Dual certificates and efficient rational sum-of-squares decompositions for polynomial optimization over compact sets

Maria Macaulay (Joint with Dávid Papp)

North Carolina State University

June 7-11, 2021

NC STATE UNIVERSITY

Supported in part by NSF DMS-1719828 and NSF DMS-1847865.

Outline

Nonnegativity certificates

- A *nonnegativity certificate* is a representation of a polynomial p that makes the nonnegativity of p on a semialgebraic set S apparent.
- Example: Weighted sums of squares
- Let S be a semialgebraic set

$$S = \{\mathbf{x} \in \mathbb{R}^U \mid g_i(\mathbf{x}) \geq 0, \text{ for all } i = 1, \ldots, m\}$$

  with each g_i a polynomial
- Denote the cone of weighted sums-of-squares (WSOS) polynomials by Σ, so

$$\Sigma = \left\{ \sum_{i=1}^{m} g_i s_i \mid s_i \text{ a sum of squares polynomial}, \deg(s_i) \leq 2d_i, i = 1, \ldots, m \right\}$$

- Any element of Σ is nonnegative on S
- If p is WSOS, then p is nonnegative on S
  - WSOS representation is a nonnegativity certificate for p.

Nonnegativity certificates

- A *nonnegativity certificate* is a representation of a polynomial p that makes the nonnegativity of p on a semialgebraic set S apparent.
- Example: Weighted sums of squares
- Let S be a semialgebraic set

$$S = \{\mathbf{x} \in \mathbb{R}^U \mid g_i(\mathbf{x}) \geq 0, \text{ for all } i = 1, \dots, m\}$$

with each g_i a polynomial

- Denote the cone of weighted sums-of-squares (WSOS) polynomials by Σ, so

$$\Sigma = \left\{ \sum_{i=1}^{m} g_i s_i \mid s_i \text{ a sum of squares polynomial, } \deg(s_i) \leq 2d_i, i = 1, \dots, m \right\}$$

- Any element of Σ is nonnegative on S
- If p is WSOS, then p is nonnegative on S
  - WSOS representation is a nonnegativity certificate for p.

Nonnegativity certificates

- A *nonnegativity certificate* is a representation of a polynomial p that makes the nonnegativity of p on a semialgebraic set S apparent.
- Example: Weighted sums of squares
- Let S be a semialgebraic set

$$S = \{\mathbf{x} \in \mathbb{R}^U \mid g_i(\mathbf{x}) \geq 0, \text{ for all } i = 1, \ldots, m\}$$

  with each g_i a polynomial
- Denote the cone of weighted sums-of-squares (WSOS) polynomials by Σ, so

$$\Sigma = \left\{ \sum_{i=1}^{m} g_i s_i \mid s_i \text{ a sum of squares polynomial, } \deg(s_i) \leq 2d_i, i = 1, \ldots, m \right\}$$

- Any element of Σ is nonnegative on S
- If p is WSOS, then p is nonnegative on S
  - WSOS representation is a nonnegativity certificate for p.

Notation and background

Proposition

There exists a linear map Λ, with Λ^* its adjoint, such that $\mathbf{s} \in \Sigma$ if and only if there exists $\mathbf{S}$ satisfying $\mathbf{S} \succcurlyeq \mathbf{0}$ satisfying $\mathbf{p} = \Lambda^*(\mathbf{S})$. The dual cone of Σ is given by $\Sigma^* = \{\mathbf{x} \mid \Lambda(\mathbf{x}) \succcurlyeq \mathbf{0}\}$.

- Example: univariate case, sums of squares over the real line: Λ maps $\mathbf{x}$ to its Hankel matrix.
- So checking if a polynomial is WSOS amounts to finding a positive semidefinite matrix.
- The matrix $\mathbf{S}$ can be taken to be a nonnegativity certificate.
- Λ notation will be used throughout.

Y. Nesterov, "Squared functional systems and optimization problems"

Notation and background

Proposition

There exists a linear map Λ, with Λ^* its adjoint, such that $\mathbf{s} \in \Sigma$ if and only if there exists $\mathbf{S}$ satisfying $\mathbf{S} \succcurlyeq \mathbf{0}$ satisfying $\mathbf{p} = \Lambda^*(\mathbf{S})$. The dual cone of Σ is given by $\Sigma^* = \{\mathbf{x} \mid \Lambda(\mathbf{x}) \succcurlyeq \mathbf{0}\}$.

- Example: univariate case, sums of squares over the real line: Λ maps $\mathbf{x}$ to its Hankel matrix.
- So checking if a polynomial is WSOS amounts to finding a positive semidefinite matrix.
- The matrix $\mathbf{S}$ can be taken to be a nonnegativity certificate.
- Λ notation will be used throughout.

Y. Nesterov, "Squared functional systems and optimization problems"

Notation and background

Proposition

There exists a linear map Λ, with Λ^* its adjoint, such that $\mathbf{s} \in \Sigma$ if and only if there exists $\mathbf{S}$ satisfying $\mathbf{S} \succcurlyeq \mathbf{0}$ satisfying $\mathbf{p} = \Lambda^*(\mathbf{S})$. The dual cone of Σ is given by $\Sigma^* = \{\mathbf{x} \mid \Lambda(\mathbf{x}) \succcurlyeq \mathbf{0}\}$.

- Example: univariate case, sums of squares over the real line: Λ maps $\mathbf{x}$ to its Hankel matrix.
- So checking if a polynomial is WSOS amounts to finding a positive semidefinite matrix.
- The matrix $\mathbf{S}$ can be taken to be a nonnegativity certificate.
- Λ notation will be used throughout.

Y. Nesterov, "Squared functional systems and optimization problems"

Notation and background

Proposition

There exists a linear map Λ, with Λ^* its adjoint, such that $\mathbf{s} \in \Sigma$ if and only if there exists $\mathbf{S}$ satisfying $\mathbf{S} \succcurlyeq \mathbf{0}$ satisfying $\mathbf{p} = \Lambda^*(\mathbf{S})$. The dual cone of Σ is given by $\Sigma^* = \{\mathbf{x} \mid \Lambda(\mathbf{x}) \succcurlyeq \mathbf{0}\}$.

- Example: univariate case, sums of squares over the real line: Λ maps $\mathbf{x}$ to its Hankel matrix.
- So checking if a polynomial is WSOS amounts to finding a positive semidefinite matrix.
- The matrix $\mathbf{S}$ can be taken to be a nonnegativity certificate.
- Λ notation will be used throughout.

Y. Nesterov, "Squared functional systems and optimization problems"

Dual certificates

- Let $g(\cdot)$, $H(\cdot)$ denote the gradient and the Hessian (respectively) of the function

$$f(\mathbf{x}) = -\ln(\det(\Lambda(\mathbf{x}))) \tag{1}$$

Theorem (M. and Papp)

Let $\mathbf{x} \in (\Sigma^*)^\circ$ be arbitrary. Then the matrix $\mathbf{S} = \mathbf{S}(\mathbf{x}, \mathbf{p})$ defined by

$$\mathbf{S}(\mathbf{x}, \mathbf{p}) \stackrel{\text{def}}{=} \Lambda(\mathbf{x})^{-1} \Lambda\Big(H(\mathbf{x})^{-1}\mathbf{p}\Big)\, \Lambda(\mathbf{x})^{-1}$$

satisfies $\Lambda^*(\mathbf{S}) = \mathbf{p}$.

- Using theorem from previous slide, if $\mathbf{S} \succcurlyeq 0$, then $\mathbf{p} \in \Sigma$.
- We say $\mathbf{x}$ is a *dual certificate* for $\mathbf{p} \in \Sigma$ if $H(\mathbf{x})^{-1}\mathbf{p} \in \Sigma^*$.

Gradient certificates

$$\mathbf{S}(\mathbf{x},\mathbf{p}) \stackrel{\text{def}}{=} \Lambda(\mathbf{x})^{-1}\Lambda\big(H(\mathbf{x})^{-1}\mathbf{p}\big)\,\Lambda(\mathbf{x})^{-1}$$

Proposition

For every $\mathbf{p} \in \Sigma^\circ$, there exists a unique $\mathbf{x} \in (\Sigma^*)^\circ$ satisfying $-g(\mathbf{x}) = \mathbf{p}$.

- If $-g(\mathbf{x}) = \mathbf{p}$, then $\mathbf{S}(\mathbf{x},\mathbf{p}) \succ 0$.
- Therefore every $\mathbf{p} \in \Sigma^\circ$ has a dual certificate
- We call $\mathbf{x}$ the *gradient certificate* of $\mathbf{p}$ if $\mathbf{x}$ is a dual certificate for $\mathbf{p}$ *and* $-g(\mathbf{x}) = \mathbf{p}$.

Y. Nesterov, "Squared functional systems and optimization problems"

Dual certificates - properties

- Can use $\mathbf{x}$ to get WSOS decomposition
  - $\mathbf{x}$ dual certificate for $\mathbf{p} \implies \mathbf{S}(\mathbf{x},\mathbf{p}) \succcurlyeq \mathbf{0}$ and $\Lambda^*(\mathbf{S}) = \mathbf{p}$
  - Then can factor $\mathbf{S}(\mathbf{x},\mathbf{p})$ (Cholesky, $\mathrm{LDL}^{\mathrm{T}}$)
  - But $\mathbf{x}$ itself is already a *nonnegativity certificate* for $\mathbf{p}$.
- Cones of certificates: denote by

$$\mathcal{C}(\mathbf{p}) \stackrel{\text{def}}{=} \{\mathbf{x} \in (\Sigma^*)^\circ \mid H(\mathbf{x})^{-1}\mathbf{p} \in \Sigma^*\}$$

$$\mathcal{P}(\mathbf{x}) \stackrel{\text{def}}{=} \{\mathbf{p} \in \Sigma \mid H(\mathbf{x})^{-1}\mathbf{p} \in \Sigma^*\}$$

Theorem (M. and Papp)

The cone $\mathcal{C}(\mathbf{p})$ (resp. $\mathcal{P}(\mathbf{x})$) is a full-dimensional cone whenever $\mathbf{p}$ (resp. $\mathbf{x}$) is in the interior of Σ (resp. Σ^*).

Dual certificates - properties

- Can use $\mathbf{x}$ to get WSOS decomposition
  - $\mathbf{x}$ dual certificate for $\mathbf{p} \implies \mathbf{S}(\mathbf{x},\mathbf{p}) \succcurlyeq \mathbf{0}$ and $\Lambda^*(\mathbf{S}) = \mathbf{p}$
  - Then can factor $\mathbf{S}(\mathbf{x},\mathbf{p})$ (Cholesky, $\mathrm{LDL}^{\mathrm{T}}$)
  - But $\mathbf{x}$ itself is already a *nonnegativity certificate* for $\mathbf{p}$.
- Cones of certificates: denote by

$$\mathcal{C}(\mathbf{p}) \stackrel{\text{def}}{=} \{\mathbf{x} \in (\Sigma^*)^\circ \mid H(\mathbf{x})^{-1}\mathbf{p} \in \Sigma^*\}$$

$$\mathcal{P}(\mathbf{x}) \stackrel{\text{def}}{=} \{\mathbf{p} \in \Sigma \mid H(\mathbf{x})^{-1}\mathbf{p} \in \Sigma^*\}$$

Theorem (M. and Papp)

The cone $\mathcal{C}(\mathbf{p})$ (resp. $\mathcal{P}(\mathbf{x})$) is a full-dimensional cone whenever $\mathbf{p}$ (resp. $\mathbf{x}$) is in the interior of Σ (resp. Σ^*).

Primal vs dual certificates

- Primal certificate: an explicit WSOS decomposition of a polynomial
  - A rewriting of the polynomial
  - A single WSOS decomposition certifies a single polynomial
  - Primal certificate: a matrix $\mathbf{S}$ with $\Lambda^*(\mathbf{S}) = \mathbf{p}$ for a polynomial $\mathbf{p}$
    - Still, a single matrix certifies a single polynomial
- Dual certificate: a vector from the dual cone which certifies a polynomial to be WSOS
  - Distinct from the polynomials they certify
  - A single dual certificate certifies a full-dimensional cone of polynomials
  - A single polynomial is certified by a full-dimensional cone of dual certificates
  - A primal certificate (WSOS decomposition, $\mathbf{S}$ matrix) can be constructed from the dual certificate

Example

- Consider the univariate polynomial p given by $p(z) = 1 - z + z^2 + z^3 - z^4$.
- Show p nonnegative on interval $[-1, 1]$: want to show coefficient vector $\mathbf{p} = (1, -1, 1, 1, -1)$ is a member of $\Sigma^{\mathbf{g}}_{1,2\mathbf{d}}$, with weights $\mathbf{g}(z) = (1, 1 - z^2)$ and degree vector $\mathbf{d} = (2, 1)$.
- $\Lambda : \mathbb{R}^5 \to \mathbb{S}^3 \oplus \mathbb{S}^2$ operator is given by

$$\Lambda(x_0, x_1, x_2, x_3, x_4) = \begin{pmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \\ x_2 & x_3 & x_4 \end{pmatrix} \oplus \begin{pmatrix} x_0 - x_2 & x_1 - x_3 \\ x_1 - x_3 & x_2 - x_4 \end{pmatrix}.$$

- The adjoint operator is given by

$$\begin{aligned}\Lambda^*(\mathbf{S}^1 \oplus \mathbf{S}^2) = (S^1_{00} + S^2_{00}, 2S^1_{01}{+}2S^2_{01}, 2S^1_{02} + S^1_{11} - S^2_{00} + S^2_{11},\\ 2S^1_{12} - 2S^2_{01}, S^1_{22} - S^2_{11}).\end{aligned}$$

- Consider $\mathbf{x} = (5, 0, 5/2, 0, 15/8)$
- Claim: $\mathbf{x}$ certifies that p is nonnegative

Example continued

- By a previous theorem, it is sufficient to verify that

$$\frac{128}{5}\Lambda\left(H(\mathbf{x})^{-1}\mathbf{p}\right) = \begin{pmatrix} 144 & -20 & 72 \\ -20 & 72 & -5 \\ 72 & -5 & 49 \end{pmatrix} \oplus \begin{pmatrix} 72 & -15 \\ -15 & 23 \end{pmatrix} \succcurlyeq \mathbf{0}.$$

- Can also compute rational matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ to certify p by plugging our certificate into the formula for the $\mathbf{S}(\mathbf{x}, \mathbf{p})$ matrix, obtaining

$$\mathbf{S}_1 = \frac{1}{40}\begin{pmatrix} 22 & -5 & -26 \\ -5 & 18 & 5 \\ -26 & 5 & 52 \end{pmatrix} \quad \text{and} \quad \mathbf{S}_2 = \frac{1}{40}\begin{pmatrix} 18 & -15 \\ -15 & 92 \end{pmatrix}.$$

- Factor these using the LDL^{T} form of Cholesky decomposition:

$$p(z) = \frac{11}{20}\left(-\frac{13z^2}{11} - \frac{5z}{22} + 1\right)^2 + \frac{371}{880}\left(z - \frac{20z^2}{371}\right)^2 + \frac{3937z^4}{7420} + \\ + \left(1 - z^2\right)\left(\frac{9}{20}\left(1 - \frac{5z}{6}\right)^2 + \frac{159z^2}{80}\right).$$

Easier-to-check sufficient conditions

- Rather than checking $\Lambda(H(\mathbf{x})^{-1}\mathbf{p}) \succcurlyeq \mathbf{0}$, evaluating formula below is sufficient:

Lemma (M. and Papp)

Let $\Lambda(\cdot) \in \mathbb{R}^{\nu \times \nu}$. Suppose $\mathbf{p} \in \Sigma^\circ$ and let $\mathbf{x} \in (\Sigma^*)^\circ$ be any vector that satisfies the inequality

$$\mathbf{p}^{\mathrm{T}} \left(\mathbf{x}\mathbf{x}^{\mathrm{T}} - (\nu - 1) H(\mathbf{x})^{-1} \right) \mathbf{p} \geq 0.$$

Then $\mathbf{x} \in \mathcal{C}(\mathbf{p})$, equivalently, $\mathbf{p} \in \mathcal{P}(\mathbf{x})$.

- Or check if the certificate is close enough to the gradient certificate:

Corollary

Let $\mathbf{x}, \mathbf{y} \in \Sigma^*$ and $\mathbf{p} \in \Sigma$, with $-g(\mathbf{y}) = \mathbf{p}$. Then if $\|H(\mathbf{x})^{1/2}(\mathbf{x} - \mathbf{y})\| < \frac{1}{2}$, $\mathbf{x}$ certifies $\mathbf{p}$.

- If $-g(\mathbf{x}) = \mathbf{p}$,
  - $\mathbf{x}$ is “central” in $\mathcal{C}(\mathbf{p})$ (result from interior-point method theory)
  - $\mathbf{p}$ is “central” in $\mathcal{P}(\mathbf{x})$

Easier-to-check sufficient conditions

- Rather than checking $\Lambda(H(\mathbf{x})^{-1}\mathbf{p}) \succcurlyeq \mathbf{0}$, evaluating formula below is sufficient:

Lemma (M. and Papp)

Let $\Lambda(\cdot) \in \mathbb{R}^{\nu\times\nu}$. Suppose $\mathbf{p} \in \Sigma^\circ$ and let $\mathbf{x} \in (\Sigma^*)^\circ$ be any vector that satisfies the inequality

$$\mathbf{p}^{\mathrm{T}}\left(\mathbf{x}\mathbf{x}^{\mathrm{T}} - (\nu-1)H(\mathbf{x})^{-1}\right)\mathbf{p} \geq 0.$$

Then $\mathbf{x} \in \mathcal{C}(\mathbf{p})$, equivalently, $\mathbf{p} \in \mathcal{P}(\mathbf{x})$.

- Or check if the certificate is close enough to the gradient certificate:

Corollary

Let $\mathbf{x}, \mathbf{y} \in \Sigma^*$ and $\mathbf{p} \in \Sigma$, with $-g(\mathbf{y}) = \mathbf{p}$. Then if $\|H(\mathbf{x})^{1/2}(\mathbf{x}-\mathbf{y})\| < \frac{1}{2}$, $\mathbf{x}$ certifies $\mathbf{p}$.

- If $-g(\mathbf{x}) = \mathbf{p}$,
  - $\mathbf{x}$ is "central" in $\mathcal{C}(\mathbf{p})$ (result from interior-point method theory)
  - $\mathbf{p}$ is "central" in $\mathcal{P}(\mathbf{x})$

Algorithm: best lower bound from all certificates

- From a previous theorem, we know that both $\mathcal{C}(\mathbf{p})$ and $\mathcal{P}(\mathbf{x})$ are full-dimensional, so
  - We can perturb $\mathbf{p}$ in any direction and still certify it with $\mathbf{x}$, and
  - We can perturb $\mathbf{x}$ in any direction and still certify $\mathbf{p}$.
- Do this iteratively to find *lower bound for* $\mathbf{p}$:
  - Perturb $\mathbf{p}$ to $\mathbf{p} - c$, with $\mathbf{p} - c$ still certified by $\mathbf{x}$
  - Perturb $\mathbf{x}$ to get close to gradient certificate of $\mathbf{p} - c$
  - Result: find (or get close to) best possible upper bound c such that $\mathbf{p} - c$ is WSOS.
- Algorithm given in later slide
  - Guaranteed to converge linearly to optimal bound
  - Requires only one Hessian computation per iteration (bottleneck)
  - Avoids (expensive) computation of gradient certificate

Algorithm: best lower bound from all certificates

- From a previous theorem, we know that both $\mathcal{C}(\mathbf{p})$ and $\mathcal{P}(\mathbf{x})$ are full-dimensional, so
  - We can perturb $\mathbf{p}$ in any direction and still certify it with $\mathbf{x}$, and
  - We can perturb $\mathbf{x}$ in any direction and still certify $\mathbf{p}$.
- Do this iteratively to find *lower bound for* $\mathbf{p}$:
  - Perturb $\mathbf{p}$ to $\mathbf{p} - c$, with $\mathbf{p} - c$ still certified by $\mathbf{x}$
  - Perturb $\mathbf{x}$ to get close to gradient certificate of $\mathbf{p} - c$
  - Result: find (or get close to) best possible upper bound c such that $\mathbf{p} - c$ is WSOS.
- Algorithm given in later slide
  - Guaranteed to converge linearly to optimal bound
  - Requires only one Hessian computation per iteration (bottleneck)
  - Avoids (expensive) computation of gradient certificate

Algorithm: best lower bound from all certificates

- From a previous theorem, we know that both $\mathcal{C}(\mathbf{p})$ and $\mathcal{P}(\mathbf{x})$ are full-dimensional, so
  - We can perturb **p** in any direction and still certify it with **x**, and
  - We can perturb **x** in any direction and still certify **p**.
- Do this iteratively to find *lower bound for* **p**:
  - Perturb **p** to $\mathbf{p} - c$, with $\mathbf{p} - c$ still certified by **x**
  - Perturb **x** to get close to gradient certificate of $\mathbf{p} - c$
  - Result: find (or get close to) best possible upper bound c such that $\mathbf{p} - c$ is WSOS.
- Algorithm given in later slide
  - Guaranteed to converge linearly to optimal bound
  - Requires only one Hessian computation per iteration (bottleneck)
  - Avoids (expensive) computation of gradient certificate

Algorithm: best lower bound from all certificates

- From a previous theorem, we know that both $\mathcal{C}(\mathbf{p})$ and $\mathcal{P}(\mathbf{x})$ are full-dimensional, so
  - We can perturb $\mathbf{p}$ in any direction and still certify it with $\mathbf{x}$, and
  - We can perturb $\mathbf{x}$ in any direction and still certify $\mathbf{p}$.
- Do this iteratively to find *lower bound for* $\mathbf{p}$:
  - Perturb $\mathbf{p}$ to $\mathbf{p} - c$, with $\mathbf{p} - c$ still certified by $\mathbf{x}$
  - Perturb $\mathbf{x}$ to get close to gradient certificate of $\mathbf{p} - c$
  - Result: find (or get close to) best possible upper bound c such that $\mathbf{p} - c$ is WSOS.
- Algorithm given in later slide
  - Guaranteed to converge linearly to optimal bound
  - Requires only one Hessian computation per iteration (bottleneck)
  - Avoids (expensive) computation of gradient certificate

Algorithm: best lower bound from all certificates

- From a previous theorem, we know that both $\mathcal{C}(\mathbf{p})$ and $\mathcal{P}(\mathbf{x})$ are full-dimensional, so
  - We can perturb $\mathbf{p}$ in any direction and still certify it with $\mathbf{x}$, and
  - We can perturb $\mathbf{x}$ in any direction and still certify $\mathbf{p}$.
- Do this iteratively to find *lower bound for* $\mathbf{p}$:
  - Perturb $\mathbf{p}$ to $\mathbf{p} - c$, with $\mathbf{p} - c$ still certified by $\mathbf{x}$
  - Perturb $\mathbf{x}$ to get close to gradient certificate of $\mathbf{p} - c$
  - Result: find (or get close to) best possible upper bound c such that $\mathbf{p} - c$ is WSOS.
- Algorithm given in later slide
  - Guaranteed to converge linearly to optimal bound
  - Requires only one Hessian computation per iteration (bottleneck)
  - Avoids (expensive) computation of gradient certificate

Algorithm: best lower bound from all certificates

- From a previous theorem, we know that both $\mathcal{C}(\mathbf{p})$ and $\mathcal{P}(\mathbf{x})$ are full-dimensional, so
  - We can perturb $\mathbf{p}$ in any direction and still certify it with $\mathbf{x}$, and
  - We can perturb $\mathbf{x}$ in any direction and still certify $\mathbf{p}$.
- Do this iteratively to find *lower bound for* $\mathbf{p}$:
  - Perturb $\mathbf{p}$ to $\mathbf{p} - c$, with $\mathbf{p} - c$ still certified by $\mathbf{x}$
  - Perturb $\mathbf{x}$ to get close to gradient certificate of $\mathbf{p} - c$
  - Result: find (or get close to) best possible upper bound c such that $\mathbf{p} - c$ is WSOS.
- Algorithm given in later slide
  - Guaranteed to converge linearly to optimal bound
  - Requires only one Hessian computation per iteration (bottleneck)
  - Avoids (expensive) computation of gradient certificate

Algorithm: best lower bound from all certificates

- From a previous theorem, we know that both $\mathcal{C}(\mathbf{p})$ and $\mathcal{P}(\mathbf{x})$ are full-dimensional, so
  - We can perturb $\mathbf{p}$ in any direction and still certify it with $\mathbf{x}$, and
  - We can perturb $\mathbf{x}$ in any direction and still certify $\mathbf{p}$.
- Do this iteratively to find *lower bound for* $\mathbf{p}$:
  - Perturb $\mathbf{p}$ to $\mathbf{p} - c$, with $\mathbf{p} - c$ still certified by $\mathbf{x}$
  - Perturb $\mathbf{x}$ to get close to gradient certificate of $\mathbf{p} - c$
  - Result: find (or get close to) best possible upper bound c such that $\mathbf{p} - c$ is WSOS.
- Algorithm given in later slide
  - Guaranteed to converge linearly to optimal bound
  - Requires only one Hessian computation per iteration (bottleneck)
  - Avoids (expensive) computation of gradient certificate

Algorithm

Algorithm 1: Compute the best WSOS lower bound and a dual certificate

input : A polynomial $\mathbf{p}$; a tolerance $\varepsilon > 0$.
parameters: An oracle for computing the barrier Hessian H for Σ; the gradient certificate $\mathbf{x}_1$ for the constant one polynomial; a radius $r \in (0, 1/4)$.
outputs : A lower bound c on the optimal WSOS lower bound c^* satisfying $c^* - c \leq \varepsilon$; a dual vector $\mathbf{x} \in (\Sigma^*)^\circ$ certifying the nonnegativity of $\mathbf{p} - c$.

Set initial $\mathbf{x}$ and c. (closed-form formula)
repeat
    Set $\mathbf{x} := 2\mathbf{x} - H(\mathbf{x})^{-1}(\mathbf{p} - c)$. (Newton step).
    Find the largest real number c_+ such that

$$\|H(\mathbf{x})^{1/2}(\mathbf{x} - H(\mathbf{x})^{-1}(\mathbf{p} - c_+\mathbf{1}))\| \leq \frac{r}{r+1}.$$

    Set $\Delta c := c_+ - c$. Set $c := c_+$.
until $\Delta c \leq \rho_r C\varepsilon$
return c *and* $\mathbf{x}$.

Efficiency

Theorem (M. and Papp)

Algorithm 1 is globally linearly convergent to $c^* = \max\{c \mid \mathbf{p} - c \in \Sigma\}$, the optimal WSOS lower bound for the polynomial $\mathbf{p}$.

- Requires $\mathcal{O}(\dim(\Sigma)^3)$ time per iteration.

Univariate polynomials

- Use Chebyshev bases
  - Represent polynomials in Chebyshev bases to help with conditioning of inverse Hessian matrices
  - Also improves rate of convergence
- Tolerance: user input, $c^* - c < \varepsilon$
  - How close the c is to the optimal weighted sums of squares bound c^*.
- The bound c returned by the algorithm is guaranteed to satisfy

$$c \leq c^* \leq c + \varepsilon$$

- Number of iterations required is $\mathcal{O}(d^5 log(\frac{\|\mathbf{p}\| d}{\varepsilon}))$
- Faster than using all-purpose semidefinite solver to find a positive semidefinite $\mathbf{S}$ matrix

Example continued

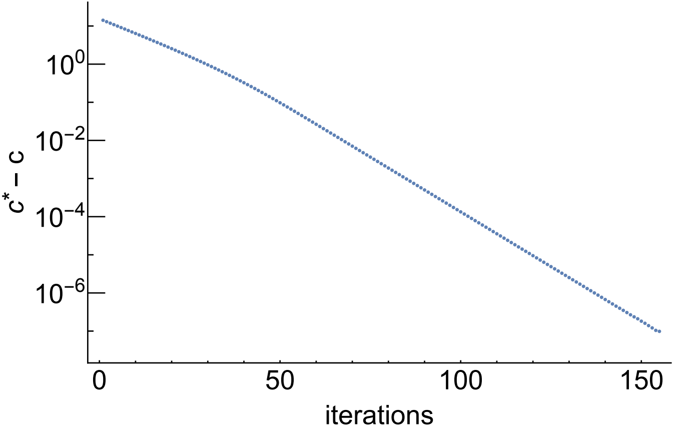
- Continue with coefficient vector in the monomial basis $\mathbf{p} = (1, -1, 1, 1, -1)$, over the interval $[-1, 1]$, represented by the weights $\mathbf{g}(z) = (1, 1 - z^2)$.
- The algorithm with inputs $\mathbf{p}$ and tolerance $\varepsilon = 10^{-7}$ in double-precision floating point arithmetic outputs the bound

$$c = 2^{-53} \cdot 7190305926654593,$$

and a certificate vector

$$\mathbf{x} = 2^{-33} \begin{pmatrix} 173493184462864992 \\ 67729650226350000 \\ -120611300436615200 \\ -161900156381728960 \\ -5796381308580693 \end{pmatrix}$$

Example continued - Plot



The convergence of the sequence of certified lower bounds computed by the algorithm to the minimum of the polynomial studied in the example, illustrating the linear convergence.

Rational certificates

- Output of Algorithm 1 gives $(\mathbf{x}, c)$ such that $\mathbf{x}$ certifies $\mathbf{p} - c \geq 0$.
- Certificate $\mathbf{x}$ is automatically a rational certificate
  - floating point number is *already* a rational number
- Can directly convert $\mathbf{x}$ to an exact rational primal certificate $\mathbf{S}(\mathbf{x}, \mathbf{p})$
- Can also round $\mathbf{x}$ to a nearby rational certificate with smaller denominators:

Lemma (M. and Papp)

Suppose that $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} \leq r < 1/2$ and choose any large enough integer denominator N to satisfy

$$\|H(\mathbf{x})^{1/2}\| \leq \frac{2}{3} \frac{N}{\sqrt{\dim(\Sigma)}}(1 - 2r).$$

Then every point $\mathbf{x}_N \in \frac{1}{N}\mathbb{Z}^{\dim(\Sigma)}$ with $\|\mathbf{x}_N - \mathbf{x}\| \leq \frac{\sqrt{\dim(\Sigma)}}{2N}$ satisfies $\|\mathbf{x}_N - \mathbf{y}\|_{\mathbf{x}_N} \leq \frac{1}{2}$.

Summary

- We use dual certificates to certify polynomials are WSOS
- Dual certificates certify entire cones of polynomials
  - Particularly useful in numerical methods
- Application includes an efficient iterative algorithm to compute the best (WSOS) lower bound for a polynomial and a rational certificate
- Full paper: `http://arxiv.org/abs/2105.11369`