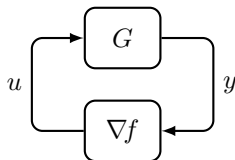
Represents a gradient-based iterative optimization algorithm. e.g.,

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

in state-space notation:

$$y = \left[ \begin{array}{cc|c} 1 + \beta & -\beta & -\alpha \\ 1 & 0 & 0 \\ \hline 1 & 0 & 0 \end{array} \right] u$$
$$u = \nabla f(y)$$

# Algorithm analysis

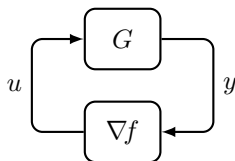


Algorithm analysis:

Find conditions on the system  $G$  that guarantees certain algorithm performance for all initial conditions  $x_0$  and all functions  $f$  in a given class.

This is a Lur'e problem!

# Algorithm analysis



## In this tutorial:

- directly construct a Lyapunov function to certify robust performance
- simple dissipation proof that naturally generalizes to other settings

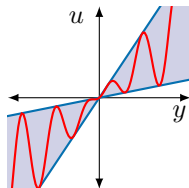
# Simple example

Iterative algorithm:

$$\left\{ \begin{array}{l} x_{k+1} = Ax_k + Bu_k \\ \begin{bmatrix} y_k \\ u_k \end{bmatrix} = Cx_k + Du_k \end{array} \right\} \quad \text{and} \quad u_k = \phi(y_k)$$

Sector-bounded nonlinearity:

$$\begin{aligned} my_k &\leq u_k \leq Ly_k \\ (u_k - my_k)(Ly_k - u_k) &\geq 0 \\ \begin{bmatrix} y_k \\ u_k \end{bmatrix}^\top \underbrace{\begin{bmatrix} -2mL & m+L \\ m+L & -2 \end{bmatrix}}_{\Pi} \begin{bmatrix} y_k \\ u_k \end{bmatrix} &\geq 0 \end{aligned}$$



Find a  $P \succ 0$  so that for all  $\{x_k, u_k, y_k\}$  satisfying the dynamics, the following *dissipation inequality* holds:

$$x_{k+1}^\top P x_{k+1} - \rho^2 x_k^\top P x_k + \begin{bmatrix} y_k \\ u_k \end{bmatrix}^\top \Pi \begin{bmatrix} y_k \\ u_k \end{bmatrix} \leq 0$$

## Simple example

Find a  $P \succ 0$  so that for all  $\{x_k, u_k, y_k\}$  satisfying the dynamics, the following *dissipation inequality* holds:

$$\underbrace{x_{k+1}^\top P x_{k+1} - \rho^2 x_k^\top P x_k}_{\text{must be } \leq 0} + \underbrace{\begin{bmatrix} y_k \\ u_k \end{bmatrix}^\top \Pi \begin{bmatrix} y_k \\ u_k \end{bmatrix}}_{\geq 0 \text{ because } u_k = \phi(y_k)} \leq 0$$

Therefore:  $x_{k+1}^\top P x_{k+1} \leq \rho^2 x_k^\top P x_k$  and  $x_k^\top P x_k \geq 0$ .

So  $V(x) = x^\top P x$  is a Lyapunov function that certifies exponential stability.

Substitute  $x_{k+1} = Ax_k + Bu_k$  and  $\begin{bmatrix} y_k \\ u_k \end{bmatrix} = Cx_k + Du_k$  and obtain:

$$\begin{bmatrix} x_k \\ u_k \end{bmatrix}^\top \left( \begin{bmatrix} A^\top P A - \rho^2 P & A^\top P B \\ B^\top P A & B^\top P B \end{bmatrix} + \begin{bmatrix} C & D \end{bmatrix}^\top \Pi \begin{bmatrix} C & D \end{bmatrix} \right) \begin{bmatrix} x_k \\ u_k \end{bmatrix} \leq 0$$

$$\begin{bmatrix} A^\top P A - \rho^2 P & A^\top P B \\ B^\top P A & B^\top P B \end{bmatrix} + \begin{bmatrix} C & D \end{bmatrix}^\top \Pi \begin{bmatrix} C & D \end{bmatrix} \preceq 0$$

## Simple example

$$\begin{bmatrix} A^T P A - \rho^2 P & A^T P B \\ B^T P A & B^T P B \end{bmatrix} + [C \ D]^T \Pi [C \ D] \preceq 0$$

Alternative form:

$$\begin{bmatrix} A & B \\ I & 0 \\ C & D \end{bmatrix}^T \begin{bmatrix} P & 0 & 0 \\ 0 & -\rho^2 P & 0 \\ 0 & 0 & \Pi \end{bmatrix} \begin{bmatrix} A & B \\ I & 0 \\ C & D \end{bmatrix} \preceq 0$$

This is a *linear matrix inequality* (LMI) in  $P \succ 0$ .

We can use a bisection search to find the smallest feasible  $\rho$ .

**Note:** This is also known as the S-procedure, or the time-domain version of the circle criterion

# Simple example

$$x_{k+1}^\top P x_{k+1} - \rho^2 x_k^\top P x_k + \begin{bmatrix} y_k \\ u_k \end{bmatrix}^\top \Pi \begin{bmatrix} y_k \\ u_k \end{bmatrix} \leq 0$$

with dynamics  $x_{k+1} = Ax_k + Bu_k$  and  $\begin{bmatrix} y_k \\ u_k \end{bmatrix} = Cx_k + Du_k$ .

**Implementation trick:** use row vectors to define *basis elements*.  
Then, express all other elements in terms of the basis.

```
% define basis (independent rows)
n = size(A,1);
x = [eye(n) zeros(n,1)];
u = [zeros(1,n) 1];

% algorithm dynamics
x1 = A*x + B*u;
yu = C*x + D*u;

% Lyapunov equation (generates LMI)
Pi = [-2*m*L m+L; m+L -2];
x1'*P*x1 - rho^2*x'*P*x + yu'*Pi*yu <= 0
```

# What's next?

- Using *lifting* to represent a more complicated function class
- Using different dissipation inequalities to change performance measure
- Numerical examples

# Smooth strongly convex functions

Consider functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  that

- are  $m$ -strongly convex:

$$f(x) - \frac{m}{2}\|x\|^2 \text{ is convex}$$

- have  $L$ -Lipschitz gradients:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n$$

# Interpolation conditions

From [Taylor, Hendrickx, Glineur, 2017]:

Consider  $\{(y_k, u_k, f_k)\}$  for  $k = 1, \dots, m$ . The following are equivalent.

**a)** There exists a smooth strongly convex  $f$  satisfying

$$f(y_k) = f_k \quad \text{and} \quad \nabla f(y_k) = u_k \quad \text{for } k = 1, \dots, m.$$

**b)** The following inequality holds for  $i, j \in \{1, \dots, m\}$ .

$$2(L - m)(f_i - f_j) - mL\|y_i - y_j\|^2 + 2(y_i - y_j)^\top(mu_i - Lu_j) - \|u_i - u_j\|^2 \geq 0.$$

# Interpolation conditions

From [Taylor, Hendrickx, Glineur, 2017]:

Consider  $\{(y_k, u_k, f_k)\}$  for  $k = 1, \dots, m$ . The following are equivalent.

**a)** There exists a smooth strongly convex  $f$  satisfying

$$f(y_k) = f_k \quad \text{and} \quad \nabla f(y_k) = u_k \quad \text{for } k = 1, \dots, m.$$

**b)** The following inequality holds for  $i, j \in \{1, \dots, m\}$ .

$$\begin{bmatrix} y_i \\ y_j \\ u_i \\ u_j \end{bmatrix}^\top \begin{bmatrix} -mL & mL & m & -L \\ mL & -mL & -m & L \\ m & -m & -1 & 1 \\ -L & L & 1 & -1 \end{bmatrix} \begin{bmatrix} y_i \\ y_j \\ u_i \\ u_j \end{bmatrix} + 2(L - m) \begin{bmatrix} 1 \\ -1 \end{bmatrix}^\top \begin{bmatrix} f_i \\ f_j \end{bmatrix} \geq 0$$

Involves *pairs of points*, and includes *function values*.

# Interpolation conditions

Lifted iterates:

$$\mathbf{y}_k = \begin{bmatrix} y_k \\ \vdots \\ y_{k-\ell} \end{bmatrix} \quad \mathbf{u}_k = \begin{bmatrix} u_k \\ \vdots \\ u_{k-\ell} \end{bmatrix} \quad \mathbf{f}_k = \begin{bmatrix} f_k \\ \vdots \\ f_{k-\ell} \end{bmatrix}$$

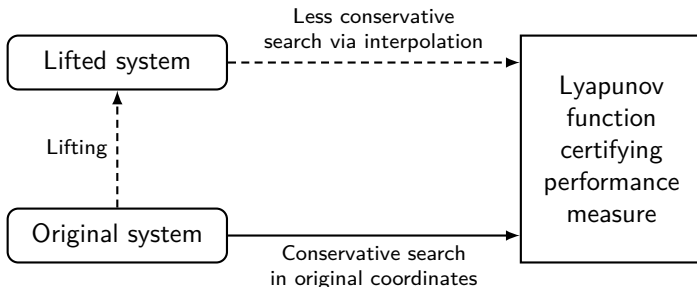
Interpolation conditions

$$\begin{bmatrix} y_i \\ y_j \\ u_i \\ u_j \end{bmatrix}^\top \begin{bmatrix} -mL & mL & m & -L \\ mL & -mL & -m & L \\ m & -m & -1 & 1 \\ -L & L & 1 & -1 \end{bmatrix} \begin{bmatrix} y_i \\ y_j \\ u_i \\ u_j \end{bmatrix} + 2(L-m) \begin{bmatrix} 1 \\ -1 \end{bmatrix}^\top \begin{bmatrix} f_i \\ f_j \end{bmatrix} \geq 0$$

Linear combination with  $\Lambda_{ij} \geq 0$  of each interpolation condition:

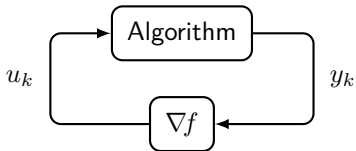
$$\underbrace{\begin{bmatrix} \mathbf{y}_k \\ \mathbf{u}_k \end{bmatrix}^\top \Pi(\Lambda) \begin{bmatrix} \mathbf{y}_k \\ \mathbf{u}_k \end{bmatrix} + \pi(\Lambda)^\top \mathbf{f}_k}_{\sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda)} \geq 0$$

# Lifting approach



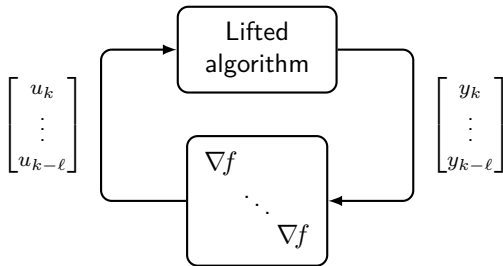
Lifting increases the number of variables but allows use of a simpler Lyapunov function.

## Original system:



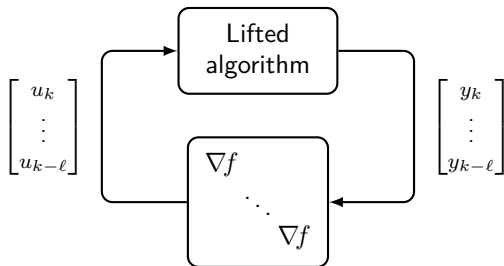
state:  $x_k$

## Lifted system:



state:  $x_k = \begin{bmatrix} x_k \\ y_{k-1} \\ \vdots \\ y_{k-l} \\ u_{k-1} \\ \vdots \\ u_{k-l} \end{bmatrix}$

# Lifted dynamics



$$\begin{bmatrix} \mathbf{x}_{k+1} \\ \mathbf{y}_k \\ \mathbf{u}_k \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{x}_k \\ u_k \end{bmatrix} \quad \text{and} \quad \mathbf{F} \mathbf{f}_{k+1} = \mathbf{F} + \mathbf{f}_k$$

# Lyapunov function

$$V(\mathbf{x}, \mathbf{f}) = \mathbf{x}^\top P \mathbf{x} + \mathbf{p}^\top \mathbf{f}$$

- Dissipation inequality

$$V(\mathbf{x}_{k+1}, \mathbf{f}_{k+1}) - \rho^2 V(\mathbf{x}_k, \mathbf{f}_k) + \sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda_1) \leq 0$$

- Positivity

$$\|\mathbf{x}_k\|^2 - V(\mathbf{x}_k, \mathbf{f}_k) + \sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda_2) \leq 0$$

- Interpolation conditions

$$\sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda) = \begin{bmatrix} \mathbf{y}_k \\ \mathbf{u}_k \end{bmatrix}^\top \Pi(\Lambda) \begin{bmatrix} \mathbf{y}_k \\ \mathbf{u}_k \end{bmatrix} + \pi(\Lambda)^\top \mathbf{f}_k$$

Proof:  $\|\mathbf{x}_k\|^2 \leq V(\mathbf{x}_k, \mathbf{f}_k) \leq \rho^2 V(\mathbf{x}_{k-1}, \mathbf{f}_{k-1}) \leq \dots \leq \rho^{2k} V(\mathbf{x}_0, \mathbf{f}_0)$

# Linear matrix inequality

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ I & 0 \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^T \begin{bmatrix} P & 0 & 0 \\ 0 & -\rho^2 P & 0 \\ 0 & 0 & \Pi(\Lambda_1) \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ I & 0 \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \preceq 0$$
$$(\mathbf{F}_+ - \rho^2 \mathbf{F})^T p + \pi(\Lambda_1) \leq 0$$
$$\begin{bmatrix} I & 0 \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^T \begin{bmatrix} -P & 0 \\ 0 & \Pi(\Lambda_2) \end{bmatrix} \begin{bmatrix} I & 0 \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \preceq 0$$
$$-\mathbf{F}^T p + \Pi(\Lambda_2) \leq 0$$

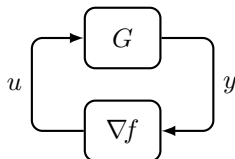
Decision variables:

- symmetric matrix  $P$
- vector  $p$
- nonnegative coefficients  $\Lambda_1$  and  $\Lambda_2$ .

Feasibility of the LMI implies that  $V(\mathbf{x}, \mathbf{f})$  is a Lyapunov function.

## Beyond convergence rate

# Certifying convergence rate

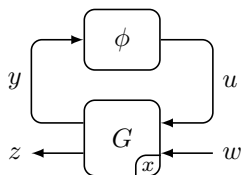


Dissipation inequality and positivity requirement:

$$\begin{aligned} V(\mathbf{x}_{k+1}, \mathbf{f}_{k+1}) - \rho^2 V(\mathbf{x}_k, \mathbf{f}_k) + \sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda_1) &\leq 0 \\ \|\mathbf{x}_k\|^2 - V(\mathbf{x}_k, \mathbf{f}_k) + \sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda_2) &\leq 0 \end{aligned}$$

$$V(\mathbf{x}_{k+1}, \mathbf{f}_{k+1}) \leq \rho^2 V(\mathbf{x}_k, \mathbf{f}_k)$$

# Robust quadratic performance



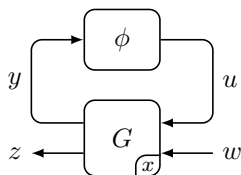
For all exogenous input signals  $w$ ,

$$V(\mathbf{x}_{k+1}, \mathbf{f}_{k+1}) - V(\mathbf{x}_k, \mathbf{f}_k) + \sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda_1) \leq \sigma_p(w_k, z_k)$$
$$\|\mathbf{x}_k\|^2 - V(\mathbf{x}_k, \mathbf{f}_k) + \sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda_2) \leq 0$$

$$\sum_{t=0}^T \sigma_p(w_k, z_k) \geq 0$$

For example, if  $\sigma_p := \gamma^2 \|w_k\|^2 - \|z_k\|^2$ , then  $G$  has a robust  $\ell_2$  gain from  $w \rightarrow z$  of  $\gamma$ .

# Robust stochastic performance

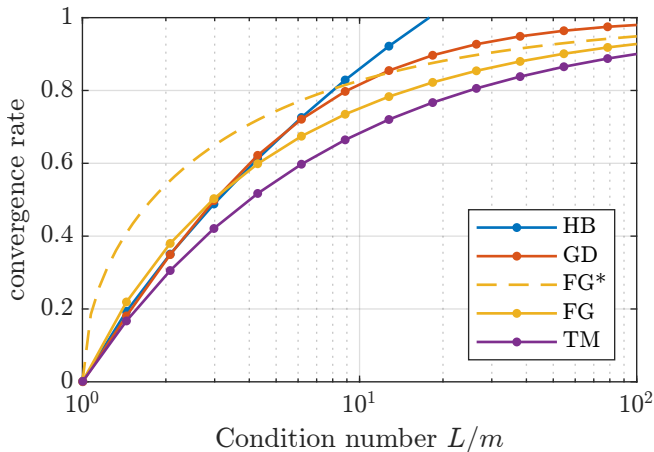


Suppose  $w_k$  is i.i.d. zero-mean random noise with covariance  $\Sigma$ .

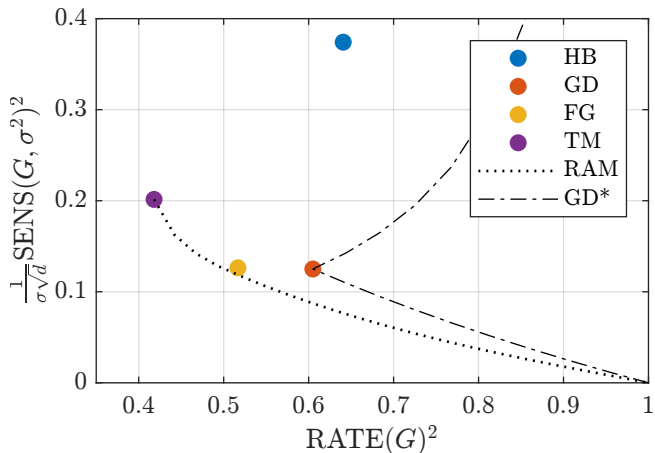
$$\begin{aligned} V(\mathbf{x}_{k+1}, \mathbf{f}_{k+1}) - V(\mathbf{x}_k, \mathbf{f}_k) + \sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda_1) + \|z_k\|^2 &\leq 0 \\ -V(\mathbf{x}_k, \mathbf{f}_k) + \sigma(\mathbf{y}_k, \mathbf{u}_k, \mathbf{f}_k, \Lambda_2) &\leq 0 \\ \text{tr}(P\mathbf{B}_w\Sigma\mathbf{B}_w^\top) &\leq \gamma^2 \end{aligned}$$

$$\limsup_{T \rightarrow \infty} \mathbf{E} \left[ \frac{1}{T} \sum_{t=0}^{T-1} \|z_k\|^2 \right] \leq \gamma^2$$

## Simulation examples



- Heavy Ball, Gradient Descent, Fast Gradient, Triple Momentum
- Identical to result obtained via IQCs [Lessard, Packard, Recht 2016], [Michalowsky, Scherer, Ebenbauer, 2021]
- Only requires  $\ell = 1$ .



- Robustness with respect to additive gradient noise
- Lifting approach [Van Scoy, Lessard, 2021] obtains same trade-off curve as IQCs [Michalowsky, Scherer, Ebenbauer, 2021]
- Requires  $\ell = 6$ .

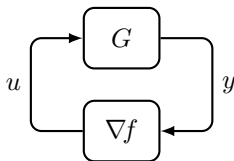
# Comparison with IQC approach

$$V(\mathbf{x}, \mathbf{f}) = \mathbf{x}^\top P \mathbf{x} + p^\top \mathbf{f}$$

- Function values are necessary for tightest bounds
- Lyapunov and IQC approach both involve solving LMI of comparable size (lifting vs. IQC dynamics)
- Lyapunov approach is possibly less conservative (more degrees of freedom), but empirically achieves same results as IQC approach
- Lyapunov certificate is more complicated (more degrees of freedom) than IQC certificate

## Approach

- lift the iterates to a higher-dimensional space
- use interpolation to find all valid inequalities
- search for “quadratic + linear” Lyapunov function in the lifted space



## Benefits

- directly construct a Lyapunov function
- simple dissipation proof that naturally generalizes to other settings
- idea also extends beyond smooth strongly convex functions

**Thank you**