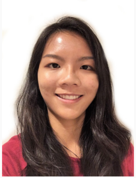
# Analysis of Optimization Algorithms via Sum-of-Squares

Sandra S. Y. Tan    Antonios Varvitsiotis (SUTD)    Vincent Y. F. Tan

National University of Singapore



# Motivation

# First-order optimization algorithms

Iterative algorithms using only (sub)gradient information

- Low computational complexity
- Ideal for large-scale problems with low-accuracy requirements

# First-order optimization algorithms

Iterative algorithms using only (sub)gradient information

- Low computational complexity
- Ideal for large-scale problems with low-accuracy requirements

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

GD: $\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \nabla f(\mathbf{x}_k)$

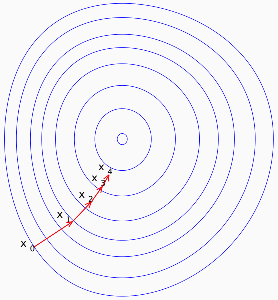



**Figure 1:** Gradient Descent (GD)

# Worst-case convergence bounds

Convergence rate of algorithm $\mathcal{A}$ wrt function class $\mathcal{F}$ *systematically*

# Worst-case convergence bounds

Convergence rate of algorithm $\mathcal{A}$ wrt function class $\mathcal{F}$ *systematically*

Find minimum contraction factor $t \in (0, 1)$ such that

$$\text{Performance Metric}(\mathbf{x}_{k+1}) \leq t \ \text{Performance Metric}(\mathbf{x}_k)$$

holds for all $f \in \mathcal{F}$ and all $\{\mathbf{x}_k\}_{k \geq 1}$ generated by $\mathcal{A}$

# Worst-case convergence bounds

Convergence rate of algorithm $\mathcal{A}$ wrt function class $\mathcal{F}$ *systematically*

Find minimum contraction factor $t \in (0, 1)$ such that

$$\text{Performance Metric}(\mathbf{x}_{k+1}) \leq t \text{ Performance Metric}(\mathbf{x}_k)$$

holds for all $f \in \mathcal{F}$ and all $\{\mathbf{x}_k\}_{k\geq 1}$ generated by $\mathcal{A}$

**Performance metric**

- Objective function accuracy $f(\mathbf{x}_k) - f(\mathbf{x}_*)$, also denoted as $f_k - f_*$
- Squared distance to optimality $\|\mathbf{x}_k - \mathbf{x}_*\|^2$
- Squared residual gradient norm $\|\nabla f(\mathbf{x}_k)\|^2$, also denoted as $\|\mathbf{g}_k\|^2$

# Example

**$L$-smooth, $\mu$-strongly convex ($(\mu, L)$-smooth) function**

- $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \quad \forall\ \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- $f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2$ is convex

# Example

$L$**-smooth,** $\mu$**-strongly convex (**$(\mu, L)$**-smooth) function**

- $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- $f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2$ is convex

**Theorem (de Klerk et al., 2017)**

*Given an $(\mu, L)$-smooth function, apply GD with exact line search:*

$$f_{k+1} - f_* \leq \left(\frac{L-\mu}{L+\mu}\right)^2 (f_k - f_*)$$

# Example

**$L$-smooth, $\mu$-strongly convex ($(\mu, L)$-smooth) function**

- $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- $f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2$ is convex

**Theorem (de Klerk et al., 2017)**

*Given an $(\mu, L)$-smooth function, apply GD with exact line search:*

$$f_{k+1} - f_* \leq \left(\frac{L-\mu}{L+\mu}\right)^2 (f_k - f_*)$$

**A sum-of-squares (SOS) proof**

$$\left(\frac{L-\mu}{L+\mu}\right)^2 (f_k - f_*) - (f_{k+1} - f_*) \geq \frac{\mu}{4}\left(\frac{\|\mathbf{q}_1\|^2}{1+\sqrt{\mu/L}} + \frac{\|\mathbf{q}_2\|^2}{1-\sqrt{\mu/L}}\right) \geq 0$$

where $\mathbf{q}_1$ and $\mathbf{q}_2$ are linear functions of $\mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_k, \mathbf{g}_{k+1}$

# Example

**$L$-smooth, $\mu$-strongly convex ($(\mu, L)$-smooth) function**

- $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- $f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2$ is convex

**Theorem (de Klerk et al., 2017)**

*Given an $(\mu, L)$-smooth function, apply GD with exact line search:*

$$f_{k+1} - f_* \leq \left(\frac{L-\mu}{L+\mu}\right)^2 (f_k - f_*)$$

**A sum-of-squares (SOS) proof**

$$\left(\frac{L-\mu}{L+\mu}\right)^2 (f_k - f_*) - (f_{k+1} - f_*) \geq \frac{\mu}{4}\left(\frac{\|\mathbf{q}_1\|^2}{1+\sqrt{\mu/L}} + \frac{\|\mathbf{q}_2\|^2}{1-\sqrt{\mu/L}}\right) \geq 0$$

where $\mathbf{q}_1$ and $\mathbf{q}_2$ are linear functions of $\mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_k, \mathbf{g}_{k+1}$

# More SOS proofs

**GD with constant step size,** $\gamma \in (0, \frac{2}{L})$ (Lessard et al., 2016)

$$\rho^2 \|\mathbf{x}_k - \mathbf{x}_*\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}_*\|^2$$
$$\geq \begin{cases} \frac{\gamma(2-\gamma(L+\mu))}{L-\mu} \|\mathbf{g}_k - \mu(\mathbf{x}_k - \mathbf{x}_*)\|^2, & \gamma \in \left(0, \frac{2}{L+\mu}\right] \\ \frac{\gamma(\gamma(L+\mu)-2)}{L-\mu} \|\mathbf{g}_k - L(\mathbf{x}_k - \mathbf{x}_*)\|^2, & \gamma \in \left[\frac{2}{L+\mu}, \frac{2}{L}\right) \end{cases}$$

where $\rho := \max\{|1-\gamma\mu|, |1-\gamma L|\}$

# More SOS proofs

**GD with constant step size,** $\gamma \in (0, \frac{2}{L})$ (Lessard et al., 2016)

$$
\begin{aligned}
&\rho^2 \|\mathbf{x}_k - \mathbf{x}_*\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}_*\|^2 \\
&\geq \begin{cases} \frac{\gamma(2-\gamma(L+\mu))}{L-\mu} \|\mathbf{g}_k - \mu(\mathbf{x}_k - \mathbf{x}_*)\|^2, & \gamma \in \left(0, \frac{2}{L+\mu}\right] \\ \frac{\gamma(\gamma(L+\mu)-2)}{L-\mu} \|\mathbf{g}_k - L(\mathbf{x}_k - \mathbf{x}_*)\|^2, & \gamma \in \left[\frac{2}{L+\mu}, \frac{2}{L}\right) \end{cases}
\end{aligned}
$$

where $\rho := \max\{|1-\gamma\mu|, |1-\gamma L|\}$

**GD with the Armijo rule**

$$
\left(1 - \frac{4\mu\epsilon(1-\epsilon)}{\eta L}\right)(f_k - f_*) - (f_{k+1} - f_*) \geq \frac{2\epsilon(1-\epsilon)}{\eta(L-\mu)} \|\mathbf{g}_k + \mu(\mathbf{x}_* - \mathbf{x}_k)\|^2,
$$

where $\epsilon \in (0,1), \eta > 1$ are the parameters of the Armijo rule

# More SOS proofs: Proximal Gradient Method (PGM)

**PGM with exact line search** (Taylor et al., 2018)

$$\left(\frac{L-\mu}{L+\mu}\right)^2 (f_k - f_*) - (f_{k+1} - f_*) \geq \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q^* \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}$$

where $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \dots)$ and $Q^*$ is a PSD matrix

# More SOS proofs: Proximal Gradient Method (PGM)

**PGM with exact line search** (Taylor et al., 2018)

$$\left(\frac{L-\mu}{L+\mu}\right)^2 (f_k - f_*) - (f_{k+1} - f_*) \geq \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q^* \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}$$

where $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \dots)$ and $Q^*$ is a PSD matrix

**PGM with constant step size,** $\gamma \in (0, \frac{2}{L})$ (Taylor et al., 2018)

$$\begin{aligned} &\rho^2 \|\mathbf{r}_k + \mathbf{s}_k\|^2 - \|\mathbf{r}_{k+1} + \bar{\mathbf{s}}_{k+1}\|^2 \\ \geq &\begin{cases} \rho^2 \|\mathbf{q}_1\|^2 + \frac{2-\gamma(L+\mu)}{\gamma(L-\mu)} \|\mathbf{q}_2\|^2, & \gamma \in \left(0, \frac{2}{L+\mu}\right] \\ \rho^2 \|\mathbf{q}_1\|^2 + \frac{\gamma(L+\mu)-2}{\gamma(L-\mu)} \|\mathbf{q}_3\|^2, & \gamma \in \left[\frac{2}{L+\mu}, \frac{2}{L}\right) \end{cases} \end{aligned}$$

where $\rho := \max\{|1-\gamma\mu|, |1-\gamma L|\}$ and $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$ are linear functions of $\mathbf{s}_k, \bar{\mathbf{s}}_{k+1}, \mathbf{r}_k, \mathbf{r}_{k+1}$

# Related Work

# Performance Estimation Problem (PEP)

Framework to derive worst-case bounds (Drori and Teboulle, 2014)

$$\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t.}\ & f \in \mathcal{F} \\
& \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right),\ \ k = 0, \ldots, N-1 \\
& f_0 - f_* \leq R
\end{aligned}$$

# Performance Estimation Problem (PEP)

Framework to derive worst-case bounds (Drori and Teboulle, 2014)

$$\begin{aligned} \max\ & f_N - f_* \\ \text{s.t.}\ & f \in \mathcal{F} \\ & \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right), \ k = 0, \ldots, N-1 \\ & f_0 - f_* \leq R \end{aligned}$$

- GD with exact line search (de Klerk et al., 2017)
- Proximal Gradient Method (PGM) with constant step size (Taylor et al., 2018)
- Proximal Point Algorithm, Conditional Gradient Method, ... (Taylor et al., 2017a); ...

# Performance Estimation Problem (PEP)

Framework to derive worst-case bounds (Drori and Teboulle, 2014)

$$
\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t. } & f \in \mathcal{F} \\
& \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right), \ k = 0, \ldots, N-1 \\
& f_0 - f_* \leq R
\end{aligned}
$$

- GD with exact line search (de Klerk et al., 2017)
- Proximal Gradient Method (PGM) with constant step size (Taylor et al., 2018)
- Proximal Point Algorithm, Conditional Gradient Method, ... (Taylor et al., 2017a); ...
- Formulated as an SDP (*exact* under certain conditions)

# Performance Estimation Problem (PEP)

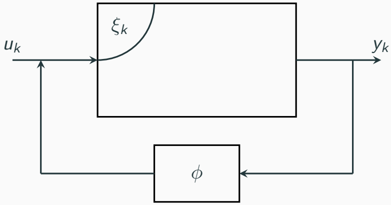
Framework to derive worst-case bounds (Drori and Teboulle, 2014)

$$\begin{aligned} \max\ & f_N - f_* \\ \text{s.t.}\ & f \in \mathcal{F} \\ & \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right),\ k = 0, \ldots, N-1 \\ & f_0 - f_* \leq R \end{aligned}$$

- GD with exact line search (de Klerk et al., 2017)
- Proximal Gradient Method (PGM) with constant step size (Taylor et al., 2018)
- Proximal Point Algorithm, Conditional Gradient Method, ... (Taylor et al., 2017a); ...
- Formulated as an SDP (*exact* under certain conditions)
- Inspiration for our current work

# Integral Quadratic Constraints (IQCs)

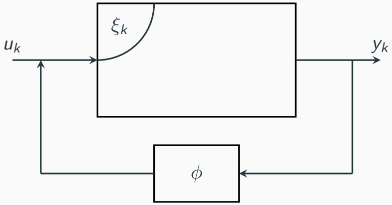
Views optimization algorithms as linear dynamical systems (LDS) with nonlinear feedback (Lessard et al., 2016)




$$
\begin{aligned}
\text{State:} \quad & \xi_{k+1} = A\xi_k + Bu_k \\
\text{Output:} \quad & y_k = C\xi_k + Du_k \\
\text{Feedback:} \quad & u_k = \phi(y_k)
\end{aligned}
$$

# Integral Quadratic Constraints (IQCs)

Views optimization algorithms as linear dynamical systems (LDS) with nonlinear feedback (Lessard et al., 2016)

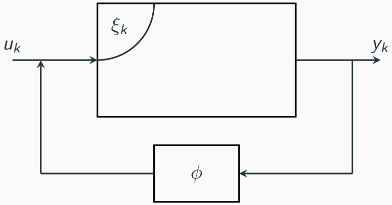



$$
\begin{aligned}
\text{State:} \quad & \xi_{k+1} = A\xi_k + Bu_k \\
\text{Output:} \quad & y_k = C\xi_k + Du_k \\
\text{Feedback:} \quad & u_k = \phi(y_k)
\end{aligned}
$$

- For first-order methods, $u_k = \phi(y_k) := \nabla f(y_k)$

# Integral Quadratic Constraints (IQCs)

Views optimization algorithms as linear dynamical systems (LDS) with nonlinear feedback (Lessard et al., 2016)

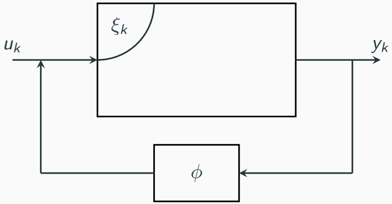



$$
\begin{aligned}
\text{State:} \quad & \xi_{k+1} = A\xi_k + Bu_k \\
\text{Output:} \quad & y_k = C\xi_k + Du_k \\
\text{Feedback:} \quad & u_k = \phi(y_k)
\end{aligned}
$$

- For first-order methods, $u_k = \phi(y_k) := \nabla f(y_k)$
- Convergence of algorithm $\equiv$ stability of dynamical system

# Integral Quadratic Constraints (IQCs)

Views optimization algorithms as linear dynamical systems (LDS) with nonlinear feedback (Lessard et al., 2016)




$$
\begin{aligned}
\text{State:}\quad & \xi_{k+1} = A\xi_k + Bu_k \\
\text{Output:}\quad & y_k = C\xi_k + Du_k \\
\text{Feedback:}\quad & u_k = \phi(y_k)
\end{aligned}
$$

- For first-order methods, $u_k = \phi(y_k) := \nabla f(y_k)$
- Convergence of algorithm $\equiv$ stability of dynamical system
- Replace $\phi$ with quadratic constraints on all instances of $(y_k, u_k)$

# The Sum-of-Squares Approach

# SDP formulation of sum-of-squares optimization

**Sum-of-squares of degree-2 polynomials**

For $\mathbf{z} = (x, y)$ and $d = 2$,

$$\begin{aligned} p(\mathbf{z}) &= (x^2 - 2)^2 + (x - 3y)^2 \\ &= x^4 - 3x^2 - 6xy + 9y^2 + 4 \end{aligned}$$

# SDP formulation of sum-of-squares optimization

**Sum-of-squares of degree-2 polynomials**

For $\mathbf{z} = (x, y)$ and $d = 2$,

$$
\begin{aligned}
p(\mathbf{z}) &= (x^2 - 2)^2 + (x - 3y)^2 \\
&= x^4 - 3x^2 - 6xy + 9y^2 + 4 \\
&= \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots & & & \\ Q_{2,1} & \ddots & & & & \\ \vdots & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix}
\end{aligned}
$$

where $Q \succeq 0$

# SDP formulation of sum-of-squares optimization

**Sum-of-squares of degree-2 polynomials**

For $\mathbf{z} = (x, y)$ and $d = 2$,

$$
\begin{aligned}
p(\mathbf{z}) &= (x^2 - 2)^2 + (x - 3y)^2 \\
&= x^4 - 3x^2 - 6xy + 9y^2 + 4 \\
&= \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots & & & \\ Q_{2,1} & \ddots & & & & \\ \vdots & & & & & \\ & & & & & \\ & & & & & \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix}
\end{aligned}
$$

where $Q \succeq 0$

- Match coefficients of $1, x, y, x^2, \ldots, y^4 \implies$ affine constraints

# SDP formulation of sum-of-squares optimization

**Sum-of-squares of degree-2 polynomials**

For $\mathbf{z} = (x, y)$ and $d = 2$,

$$\begin{aligned} p(\mathbf{z}) &= (x^2 - 2)^2 + (x - 3y)^2 \\ &= x^4 - 3x^2 - 6xy + 9y^2 + 4 \\ &= \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots & & & \\ Q_{2,1} & \ddots & & & & \\ \vdots & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix} \end{aligned}$$

where $Q \succeq 0$

- Match coefficients of $1, x, y, x^2, \ldots, y^4 \implies$ affine constraints

$$1: \quad 4 = Q_{1,1}$$

# SDP formulation of sum-of-squares optimization

**Sum-of-squares of degree-2 polynomials**

For $\mathbf{z} = (x, y)$ and $d = 2$,

$$\begin{aligned} p(\mathbf{z}) &= (x^2 - 2)^2 + (x - 3y)^2 \\ &= x^4 - 3x^2 - 6xy + 9y^2 + 4 \\ &= \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots & & & \\ Q_{2,1} & \ddots & & & & \\ \vdots & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix} \end{aligned}$$

where $Q \succeq 0$

- Match coefficients of $1, x, y, x^2, \ldots, y^4 \implies$ affine constraints

$$\begin{aligned} 1 :& \quad 4 = Q_{1,1} \\ x :& \quad 0 = Q_{1,2} + Q_{2,1} = 2Q_{1,2} \end{aligned}$$

# SDP formulation of sum-of-squares optimization

**Sum-of-squares of degree-2 polynomials**

For $\mathbf{z} = (x, y)$ and $d = 2$,

$$\begin{aligned} p(\mathbf{z}) &= (x^2 - 2)^2 + (x - 3y)^2 \\ &= x^4 - 3x^2 - 6xy + 9y^2 + 4 \\ &= \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots & & & \\ Q_{2,1} & \ddots & & & & \\ \vdots & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \\ x^2 \\ xy \\ y^2 \end{pmatrix} \end{aligned}$$

where $Q \succeq 0$

- Match coefficients of $1, x, y, x^2, \ldots, y^4 \implies$ affine constraints
- Affine constraints + $Q \succeq 0$ condition $\implies$ SDP feasibility problem

$$\min 0 \;\; \text{s.t.} \;\; Q \succeq 0, \;\text{affine constraints}$$

# Certifying constrained nonnegativity

**Closed semi-algebraic set**

$$K = \{\mathbf{z} : \; h_i(\mathbf{z}) \geq 0, \; v_j(\mathbf{z}) = 0 \quad \forall \, i, j\}$$

for some polynomials $h_i(\mathbf{z}), v_j(\mathbf{z})$'s

# Certifying constrained nonnegativity

**Closed semi-algebraic set**

$$K = \{\mathbf{z} : \; h_i(\mathbf{z}) \geq 0, \; v_j(\mathbf{z}) = 0 \quad \forall \, i, j\}$$

for some polynomials $h_i(\mathbf{z}), v_j(\mathbf{z})$'s

**"Constrained" nonnegativity**

To certify that $p(\mathbf{z}) \geq 0$ for all $\mathbf{z} \in K$ (Parrilo, 2003; Lasserre, 2007),

$$p(\mathbf{z}) = \sigma_0(\mathbf{z}) + \sum_i \sigma_i(\mathbf{z}) h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z})$$

# Certifying constrained nonnegativity

**Closed semi-algebraic set**

$$K = \{\mathbf{z} : \; h_i(\mathbf{z}) \geq 0, \; v_j(\mathbf{z}) = 0 \quad \forall \; i, j\}$$

for some polynomials $h_i(\mathbf{z}), v_j(\mathbf{z})$'s

**"Constrained" nonnegativity**
To certify that $p(\mathbf{z}) \geq 0$ for all $\mathbf{z} \in K$ (Parrilo, 2003; Lasserre, 2007),

$$p(\mathbf{z}) = \underbrace{\sigma_0(\mathbf{z})}_{\text{SOS}} + \sum_i \underbrace{\sigma_i(\mathbf{z})}_{\text{SOS}} h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z})$$

where a sum-of-squares (SOS) polynomial: $\sigma(\mathbf{z}) = \sum_k q_k^2(\mathbf{z})$

# Certifying constrained nonnegativity

**Closed semi-algebraic set**

$$K = \{\mathbf{z} : \; h_i(\mathbf{z}) \geq 0, \; v_j(\mathbf{z}) = 0 \quad \forall \; i, j\}$$

for some polynomials $h_i(\mathbf{z}), v_j(\mathbf{z})$'s

**"Constrained" nonnegativity**

To certify that $p(\mathbf{z}) \geq 0$ for all $\mathbf{z} \in K$ (Parrilo, 2003; Lasserre, 2007),

$$p(\mathbf{z}) = \underbrace{\sigma_0(\mathbf{z})}_{\text{SOS}} + \sum_i \underbrace{\sigma_i(\mathbf{z})}_{\text{SOS}} \underbrace{h_i(\mathbf{z})}_{\geq 0} + \sum_j \theta_j(\mathbf{z}) \underbrace{v_j(\mathbf{z})}_{=0}$$

# Certifying constrained nonnegativity

**Closed semi-algebraic set**

$$K = \{\mathbf{z} : \; h_i(\mathbf{z}) \geq 0, \; v_j(\mathbf{z}) = 0 \quad \forall \, i, j\}$$

for some polynomials $h_i(\mathbf{z}), v_j(\mathbf{z})$'s

**"Constrained" nonnegativity**

To certify that $p(\mathbf{z}) \geq 0$ for all $\mathbf{z} \in K$ (Parrilo, 2003; Lasserre, 2007),

$$p(\mathbf{z}) = \underbrace{\sigma_0(\mathbf{z})}_{\text{SOS}} + \sum_i \underbrace{\sigma_i(\mathbf{z})}_{\text{SOS}} \underbrace{h_i(\mathbf{z})}_{\geq 0} + \sum_j \theta_j(\mathbf{z}) \underbrace{v_j(\mathbf{z})}_{=0}$$

- Fix degrees of the $\sigma_i(\mathbf{z})$'s and $\theta_j(\mathbf{z})$'s
- Construct and solve the corresponding SDP feasibility problem

# Analyzing optimization algorithms

$$\begin{aligned}
\min \quad & t \\
\text{s.t.} \quad & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0 \\
& t \in (0,1) \\
& f \in \mathcal{F} \\
& \mathbf{x}_{k+1} = \mathcal{A}(\mathbf{x}_k, f_k, \mathbf{g}_k)
\end{aligned}$$

where $f_k = f(\mathbf{x}_k)$ and $\mathbf{g}_k \in \partial f(\mathbf{x}_k)$ for all $k$

# First relaxation step

$$
\begin{aligned}
\min \quad & t \\
\text{s.t.} \quad & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0 \\
& t \in (0, 1) \\
& f \in \mathcal{F} \\
& \mathbf{x}_{k+1} = \mathcal{A}(\mathbf{x}_k, f_k, \mathbf{g}_k)
\end{aligned}
$$

where $f_k = f(\mathbf{x}_k)$ and $\mathbf{g}_k \in \partial f(\mathbf{x}_k)$ for all $k$

1. Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

# First relaxation step

$$\begin{aligned}
\min \quad & t \\
\text{s.t.} \quad & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0 \\
& t \in (0,1) \\
& h_i(\mathbf{z}) \geq 0, \ \forall i \\
& v_j(\mathbf{z}) = 0, \ \forall j
\end{aligned}$$

where $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_*, \mathbf{g}_k, \mathbf{g}_{k+1}) \in \mathbb{R}^{6n+3}$

1. Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

# First relaxation step

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0 \\ & t \in (0, 1) \\ & h_i(\mathbf{z}) \geq 0, \ \forall i \\ & v_j(\mathbf{z}) = 0, \ \forall j \end{aligned}$$

where $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_*, \mathbf{g}_k, \mathbf{g}_{k+1}) \in \mathbb{R}^{6n+3}$

1. Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

E.g. $f$ is $L$-smooth $\implies \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \quad \forall\, \mathbf{x}, \mathbf{y}$

# First relaxation step

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0 \\ & t \in (0,1) \\ & h_i(\mathbf{z}) \geq 0, \ \forall i \\ & v_j(\mathbf{z}) = 0, \ \forall j \end{aligned}$$

where $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_*, \mathbf{g}_k, \mathbf{g}_{k+1}) \in \mathbb{R}^{6n+3}$

1. Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

E.g. $f$ is $L$-smooth $\implies \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\| \quad \forall\ \mathbf{x}, \mathbf{y}$
$\implies h_1(\mathbf{z}) = L^2\|\mathbf{x}_k - \mathbf{x}_*\|^2 - \|\mathbf{g}_k - \mathbf{g}_*\|^2 \geq 0$

# First relaxation step

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0 \\ & t \in (0,1) \\ & \left.\begin{aligned} & h_i(\mathbf{z}) \geq 0, \ \forall i \\ & v_j(\mathbf{z}) = 0, \ \forall j \end{aligned}\right\} K \end{aligned}$$

where $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_*, \mathbf{g}_k, \mathbf{g}_{k+1}) \in \mathbb{R}^{6n+3}$

1. Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

# First relaxation step

$$\begin{aligned}
\min \quad & t \\
\text{s.t.} \quad & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0 \\
& t \in (0,1) \\
& \left.\begin{aligned} & h_i(\mathbf{z}) \geq 0, \ \forall i \\ & v_j(\mathbf{z}) = 0, \ \forall j \end{aligned}\right\} K
\end{aligned}$$

where $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_*, \mathbf{g}_k, \mathbf{g}_{k+1}) \in \mathbb{R}^{6n+3}$

1. Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

# First relaxation step

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0 \quad \forall\, \mathbf{z} \in K \\ & t \in (0,1) \end{aligned}$$

where $K = \{\mathbf{z} : \ h_i(\mathbf{z}) \geq 0,\ v_j(\mathbf{z}) = 0 \quad \forall\ i,j\}$

# Second relaxation step

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0 \quad \forall\, \mathbf{z} \in K \\ & t \in (0, 1) \end{aligned}$$

where $K = \{\mathbf{z} : \; h_i(\mathbf{z}) \geq 0, \; v_j(\mathbf{z}) = 0 \quad \forall\, i, j\}$

2. Relax the constraint that $p(\mathbf{z}) := t(f_k - f_*) - (f_{k+1} - f_*)$ is nonnegative over $K$ with an SOS cert

# Second relaxation step

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & p(\mathbf{z}) = \sigma_0(\mathbf{z}) + \sum_i \sigma_i(\mathbf{z}) h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z}) \\ & t \in (0, 1) \end{aligned}$$

where $K = \{\mathbf{z} : \; h_i(\mathbf{z}) \geq 0, \; v_j(\mathbf{z}) = 0 \quad \forall \, i, j\}$

2. Relax the constraint that $p(\mathbf{z}) := t(f_k - f_*) - (f_{k+1} - f_*)$ is nonnegative over $K$ with an SOS cert

# Obtaining the bound

1. Identify polynomial (in)equalities necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

$$\begin{aligned} & f \in \mathcal{F} \\ & \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right) \end{aligned} \implies \begin{aligned} & h_i(\mathbf{z}) \geq 0, \ \forall i \\ & v_j(\mathbf{z}) = 0, \ \forall j \end{aligned}$$

# Obtaining the bound

1. Identify polynomial (in)equalities necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

$$\begin{array}{l} f \in \mathcal{F} \\ \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right) \end{array} \implies \begin{array}{l} h_i(\mathbf{z}) \geq 0, \ \forall i \\ v_j(\mathbf{z}) = 0, \ \forall j \end{array}$$

2. Relax nonnegativity constraint with an SOS cert, i.e., relax

$$p(\mathbf{z}) \geq 0 \ \forall \ \mathbf{z} \in K \ \text{ to } \ p(\mathbf{z}) = \sigma_0(\mathbf{z}) + \sum_i \sigma_i(\mathbf{z}) h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z})$$

for SOS polynomials $\sigma_i(\mathbf{z})$ and arbitrary polynomials $\theta_j(\mathbf{z})$

# Obtaining the bound

1. Identify polynomial (in)equalities necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

$$\begin{array}{l} f \in \mathcal{F} \\ \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right) \end{array} \implies \begin{array}{l} h_i(\mathbf{z}) \geq 0, \ \forall i \\ v_j(\mathbf{z}) = 0, \ \forall j \end{array}$$

2. Relax nonnegativity constraint with an SOS cert, i.e., relax

$$p(\mathbf{z}) \geq 0 \ \forall \ \mathbf{z} \in K \ \text{ to } \ p(\mathbf{z}) = \sigma_0(\mathbf{z}) + \sum_i \sigma_i(\mathbf{z}) h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z})$$

   for SOS polynomials $\sigma_i(\mathbf{z})$ and arbitrary polynomials $\theta_j(\mathbf{z})$

3. Solve the SDP multiple times for the $n = 1$ case and 'guess' the analytic expression of the optimal contraction factor

$$t^* = \min \left\{ t : t \in (0, 1), \ \text{SOS constraints} \right\}$$

# Obtaining the bound

1. Identify polynomial (in)equalities necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

$$\begin{array}{l} f \in \mathcal{F} \\ \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right) \end{array} \implies \begin{array}{l} h_i(\mathbf{z}) \geq 0, \ \forall i \\ v_j(\mathbf{z}) = 0, \ \forall j \end{array}$$

2. Relax nonnegativity constraint with an SOS cert, i.e., relax

$$p(\mathbf{z}) \geq 0 \ \forall \ \mathbf{z} \in K \text{ to } p(\mathbf{z}) = \sigma_0(\mathbf{z}) + \sum_i \sigma_i(\mathbf{z}) h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z})$$

   for SOS polynomials $\sigma_i(\mathbf{z})$ and arbitrary polynomials $\theta_j(\mathbf{z})$

3. Solve the SDP multiple times for the $n = 1$ case and 'guess' the analytic expression of the optimal contraction factor

$$t^* = \min\left\{t : t \in (0, 1), \text{ SOS constraints}\right\}$$

4. Generalize this to a feasible SDP solution for the multivariate case

# Example

# Function class

Class of $(\mu, L)$-smooth functions (denoted $\mathcal{F}_{\mu,L}(\mathbb{R}^n)$)

# Function class

Class of $(\mu, L)$-smooth functions (denoted $\mathcal{F}_{\mu,L}(\mathbb{R}^n)$)

**Theorem ($\mathcal{F}_{\mu,L}$-interpolability (Taylor et al., 2017b))**

*Given a set $\{(\mathbf{x}_i, f_i, \mathbf{g}_i)\}_{i\in I}$, $\exists$ a $(\mu, L)$-smooth function $f$ where $f_i = f(\mathbf{x}_i)$ and $\mathbf{g}_i \in \partial f(\mathbf{x}_i)$ $\forall$ $i$ iff*

$$f_i - f_j - \mathbf{g}_j^\top(\mathbf{x}_i - \mathbf{x}_j) \geq \frac{L}{2(L-\mu)}\left[\frac{1}{L}\|\mathbf{g}_i - \mathbf{g}_j\|^2 + \mu\|\mathbf{x}_i - \mathbf{x}_j\|^2 - 2\frac{\mu}{L}(\mathbf{g}_j - \mathbf{g}_i)^\top(\mathbf{x}_j - \mathbf{x}_i)\right] \; \forall \; i \neq j$$

# Function class

Class of $(\mu, L)$-smooth functions (denoted $\mathcal{F}_{\mu,L}(\mathbb{R}^n)$)

**Theorem ($\mathcal{F}_{\mu,L}$-interpolability (Taylor et al., 2017b))**

*Given a set* $\{(\mathbf{x}_i, f_i, \mathbf{g}_i)\}_{i\in I}$, $\exists$ *a* $(\mu, L)$*-smooth function* $f$ *where* $f_i = f(\mathbf{x}_i)$ *and* $\mathbf{g}_i \in \partial f(\mathbf{x}_i)$ $\forall$ $i$ *iff*

$$f_i - f_j - \mathbf{g}_j^\top (\mathbf{x}_i - \mathbf{x}_j) \geq \frac{L}{2(L-\mu)}\left[\frac{1}{L}\|\mathbf{g}_i - \mathbf{g}_j\|^2 + \mu\|\mathbf{x}_i - \mathbf{x}_j\|^2 - 2\frac{\mu}{L}(\mathbf{g}_j - \mathbf{g}_i)^\top(\mathbf{x}_j - \mathbf{x}_i)\right] \; \forall \; i \neq j$$

E.g. if we set $i = k$ and $j = k+1$:

$$f_k - f_{k+1} - \mathbf{g}_{k+1}^\top (\mathbf{x}_k - \mathbf{x}_{k+1}) - \frac{L}{2(L-\mu)}\left[\frac{1}{L}\|\mathbf{g}_k - \mathbf{g}_{k+1}\|^2 + \mu\|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2 - 2\frac{\mu}{L}(\mathbf{g}_{k+1} - \mathbf{g}_k)^\top(\mathbf{x}_{k+1} - \mathbf{x}_k)\right] \geq 0$$

# Algorithm

**Gradient Descent (GD) with exact line search (ELS)**

$$\gamma_k = \arg\min_{\gamma>0} f\left(\mathbf{x}_k - \gamma \mathbf{g}_k\right)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \mathbf{g}_k$$

where $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$

# Algorithm

**Gradient Descent (GD) with exact line search (ELS)**

$$\gamma_k = \arg\min_{\gamma>0} f\left(\mathbf{x}_k - \gamma \mathbf{g}_k\right)$$
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \mathbf{g}_k$$

where $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$

**Derived constraints**

$$\mathbf{g}_{k+1}^\top \left(\mathbf{x}_{k+1} - \mathbf{x}_k\right) = 0$$
$$\mathbf{g}_{k+1}^\top \mathbf{g}_k = 0$$

# Applying the SOS approach

**Necessary constraints**

- 6 $\mathcal{F}_{\mu,L}$-interpolability constraints corresponding to $k, k+1, *$ ($h_i(\mathbf{z}) \geq 0$)
- 2 algorithmic constraints ($v_j(\mathbf{z}) = 0$)

$$\begin{aligned} \mathbf{g}_{k+1}^\top (\mathbf{x}_{k+1} - \mathbf{x}_k) &= 0 \\ \mathbf{g}_{k+1}^\top \mathbf{g}_k &= 0 \end{aligned}$$

# Applying the SOS approach

**Necessary constraints**

- 6 $\mathcal{F}_{\mu,L}$-interpolability constraints corresponding to $k, k+1, *$ ($h_i(\mathbf{z}) \geq 0$)
- 2 algorithmic constraints ($v_j(\mathbf{z}) = 0$)

**The SDP**

$$\begin{aligned}
&\min \ t \\
&\text{s.t. } p(\mathbf{z}) = \sigma_0(\mathbf{z}) + \sum_i \sigma_i(\mathbf{z}) h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z}) \\
&\qquad t \in (0,1) \\
&\qquad \sigma_0(\mathbf{z}) : \text{SOS of linear polynomials} \\
&\qquad \sigma_i(\mathbf{z}), i \geq 1 : \text{SOS of degree } 0, \ \theta_j(\mathbf{z}) : \text{degree } 0
\end{aligned}$$

# Applying the SOS approach

**Necessary constraints**

- 6 $\mathcal{F}_{\mu,L}$-interpolability constraints corresponding to $k, k+1, *$ $(h_i(\mathbf{z}) \geq 0)$
- 2 algorithmic constraints $(v_j(\mathbf{z}) = 0)$

**The SDP**

$$\begin{aligned} &\min \ t \\ &\text{s.t. } p(\mathbf{z}) = \sigma_0(\mathbf{z}) + \sum_i \sigma_i(\mathbf{z}) h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z}) \\ &\qquad t \in (0,1) \\ &\qquad \sigma_0(\mathbf{z}) : \text{SOS of linear polynomials} \\ &\qquad \sigma_i(\mathbf{z}), i \geq 1 : \text{SOS of degree } 0, \ \theta_j(\mathbf{z}) : \text{degree } 0 \end{aligned}$$

**Necessary constraints**

- 6 $\mathcal{F}_{\mu,L}$-interpolability constraints corresponding to $k, k+1, *$ $(h_i(\mathbf{z}) \geq 0)$
- 2 algorithmic constraints $(v_j(\mathbf{z}) = 0)$

**The SDP**

$$\begin{aligned} &\min\ t \\ &\text{s.t. } p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i(\mathbf{z}) h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z}) \\ &\qquad t \in (0,1) \\ &\qquad Q \succeq 0 \\ &\qquad \sigma_i(\mathbf{z}), i \geq 1 : \text{SOS of degree } 0,\ \theta_j(\mathbf{z}) : \text{degree } 0 \end{aligned}$$

# Applying the SOS approach

**Necessary constraints**

- 6 $\mathcal{F}_{\mu,L}$-interpolability constraints corresponding to $k, k+1, *$ $(h_i(\mathbf{z}) \geq 0)$
- 2 algorithmic constraints $(v_j(\mathbf{z}) = 0)$

**The SDP**

$$\begin{aligned} \min\ & t \\ \text{s.t. } & p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i(\mathbf{z}) h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z}) \\ & t \in (0,1) \\ & Q \succeq 0 \\ & \sigma_i(\mathbf{z}), i \geq 1 : \text{SOS of degree } 0,\ \theta_j(\mathbf{z}) : \text{degree } 0 \end{aligned}$$

# Applying the SOS approach

**Necessary constraints**

- 6 $\mathcal{F}_{\mu,L}$-interpolability constraints corresponding to $k, k+1, *$ $(h_i(\mathbf{z}) \geq 0)$
- 2 algorithmic constraints $(v_j(\mathbf{z}) = 0)$

**The SDP**

$$\begin{aligned}
&\min \ t \\
&\text{s.t. } p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z}) \\
&\qquad t \in (0,1) \\
&\qquad Q \succeq 0 \\
&\qquad \sigma_i \geq 0, i \geq 1, \ \theta_j(\mathbf{z}) : \text{degree } 0
\end{aligned}$$

# Applying the SOS approach

## Necessary constraints

- 6 $\mathcal{F}_{\mu,L}$-interpolability constraints corresponding to $k, k+1, *$ $(h_i(\mathbf{z}) \geq 0)$
- 2 algorithmic constraints $(v_j(\mathbf{z}) = 0)$

## The SDP

$$\begin{aligned} \min\ & t \\ \text{s.t.}\ & p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^{\top} Q \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j(\mathbf{z}) v_j(\mathbf{z}) \\ & t \in (0,1) \\ & Q \succeq 0 \\ & \sigma_i \geq 0, i \geq 1,\ \theta_j(\mathbf{z}) : \text{degree } 0 \end{aligned}$$

# Applying the SOS approach

**Necessary constraints**

- 6 $\mathcal{F}_{\mu,L}$-interpolability constraints corresponding to $k, k+1, *$ $(h_i(\mathbf{z}) \geq 0)$
- 2 algorithmic constraints $(v_j(\mathbf{z}) = 0)$

**The SDP**

$$\begin{aligned} &\min\ t \\ &\text{s.t. } p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z}) \\ &\qquad t \in (0,1) \\ &\qquad Q \succeq 0 \\ &\qquad \sigma_i \geq 0, i \geq 1,\ \theta_j \in \mathbb{R} \end{aligned}$$

# Deriving the affine constraints

$$
\begin{aligned}
&t(f_k - f_*) - (f_{k+1} - f_*) \\
&= \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,1} & \ddots & \\ \vdots & & \end{pmatrix} \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})
\end{aligned}
$$

# Deriving the affine constraints

$$
\begin{aligned}
&t(f_k - f_*) - (f_{k+1} - f_*) \\
&= \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,1} & \ddots & \\ \vdots & & \end{pmatrix} \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})
\end{aligned}
$$

Matching coefficients for e.g. $f_*$ term:

# Deriving the affine constraints

$$
\begin{aligned}
&t(f_k - f_*) - (f_{k+1} - f_*) \\
&= \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,1} & \ddots & \\ \vdots & & \end{pmatrix} \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})
\end{aligned}
$$

Matching coefficients for e.g. $f_*$ term:

$$1 - t =$$

# Deriving the affine constraints

$$
\begin{aligned}
&t(f_k - f_*) - (f_{k+1} - f_*) \\
&= \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,1} & \ddots & \\ \vdots & & \end{pmatrix} \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})
\end{aligned}
$$

Matching coefficients for e.g. $f_*$ term:

$$1 - t = 2Q_{1,2}$$

# Deriving the affine constraints

$$t(f_k - f_*) - (f_{k+1} - f_*)$$

$$= \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,1} & \ddots & \\ \vdots & & \end{pmatrix} \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})$$

Matching coefficients for e.g. $f_*$ term:

$$1 - t = 2Q_{1,2} - \sigma_2$$

Constraint $h_2(\mathbf{z}) \geq 0$:

$$f_k - f_* - \frac{L}{2(L-\mu)} \left[ \frac{1}{L} \|\mathbf{g}_k\|^2 + \mu \|\mathbf{x}_k - \mathbf{x}_*\|^2 + 2\frac{\mu}{L} \mathbf{g}_k^\top (\mathbf{x}_* - \mathbf{x}_k) \right] \geq 0$$

# Deriving the affine constraints

$$
\begin{aligned}
&t(f_k - f_*) - (f_{k+1} - f_*) \\
&= \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix}^\top \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,1} & \ddots & \\ \vdots & & \end{pmatrix} \begin{pmatrix} 1 \\ f_* \\ \vdots \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})
\end{aligned}
$$

Matching coefficients for e.g. $f_*$ term:

$$
1 - t = 2Q_{1,2} - \sigma_2 - \sigma_4 + \sigma_5 + \sigma_6
$$

# Deriving the affine constraints

$$t(f_k - f_*) - (f_{k+1} - f_*) = \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})$$

$$\begin{aligned}
1 : \quad & 0 = Q_{1,1} \\
f_* : \quad & 1 - t = 2Q_{1,2} - \sigma_2 - \sigma_4 + \sigma_5 + \sigma_6 \\
f_k : \quad & t = 2Q_{1,3} + \sigma_1 + \sigma_2 - \sigma_3 - \sigma_5 \\
f_{k+1} : \quad & -1 = 2Q_{1,4} - \sigma_1 + \sigma_3 + \sigma_4 - \sigma_6 \\
& \vdots
\end{aligned}$$

# Solving the SDP

$$
\begin{aligned}
\min \;\; & t \\
\text{s.t.} \;\; & p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^{\top} Q \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z}) \\
& t \in (0, 1) \\
& Q \succeq 0 \\
& \sigma_i \geq 0, \;\; \theta_j \in \mathbb{R}
\end{aligned}
$$

# Solving the SDP

$$\min\ t$$

$$\text{s.t. } p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})$$

$$t \in (0,1)$$

$$Q \succeq 0$$

$$\sigma_i \geq 0,\ \theta_j \in \mathbb{R}$$

**Guessing the optimal solution**

$$t^* = \left(\frac{L-\mu}{L+\mu}\right)^2,\ Q^* = \left(\begin{array}{c|c} 0_{4\times 4} & 0_{4\times 5} \\ \hline 0_{5\times 4} & Q^*_{5\times 5} \end{array}\right),\ \sigma_1^* = \frac{L-\mu}{L+\mu},\ \ldots$$

$$Q^*_{5\times 5} = \begin{pmatrix} \frac{2L^2\mu^2}{(L+\mu)^2(L-\mu)} & -\frac{L\mu^2}{(L+\mu)^2} & -\frac{L\mu^2}{(L+\mu)(L-\mu)} & \frac{L\mu}{(L+\mu)^2} & \frac{L\mu}{(L+\mu)(L-\mu)} \\ & \frac{L\mu(L+3\mu)}{2(L+\mu)^2} & -\frac{L\mu}{2(L+\mu)} & -\frac{\mu(3L+\mu)}{2(L+\mu)^2} & -\frac{\mu}{2(L+\mu)} \\ & & \frac{L\mu}{2(L-\mu)} & \frac{\mu}{2(L+\mu)} & -\frac{\mu}{2(L-\mu)} \\ & & & \frac{L+3\mu}{2(L+\mu)^2} & \frac{1}{2(L+\mu)} \\ & & & & \frac{1}{2(L-\mu)} \end{pmatrix}$$

# Multivariate case

Recall $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_*, \mathbf{g}_k, \mathbf{g}_{k+1})$

Let $\mathbf{z} = (\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n)$ where

$$\mathbf{z}_0 = (f_*, f_k, f_{k+1}) \text{ and } \mathbf{z}_\ell = (x_*(\ell), x_k(\ell), x_{k+1}(\ell), g_*(\ell), g_k(\ell), g_{k+1}(\ell))$$

# Multivariate case

Recall $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_*, \mathbf{g}_k, \mathbf{g}_{k+1})$

Let $\mathbf{z} = (\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n)$ where

$$\mathbf{z}_0 = (f_*, f_k, f_{k+1}) \text{ and } \mathbf{z}_\ell = (x_*(\ell), x_k(\ell), x_{k+1}(\ell), g_*(\ell), g_k(\ell), g_{k+1}(\ell))$$

Require polynomials $p(\mathbf{z})$, $h_i(\mathbf{z})$ and $v_j(\mathbf{z})$ satisfy

$$\mathrm{pol}(\mathbf{z}) = \mathrm{pol}^0(\mathbf{z}_0) + \sum_{\ell=1}^{n} \mathrm{pol}^1(\mathbf{z}_\ell)$$

# Multivariate case

Recall $\mathbf{z} = (f_*, f_k, f_{k+1}, \mathbf{x}_*, \mathbf{x}_k, \mathbf{x}_{k+1}, \mathbf{g}_*, \mathbf{g}_k, \mathbf{g}_{k+1})$

Let $\mathbf{z} = (\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_n)$ where

$$\mathbf{z}_0 = (f_*, f_k, f_{k+1}) \text{ and } \mathbf{z}_\ell = (x_*(\ell), x_k(\ell), x_{k+1}(\ell), g_*(\ell), g_k(\ell), g_{k+1}(\ell))$$

Require polynomials $p(\mathbf{z})$, $h_i(\mathbf{z})$ and $v_j(\mathbf{z})$ satisfy

$$\mathrm{pol}(\mathbf{z}) = \mathrm{pol}^0(\mathbf{z}_0) + \sum_{\ell=1}^{n} \mathrm{pol}^1(\mathbf{z}_\ell)$$

- Fulfilled when polynomials are linear in $f$'s and in the inner products of $\mathbf{x}$'s and $\mathbf{g}$'s e.g.

$$f_k - f_* - \frac{L}{2(L-\mu)}\left[\frac{1}{L}\|\mathbf{g}_k\|^2 + \mu\|\mathbf{x}_k - \mathbf{x}_*\|^2 + 2\frac{\mu}{L}\mathbf{g}_k^\top(\mathbf{x}_{k+1} - \mathbf{x}_k)\right] \geq 0$$

# Multivariate case

**Univariate case**

$$p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \end{pmatrix}^{\top} \begin{pmatrix} 0 & & \\ & Q_0 & \\ & & Q_1 \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})$$

**Univariate case**

$$p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \end{pmatrix}^\top \begin{pmatrix} 0 & & \\ & Q_0 & \\ & & Q_1 \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})$$

**Multivariate case**

$$p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_n \end{pmatrix}^\top \begin{pmatrix} 0 & & \\ & Q_0 & \\ & & Q_1 \otimes I_n \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_n \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})$$

**Univariate case**

$$p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \end{pmatrix}^\top \begin{pmatrix} 0 & & \\ & Q_0 & \\ & & Q_1 \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})$$

**Multivariate case**

$$p(\mathbf{z}) = \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_n \end{pmatrix}^\top \begin{pmatrix} 0 & & \\ & Q_0 & \\ & & Q_1 \otimes I_n \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{z}_0 \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_n \end{pmatrix} + \sum_i \sigma_i h_i(\mathbf{z}) + \sum_j \theta_j v_j(\mathbf{z})$$

SOS cert for univariate case $\implies$ SOS cert for multivariate case

# Obtaining the SOS certificate

$$t^*(f_k - f_*) - (f_{k+1} - f_*) \overset{(1)}{=} \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q^* \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i^* h_i(\mathbf{z}) + \sum_j \theta_j^* v_j(\mathbf{z})$$

1. Verifying the solution to be feasible

# Obtaining the SOS certificate

$$t^*(f_k - f_*) - (f_{k+1} - f_*) \overset{(1)}{=} \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q^* \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i^* h_i(\mathbf{z}) + \sum_j \theta_j^* v_j(\mathbf{z})$$
$$\overset{(2)}{\geq} \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q^* \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}$$

1. Verifying the solution to be feasible
2. Since $\sigma_i^*, h_i(\mathbf{z}) \geq 0$ and $v_j(\mathbf{z}) = 0$

# Obtaining the SOS certificate

$$
\begin{aligned}
t^*(f_k - f_*) - (f_{k+1} - f_*) &\overset{(1)}{=} \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q^* \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} + \sum_i \sigma_i^* h_i(\mathbf{z}) + \sum_j \theta_j^* v_j(\mathbf{z}) \\
&\overset{(2)}{\geq} \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix}^\top Q^* \begin{pmatrix} 1 \\ \mathbf{z} \end{pmatrix} \\
&\overset{(3)}{=} \frac{\mu}{4} \left( \frac{\|\mathbf{q}_1\|^2}{1 + \sqrt{\mu/L}} + \frac{\|\mathbf{q}_2\|^2}{1 - \sqrt{\mu/L}} \right)
\end{aligned}
$$

1. Verifying the solution to be feasible
2. Since $\sigma_i^*, h_i(\mathbf{z}) \geq 0$ and $v_j(\mathbf{z}) = 0$
3. Expanding the quadratic term into its SOS decomposition

# Revisiting the PEP

# The PEP as an SDP

$$\begin{aligned} \max\ & f_N - f_* \\ \text{s.t.}\ & f \in \mathcal{F} \\ & \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right),\ \ k = 0, \ldots, N-1 \\ & f_0 - f_* \le R \end{aligned}$$

# The PEP as an SDP

$$
\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t. } & f \in \mathcal{F} \\
& \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right), \ \ k = 0, \ldots, N-1 \\
& f_0 - f_* \leq R
\end{aligned}
$$

- Assume black-box model using oracle $\mathcal{O}_f$

# The PEP as an SDP

$$
\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t. } & \exists f \in \mathcal{F} \text{ s.t. } \mathcal{O}_f(\mathbf{x}_i) = \{f_i, \mathbf{g}_i\}\ \forall\ i \\
& \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right),\ \ k = 0, \ldots, N-1 \\
& f_0 - f_* \leq R
\end{aligned}
$$

- Assume black-box model using oracle $\mathcal{O}_f$

# The PEP as an SDP

$$
\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t. } & \exists f \in \mathcal{F} \text{ s.t. } \mathcal{O}_f(\mathbf{x}_i) = \{f_i, \mathbf{g}_i\} \ \forall\ i \\
& \mathbf{x}_{k+1} = \mathcal{A}\left(\mathbf{x}_k, f_k, \mathbf{g}_k\right),\ \ k = 0, \ldots, N-1 \\
& f_0 - f_* \leq R
\end{aligned}
$$

- Assume black-box model using oracle $\mathcal{O}_f$
- Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

# The PEP as an SDP

$$
\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t. } & h_i(\mathbf{z}) \geq 0,\ \forall i \\
& v_j(\mathbf{z}) = 0,\ \forall j \\
& f_0 - f_* \leq R
\end{aligned}
$$

- Assume black-box model using oracle $\mathcal{O}_f$
- Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$

# The PEP as an SDP

$$
\begin{aligned}
\max \ & f_N - f_* \\
\text{s.t.} \ & h_i(\mathbf{z}) \geq 0, \ \forall i \\
& v_j(\mathbf{z}) = 0, \ \forall j \\
& f_0 - f_* \leq R
\end{aligned}
$$

- Assume black-box model using oracle $\mathcal{O}_f$
- Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$
  - Linear in $f$'s and in the inner products of $\mathbf{x}$'s and $\mathbf{g}$'s $\implies$ Gram matrices

# The PEP as an SDP

$$
\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t.}\ & h_i(\mathbf{z}) \geq 0,\ \forall i \\
& v_j(\mathbf{z}) = 0,\ \forall j \\
& f_0 - f_* \leq R
\end{aligned}
$$

- Assume black-box model using oracle $\mathcal{O}_f$
- Identify polynomial inequalities $h_i(\mathbf{z}) \geq 0$ and equalities $v_j(\mathbf{z}) = 0$ necessarily satisfied given $\mathcal{F}$ and $\mathcal{A}$
  - Linear in $f$'s and in the inner products of $\mathbf{x}$'s and $\mathbf{g}$'s $\implies$ Gram matrices
- Formulated as an SDP

# Relation between PEP and SOS

$$
\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t. } & \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \\
& f \in \mathcal{F} \\
& f_0 - f_* \leq R
\end{aligned}
\qquad \text{(PEP)}
$$

# Relation between PEP and SOS

$$\begin{aligned} \max\ & f_N - f_* \\ \text{s.t.}\ & \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \\ & f \in \mathcal{F} \\ & f_0 - f_* \le R \end{aligned} \tag{PEP}$$

$\downarrow$

$N$-step PEP-SDP

# Relation between PEP and SOS

$$\begin{aligned} \max\ & f_N - f_* \\ \text{s.t. } & \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \\ & f \in \mathcal{F} \\ & f_0 - f_* \le R \end{aligned} \tag{PEP}$$

$\downarrow$

$N$-step PEP-SDP

$\downarrow$

1-step PEP-SDP

# Relation between PEP and SOS

$$
\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t. } & \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \\
& f \in \mathcal{F} \\
& f_0 - f_* \leq R
\end{aligned}
\tag{PEP}
$$

$$
\begin{aligned}
\min\ & t \\
\text{s.t. } & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0, \\
& \forall\ \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \text{ and} \\
& f \in \mathcal{F}
\end{aligned}
\tag{SOS}
$$

$\downarrow$

$N$-step PEP-SDP

$\downarrow$

1-step PEP-SDP

# Relation between PEP and SOS

$$
\begin{aligned}
\max\ & f_N - f_* \\
\text{s.t.}\ & \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \\
& f \in \mathcal{F} \\
& f_0 - f_* \leq R
\end{aligned}
\tag{PEP}
$$

$$
\begin{aligned}
\min\ & t \\
\text{s.t.}\ & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0, \\
& \quad \forall\ \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \text{ and} \\
& \quad f \in \mathcal{F}
\end{aligned}
\tag{SOS}
$$

$\downarrow$

$N$-step PEP-SDP

$\downarrow$

1-step PEP-SDP

$\downarrow$

Deg-$d$ SOS-SDP

# Relation between PEP and SOS

$$\begin{aligned} \max\ & f_N - f_* \\ \text{s.t.}\ & \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \\ & f \in \mathcal{F} \\ & f_0 - f_* \leq R \end{aligned} \tag{PEP}$$

$$\begin{aligned} \min\ & t \\ \text{s.t.}\ & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0, \\ & \forall\ \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \text{ and} \\ & f \in \mathcal{F} \end{aligned} \tag{SOS}$$

$\downarrow$

$N$-step PEP-SDP

$\downarrow$

1-step PEP-SDP

$\downarrow$

Deg-$d$ SOS-SDP

$\downarrow$

Deg-1 SOS-SDP

# Relation between PEP and SOS

$$\begin{aligned} \max\ & f_N - f_* \\ \text{s.t. } & \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \\ & f \in \mathcal{F} \\ & f_0 - f_* \leq R \end{aligned} \qquad \text{(PEP)}$$

$$\begin{aligned} \min\ & t \\ \text{s.t. } & t(f_k - f_*) - (f_{k+1} - f_*) \geq 0, \\ & \forall\ \{\mathbf{x}_i\} \text{ generated by } \mathcal{A} \text{ and} \\ & f \in \mathcal{F} \end{aligned} \qquad \text{(SOS)}$$

$\downarrow$ $\qquad\qquad$ $\downarrow$

$N$-step PEP-SDP $\qquad\qquad$ Deg-$d$ SOS-SDP

$\downarrow$ $\qquad\qquad$ $\downarrow$

1-step PEP-SDP $\Longleftarrow$ Dual $\Longrightarrow$ Deg-1 SOS-SDP

# Results

# Summary of Results

Form of bounds: Performance metric($\mathbf{x}_{k+1}$) $\leq t^*$ Performance metric($\mathbf{x}_k$)

Set $\epsilon$ and $\eta$ to be parameters related to the Armijo rule, and

$$\rho_\gamma := \max\{|1-\gamma\mu|, |1-\gamma L|\}$$

| **Method** | **Step size $\gamma$** | **Metric** | **Rate $t^*$** | **Known?** |
|---|---|---|---|---|
| GD | ELS | $f_k - f_*$ | $\left(\frac{L-\mu}{L+\mu}\right)^2$ | de Klerk et al. (2017) |

# Summary of Results

Form of bounds: Performance metric($\mathbf{x}_{k+1}$) $\leq t^*$ Performance metric($\mathbf{x}_k$)

Set $\epsilon$ and $\eta$ to be parameters related to the Armijo rule, and

$$\rho_\gamma := \max\{|1-\gamma\mu|, |1-\gamma L|\}$$

| **Method** | **Step size $\gamma$** | **Metric** | **Rate $t^*$** | **Known?** |
|---|---|---|---|---|
| GD | ELS | $f_k - f_*$ | $\left(\frac{L-\mu}{L+\mu}\right)^2$ | de Klerk et al. (2017) |
| | $\gamma \in (0, \frac{2}{L})$ | $\lVert\mathbf{x}_k - \mathbf{x}_*\rVert^2$ | $\rho_\gamma^2$ | Lessard et al. (2016) |

# Summary of Results

Form of bounds: Performance metric($\mathbf{x}_{k+1}$) $\leq t^*$ Performance metric($\mathbf{x}_k$)

Set $\epsilon$ and $\eta$ to be parameters related to the Armijo rule, and

$$\rho_\gamma := \max\{|1-\gamma\mu|, |1-\gamma L|\}$$

| **Method** | **Step size $\gamma$** | **Metric** | **Rate $t^*$** | **Known?** |
|---|---|---|---|---|
| GD | ELS | $f_k - f_*$ | $\left(\frac{L-\mu}{L+\mu}\right)^2$ | de Klerk et al. (2017) |
| | $\gamma \in (0, \frac{2}{L})$ | $\|\mathbf{x}_k - \mathbf{x}_*\|^2$ | $\rho_\gamma^2$ | Lessard et al. (2016) |
| | Armijo rule | $f_k - f_*$ | $1 - \frac{4\mu\epsilon(1-\epsilon)}{\eta L}$ | New |

# GD with Armijo-terminated Line Search

Our result $t_{\mathrm{new}}$ vs. known contraction factors $t_{\mathrm{LY}}$ (Luenberger and Ye, 2016) and $t_{\mathrm{nemi}}$ (Nemirovski, 1999):

$$t_{\mathrm{new}} = 1 - \frac{4\epsilon(1-\epsilon)}{\eta\kappa} \quad t_{\mathrm{LY}} = 1 - \frac{2\epsilon}{\eta\kappa} \quad t_{\mathrm{nemi}} = \frac{\kappa - (2-\epsilon^{-1})(1-\epsilon)\eta^{-1}}{\kappa + (\epsilon^{-1}-1)\eta^{-1}}$$

# GD with Armijo-terminated Line Search

Our result $t_{\text{new}}$ vs. known contraction factors $t_{\text{LY}}$ (Luenberger and Ye, 2016) and $t_{\text{nemi}}$ (Nemirovski, 1999):

$$t_{\text{new}} = 1 - \frac{4\epsilon(1-\epsilon)}{\eta\kappa} \quad t_{\text{LY}} = 1 - \frac{2\epsilon}{\eta\kappa} \quad t_{\text{nemi}} = \frac{\kappa - (2-\epsilon^{-1})(1-\epsilon)\eta^{-1}}{\kappa + (\epsilon^{-1}-1)\eta^{-1}}$$

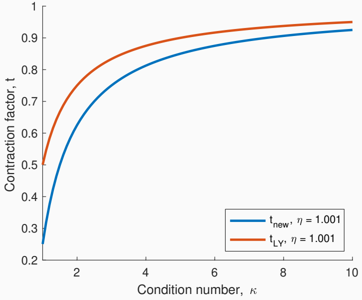



**Figure 2:** $t_{\text{new}}$ and $t_{\text{LY}}$ for $\epsilon = 0.25$

Condition number, $\kappa = L/\mu$

# GD with Armijo-terminated Line Search

Our result $t_{\text{new}}$ vs. known contraction factors $t_{\text{LY}}$ (Luenberger and Ye, 2016) and $t_{\text{nemi}}$ (Nemirovski, 1999):

$$t_{\text{new}} = 1 - \frac{4\epsilon(1-\epsilon)}{\eta\kappa} \quad t_{\text{LY}} = 1 - \frac{2\epsilon}{\eta\kappa} \quad t_{\text{nemi}} = \frac{\kappa - (2-\epsilon^{-1})(1-\epsilon)\eta^{-1}}{\kappa + (\epsilon^{-1}-1)\eta^{-1}}$$

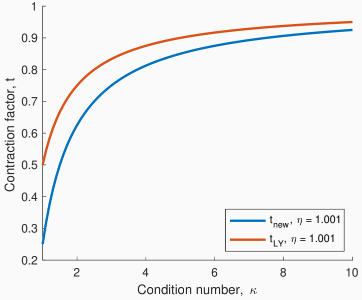



**Figure 2:** $t_{\text{new}}$ and $t_{\text{LY}}$ for $\epsilon = 0.25$

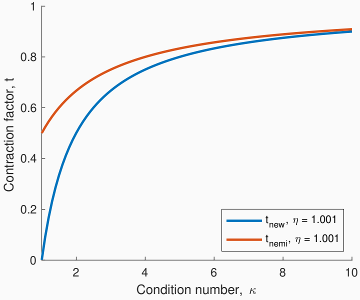



**Figure 3:** $t_{\text{new}}$ and $t_{\text{nemi}}$ for $\epsilon = 0.5$

Condition number, $\kappa = L/\mu$

**Composite convex minimization problem**

$$\min_{\mathbf{x}\in\mathbb{R}^n} \{f(\mathbf{x}) := a(\mathbf{x}) + b(\mathbf{x})\},$$

where $a \in \mathcal{F}_{\mu,L}$ and $b$ is closed, convex and proper.

# A new function class: Composite functions

**Composite convex minimization problem**

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ f(\mathbf{x}) := a(\mathbf{x}) + b(\mathbf{x}) \right\},$$

where $a \in \mathcal{F}_{\mu,L}$ and $b$ is closed, convex and proper.

**Further assumptions**

Assume that the proximal operator of $b$,

$$\mathbf{prox}_{\gamma b}(\mathbf{x}) := \arg\min_{\mathbf{y}\in\mathbb{R}^n} \left\{ \gamma b(\mathbf{y}) + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 \right\},$$

exists at every $\mathbf{x}$

# Algorithm

**Proximal gradient method (PGM) with constant step size** $\gamma > 0$

$$\mathbf{x}_{k+1} = \mathbf{prox}_{\gamma b}\left(\mathbf{x}_k - \gamma \nabla a(\mathbf{x}_k)\right),$$

where $0 \leq \gamma \leq \frac{2}{L}$

# Algorithm

**Proximal gradient method (PGM) with constant step size** $\gamma > 0$

$$\mathbf{x}_{k+1} = \mathbf{prox}_{\gamma b}\left(\mathbf{x}_k - \gamma \nabla a(\mathbf{x}_k)\right),$$

where $0 \leq \gamma \leq \frac{2}{L}$

**PGM with exact line search**

$$\gamma_k = \arg\min_{\gamma > 0} f\left[\mathbf{prox}_{\gamma b}\left(\mathbf{x}_k - \gamma \nabla a(\mathbf{x}_k)\right)\right]$$
$$\mathbf{x}_{k+1} = \mathbf{prox}_{\gamma b}\left(\mathbf{x}_k - \gamma_k \nabla a(\mathbf{x}_k)\right)$$

# Summary of Results

Form of bounds: Performance metric($\mathbf{x}_{k+1}$) $\leq t^*$ Performance metric($\mathbf{x}_k$)

Let $\mathbf{g}_k$ denotes a (sub)gradient of $f$ at $\mathbf{x}_k$, and

$$\rho_\gamma := \max\{|1-\gamma\mu|, |1-\gamma L|\}$$

| Method | Step size $\gamma$ | Metric | Rate $t^*$ | Known? |
|---|---|---|---|---|
| PGM | $\gamma \in (0, \frac{2}{L})$ | $\|\mathbf{g}_k\|^2$ | $\rho_\gamma^2$ | Taylor et al. (2018) |

# Summary of Results

Form of bounds: Performance metric($\mathbf{x}_{k+1}$) $\leq t^*$ Performance metric($\mathbf{x}_k$)

Let $\mathbf{g}_k$ denotes a (sub)gradient of $f$ at $\mathbf{x}_k$, and

$$\rho_\gamma := \max\{|1-\gamma\mu|, |1-\gamma L|\}$$

| **Method** | **Step size $\gamma$** | **Metric** | **Rate $t^*$** | **Known?** |
|---|---|---|---|---|
| PGM | $\gamma \in (0, \frac{2}{L})$ | $\|\mathbf{g}_k\|^2$ | $\rho_\gamma^2$ | Taylor et al. (2018) |
| | ELS | $f_k - f_*$ | $\left(\frac{L-\mu}{L+\mu}\right)^2$ | Taylor et al. (2018) |

# Conclusion

# Conclusion

Introduced an SDP hierarchy for bounding convergence rates via SOS certificates
$\Rightarrow$ First level coincides with the PEP

**Future work**

- Other function classes
- Other algorithms
- Understanding when SOS certificates cannot be found

# Conclusion

Introduced an SDP hierarchy for bounding convergence rates via SOS certificates
$\Rightarrow$ First level coincides with the PEP

**Future work**

- Other function classes
- Other algorithms
- Understanding when SOS certificates cannot be found

**Links**

arXiv paper: `http://arxiv.org/abs/1906.04648`
github codes: `https://github.com/sandratsy/SumsOfSquares`

Thank you!

# References

de Klerk, E., Glineur, F., and Taylor, A. B. (2017). On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions. *Optimization Letters*, 11(7):1185–1199.

Drori, Y. and Teboulle, M. (2014). Performance of first-order methods for smooth convex minimization: A novel approach. *Mathematical Programming*, 145(1):451–482.

Lasserre, J. (2007). A sum of squares approximation of nonnegative polynomials. *SIAM Review*, 49(4):651–669.

Lessard, L., Recht, B., and Packard, A. (2016). Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM Journal on Optimization*, 26(1):57–95.

# References II


Luenberger, D. G. and Ye, Y. (2016). *Linear and Nonlinear Programming*. Springer International Publishing, 4th edition.

Nemirovski, A. (1999). Optimization II: Numerical methods for nonlinear continuous optimization. `https://www2.isye.gatech.edu/~nemirovs/Lect_OptII.pdf`.

Parrilo, P. A. (2003). Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96(2):293–320.

Taylor, A., Hendrickx, J., and Glineur, F. (2017a). Exact worst-case performance of first-order methods for composite convex optimization. *SIAM Journal on Optimization*, 27(3):1283–1313.

Taylor, A. B., Hendrickx, J. M., and Glineur, F. (2017b). Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161(1):307–345.

# References III

Taylor, A. B., Hendrickx, J. M., and Glineur, F. (2018). Exact worst-case convergence rates of the proximal gradient method for composite convex minimization. *Journal of Optimization Theory and Applications*, 178(2):455–476.