Formal Non-linear Optimization via Templates and Sum-of-Squares

Joint Work with B. Werner, S. Gaubert and X. Allamigeon

Third year PhD Victor MAGRON

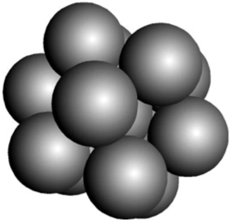
LIX/INRIA, École Polytechnique

TYPES 2013 Tuesday April 23 rd

Kepler Conjecture (1611):

The maximal density of sphere packings in 3D-space is $\frac{\pi}{\sqrt{18}}$

- It corresponds to the way people would intuitively stack oranges, as a pyramid shape
- The proof of T. Hales (1998) consists of thousands of non-linear inequalities
- Many recent efforts have been done to give a formal proof of these inequalities: Flyspeck Project
- Motivation: get positivity certificates and check them with Proof assistants like Coq



Contents

Inequalities issued from Flyspeck non-linear part involve:

1. Multivariate Polynomials:
   $x_1x_4(-x_1+x_2+x_3-x_4+x_5+x_6)+x_2x_5(x_1-x_2+x_3+x_4-x_5+x_6)+x_3x_6(x_1+x_2-x_3+x_4+x_5-x_6)-x_2(x_3x_4+x_1x_6)-x_5(x_1x_3+x_4x_6)$
2. Semi-Algebraic functions algebra $\mathcal{A}$: composition of polynomials with $|\cdot|, \sqrt{}, +, -, \times, /, \sup, \inf, \cdots$
3. Transcendental functions $\mathcal{T}$: composition of semi-algebraic functions with $\boldsymbol{arctan}$, $\boldsymbol{exp}$, $+, -, \times, \cdots$

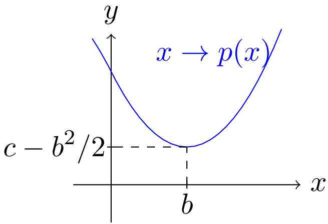
Lemma from Flyspeck (inequality ID 6096597438)

$\forall x \in [3, 64], 2\pi - 2(x \arcsin(\cos(0.797)\sin(\pi/x)) - (0.591 - 0.0331x + 1.506) \geq 0$

Certification Framework: who does what?

Polynomial Optimization (POP): $\min_{x \in \mathbb{R}} p(x) = 1/2x^2 - bx + c$

1. A program written in OCaml/C provides the Sum-of-Squares decomposition: $1/2(x-b)^2$
2. A program written in Coq checks: $\forall x \in \mathbb{R}, p(x) = 1/2(x-b)^2 + c - b^2/2$



- Sceptical approach: obtain *certificates* of positivity with efficient oracles and check them formally
- Questions: How to obtain the certificates? How to deal with non-polynomial case?

The Polynomial Case

- General POP $\min_{\mathbf{x} \in K} p(\mathbf{x})$ with K the compact set of constraints:

$$K = \{\mathbf{x} \in \mathbb{R}^n : g_1(\mathbf{x}) \geq 0, \cdots, g_m(\mathbf{x}) \geq 0\}$$

- Let $\Sigma[\mathbf{x}]$ be the cone of Sum-of-Squares (SOS) and consider the restriction $\Sigma_d[\mathbf{x}]$ to polynomials of degree at most $2d$:

$$\Sigma_d[\mathbf{x}] = \left\{ \sum_i q_i(\mathbf{x})^2, \text{ with } q_i \in \mathbb{R}_d[\mathbf{x}] \right\}$$

- Let $g_0 := 1$ and $M(\mathbf{g})$ be the quadratic module generated by $g_1, \cdots, g_m$:

$$M(\mathbf{g}) = \left\{ \sum_{j=0}^{m} \sigma_j(\mathbf{x}) g_j(\mathbf{x}), \text{ with } \sigma_j \in \Sigma[\mathbf{x}] \right\}$$

- Certificates for positive polynomials: Sum-of-Squares

The Polynomial Case: Putinar Theorem

$$M(\mathbf{g}) = \left\{ \sum_{j=0}^{m} \sigma_j(\mathbf{x}) g_j(\mathbf{x}), \text{ with } \sigma_j \in \Sigma[\mathbf{x}] \right\}$$

Proposition (Putinar)

Suppose $\mathbf{x} \in [\mathbf{a}, \mathbf{b}]$. $p(\mathbf{x}) - p^* > 0$ on $K \implies (p(\mathbf{x}) - p^*) \in M(\mathbf{g})$

- But the search space for $\sigma_0, \cdots, \sigma_m$ is infinite so consider the truncated module $M_d(\mathbf{g})$:

$$M_d(\mathbf{g}) = \left\{ \sum_{j=0}^{m} \sigma_j(\mathbf{x}) g_j(\mathbf{x}), \text{ with } \sigma_j \in \Sigma[\mathbf{x}], (\sigma_j g_j) \in \mathbb{R}_{2d}[\mathbf{x}] \right\}$$

- $M_0(\mathbf{g}) \subset M_1(\mathbf{g}) \subset M_2(\mathbf{g}) \subset \cdots \subset M(\mathbf{g})$
- Hence, we consider the hierarchy of SOS *relaxation* programs: $\mu_k := \sup_{\mu, \sigma_0, \cdots, \sigma_m} \left\{ \mu : (p(\mathbf{x}) - \mu) \in M_k(\mathbf{g}) \right\}$

The Polynomial Case: Examples

- $\min_{\mathbf{x}\in[4,6.3504]^6} \Delta(\mathbf{x}) = x_1x_4(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6) + x_2x_5(x_1 - x_2 + x_3 + x_4 - x_5 + x_6) + x_3x_6(x_1 + x_2 - x_3 + x_4 + x_5 - x_6) - x_2(x_3x_4 + x_1x_6) - x_5(x_1x_3 + x_4x_6) = \mu_2 = 128$
- $\Delta(\mathbf{x}) - \mu_2 = \sigma_0(\mathbf{x}) + \sum_{j=1}^{6} \sigma_j(\mathbf{x})(6.3504 - x_j)(x_j - 4)$ with $\sigma_0 \in \Sigma_2[\mathbf{x}], \sigma_j \in \Sigma_1[\mathbf{x}]$
- Also works for Semi-algebraic functions with *lifting* variables:

  $f := \Delta\mathbf{x} - \sqrt{x_1^2 + x_2^2}$

  Define $K = \{(\mathbf{x}, z) \in \mathbb{R}^{n+1} : \mathbf{x} \in [4, 6.3504]^6, z^2 \geq x_1^2 + x_2^2, x_1^2 + x_2^2, z^2 \leq x_1^2 + x_2^2, x_1^2 + x_2^2, z \geq 0\}$

  $$\min_{\mathbf{x}\in[4,6.3504]^6} f(\mathbf{x}) = \min_{(\mathbf{x},z)\in K} (\Delta(\mathbf{x}) - z) \text{ (POP)}$$

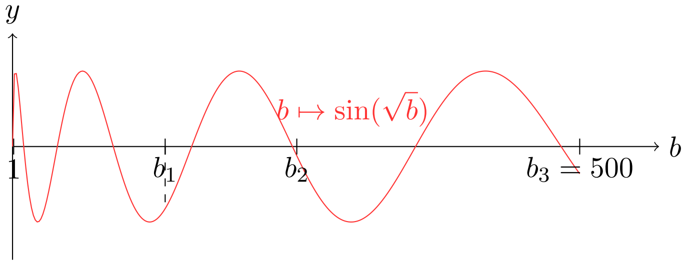
Non-Polynomial Optimization: an Example

Example: $\min_{\mathbf{x} \in [1,500]^n} f(\mathbf{x}) = - \sum_{i=1}^{n} (x_i + x_{i+1}) \sin(\sqrt{x_i})$

- Classical idea: approximate $\sin(\sqrt{\cdot})$ by a degree-d Taylor Polynomial f_d, solve $\min_{\mathbf{x} \in [1,500]^n} - \sum_{i=1}^{n} (x_i + x_{i+1}) f_d(x_i)$ (POP)
- Lack of accuracy if d is not large enough
- No free lunch: the complexity to solve POP with Sum-of-Squares of degree $2d$ is $O(n^{2d})$
- Alternative: deal with the complexity issue with low degree approximations: Templates method

Non-Polynomial Optimization via Templates

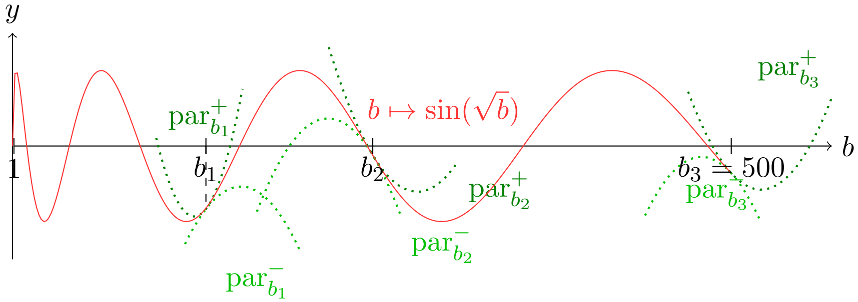
- Consider the univariate function $\hat{f} : b \mapsto \sin(\sqrt{b})$ on $I = [1, 500]$



- Pick several points $b_j \in I$
- $\hat{f}$ is semi-convex: there exists a constant $c_j > 0$ s.t. $b \mapsto \hat{f}(b) + c_j/2(b - b_j)^2$ is convex
- By convexity, $\forall b \in I, \hat{f}(b) \geq -c_j/2(b-b_j)^2 + \hat{f}'(b_j)(b-b_j) + \hat{f}(b_j) = \mathrm{par}^-_{b_j}(b)$

Non-Polynomial Optimization via Templates

- $\forall j, \hat{f} \geq \mathrm{par}_{b_j}^- \implies \hat{f} \geq \max_j \left\{ \mathrm{par}_{b_j}^- \right\}$: **Max-Plus underestimator**
- $\forall j, \hat{f} \leq \mathrm{par}_{b_j}^+ \implies \hat{f} \leq \min_j \left\{ \mathrm{par}_{b_j}^+ \right\}$: **Max-Plus overestimator**



Templates based on Max-plus Semi-algebraic Estimators for $b \mapsto \sin(\sqrt{b})$:

$$\max_{j \in \{1,2,3\}} \{\mathrm{par}_{b_j}^-(x_i)\} \leq \sin\sqrt{x_i} \leq \min_{j \in \{1,2,3\}} \{\mathrm{par}_{b_j}^+(x_i)\}$$

Non-Polynomial Optimization via Templates: Lifting

- Use a lifting variable z_i to represent $x_i \mapsto \sin(\sqrt{x_i})$
- For each i, pick points b_{ji}
- With 3 points b_{ji}, we solve the POP:

$$\begin{cases} \min\limits_{\mathbf{x} \in [1,500]^n, \mathbf{z} \in [-1,1]^n} & -\sum_{i=1}^{n} (x_i + x_{i+1}) z_i \\ \text{s.t.} & z_i \leq \text{par}^{+}_{b_{ji}}(x_i), j \in \{1, 2, 3\} \end{cases}$$

- POP with $n + n$ variables ($n_{\text{lifting}} = n$ variables), with Sum-of-Squares of degree $2d$: $O((2n)^{2d})$ complexity

Full Lifting Templates / Lifting Free Templates

- Other choice: lifting variable y_i to represent $x_i \mapsto \sqrt{x_i}$ and lifting variable z_i to represent $x_i \mapsto \sin(x_i)$

$$\begin{cases} \min\limits_{\mathbf{x}\in[1,500]^n, \mathbf{y}\in[1,\sqrt{500}]^n, \mathbf{z}\in[-1,1]^n} & -\sum_{i=1}^{n}(x_i + x_{i+1}) z_i \\ \text{s.t.} & z_i \leq \mathrm{par}^{+}_{a_{ji}}(y_i), j \in \{1, 2, 3\} \\ & y_i^2 = x_i \end{cases}$$

- POP with $n + 2n$ variables ($n_{\text{lifting}} = 2n$ variables), with Sum-of-Squares of degree $2d$: $O((3n)^{2d})$ complexity
- Taylor approximations: templates with n variables ($n_{\text{lifting}} = 0$ variables)

Templates and SOS: the Algorithm

Algorithm `template_optim`:

Input: tree t, box K, number of lifting variables n_{lifting}

1: **if** t is semi-algebraic **then**
2: Define lifting variables and solve the resulting POP
3: **else if** $\texttt{bop} := \texttt{root}\ (t)$ is a binary operation with children c_1 and c_2 **then**
4: Apply `template_optim` recursively to c_1, c_2
5: Compose the results
6: **else if** $r := \texttt{root}(t)$ is univariate transcendental function with child c **then**
7: Apply `template_optim` recursively to c
8: Build estimators for a sub-tree of t with up to n_{lifting} variables
9: Solve the resulting POP
10: **end**

Templates and SOS: Results for the Example

$$\min_{\mathbf{x} \in [1,500]^n} f(\mathbf{x}) = -\sum_{i=1}^{n} (x_i + \epsilon x_{i+1}) \sin(\sqrt{x_i})$$

n	lower bound	n_{lifting}	#boxes	time
$10 (\epsilon = 0)$	$-430n$	$2n$	16	$40\ s$
$10 (\epsilon = 0)$	$-430n$	0	827	$177\ s$
$1000 (\epsilon = 1)$	$-967n$	$2n$	1	$543\ s$
$1000 (\epsilon = 1)$	$-968n$	n	1	$272\ s$

Templates and SOS: Results for Flyspeck

- $n = 6$ variables, SOS of degree $2k = 4$
- $n_{\mathcal{T}}$ univariate transcendental functions
- #boxes sub-problems

Inequality id	$n_{\mathcal{T}}$	n_{lifting}	#boxes	time
9922699028	1	9	47	$241\,s$
9922699028	1	3	39	$190\,s$
3318775219	1	9	338	$26\,min$
7726998381	3	15	70	$43\,min$
7394240696	3	15	351	$1.8\,h$
4652969746_1	6	15	81	$1.3\,h$
OXLZLEZ 6346351218_2_0	6	24	200	$5.7\,h$

Towards Formal Non-Linear Optimization

- Use Sparsity/Symmetries for a positive domino effect:
  1. on the global optimization oracle to decrease the $O(n^{2d})$ complexity
  2. to check Sum-of-Squares with `field` tactic
- Formal proofs for Max-Plus estimators: certify rigorous under/over estimators for univariate transcendental functions

End

Thank you for your attention!