Sums of squares:
from algebra to analysis

Francis Bach

INRIA - Ecole Normale Supérieure, Paris, France

Joint work with Alessandro Rudi,
Ulysse Marteau-Ferey, and Blake Woodworth

Brainpop online seminar - February 15, 2023

Sums of squares: from algebra to analysis

One-minute summary

- **Minimization of continuous functions on** $[0,1]^d$
  - From polynomials to trigonometric polynomials
  - Simpler "more intuitive" sum-of-squares formulations

Sums of squares: from algebra to analysis

One-minute summary

- **Minimization of continuous functions on** $[0,1]^d$
  - From polynomials to trigonometric polynomials
  - Simpler "more intuitive" sum-of-squares formulations

- **From bound on degree to smoothness**
  - Allows for explicit convergence rates
    (up to exponential in the degree of the finite hierarchy)
  - Allows for zero-th order oracle with kernel methods

Optimization of trigonometric polynomials

- **Trigonometric polynomials:** $f(x) = \sum_{\omega \in \mathbb{Z}^d} \hat{f}(\omega) e^{2i\pi\omega^\top x}$

Optimization of trigonometric polynomials

- **Trigonometric polynomials:** $f(x) = \sum_{\omega \in \mathbb{Z}^d} \hat{f}(\omega) e^{2i\pi\omega^\top x}$
  - Fourier series $\hat{f}(\omega) = \int_{[0,1]^d} f(x) e^{-2i\pi\omega^\top x} dx \in \mathbb{C}$
  - Real values for $f \;\Leftrightarrow\; \forall \omega \in \mathbb{Z}^d,\ \hat{f}(-\omega) = \hat{f}(\omega)^*$
    (polynomial in $\cos 2\pi x_j$ and $\sin 2\pi x_j$, $j \in \{1, \dots, d\}$)
  - Degree $= \max\big\{\|\omega\|_\infty,\ \hat{f}(\omega) \neq 0\big\}$

Optimization of trigonometric polynomials

- **Trigonometric polynomials:** $f(x) = \sum_{\omega \in \mathbb{Z}^d} \hat{f}(\omega) e^{2i\pi\omega^\top x}$

- **Representation as quadratic forms**
  - Feature map $\varphi : [0,1]^d \to \mathbb{C}^m$: $\varphi(x)_\omega = \hat{q}(\omega) e^{2i\pi\omega^\top x}$, for $\omega \in \Omega$
  - If $\Omega = \{\omega \in \mathbb{Z}^d, \ \|\omega\|_\infty \leqslant r\}$, then $m = |\Omega| = (2r+1)^d$
  - Normalization: $\|\varphi(x)\|^2 = \sum_{\omega \in \Omega} |\hat{q}(\omega)|^2 = 1$

Optimization of trigonometric polynomials

- **Trigonometric polynomials:** $f(x) = \sum_{\omega \in \mathbb{Z}^d} \hat{f}(\omega) e^{2i\pi\omega^\top x}$

- **Representation as quadratic forms**
  - Feature map $\varphi : [0,1]^d \to \mathbb{C}^m$: $\varphi(x)_\omega = \hat{q}(\omega) e^{2i\pi\omega^\top x}$, for $\omega \in \Omega$
  - If $\Omega = \{\omega \in \mathbb{Z}^d, \ \|\omega\|_\infty \leqslant r\}$, then $m = |\Omega| = (2r+1)^d$
  - Normalization: $\|\varphi(x)\|^2 = \sum_{\omega \in \Omega} |\hat{q}(\omega)|^2 = 1$
  - With $F \in \mathbb{C}^{m \times m}$ Hermitian

$$f(x) = \varphi(x)^* F \varphi(x) = \sum_{\omega, \omega' \in \Omega} F_{\omega\omega'} \hat{q}(\omega) \hat{q}(\omega')^* \cdot e^{2i\pi(\omega - \omega')^\top x}$$

  - Represents all trigonometric polynomials of degree $2r$
  - F not uniquely defined

Optimization of trigonometric polynomials

- **Generic problem on** $\mathcal{X} = [0,1]^d$: $\displaystyle\min_{x \in \mathcal{X}} f(x) = \varphi(x)^* F \varphi(x)$
  - Normalized feature map $\varphi : \mathcal{X} \to \mathbb{C}^m$ such that $\|\varphi(x)\|^2 = 1$

Optimization of trigonometric polynomials

- **Generic problem on** $\mathcal{X} = [0,1]^d$: $\min_{x \in \mathcal{X}} f(x) = \varphi(x)^* F \varphi(x)$
  - Normalized feature map $\varphi : \mathcal{X} \to \mathbb{C}^m$ such that $\|\varphi(x)\|^2 = 1$

- **Sum-of-squares relaxations**
  - Lasserre (2001); Parrilo (2003)
  - Books (Lasserre, 2010; Parrilo et al., 2013; Dumitrescu, 2007; Henrion et al., 2020)
  - Review paper (Laurent, 2009)

Optimization of trigonometric polynomials

- **Generic problem on** $\mathcal{X} = [0,1]^d$: $\min_{x \in \mathcal{X}} f(x) = \varphi(x)^* F \varphi(x)$
  - Normalized feature map $\varphi : \mathcal{X} \to \mathbb{C}^m$ such that $\|\varphi(x)\|^2 = 1$

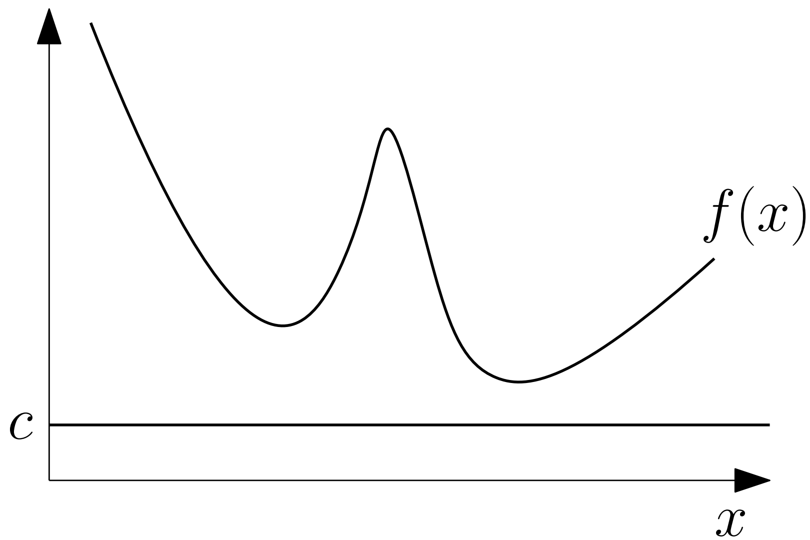
- **Sum-of-squares relaxations**
  - Lasserre (2001); Parrilo (2003)
  - Books (Lasserre, 2010; Parrilo et al., 2013; Dumitrescu, 2007; Henrion et al., 2020)
  - Review paper (Laurent, 2009)

- **Simplification**
  - Assumption: $\mathcal{X}$ is a (very) "simple" set
  - From polynomials to trigonometric polynomials (will be lifted)

Convex relaxation: the SOS view

- **Exact reformulation of minimization problem**

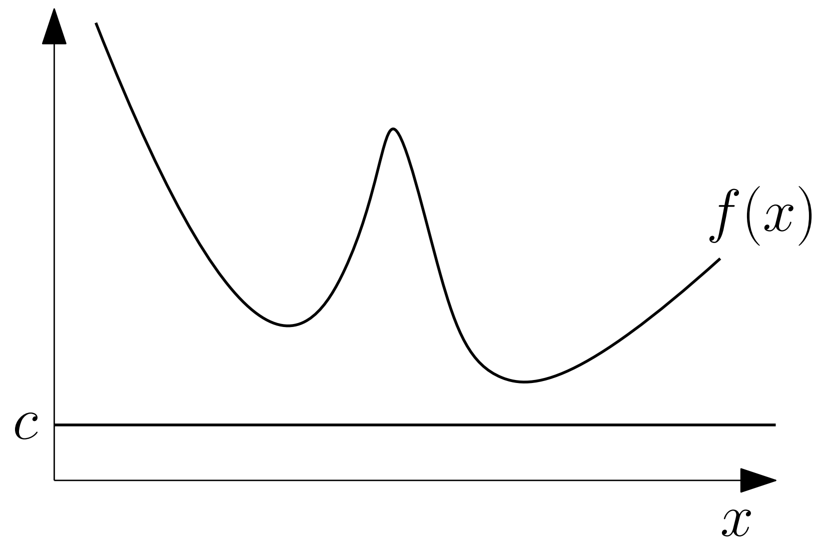
$$\min_{x \in \mathcal{X}} f(x) = \max_{c \in \mathbb{R}} c \quad \text{such that} \quad \forall x \in \mathcal{X},\ f(x) - c \geqslant 0$$



Convex relaxation: the SOS view

- **Exact reformulation of minimization problem**

$$\min_{x \in \mathcal{X}} f(x) = \max_{c \in \mathbb{R}} c \quad \text{such that} \quad \forall x \in \mathcal{X}, \ f(x) - c \geqslant 0$$



- **SOS relaxation:** replace $f(x) - c \geqslant 0$ by $f(x) - c = \varphi(x)^* A \varphi(x)$ with A Hermitian positive semi-definite ($A \succcurlyeq 0$)
  - If $A = \sum_{i=1}^m \lambda_i u_i u_i^*$, then $\varphi(x)^* A \varphi(x) = \sum_{i=1}^m \left| \lambda_i^{1/2} u_i^* \varphi(x) \right|^2$

Convex relaxation: the SOS view

- **Relaxed problem for minimizing** $f(x) = \varphi(x)^* F \varphi(x)$ **:**

$$\max_{c \in \mathbb{R},\ A \succcurlyeq 0} \ c \quad \text{such that} \quad \forall x \in \mathcal{X},\ f(x) - c = \varphi(x)^* A \varphi(x)$$

Convex relaxation: the SOS view

- **Relaxed problem for minimizing** $f(x) = \varphi(x)^* F \varphi(x)$ **:**

$$\max_{c \in \mathbb{R},\ A \succcurlyeq 0} c \quad \text{such that} \quad \forall x \in \mathcal{X},\ f(x) - c = \varphi(x)^* A \varphi(x)$$

$$= \max_{c \in \mathbb{R},\ A \succcurlyeq 0} c \quad \text{such that} \quad \forall x \in \mathcal{X},\ \mathrm{tr}\big[\varphi(x)\varphi(x)^*(F - cI - A)\big] = 0$$

Convex relaxation: the SOS view

- **Relaxed problem for minimizing $f(x) = \varphi(x)^* F \varphi(x)$:**

$$\max_{c\in\mathbb{R},\ A\succcurlyeq 0} c \quad \text{such that} \quad \forall x \in \mathcal{X},\ f(x) - c = \varphi(x)^* A \varphi(x)$$

$$= \max_{c\in\mathbb{R},\ A\succcurlyeq 0} c \quad \text{such that} \quad \forall x \in \mathcal{X},\ \mathrm{tr}\left[\varphi(x)\varphi(x)^*(F - cI - A)\right] = 0$$

$$= \max_{c\in\mathbb{R},\ A\succcurlyeq 0} c \quad \text{such that} \quad F - cI - A + Y = 0, \quad \text{with } Y \in \mathcal{V}^\perp$$

where $\mathcal{V} = \mathrm{span}\big(\{\varphi(x)\varphi(x)^*,\ x \in \mathcal{X}\}\big)$

$\mathcal{V}$ = multivariate Toeplitz matrices

Convex relaxation: the SOS view

- **Relaxed problem for minimizing** $f(x) = \varphi(x)^* F \varphi(x)$ **:**

$$\max_{c \in \mathbb{R},\ A \succcurlyeq 0} \ c \quad \text{such that} \quad \forall x \in \mathcal{X},\ f(x) - c = \varphi(x)^* A \varphi(x)$$

$$= \max_{c \in \mathbb{R},\ A \succcurlyeq 0} \ c \quad \text{such that} \quad \forall x \in \mathcal{X},\ \operatorname{tr}\big[\varphi(x)\varphi(x)^*(F - cI - A)\big] = 0$$

$$= \max_{c \in \mathbb{R},\ A \succcurlyeq 0} \ c \quad \text{such that} \quad F - cI - A + Y = 0, \quad \text{with } Y \in \mathcal{V}^\perp$$

where $\mathcal{V} = \operatorname{span}\big(\{\varphi(x)\varphi(x)^*,\ x \in \mathcal{X}\}\big)$
$\mathcal{V}$ = multivariate Toeplitz matrices

- **Optimizing over** c **and** A**:**

$$\max_{Y \in \mathcal{V}^\perp} \ \lambda_{\min}(F + Y)$$

  – Link with spectral relaxation ($Y = 0$)

Convex relaxation: the moment view

- **Dual exact reformulation of minimization problem**

$$\min_{\mu \in \mathcal{P}(\mathcal{X})} \int_{\mathcal{X}} f(x) d\mu(x) = \operatorname{tr}\Big[F\Big(\int_{\mathcal{X}} \varphi(x)\varphi(x)^* d\mu(x)\Big)\Big]$$

  – with $\mathcal{P}(\mathcal{X})$ = set of probability measures on $\mathcal{X}$

Convex relaxation: the moment view

- **Dual exact reformulation of minimization problem**

$$\min_{\mu \in \mathcal{P}(\mathcal{X})} \int_{\mathcal{X}} f(x) d\mu(x) = \operatorname{tr}\Big[F\Big(\int_{\mathcal{X}} \varphi(x)\varphi(x)^* d\mu(x)\Big)\Big]$$

  – with $\mathcal{P}(\mathcal{X})$ = set of probability measures on $\mathcal{X}$

- **Equivalent reformulation**: $\min_{\Sigma \in \mathcal{K}} \operatorname{tr}[F\Sigma]$

  – with $\mathcal{K}$ closure of convex hull of $\{\varphi(x)\varphi(x)^*, \ x \in \mathcal{X}\}$

Convex relaxation: the moment view

- **Dual exact reformulation of minimization problem**

$$\min_{\mu \in \mathcal{P}(\mathcal{X})} \int_{\mathcal{X}} f(x) d\mu(x) = \operatorname{tr}\Big[F\Big(\int_{\mathcal{X}} \varphi(x)\varphi(x)^* d\mu(x)\Big)\Big]$$

  – with $\mathcal{P}(\mathcal{X})$ = set of probability measures on $\mathcal{X}$

- **Equivalent reformulation**: $\min_{\Sigma \in \mathcal{K}} \operatorname{tr}[F\Sigma]$

  – with $\mathcal{K}$ closure of convex hull of $\{\varphi(x)\varphi(x)^*, \ x \in \mathcal{X}\}$

- **Relaxation using outer approximation** $\widehat{\mathcal{K}} \supset \mathcal{K}$

  – Preserve affine hull and add positivity constraint

$$\widehat{\mathcal{K}} = \big\{\Sigma \in \mathbb{C}^{m \times m}, \ \Sigma \in \mathcal{V}, \ \operatorname{tr}[\Sigma] = 1, \ \Sigma \succcurlyeq 0\big\}$$

Tightness of SOS relaxations

- **Two equivalent views**

(1) Are all non-negative functions sums-of-squares?
(2) Is $\widehat{\mathcal{K}} = \mathcal{K}$?

Tightness of SOS relaxations

- **Two equivalent views**

(1) Are all non-negative functions sums-of-squares?
(2) Is $\widehat{\mathcal{K}} = \mathcal{K}$?

- **Univariate polynomials ($d = 1$)**
  - Tight relaxation (Fejér, 1916; Riesz, 1916; Nesterov, 2000)
  - Elementary proof based on polynomial factorization
  - NB: spectral relaxation only converges at $O(1/s)$ with $s =$ degree (Grenander and Szegö, 1958)

Tightness of SOS relaxations

- **Two equivalent views**

(1) Are all non-negative functions sums-of-squares?
(2) Is $\widehat{\mathcal{K}} = \mathcal{K}$?

- **Univariate polynomials ($d = 1$)**
  - Tight relaxation (Fejér, 1916; Riesz, 1916; Nesterov, 2000)
  - Elementary proof based on polynomial factorization
  - NB: spectral relaxation only converges at $O(1/s)$ with $s =$ degree (Grenander and Szegö, 1958)

- **What about multivariate polynomials ($d > 1$)?**
  - Bad and good news...

Tightness of SOS relaxations
Multivariate trigonometric polynomials

- **Not all non-negative trigonometric polynomials are SOSs**
  - Generic construction (Naftalovich and Schreiber, 1985)
  - Based on Motzkin counter-example

$$f(x) = M(1 - \cos 2\pi x_1, 1 - \cos 2\pi x_2, 1 - \cos 2\pi x_3)$$

with $M(y_1, y_2, y_3) = y_1^2 y_2 + y_1 y_2^2 + y_3^3 - 3 y_1 y_2 y_3$

Tightness of SOS relaxations
Multivariate trigonometric polynomials

- **Not all non-negative trigonometric polynomials are SOSs**
  - Generic construction (Naftalovich and Schreiber, 1985)
  - Based on Motzkin counter-example

$$f(x) = M(1-\cos 2\pi x_1, 1-\cos 2\pi x_2, 1-\cos 2\pi x_3)$$

$$\text{with } M(y_1,y_2,y_3) = y_1^2 y_2 + y_1 y_2^2 + y_3^3 - 3y_1y_2y_3$$

- **All strictly positive polynomials are sums-of-squares**
  - See Putinar (1992); Megretski (2003)
  - Degrees not known a priori
  - Allows for hierarchies
  - NB: always finite convergence for $d = 2$ (Scheiderer, 2006)

Trigonometric polynomial hierarchies

- **Goal: minimize degree $2r$ trigonometric polynomial f**
  - Define $\varphi^{(s)} : [0,1]^d \to \mathbb{C}^{(2s+1)^d}$ with all Fourier exponentials of degree less than $s \geqslant r$
  - Represent f as quadratic form $f(x) = \varphi^{(s)}(x)^*(F^{(s)})\varphi^{(s)}(x)$
  - Solve the primal/dual pair of SOS relaxations, with values $c_*^{(s)}$

$$c_*^{(s)} \to \min_{x\in[0,1]^d} f(x) \quad \text{when } s \to +\infty$$

Trigonometric polynomial hierarchies

- **Goal: minimize degree $2r$ trigonometric polynomial f**
  - Define $\varphi^{(s)} : [0,1]^d \to \mathbb{C}^{(2s+1)^d}$ with all Fourier exponentials of degree less than $s \geqslant r$
  - Represent f as quadratic form $f(x) = \varphi^{(s)}(x)^*(F^{(s)})\varphi^{(s)}(x)$
  - Solve the primal/dual pair of SOS relaxations, with values $c_*^{(s)}$

$$c_*^{(s)} \to \min_{x\in[0,1]^d} f(x) \text{ when } s \to +\infty$$

- **How fast?**
  - Finite convergence often observed, and provable for locally well-behaved problems (Nie, 2014), with no rate
  - Existing bounds in $O(1/s^2)$ for other special cases (Fang and Fawzi, 2021; Laurent and Slot, 2022; Slot, 2022)

From trigonometric polynomials to polynomials

- **Representation of non-negative polynomials on** $[-1, 1]$
  - Given a polynomial P on $[-1, 1]$ of degree $2r$
  - Define $f(y) = P(\cos 2\pi y)$ a trigonometric polynomial on $[0, 1]$
  - f is non-negative if and only if $f(y) = \left| \sum_{|\omega| \leqslant r} \hat{g}(\omega) e^{2i\pi\omega y} \right|^2$

From trigonometric polynomials to polynomials

- **Representation of non-negative polynomials on** $[-1,1]$
  - Given a polynomial P on $[-1,1]$ of degree $2r$
  - Define $f(y) = P(\cos 2\pi y)$ a trigonometric polynomial on $[0,1]$
  - f is non-negative if and only if $f(y) = \Big| \sum_{|\omega| \leqslant r} \hat{g}(\omega) e^{2i\pi\omega y} \Big|^2$

- **Chebyshev polynomials** for $\omega > 0$
  - $\cos 2\pi\omega y = T_\omega(\cos 2\pi y)$ and $\sin 2\pi\omega y = U_{\omega-1}(\cos 2\pi y) \cdot \sin 2\pi y$
  - Can expand $f(y) = Q(\cos 2\pi y)^2 + R(\cos 2\pi y)^2 \cdot \sin^2 2\pi y$
  - With $\sin^2 2\pi y = 1 - \cos^2 2\pi y$, we have:

$$\forall x \in [-1,1], \quad f(x) = Q(x)^2 + R(x)^2 \cdot (1 - x^2)$$

  - Classical "Putinar" representation

From trigonometric polynomials to polynomials

- **Representation of non-negative polynomials on** $[-1,1]$
  - Given a polynomial P on $[-1,1]$ of degree $2r$
  - Define $f(y) = P(\cos 2\pi y)$ a trigonometric polynomial on $[0,1]$
  - f is non-negative if and only if $f(y) = \big| \sum_{|\omega| \leqslant r} \hat{g}(\omega) e^{2i\pi\omega y} \big|^2$

- **Chebyshev polynomials** for $\omega > 0$
  - $\cos 2\pi\omega y = T_\omega(\cos 2\pi y)$ and $\sin 2\pi\omega y = U_{\omega-1}(\cos 2\pi y) \cdot \sin 2\pi y$
  - Can expand $f(y) = Q(\cos 2\pi y)^2 + R(\cos 2\pi y)^2 \cdot \sin^2 2\pi y$
  - With $\sin^2 2\pi y = 1 - \cos^2 2\pi y$, we have:

$$\forall x \in [-1,1], \quad f(x) = Q(x)^2 + R(x)^2 \cdot (1 - x^2)$$

  - Classical "Putinar" representation

- **Extension to** $[-1,1]^d$: Schmüdgen (2017) representation

Convergence bounds with no assumptions

- **Theorem** (Bach and Rudi, 2022)
  - Assume $s \geqslant 3r$, and define $\|f\|_{\mathrm{F}} = \sum_{\omega \in \mathbb{Z}^d} |\hat{f}(\omega)|$

$$0 \leqslant \min_{x \in [0,1]^d} f(x) - c_*^{(s)} \leqslant \|f - f_*\|_{\mathrm{F}} \cdot \left[\left(1 - \frac{6r^2}{s^2}\right)^{-d} - 1 \right] \sim 6 \|f - f_*\|_{\mathrm{F}} \cdot \frac{r^2 d}{s^2}$$

  - Proof based on Fang and Fawzi (2021)
  - Essentially the same result as Laurent and Slot (2022) with different notations and better constants

Convergence bounds with no assumptions

- **Theorem** (Bach and Rudi, 2022)
  - Assume $s \geqslant 3r$, and define $\|f\|_{\mathrm{F}} = \sum_{\omega \in \mathbb{Z}^d} |\hat{f}(\omega)|$

$$0 \leqslant \min_{x \in [0,1]^d} f(x) - c_*^{(s)} \leqslant \|f - f_*\|_{\mathrm{F}} \cdot \left[\left(1 - \frac{6r^2}{s^2}\right)^{-d} - 1 \right] \sim 6 \|f - f_*\|_{\mathrm{F}} \cdot \frac{r^2 d}{s^2}$$

  - Proof based on Fang and Fawzi (2021)
  - Essentially the same result as Laurent and Slot (2022) with different notations and better constants

- **Discussion**
  - Spectral relaxation only achieves $O(1/s)$
  - Is it optimal without further assumptions?
  - Can it be improved with further assumptions?

From bound on degree to smoothness

- **From algebra to analysis**
  - Trigonometric polynomials are C^∞ functions
  - Smoothness of f typically characterized by decay of $\hat{f}(\omega)$ for $\|\omega\| \to +\infty$
  - Support of Fourier series not precise enough

From bound on degree to smoothness

- **From algebra to analysis**
  - Trigonometric polynomials are C^∞ functions
  - Smoothness of f typically characterized by decay of $\hat{f}(\omega)$ for $\|\omega\| \to +\infty$
  - Support of Fourier series not precise enough

- **Using local optimality conditions**
  - Assumptions: ( ) f attains its minimum at a single point
    ( ) f is twice differentiable and $f''(x_*)$ invertible
  - Can be relaxed (Marteau-Ferey, Bach, and Rudi, 2022)

Decomposing non-negative C^p functions as sums-of-squares

- **Theorem** (Rudi, Marteau-Ferey, and Bach, 2020):
  - Assumptions: $f : [0,1]^d \to \mathbb{R}$ is C^p (p-th continuous derivatives)
    f has a unique minimum x_* located in $(0,1)^d$
    $f''(x_*)$ invertible

Decomposing non-negative C^p functions as sums-of-squares

- **Theorem** (Rudi, Marteau-Ferey, and Bach, 2020):
  - Assumptions: $f : [0,1]^d \to \mathbb{R}$ is C^p (p-th continuous derivatives)
    f has a unique minimum x_* located in $(0,1)^d$
    $f''(x_*)$ invertible
  - There exist $d+1$ functions $g_1, \dots, g_{d+1}$ in C^{p-2} such that

$$\forall x \in [0,1]^d, \; f(x) - f(x_*) = \sum_{i=1}^{d+1} g_i(x)^2$$

Decomposing non-negative C^p functions as sums-of-squares

- **Theorem** (Rudi, Marteau-Ferey, and Bach, 2020):
  - Assumptions: $f:[0,1]^d \to \mathbb{R}$ is C^p (p-th continuous derivatives)
    f has a unique minimum x_* located in $(0,1)^d$
    $f''(x_*)$ invertible
  - There exist $d+1$ functions $g_1,\dots,g_{d+1}$ in C^{p-2} such that

$$\forall x \in [0,1]^d,\ f(x) - f(x_*) = \sum_{i=1}^{d+1} g_i(x)^2$$

- **Proof technique**
  - Around x_*, Taylor formula with integral remainder $\Rightarrow$ d functions
  - Away from x_*, use the square root
  - Use partitions of unity to glue them

Consequence on convergence rate of hierarchies (Woodworth, Bach, and Rudi, 2022)

- **A trigonometric polynomial is a C^∞ function!**

$$f(x) - f(x_*) = \sum_{i=1}^{d+1} g_i(x)^2$$

  - With g_i's all C^∞
  - Let $\bar{g}_i(x) = \sum_{\|\omega\|_\infty \leqslant s} \hat{g}_i(\omega) e^{2i\pi\omega^\top x}$ (truncated version)
  - Property: for any order p, $\|g_i - \bar{g}_i\|_{\mathrm{F}} \leqslant \frac{c_p(g_i)}{s^p}$

Consequence on convergence rate of hierarchies (Woodworth, Bach, and Rudi, 2022)

- **A trigonometric polynomial is a C^∞ function!**

$$f(x) - f(x_*) = \sum_{i=1}^{d+1} g_i(x)^2$$

  - With g_i's all C^∞
  - Let $\bar{g}_i(x) = \sum_{\|\omega\|_\infty \leqslant s} \hat{g}_i(\omega) e^{2i\pi\omega^\top x}$ (truncated version)
  - Property: for any order p, $\|g_i - \bar{g}_i\|_{\mathrm{F}} \leqslant \frac{c_p(g_i)}{s^p}$

- **Lemma:** $\displaystyle \Big\| f - f(x_*) - \sum_{i=1}^{d+1} \bar{g}_i^2 \Big\|_{\mathrm{F}} \leqslant \sum_{i=1}^{d+1} \|g_i\|_{\mathrm{F}} \cdot \|g_i - \bar{g}_i\|_{\mathrm{F}}$

Consequence on convergence rate of hierarchies (Woodworth, Bach, and Rudi, 2022)

- **A trigonometric polynomial is a C^∞ function!**

$$f(x) - f(x_*) = \sum_{i=1}^{d+1} g_i(x)^2$$

  – With g_i's all C^∞
  – Let $\bar{g}_i(x) = \sum_{\|\omega\|_\infty \leqslant s} \hat{g}_i(\omega) e^{2i\pi\omega^\top x}$ (truncated version)
  – Property: for any order p, $\|g_i - \bar{g}_i\|_{\mathrm{F}} \leqslant \frac{c_p(g_i)}{s^p}$

- **Lemma:** $\displaystyle \Big\| f - f(x_*) - \sum_{i=1}^{d+1} \bar{g}_i^2 \Big\|_{\mathrm{F}} \leqslant \sum_{i=1}^{d+1} \|g_i\|_{\mathrm{F}} \cdot \|g_i - \bar{g}_i\|_{\mathrm{F}}$

- **Consequence**: For any p, up to a uniform error less than $\displaystyle \frac{c_p'(f)}{s^p}$, $f - f(x_*)$ is a sum of squares of polynomials of degree s

Exponential convergence rates

- **Theorem** (Bach and Rudi, 2022)
  - Assume unique minimizer with positive definite Hessian
  - For any $\xi \in (0, 1/2]$:

$$0 \leqslant \min_{x \in [0,1]^d} f(x) - c_*^{(s)} \leqslant \triangle_1 \exp\Big(- \Big(\frac{s}{\triangle_2}\Big)^{1+\xi}\Big),$$

  - Explicit dependence of $\triangle_1$ and $\triangle_2$ on all problem constants

Exponential convergence rates

- **Theorem** (Bach and Rudi, 2022)
  - Assume unique minimizer with positive definite Hessian
  - For any $\xi \in (0, 1/2]$:

$$0 \leqslant \min_{x \in [0,1]^d} f(x) - c_*^{(s)} \leqslant \triangle_1 \exp\Big(- \Big(\frac{s}{\triangle_2}\Big)^{1+\xi} \Big),$$

  - Explicit dependence of $\triangle_1$ and $\triangle_2$ on all problem constants

- **Proof technique**
  - Explicit control of the constants $c_p(g_i)$ and $c_p'(f)$
  - Bounding all derivatives of (matrix) square roots (Del Moral and Niclas, 2018) and partitions of unity (Israel, 2015)
  - Extensive use of Faà di Bruno's formula

Towards zero-th order oracles

- **Traditional SOS relaxations**
  - (trigonometric) polynomial f given by its coefficients

Towards zero-th order oracles

- **Traditional SOS relaxations**
  - (trigonometric) polynomial f given by its coefficients
- **Using zero-th order oracle for f or $\hat{f}$ for smooth functions**

Towards zero-th order oracles

- **Traditional SOS relaxations**
  - (trigonometric) polynomial f given by its coefficients
- **Using zero-th order oracle for f or $\hat{f}$ for smooth functions**
- **Option 1:** Compute approximation by (trigonometric) polynomial and optimize using SOS (see Novak, 2006)
  - Optimal in terms of number of calls to zero-th order oracle

Towards zero-th order oracles

- **Traditional SOS relaxations**
  - (trigonometric) polynomial f given by its coefficients
- **Using zero-th order oracle for f or $\hat{f}$ for smooth functions**
- **Option 1:** Compute approximation by (trigonometric) polynomial and optimize using SOS (see Novak, 2006)
  - Optimal in terms of number of calls to zero-th order oracle
- **Option 2:** Approximate and optimize simultaneously
  - Efficient algorithms (Rudi, Marteau-Ferey, and Bach, 2020)
  - Certificates of optimality (Woodworth, Bach, and Rudi, 2022)

Using function values with trigonometric polynomials

- **SOS relaxation:**

$$
\begin{aligned}
& \min_{\Sigma \in \mathbb{C}^{d \times d}} \operatorname{tr}[F\Sigma] \ \text{ such that } \ \Sigma \in \mathcal{V}, \ \operatorname{tr}[\Sigma] = 1, \ \Sigma \succcurlyeq 0 \\
= \ & \max_{Y \in \mathcal{V}^\perp} \lambda_{\min}(F + Y)
\end{aligned}
$$

  - where $\mathcal{V} = \operatorname{span}\big(\{\varphi(x)\varphi(x)^*, \ x \in \mathcal{X}\}\big)$
  - $\mathcal{V}$ may be cumbersome to characterize computationally

Using function values with trigonometric polynomials

- **SOS relaxation:**

$$\begin{aligned} & \min_{\Sigma \in \mathbb{C}^{d\times d}} \operatorname{tr}[F\Sigma] \ \text{ such that } \ \Sigma \in \mathcal{V}, \ \operatorname{tr}[\Sigma] = 1, \ \Sigma \succcurlyeq 0 \\ = \ & \max_{Y \in \mathcal{V}^\perp} \lambda_{\min}(F + Y) \end{aligned}$$

  - where $\mathcal{V} = \operatorname{span}\big(\{\varphi(x)\varphi(x)^*, \ x \in \mathcal{X}\}\big)$
  - $\mathcal{V}$ may be cumbersome to characterize computationally

- **Replace** $\mathcal{V}$ **by** $\operatorname{span}\big(\{\varphi(x_i)\varphi(x_i)^*, i \in \{1, \dots, n\}\}\big)$
  - Generating family obtained by random samples $x_1, \dots, x_n$ in $\mathcal{X}$ (Cifuentes and Parrilo, 2017)

$$\max_{c \in \mathbb{R}, \ A \succcurlyeq 0} \ c \ \text{ such that } \ \forall i \in \{1, \dots, n\}, \ f(x_i) - c = \varphi(x_i)^* A \varphi(x_i)$$

Infinite expansions
(Rudi, Marteau-Ferey, and Bach, 2020)

- **Feature map** $\varphi : [0,1]^d \to \mathbb{C}^{|\Omega|}$: $\varphi(x)_\omega = \hat{q}(\omega) e^{2i\pi\omega^\top x}$, for $\omega \in \Omega$
  - Contraint $\sum_{\omega\in\Omega} |\hat{q}(\omega)|^2 = 1$
  - What if $\Omega = \mathbb{Z}^d$?

Infinite expansions (Rudi, Marteau-Ferey, and Bach, 2020)

- **Feature map** $\varphi : [0,1]^d \to \mathbb{C}^{|\Omega|}$: $\varphi(x)_\omega = \hat{q}(\omega) e^{2i\pi\omega^\top x}$, for $\omega \in \Omega$
  - Contraint $\sum_{\omega\in\Omega} |\hat{q}(\omega)|^2 = 1$
  - What if $\Omega = \mathbb{Z}^d$?

- **"Tightness" of relaxation if** $\forall \omega \in \mathbb{Z}^d, \ \hat{q}(\omega) > 0$

$$\sup_{c\in\mathbb{R}, \ A\succcurlyeq 0} c \quad \text{such that} \quad \forall x \in \mathcal{X}, \ f(x) - c = \varphi(x)^* A \varphi(x)$$

  - Attained with a finite rank operator A under local optimality conditions (isolated minimizers with invertible Hessians)
  - Still hard to solve ($\mathcal{X}$ dense *and* A infinite-dimensional)

Efficient sampling algorithms

- **Sampling and regularization:**

$$\max_{c\in\mathbb{R},\ A\succcurlyeq 0} c - \lambda\,\mathrm{tr}(A) \text{ such that } \forall i \in \{1,\dots,n\},\ f(x_i) - c = \varphi(x_i)^* A \varphi(x_i)$$

Efficient sampling algorithms

- **Sampling and regularization:**

$$\max_{c\in\mathbb{R},\ A\succcurlyeq 0} c - \lambda\,\mathrm{tr}(A) \text{ such that } \forall i \in \{1,\dots,n\},\ f(x_i) - c = \varphi(x_i)^* A \varphi(x_i)$$

  - Leads to provable *a priori* performance guarantees (with correct λ)
  - Up to logarithms and a few constants $\varepsilon \propto n^{-p/d}$ for C^p functions
  - See Rudi, Marteau-Ferey, and Bach (2020) for details

Efficient sampling algorithms

- **Sampling and regularization:**

$$\max_{c\in\mathbb{R},\ A\succcurlyeq 0} c - \lambda \operatorname{tr}(A) \text{ such that } \forall i \in \{1,\dots,n\},\ f(x_i) - c = \varphi(x_i)^* A \varphi(x_i)$$

  - Leads to provable *a priori* performance guarantees (with correct λ)
  - Up to logarithms and a few constants $\varepsilon \propto n^{-p/d}$ for C^p functions
  - See Rudi, Marteau-Ferey, and Bach (2020) for details

- **Finite-dimensional algorithm through representer theorem**
  - Can restrict search to $A = \sum_{j,k=1}^n B_{ik}\varphi(x_j)\varphi(x_k)^*$ with $B \in \mathbb{R}^{n\times n}$ and $B \succcurlyeq 0$
  - Only need access to $\varphi(x_j)^*\varphi(x_k) = \sum_{\omega\in\mathbb{Z}^d} |\hat{q}(\omega)|^2 e^{2i\pi\omega^\top(x_j - x_k)}$
  - See Marteau-Ferey, Bach, and Rudi (2020) for details

Conclusion

- **Sum-of-squares relaxations in the Fourier domain**
  - From trigonometric polynomials to C^∞ functions
  - Exponential convergence rates for polynomial hierarchies
  - Extension to zero-th order oracles and infinite expansions

Conclusion

- **Sum-of-squares relaxations in the Fourier domain**
  - From trigonometric polynomials to C^∞ functions
  - Exponential convergence rates for polynomial hierarchies
  - Extension to zero-th order oracles and infinite expansions
- **Improvements**
  - Certificates of optimality (Woodworth, Bach, and Rudi, 2022)
  - Constrained problems (going beyond simple sets $\mathcal{X}$)

Conclusion

- **Sum-of-squares relaxations in the Fourier domain**
  - From trigonometric polynomials to C^∞ functions
  - Exponential convergence rates for polynomial hierarchies
  - Extension to zero-th order oracles and infinite expansions

- **Improvements**
  - Certificates of optimality (Woodworth, Bach, and Rudi, 2022)
  - Constrained problems (going beyond simple sets $\mathcal{X}$)

- **SOS relaxations beyond optimization**
  - Optimal control (Berthier, Carpentier, Rudi, and Bach, 2021)
  - Optimal transport (Vacher, Muzellec, Rudi, Bach, and Vialard, 2021)
  - Log-partition functions and variational inference (Bach, 2022a,b)

Log-partition functions and variational inference

- **Log-partition function**: given $f : \mathcal{X} \to \mathbb{R}$ and a distribution q on $\mathcal{X}$

$$-\varepsilon \log \int_{\mathcal{X}} e^{-f(x)/\varepsilon} dq(x) = \inf_{p \text{ probability}} \int_{\mathcal{X}} f(x) dp(x) + \varepsilon D(p\|q)$$

  with $D(p\|q) = \displaystyle\int_{\mathcal{X}} \log \Big(\frac{dp}{dq}(x) \Big) dp(x)$ Kullback-Leibler divergence

  - Used within variational inference (Wainwright and Jordan, 2008)
  - Duality between maximum entropy and maximum likelihood

Log-partition functions and variational inference

- **Log-partition function**: given $f : \mathcal{X} \to \mathbb{R}$ and a distribution q on $\mathcal{X}$

$$-\varepsilon \log \int_{\mathcal{X}} e^{-f(x)/\varepsilon} dq(x) = \inf_{p \text{ probability}} \int_{\mathcal{X}} f(x) dp(x) + \varepsilon D(p\|q)$$

with $D(p\|q) = \displaystyle\int_{\mathcal{X}} \log\Big(\frac{dp}{dq}(x)\Big) dp(x)$ Kullback-Leibler divergence

  - Used within variational inference (Wainwright and Jordan, 2008)
  - Duality between maximum entropy and maximum likelihood

- **Von Neumann relative entropy**

$$D(\Sigma_p\|\Sigma_q) = \mathrm{tr}[\Sigma_p(\log \Sigma_p - \log \Sigma_q)]$$

  - With $\Sigma_p = \displaystyle\int_{\mathcal{X}} \varphi(x)\varphi(x)^* dp(x)$ and $\Sigma_q = \displaystyle\int_{\mathcal{X}} \varphi(x)\varphi(x)^* dq(x)$
  - Always a lower bound on $D(p\|q)$ (Bach, 2022a)

References

Francis Bach. Information theory with kernel methods. *arXiv preprint arXiv:2202.08545*, 2022a.

Francis Bach. Sum-of-squares relaxations for information theory and variational inference. *arXiv preprint arXiv:2206.13285*, 2022b.

Francis Bach and Alessandro Rudi. Exponential convergence of sum-of-squares hierarchies for trigonometric polynomials. Technical Report 2211.04889, ArXiv, 2022.

Eloïse Berthier, Justin Carpentier, Alessandro Rudi, and Francis Bach. Infinite-dimensional sums-of-squares for optimal control. Technical Report 2110.07396, arXiv, 2021.

Diego Cifuentes and Pablo A. Parrilo. Sampling algebraic varieties for sum of squares programs. *SIAM Journal on Optimization*, 27(4):2381–2404, 2017.

Pierre Del Moral and Angele Niclas. A Taylor expansion of the square root matrix function. *Journal of Mathematical Analysis and Applications*, 465(1):259–266, 2018.

Bogdan Dumitrescu. *Positive trigonometric polynomials and signal processing applications*, volume 103. Springer, 2007.

Kun Fang and Hamza Fawzi. The sum-of-squares hierarchy on the sphere and applications in quantum information theory. *Mathematical Programming*, 190(1):331–360, 2021.

Leopold Fejér. Uber trigonometrische Polynome. *Journal für die reine und angewandte Mathematik*, (146):55–82, 1916.

Ulf Grenander and Gabor Szegö. *Toeplitz forms and their applications*. Univ of California Press, 1958.

Didier Henrion, Milan Korda, and Jean Bernard Lasserre. *Moment-sos Hierarchy, The: Lectures In Probability, Statistics, Computational Geometry, Control And Nonlinear PDEs*, volume 4. World Scientific, 2020.

Arie Israel. The eigenvalue distribution of time-frequency localization operators. Technical Report 1502.04404, arXiv, 2015.

Jean-Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

Jean-Bernard Lasserre. *Moments, Positive Polynomials and their Applications*, volume 1. World Scientific, 2010.

Jean-Bernard Lasserre, Didier Henrion, Christophe Prieur, and Emmanuel Trélat. Nonlinear optimal control via occupation measures and lmi-relaxations. *SIAM Journal on Control and Optimization*, 47(4):1643–1666, 2008.

Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of Algebraic Geometry*, pages 157–270. Springer, 2009.

Monique Laurent and Lucas Slot. An effective version of Schmüdgen's Positivstellensatz for the hypercube. *Optimization Letters*, pages 1–16, 2022.

Daniel Liberzon. *Calculus of Variations and Optimal Control Theory*. Princeton University Press, 2011.

Ulysse Marteau-Ferey, Francis Bach, and Alessandro Rudi. Non-parametric models for non-negative functions. *Advances in Neural Information Processing Systems*, 33, 2020.

Ulysse Marteau-Ferey, Francis Bach, and Alessandro Rudi. Second order conditions to decompose smooth functions as sums of squares. *arXiv preprint arXiv:2202.13729*, 2022.

Alexandre Megretski. Positivity of trigonometric polynomials. In *International Conference on Decision and Control*, volume 4, pages 3814–3817, 2003.

Aaron Naftalovich and M. Schreiber. Trigonometric polynomials and sums of squares. In *Number Theory*, pages 225–238. Springer, 1985.

Yurii Nesterov. Squared functional systems and optimization problems. In *High Performance Optimization*, pages 405–440. Springer, 2000.

Jiawang Nie. Optimality conditions and finite convergence of Lasserre's hierarchy. *Mathematical Programming*, 146(1-2):97–121, 2014.

Erich Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis*, volume 1349. Springer, 2006.

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

Pablo A Parrilo, Grigoriy Blekherman, and Rekha R. Thomas. *Semidefinite optimization and convex algebraic geometry*. SIAM Society for Industrial and Applied Mathematics., 2013.

Mihai Putinar. Sur la complexification du problème des moments. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, 314(10):743–745, 1992.

Friedrich Riesz. Uber ein Problem des Herrn Carathéodory. *Journal für die reine und angewandte Mathematik*, (146):83–87, 1916.

Alessandro Rudi, Ulysse Marteau-Ferey, and Francis Bach. Finding global minima via kernel approximations. Technical Report 2012.11978, arXiv, 2020.

Claus Scheiderer. Sums of squares on real algebraic surfaces. *Manuscripta Mathematica*, 119(4):

395–410, 2006.

Konrad Schmüdgen. *The Moment Problem*. Springer, 2017.

Lucas Slot. Sum-of-squares hierarchies for polynomial optimization and the Christoffel–Darboux kernel. *SIAM Journal on Optimization*, 32(4):2612–2635, 2022.

Adrien Vacher, Boris Muzellec, Alessandro Rudi, Francis Bach, and Francois-Xavier Vialard. A dimension-free computational upper-bound for smooth optimal transport estimation. In *Conference on Learning Theory*, pages 4143–4173, 2021.

Richard Vinter. Convex duality and nonlinear optimal control. *SIAM Journal on Control and Optimization*, 31(2):518–538, 1993.

Martin J. Wainwright and Michael I. Jordan. *Graphical Models, Exponential Families, and Variational Inference*. Now Publishers Inc., 2008.

Jonathan Weed and Quentin Berthet. Estimation of smooth densities in Wasserstein distance. In *Conference on Learning Theory*, pages 3118–3119. PMLR, 2019.

Blake Woodworth, Francis Bach, and Alessandro Rudi. Non-convex optimization with certificates and fast rates through kernel sums of squares. *arXiv preprint arXiv:2204.04970*, 2022.

Smooth optimal transport

- **Primal formulation**: $\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y) d\gamma(x,y)$
  - $\Gamma(\mu,\nu)$ set of probability distributions with marginals μ and ν

- **Dual formulation**: $\sup_{u,v \in C(\mathbb{R}^n)} \int_{\mathcal{X}} u(x) d\mu(x) + \int_{\mathcal{Y}} v(y) d\mu(y)$

  such that $\quad \forall (x,y) \in \mathcal{X} \times \mathcal{Y},\ c(x,y) - u(x) + v(y) \geqslant 0$

Smooth optimal transport

- **Primal formulation**: $\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y) d\gamma(x,y)$
  - $\Gamma(\mu,\nu)$ set of probability distributions with marginals μ and ν

- **Dual formulation**: $\sup_{u,v \in C(\mathbb{R}^n)} \int_{\mathcal{X}} u(x) d\mu(x) + \int_{\mathcal{Y}} v(y) d\mu(y)$

  such that $\quad \forall (x,y) \in \mathcal{X} \times \mathcal{Y},\ c(x,y) - u(x) + v(y) \geqslant 0$

- **Estimation from i.i.d. samples from smooth densities for μ and ν**
  - Rate: from $O(n^{-1/d})$ to $O(n^{-p/d})$ (Weed and Berthet, 2019)
  - No polynomial-time algorithm

Smooth optimal transport

- **Primal formulation**: $\displaystyle \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\gamma(x,y)$
  - $\Gamma(\mu,\nu)$ set of probability distributions with marginals μ and ν

- **Dual formulation**: $\displaystyle \sup_{u,v \in C(\mathbb{R}^n)} \int_{\mathcal{X}} u(x)d\mu(x) + \int_{\mathcal{Y}} v(y)d\mu(y)$

  such that $\quad \forall (x,y) \in \mathcal{X}\times\mathcal{Y},\ c(x,y) - u(x) + v(y) \geqslant 0$

- **Estimation from i.i.d. samples from smooth densities for μ and ν**
  - Rate: from $O(n^{-1/d})$ to $O(n^{-p/d})$ (Weed and Berthet, 2019)
  - No polynomial-time algorithm

- **Kernel sums of squares**: "polynomial"-time algorithm
  - Vacher, Muzellec, Rudi, Bach, and Vialard (2021)

Optimal control / reinforcement learning

- **Optimal control** (Liberzon, 2011)

$$V^*(t_0, x_0) = \inf_{u:[t_0,T]\to\mathcal{U}} \int_{t_0}^{T} L(t, x(t), u(t))\mathsf{d}t + M(x(T))$$

$$\forall t \in [t_0, T],\ \dot{x}(t) = f(t, x(t), u(t)), \quad x(t_0) = x_0.$$

Optimal control / reinforcement learning

- **Optimal control** (Liberzon, 2011)

$$V^*(t_0, x_0) = \inf_{u:[t_0,T]\to\mathcal{U}} \int_{t_0}^{T} L(t, x(t), u(t))\mathsf{d}t + M(x(T))$$
$$\forall t \in [t_0, T],\ \dot{x}(t) = f(t, x(t), u(t)), \quad x(t_0) = x_0.$$

- Subsolution of **Hamilton-Jacobi-Bellman** equation (Vinter, 1993)

$$\sup_{V:[0,T]\times\mathcal{X}\to\mathbb{R}} \int V(0, x_0)\mathsf{d}\mu_0(x_0)$$
$$\forall (t, x, u), \quad \frac{\partial V}{\partial t}(t, x) + L(t, x, u) + \nabla V(t, x)^\top f(t, x, u) \geqslant 0$$
$$\forall x, \quad V(T, x) = M(x).$$

Optimal control / reinforcement learning

- Subsolution of **Hamilton-Jacobi-Bellman** equation (Vinter, 1993)

$$\sup_{V:[0,T]\times\mathcal{X}\to\mathbb{R}} \int V(0,x_0)\mathsf{d}\mu_0(x_0)$$

$$\begin{aligned}\forall (t,x,u), \quad & \frac{\partial V}{\partial t}(t,x) + L(t,x,u) + \nabla V(t,x)^\top f(t,x,u) \geqslant 0 \\ \forall x, \quad & V(T,x) = M(x).\end{aligned}$$

Optimal control / reinforcement learning

- Subsolution of **Hamilton-Jacobi-Bellman** equation (Vinter, 1993)

$$\sup_{V:[0,T]\times\mathcal{X}\to\mathbb{R}} \int V(0,x_0)\mathsf{d}\mu_0(x_0)$$

$$\begin{aligned} \forall (t,x,u), \quad & \frac{\partial V}{\partial t}(t,x) + L(t,x,u) + \nabla V(t,x)^\top f(t,x,u) \geqslant 0 \\ \forall x, \quad & V(T,x) = M(x). \end{aligned}$$

- **Polynomial sums-of-squares**
  - Lasserre, Henrion, Prieur, and Trélat (2008)

Optimal control / reinforcement learning

- Subsolution of **Hamilton-Jacobi-Bellman** equation (Vinter, 1993)

$$\sup_{V:[0,T]\times\mathcal{X}\to\mathbb{R}} \int V(0,x_0)\mathsf{d}\mu_0(x_0)$$

$$\begin{aligned}
\forall (t,x,u), \quad & \frac{\partial V}{\partial t}(t,x) + L(t,x,u) + \nabla V(t,x)^\top f(t,x,u) \geqslant 0 \\
\forall x, \quad & V(T,x) = M(x).
\end{aligned}$$

- **Polynomial sums-of-squares**
  - Lasserre, Henrion, Prieur, and Trélat (2008)

- **Extension to kernel sums-of-squares**
  - Berthier, Carpentier, Rudi, and Bach (2021)
  - Allows some form of modelling