OSE SEMINAR 2012

Two approaches to underestimating quadratic functions

Anders Skjäl

CENTER OF EXCELLENCE IN
OPTIMIZATION AND SYSTEMS ENGINEERING
ÅBO AKADEMI UNIVERSITY

ÅBO, NOVEMBER 29 2012

OSE OPTIMIZATION AND SYSTEMS ENGINEERING

Åbo Akademi University

The Application

A nonconvex mathematical programming problem:

$$\begin{array}{lll} \min & f_0(\boldsymbol{x}) & \\ \text{s.t.} & f_m(\boldsymbol{x}) \leq 0 & i = 1, 2, \ldots, M \\ & \boldsymbol{x}^{\boldsymbol{L}} \leq x \leq \boldsymbol{x}^{\boldsymbol{U}} & \end{array}$$

The Application

A nonconvex mathematical programming problem:

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) & \\ \text{s.t.} \quad & f_m(\boldsymbol{x}) \leq 0 & i = 1, 2, \ldots, M \\ & \boldsymbol{x}^{\boldsymbol{L}} \leq x \leq \boldsymbol{x}^{\boldsymbol{U}} & \end{aligned}$$

Many global optimization algorithms use convex underestimation

- branch-and-bound methods
- lower bounds or proof of infeasibility

A Comparison

I will describe two underestimation methods:

- An αBB variant (Skjäl and Westerlund, 2012)
  - smooth ($\mathcal{C}^2$) functions
  - perturbations

A Comparison

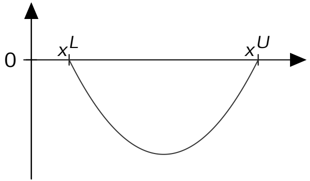
I will describe two underestimation methods:

- An αBB variant (Skjäl and Westerlund, 2012)
  - smooth ($\mathcal{C}^2$) functions
  - perturbations
- An underestimation method with roots in algebraic geometry (Jean B. Lasserre and Tung Phan Thanh, 2012)
  - polynomials
  - underestimator of a specified degree

αBB

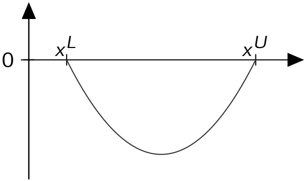
Perturbations

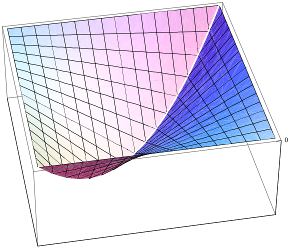
- All αBB methods use perturbations
- $-\alpha_i(x_i - x_i^L)(x_i^U - x_i)$



Perturbations

- All αBB methods use perturbations
- $-\alpha_i(x_i - x_i^L)(x_i^U - x_i)$
- $\beta_{ij}x_ix_j - \mathrm{conc}(\beta_{ij}x_ix_j)$

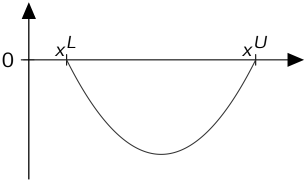


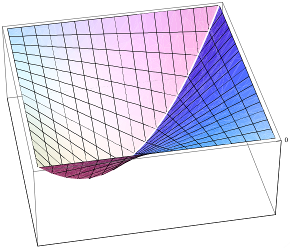


Perturbations

- All αBB methods use perturbations
- $-\alpha_i(x_i - x_i^L)(x_i^U - x_i)$
- $\beta_{ij} x_i x_j - \mathrm{conc}(\beta_{ij} x_i x_j)$
- underestimation: ok
- convexity: use a sufficient condition for positive semidefiniteness

$$\nabla^2 f(\boldsymbol{x}) + H^P \succeq 0, \; \forall \boldsymbol{x} \in [\boldsymbol{x}^L, \boldsymbol{x}^U]$$



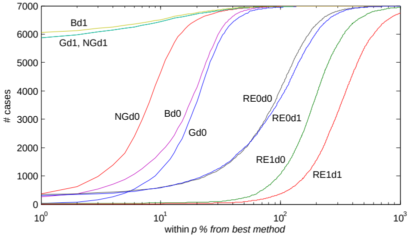


Parameter Calculation

- Adjiman, Dallwig, Floudas & Neumaier, 1998
  - diagonal
  - many calculation methods (*e.g.* "Gerschgorin")

Parameter Calculation

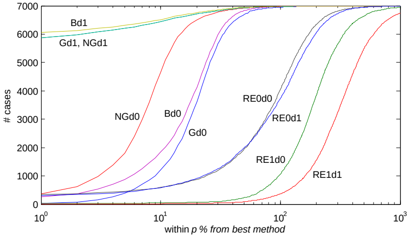
- Adjiman, Dallwig, Floudas & Neumaier, 1998
  - diagonal
  - many calculation methods (*e.g.* "Gerschgorin")
- Skjäl & Westerlund, manuscript
  - additional methods, nondiagonal



(maximum error)

Parameter Calculation

- Adjiman, Dallwig, Floudas & Neumaier, 1998
  - diagonal
  - many calculation methods (*e.g.* "Gerschgorin")
- Skjäl & Westerlund, manuscript
  - additional methods, nondiagonal
- The scaled diagonal Gerschgorin method is recommended for general purposes
  - calculation: $O(n^2)$



(maximum error)

Quadratic Functions

- Quadratic functions have constant second derivatives
- No need for interval approximations
- Convexity of the perturbed function is equivalent to

$$H + H^P = H + \begin{bmatrix} 2\alpha_1 & \beta_{1,2} & \cdots & \beta_{1,n} \\ \beta_{1,2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \beta_{n-1,n} \\ \beta_{1,n} & \cdots & \beta_{n-1,n} & 2\alpha_n \end{bmatrix} \succeq 0$$

Quadratic Functions

- Quadratic functions have constant second derivatives
- No need for interval approximations
- Convexity of the perturbed function is equivalent to

$$H + H^P = H + \begin{bmatrix} 2\alpha_1 & \beta_{1,2} & \cdots & \beta_{1,n} \\ \beta_{1,2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \beta_{n-1,n} \\ \beta_{1,n} & \cdots & \beta_{n-1,n} & 2\alpha_n \end{bmatrix} \succeq 0$$

- The best perturbation(s) minimize some error measure
  - in literature: the maximum underestimation error
  - a new choice: the average error

Error Measures

- Maximum underestimation error (L^∞-norm)

$$\sum_i \frac{1}{4}(x_i^U - x_i^L)^2 \alpha_i + \sum_i \sum_{j>i} \frac{1}{4} (x_i^U - x_i^L)(x_j^U - x_j^L) |\beta_{ij}|$$

Error Measures

- Maximum underestimation error (L^∞-norm)

$$\sum_i \frac{1}{4}(x_i^U - x_i^L)^2\alpha_i + \sum_i \sum_{j>i} \frac{1}{4}(x_i^U - x_i^L)(x_j^U - x_j^L)\,|\beta_{ij}|$$

- Average error (normalized L^1-norm)
  - α_i weight

$$\frac{\int_{[\mathbf{x}^L,\mathbf{x}^U]} (x_i - x_i^L)(x_i^U - x_i)\,\mathrm{d}\mathbf{x}}{\int_{[\mathbf{x}^L,\mathbf{x}^U]} \mathrm{d}\mathbf{x}} = \frac{1}{6}(x_i^U - x_i^L)^2$$

Error Measures

- Maximum underestimation error (L^∞-norm)

$$\sum_i \frac{1}{4}(x_i^U - x_i^L)^2 \alpha_i + \sum_i \sum_{j>i} \frac{1}{4}(x_i^U - x_i^L)(x_j^U - x_j^L)\,|\beta_{ij}|$$

- Average error (normalized L^1-norm)
  - α_i weight

$$\frac{\int_{[\boldsymbol{x}^L,\boldsymbol{x}^U]} (x_i - x_i^L)(x_i^U - x_i)\,\mathrm{d}\boldsymbol{x}}{\int_{[\boldsymbol{x}^L,\boldsymbol{x}^U]} \mathrm{d}\boldsymbol{x}} = \frac{1}{6}(x_i^U - x_i^L)^2$$

  - $|\beta_{ij}|$ weight, symbolical integration with Mathematica

$$\frac{1}{12}(x_i^U - x_i^L)(x_j^U - x_j^L)$$

Lasserre & Thanh's Method

Positive Polynomials on $\mathbb{R}^n$

- If a function can be decomposed as a sum of squares it is nonnegative

$$x^2 - 4xy + 5y^2 - 2yz + z^2 = (x - 2y)^2 + (y - z)^2 \geq 0$$

Positive Polynomials on $\mathbb{R}^n$

Hilbert's Seventeenth Problem

Can any nonnegative polynomial be represented as a sum of squares of rational functions?

- If a function can be decomposed as a sum of squares it is nonnegative

$$x^2 - 4xy + 5y^2 - 2yz + z^2 = (x - 2y)^2 + (y - z)^2 \geq 0$$

Positive Polynomials on $\mathbb{R}^n$

Hilbert's Seventeenth Problem

Can any nonnegative polynomial be represented as a sum of squares of rational functions?

- If a function can be decomposed as a sum of squares it is nonnegative

$$x^2 - 4xy + 5y^2 - 2yz + z^2 = (x - 2y)^2 + (y - z)^2 \geq 0$$

- Hilbert showed that a nonnegative polynomial is not in general a sum of squares of polynomials
- Motzkin gave the first example (1966)

$$z^6 + x^4y^2 + x^2y^4 - 3x^2y^2z^2$$

Positive Polynomials on Semialgebraic Sets

- A set is called semialgebraic if it is described by polynomial inequalities

$$\{\boldsymbol{x} \in \mathbb{R}^n : p_1(x) \geq 0, \ldots, p_m(x) \geq 0\}$$

- Let Σ denote all sums of squares of polynomials in $\boldsymbol{x}$; a convex cone

Positive Polynomials on Semialgebraic Sets

Putinar's Positivstellensatz

Let K be a **compact semialgebraic set.** Assume that $p_1 \dots p_m$ have even degrees and that their highest degree homogenous parts have no common zeroes in $\mathbb{R}^n$, except 0. Then **any positive polynomial** p **on** K **belongs to** the cone $\Sigma + p_1\Sigma + \ldots + p_m\Sigma$.

- A set is called semialgebraic if it is described by polynomial inequalities

$$\{\boldsymbol{x} \in \mathbb{R}^n : p_1(x) \geq 0, \ldots, p_m(x) \geq 0\}$$

- Let Σ denote all sums of squares of polynomials in $\boldsymbol{x}$; a convex cone

Underestimation and Convexity

- Lasserre and Thanh use the Positivstellensatz for both properties
  - underestimation:

$$f(x) - h(x) \geq 0, \quad \forall \boldsymbol{x} \in [\boldsymbol{x^L}, \boldsymbol{x^U}]$$

Underestimation and Convexity

- Lasserre and Thanh use the Positivstellensatz for both properties
  - underestimation:

$$f(\boldsymbol{x}) - h(\boldsymbol{x}) \geq 0, \quad \forall \boldsymbol{x} \in [\boldsymbol{x}^{\boldsymbol{L}}, \boldsymbol{x}^{\boldsymbol{U}}]$$

  - convexity:

$$\boldsymbol{y}^T \nabla^2 h(\boldsymbol{x})\, \boldsymbol{y} \geq 0, \quad \forall \boldsymbol{x} \in [\boldsymbol{x}^{\boldsymbol{L}}, \boldsymbol{x}^{\boldsymbol{U}}], \forall \boldsymbol{y} \in \mathbb{R}^n : \|y\| \leq 1$$

Underestimation and Convexity

- Lasserre and Thanh use the Positivstellensatz for both properties
  - underestimation:

$$f(\boldsymbol{x}) - h(\boldsymbol{x}) \geq 0, \quad \forall \boldsymbol{x} \in [\boldsymbol{x^L}, \boldsymbol{x^U}]$$

  - convexity:

$$\boldsymbol{y}^T \nabla^2 h(\boldsymbol{x})\, \boldsymbol{y} \geq 0, \quad \forall \boldsymbol{x} \in [\boldsymbol{x^L}, \boldsymbol{x^U}], \forall \boldsymbol{y} \in \mathbb{R}^n : \|y\| \leq 1$$

  - the sets are compact semialgebraic

$$\left\{\boldsymbol{x} \in \mathbb{R}^n : (x_i - x_i^L)(x_i^U - x_i) \geq 0, i = 1, \ldots, n\right\}$$

$$\cap \left\{\boldsymbol{y} \in \mathbb{R}^n : 1 - \sum_i y_i^2 \geq 0\right\}$$

Finite-Dimensional Approximation

- The underestimator degree is fixed, $\deg(h) = d$
- The sum-of-squares cones are restricted

$$\Sigma_k := \{p \in \Sigma : \deg(p) \leq 2k\}$$

Finite-Dimensional Approximation

- The underestimator degree is fixed, $\deg(h) = d$
- The sum-of-squares cones are restricted

$$\Sigma_k := \{p \in \Sigma : \deg(p) \leq 2k\}$$

- Lasserre & Thanh proved convergence properties as $k \to \infty$

Finite-Dimensional Approximation

- The underestimator degree is fixed, $\deg(h) = d$
- The sum-of-squares cones are restricted

$$\Sigma_k := \{p \in \Sigma : \deg(p) \leq 2k\}$$

- Lasserre & Thanh proved convergence properties as $k \to \infty$
- Elements in Σ_k kan be represented as positive semidefinite matrices

$$(x-2y)^2+(y-1)^2 = \begin{bmatrix} 1 & x & y \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ y \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ x \\ y \end{bmatrix}^T \underbrace{\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ -1 & -2 & 5 \end{bmatrix}}_{\succeq 0} \begin{bmatrix} 1 \\ x \\ y \end{bmatrix}$$

Constraints

- The underestimation condition takes the form:

$$f(\boldsymbol{x}) - h(\boldsymbol{x}) = \sigma_0(\boldsymbol{x}) + \sum_{i=1}^{n} \sigma_i(\boldsymbol{x})\,(x_i - x_i^L)(x_i^U - x_i), \qquad \forall \boldsymbol{x}$$

$$\sigma_0 \in \Sigma_k$$

$$\sigma_i \in \Sigma_{k-1},\ i = 1, \ldots, n$$

Constraints

- The underestimation condition takes the form:

$$f(\boldsymbol{x}) - h(\boldsymbol{x}) = \sigma_0(\boldsymbol{x}) + \sum_{i=1}^{n} \sigma_i(\boldsymbol{x})\,(x_i - x_i^L)(x_i^U - x_i), \qquad \forall \boldsymbol{x}$$

$$\sigma_0 \in \Sigma_k$$

$$\sigma_i \in \Sigma_{k-1},\ i = 1, \dots, n$$

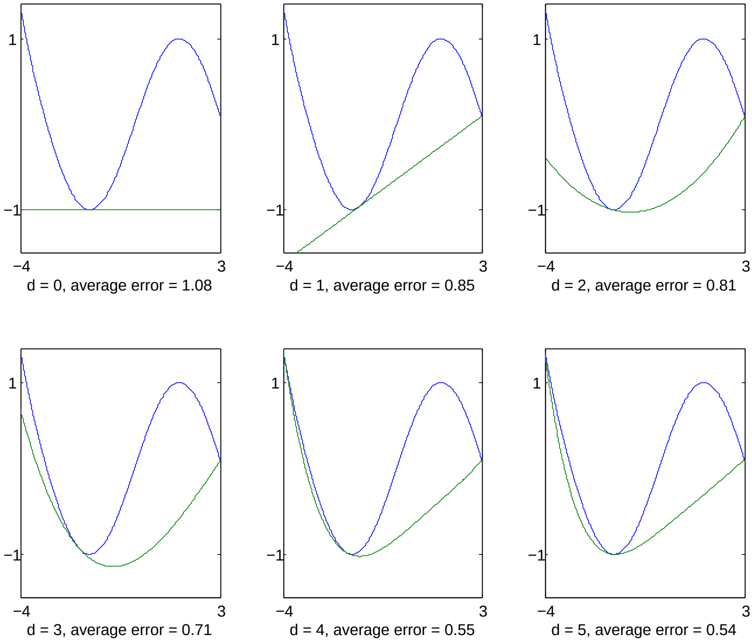
- Rewritten in the monomial basis we get $\binom{n+2k}{n}$ linear constraints

$$f_\alpha - h_\alpha = \sum_{i=0}^{n} \left\langle Z_j, C_\alpha^j \right\rangle$$

involving $n+1$ semidefinite variable matrices

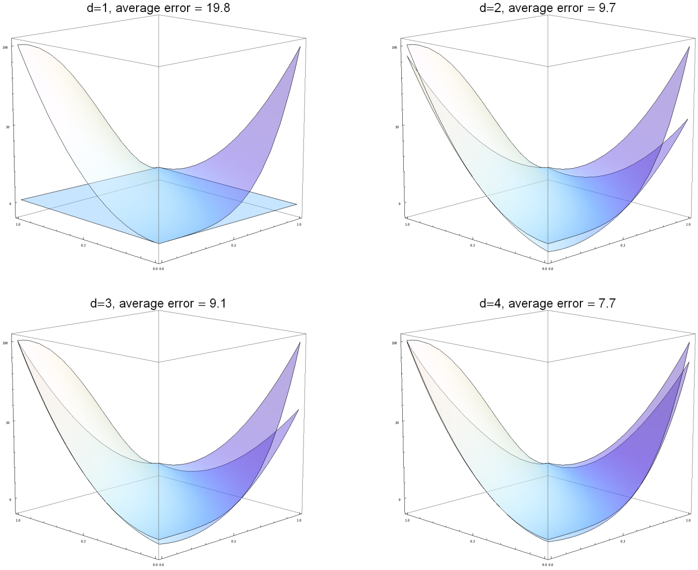
$$Z_j \succeq 0,\ j = 0, 1, \dots, n$$

Example - 1D



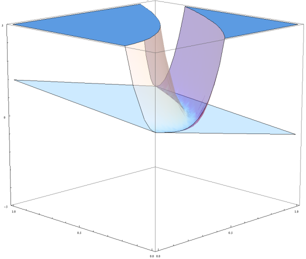
1
−1
−4
3
d = 0, average error = 1.08
d = 1, average error = 0.85
d = 2, average error = 0.81
d = 3, average error = 0.71
d = 4, average error = 0.55
d = 5, average error = 0.54

Example - 2D

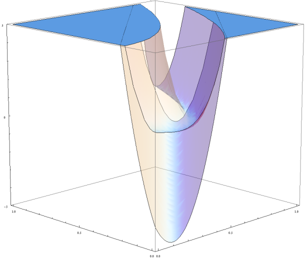

d=1, average error = 19.8
d=2, average error = 9.7
d=3, average error = 9.1
d=4, average error = 7.7

Example - 2D (detail)

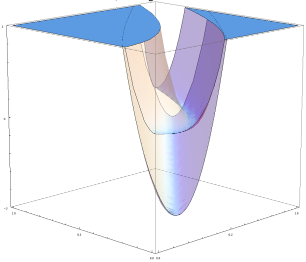
d=1, average error = 19.8



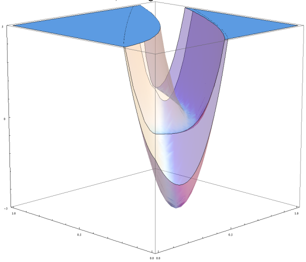
d=2, average error = 9.7



d=3, average error = 9.1



d=4, average error = 7.7



L&T, Quadratic Case

- Lasserre & Thanh's constraints simplify when $\deg(f) = 2, d = 2, k = 1$

$$f(\boldsymbol{x}) - (b + \boldsymbol{a}'\boldsymbol{x} + \boldsymbol{x}'A\boldsymbol{x}) = \sum_{i=1}^{n} \sigma_i (x_i - x_i^L)(x_i^U - x_i) + [\boldsymbol{x}'\ 1]\, C\, [\boldsymbol{x}'\ 1]'$$

$$A \succeq 0,\ C \succeq 0$$

$$\sigma_i \geq 0, \quad \forall i = 1, \ldots, n$$

L&T, Quadratic Case

- Lasserre & Thanh's constraints simplify when $\deg(f) = 2$, $d = 2$, $k = 1$

$$f(\boldsymbol{x}) - (b + \boldsymbol{a}'\boldsymbol{x} + \boldsymbol{x}'A\boldsymbol{x}) = \sum_{i=1}^{n} \sigma_i (x_i - x_i^L)(x_i^U - x_i) + [\boldsymbol{x}'\ 1]\, C\, [\boldsymbol{x}'\ 1]'$$

$$A \succeq 0,\ C \succeq 0$$

$$\sigma_i \geq 0, \quad \forall i = 1, \ldots, n$$

- Note the similarity with α perturbations
  $\Rightarrow$ average error better ($\leq$) than diagonal αBB methods

L&T, Quadratic Case

- Lasserre & Thanh's constraints simplify when $\deg(f) = 2, d = 2, k = 1$

$$f(\boldsymbol{x}) - (b + \boldsymbol{a}'\boldsymbol{x} + \boldsymbol{x}'A\boldsymbol{x}) = \sum_{i=1}^{n} \sigma_i (x_i - x_i^L)(x_i^U - x_i) + [\boldsymbol{x}'\ 1]\, C\, [\boldsymbol{x}'\ 1]'$$

$$A \succeq 0,\ C \succeq 0$$

$$\sigma_i \geq 0, \quad \forall i = 1, \ldots, n$$

- Note the similarity with α perturbations
  $\Rightarrow$ average error better ($\leq$) than diagonal αBB methods
- Nondiagonal αBB was better on a test suite of 300 generated quadratic functions
  - lower average error in all cases, higher minimum in 279 cases

Properties

- Similar calculation complexity in the quadratic case

Properties

- Similar calculation complexity in the quadratic case
- L&T
  - better for general polynomials
  - a hierarchy of underestimators
  - attractive theoretical convergence

Properties

- Similar calculation complexity in the quadratic case
- L&T
  - better for general polynomials
  - a hierarchy of underestimators
  - attractive theoretical convergence
- αBB
  - only requires smoothness
  - relatively fast
  - introduces additional (linear) constraints and variables
  - slightly tighter and faster in the quadratic case

Properties

- Similar calculation complexity in the quadratic case
- L&T
  - better for general polynomials
  - a hierarchy of underestimators
  - attractive theoretical convergence
- αBB
  - only requires smoothness
  - relatively fast
  - introduces additional (linear) constraints and variables
  - slightly tighter and faster in the quadratic case
- Conclusion: your best choice is problem-dependent!

References

Claire S. Adjiman, S. Dallwig, Christodoulos A. Floudas, and A. Neumaier.
A global optimization method, αBB, for general twice-differentiable constrained NLPs – I. Theoretical advances.
Computers & Chemical Engineering, 22(9):1137–1158, 1998.

Jean B. Lasserre and Tung Phan Thanh.
Convex underestimators of polynomials.
Journal of Global Optimization, 2012.

A. Skjäl, T. Westerlund, R. Misener, and Christodoulos A. Floudas.
A generalization of the classical αBB convex underestimation via diagonal and non-diagonal quadratic terms.
Journal of Optimization Theory and Applications, 154(2):462–490, 2012.

Thank you for listening!

Thank you for listening!

Questions?

L&T, Function Form

$$
\begin{cases}
\max\limits_{h \in \mathbb{R}[\mathbf{x}]_d, \sigma_j, \theta_\ell} & \int\limits_{\mathbf{B}} h \, d\lambda \\
 & f(\mathbf{x}) = h(\mathbf{x}) + \sum_{j=0}^{n} \sigma_j(\mathbf{x}) g_j(\mathbf{x}) \quad \forall \mathbf{x} \\
\text{s.t.} & \mathbf{T}h(\mathbf{x}, \mathbf{y}) = \sum_{j=0}^{n} \theta_j(\mathbf{x}, \mathbf{y}) g_j(\mathbf{x}) \\
 & \qquad + \theta_{n+1}(\mathbf{x}, \mathbf{y}) g_{n+1}(\mathbf{x}, \mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \\
 & \sigma_0 \in \Sigma[\mathbf{x}]_k, \ \sigma_j \in \Sigma[\mathbf{x}]_{k-1}, \ j \geq 1 \\
 & \theta_0 \in \Sigma[\mathbf{x}, \mathbf{y}]_k, \theta_j \in \Sigma[\mathbf{x}, \mathbf{y}]_{k-1}, \ j \geq 1,
\end{cases}
$$

L&T, SDP Form

$$
\begin{cases}
\max\limits_{h \in \mathbb{R}[\mathbf{x}]_d, \mathbf{Z}_j, \Theta_\ell} & \sum\limits_{\alpha \in \mathbb{N}_d^n} h_\alpha \gamma_\alpha \\
& f_\alpha = h_\alpha + \sum\limits_{j=0}^{n} \langle \mathbf{Z}^j, \mathbf{C}_\alpha^j \rangle, \qquad \forall \alpha \in \mathbb{N}_{2k}^n \\
\text{s.t.} & (\mathbf{T}h)_{\alpha\beta} = \sum\limits_{\ell=0}^{n+1} \langle \Theta_\ell, \Delta_{\alpha\beta}^\ell \rangle, \qquad \forall (\alpha, \beta) \in \mathbb{N}_{2k}^{2n} \\
& \mathbf{Z}^j, \ \Theta^\ell \succeq 0, \quad j = 0, \ldots, n; \ \ell = 0, \ldots, n+1,
\end{cases}
$$