Conditional Moment Relaxations and Sums-of-AM/GM-Exponentials

Riley Murray

California Institute of Technology

MIT Virtual Seminar on Optimization and Related Areas
17 April 2020
The functions of interest

**polynomials**

Parameters $a_i$ in $\mathbb{N}^n$, $c_i$ in $\mathbb{R}$.

Using $x^{a_i} = \prod_{j=1}^{n} x_j^{a_{ij}}$,

$$x \mapsto \sum_{i=1}^{m} c_i x^{a_i}.$$ 

Care about degree: $\max_i \|a_i\|_1$.

For historical and modeling reasons, signomials are often written in *geometric form*

$$y \mapsto \sum_{i=1}^{m} c_i y^{a_i}$$

where $y \in \mathbb{R}_{++}^n$ has the correspondence $y_i = \exp(x_i)$. **We use the exponential form!**

**signomials**

Parameters $a_i$ in $\mathbb{R}^n$, $c_i$ in $\mathbb{R}$.

In “exponential form”,

$$x \mapsto \sum_{i=1}^{m} c_i \exp(a_i \cdot x).$$

Care about number of terms: $m$. 

Riley Murray
The signomial

\[ f(x) = \sum_{i=1}^{m} c_i \exp(a_i \cdot x) \]

is called a posynomial when all \( c_i \geq 0 \).

**Geometric programs (GPs):**

\[ \inf_{x \in \mathbb{R}^n} \{ f(x) : g_i(x) \leq 1 \ \forall \ i \in [k] \} \]

where \( f \) and \( \{g_i\}_{i=1}^{k} \) are posynomials.

Study of GPs initiated by Zener, Duffin, and Peterson (1967). Exponential-form GPs are convex & poly-time solvable via IPMs [1].

Optimization-based engineering design: electrical [2, 3, 4], structural [5, 6], environmental [7], and aeronautical [8, 9].

Epidemiological process control [10, 11, 12], power control and storage [13, 14], self-driving cars [15], gas network operation [16].

Additional applications in healthcare [17], biology [18], economics [19, 20, 21], and statistics [22, 23].
A signomial program (SP) is an optimization problem stated with signomials, e.g.

$$\inf_{x \in \mathbb{R}^n} \left\{ f(x) : g_i(x) \leq 0 \text{ for all } i \text{ in } [k] \right\}.$$ 

Major applications in aircraft design [24, 25, 26, 27, 28] and structural engineering [29, 30, 31, 32]. Additional applications in EE [33], communications [34], and ML [35].
Motivation.

Mathematical Preliminaries.

Sums-of-AM/GM-Exponentials.

Sparsity preservation.

A hierarchy.

Extreme rays.

Conclusion.
The AM/GM-inequality

If $u, \lambda \in \mathbb{R}^m$ are positive and $\mathbf{1}^T \lambda = 1$, then

$$u^\lambda \leq \lambda^T u.$$  

Proof. If $v = \log u$, then $u^\lambda = \exp(\lambda^Tv) \leq \sum_{i=1}^{m} \lambda_i \exp v_i = \lambda^T u$. \hfill \Box

A recent history of using the AM/GM inequality to certify function nonnegativity:

- 1978 and 1989: Reznick [36, 37].
- 2009: Pébay, Rojas and Thompson [38].
- 2012: Paneta, Koepppl, and Craciun [41], and August, Craciun, and Koepppl [42].
- 2016: Iliman and de Wolff [43].

When used for computation, exponents \(\{a_i\}_{i=1}^{m}\) were presumed to be highly structured.

E.g. \(\text{conv}\{a_i\}_{i=1}^{m}\) has \(m - 1\) extreme points, 1 point in its relative interior.
Definitions from convex analysis

A set convex set $K$ is called a cone if

$$x \in K \implies \lambda x \in K \quad \text{for all } \lambda \geq 0;$$

the dual cone to $K$ is

$$K^\dagger = \{ y : y^\top x \geq 0 \text{ for all } x \text{ in } K \}.\$$

and we have $(K^\dagger)^\dagger = \text{cl } K$

A convex set $X$ induces a support function

$$\sigma_X(\lambda) = \sup\{ \lambda^\top x : x \text{ in } X \}.$$ 

The relative entropy function continuously extends

$$D(u, v) = \sum_{i=1}^{m} u_i \log(u_i/v_i) \quad \text{to } \mathbb{R}_+^m \times \mathbb{R}_+^m.$$ 

Important: if you evaluate $D(\cdot, \cdot)$ outside $\mathbb{R}_+^m \times \mathbb{R}_+^m$, you get $+\infty$. 
A trick with convex duality

Start with a **primal** problem

\[
\text{Val}(c) = \inf_{x} \{ c^T x : A x = b, x \geq 0 \}. 
\]

Obtain a **dual** problem

\[
\text{Val}(c) = \sup_{\mu} \{ -b^T \mu : A^T \mu + c \geq 0 \}. 
\]

We will encounter constraints like

\[
\text{Val}(c) + L \geq 0.
\]

Write such a constraint as: *there exists* a \( \mu \) where

\[
A^T \mu + c \geq 0 \quad \text{and} \quad b^T \mu \leq L.
\]
Nonnegativity and optimization

We'll work with sets $X \subset \mathbb{R}^n$. Speaking abstractly, for any $f : \mathbb{R}^n \to \mathbb{R}$

$$f_X^* = \inf \{ f(x) : x \text{ in } X \}$$

$$= \sup \{ \gamma : f - \gamma \text{ is nonnegative over } X \}.$$ 

Make this more concrete. For signomials:

$$C_{\text{NNS}}(A, X) = \left\{ c : \sum_{i=1}^{m} c_i \exp(a_i \cdot x) \geq 0 \forall x \in X \right\}, \quad A = \begin{bmatrix} 0 & \cdots \\ - & a_2 \\ \vdots \\ - & a_m \end{bmatrix} \in \mathbb{R}^{m \times n}.$$ 

So for $f(x) = \sum_{i=1}^{m} c_i \exp(a_i \cdot x)$,

$$f_X^* = \sup \{ \gamma : c - \gamma e_1 \in C_{\text{NNS}}(A, X) \}$$

– where $e_1$ is the $1^{\text{st}}$ standard basis vector in $\mathbb{R}^m$. 

Riley Murray
Duality and moment relaxations

Abbreviate \( \exp(Ax) \in \mathbb{R}^m \) elementwise, and express

\[
C_{\text{NNS}}(A, X) = \{ c : c^\top \exp(Ax) \geq 0 \, \forall \, x \in X \}.
\]

The definition of “dual cone” requires

\[
C_{\text{NNS}}(A, X)^\dagger = \{ v : c^\top v \geq 0 \, \forall \, c \in C_{\text{NNS}}(A, X) \}.
\]

So we end up getting

\[
C_{\text{NNS}}(A, X)^\dagger = \text{co} \{ \exp(Ax) : x \in X \} - \text{a “moment cone.”}
\]

\[
\text{conv} \{ \exp(Ax) : x \in X \} = \left\{ \mathbb{E}_x [\exp(Ax)] : x \sim F, \text{ supp } F \subset X \right\}
\]

Get moment relaxations from conic duality

\[
\sup_{\gamma} \{ \gamma : c - \gamma e_1 \in C_{\text{NNS}}(A, X) \} = \inf \left\{ c^\top v : v \in C_{\text{NNS}}(A, X)^\dagger \right\}.
\]
Motivation.
Mathematical Preliminaries.
Sums-of-AM/GM-Exponentials.
Sparsity preservation.
A hierarchy.
Extreme rays.
Conclusion.
**X-AGE functions**

**Definition.** An **X-AGE** function is an $X$-nonnegative signomial, which has at most one negative coefficient. Generalizes $X = \mathbb{R}^n$ from [44]; see [45].

Consider $f(x) = \sum_{i=1}^{m} c_i \exp(a_i^T x)$. If $c \in \mathbb{R}^m$ has $c \not\in_k = (c_i)_{i \in [m] \setminus k} \geq 0$, then

$$f(x) \geq 0 \iff \sum_{i=1}^{m} c_i \exp([a_i - a_k] \cdot x) \geq 0 \iff \sum_{i \neq k} c_i \exp([a_i - a_k] \cdot x) + c_k \geq 0.$$  

**Theorem (M., Chandrasekaran, & Wierman (2019))**

If $X$ is a convex set, then the conditions

$$c \not\in_k \geq 0 \quad \text{and} \quad c \in C_{\text{NNS}}(A, X)$$

are equivalent to the existence of some $\nu \in \mathbb{R}^m$ satisfying

$$1^T \nu = 0 \quad \text{and} \quad \sigma_X (-A^T \nu) + D(\nu \not\in_k, e c \not\in_k) \leq c_k.$$
A signomial is $X$-SAGE if it can be written as a sum of appropriate $X$-AGE functions.

The cone of coefficients

$$C_{SAGE}(A, X) = \{ c : \{ (\nu^{(k)}, c^{(k)}) \}_{k=1}^m \text{ satisfy } c = \sum_{k=1}^m c^{(k)}, \ 1^\top \nu^{(k)} = 0, \text{ and } \sigma_X \left( -A \nu^{(k)} \right) + D \left( \nu^{(k)}_\setminus_k, c^{(k)}_\setminus_k \right) \leq c^{(k)}_k \ \forall k \in [m] \}$$

is contained within $C_{NNS}(A, X)$.

Consider $f(x) = \sum_{i=1}^m c_i \exp(a_i \cdot x)$ with $a_1 = 0$:

$$f_X^* = \sup \{ \gamma : c - \gamma e_1 \text{ in } C_{NNS}(A, X) \}$$

$$\geq \sup \{ \gamma : c - \gamma e_1 \text{ in } C_{SAGE}(A, X) \} =: f_{X}^{SAGE}.$$

MOSEK + sageopt = off-the-shelf software for computing $f_{X}^{SAGE}$.
Conditional moment relaxations via SAGE

Consider \( f(x) = \sum_{i=1}^{m} c_i \exp(a_i \cdot x) \) with \( a_1 = 0 \). Applying conic duality ...

\[
sup\{ \gamma : c - \gamma e_1 \text{ in } C_{\text{SAGE}}(A, X) \} = f_X^{\text{SAGE}} = \inf \left\{ c^\top v : v \text{ in } C_{\text{SAGE}}(A, X)^\dagger \text{ satisfies } v \cdot e_1 = 1 \right\}
\]

Conic duality reverses inclusions

\[
C_{\text{NNS}}(A, X)^\dagger \subset C_{\text{SAGE}}(A, X)^\dagger.
\]

The dual \( X\)-SAGE cone is

\[
C_{\text{SAGE}}(A, X)^\dagger = \text{cl}\{v : \text{some } z_1, \ldots, z_m \text{ in } \mathbb{R}^n \text{ satisfy } \\
v_k \log(v/v_k) \geq [A - 1a_k]z_k \\
\text{and } z_k/v_k \in X \text{ for all } k \text{ in } [m]\}.
\]

The dual helps with solution recovery. Useful even when \( f_X^{\text{SAGE}} < f_X^* \)!
An example in $\mathbb{R}^3$

Minimize

$$f(x) = \exp(x_1 - x_2)/2 - \exp x_1 - 5 \exp(-x_2)$$

over

$$X = \{ x : (70, 1, 0.5) \leq \exp x \leq (150, 30, 21) \}
\exp(x_2 - x_3)/100 + \exp x_2/100 + \exp(x_1 + x_3)/2000 \leq 1 \}.$$

Compute $f_X^{\text{SAGE}} = -147.85713 \leq f^*_X$, and recover feasible

$$\tilde{x} = (5.01063529, 3.40119660, -0.48450710)$$

satisfying $f(\tilde{x}) = -147.66666$. 
Motivation.
Mathematical Preliminaries.
Sums-of-AM/GM-Exponentials.
Sparsity preservation.
A hierarchy.
Extreme rays.
Conclusion.
Theorem (M., Chandrasekaran, & Wierman)

Fix a vector $c \in \mathbb{R}^m$ with nonempty $N = \{i : c_i < 0\}$.

If $c \in C_{SAGE}(A, X)$, then there exist $X$-AGE vectors $\{c^{(i)}\}_{i \in N}$ where $c_i^{(i)} = c_i$ and

$$c = \sum_{i \in N} c^{(i)}.$$

This is true even if $X$ is not convex.

Proven formally for $X = \mathbb{R}^n$ in [46].

Let $K \supset \mathbb{R}_+^m$ induce $C_i \doteq \{c \in K : c_{\setminus i} \geq 0\}$.

1. Show that for $j \in N$, can eliminate $c^{(j)} \in C_j$ from an existing decomposition.

2. Show that conic combinations of $\{c^{(i)} \in C_i\}_{i \in N}$ can reduce to $c_i^{(i)} = c_i$. 
Sparsity preservation: a univariate example

\[ f(x) = e^{-3x} + e^{-2x} + 4e^x + e^{2x} - 4e^{-x} - 1 - e^{3x} \text{ over } x \leq 0 \]

\[ f_1(x) = 0.88 \cdot e^{-3x} + 0.82 \cdot e^{-2x} + 2.69 \cdot e^x + 0.12 \cdot e^{2x} - 4 \cdot e^{-x} \]
\[ f_2(x) = 0.10 \cdot e^{-3x} + 0.15 \cdot e^{-2x} + 0.90 \cdot e^x + 0.12 \cdot e^{2x} - 1 \]
\[ f_3(x) = 0.02 \cdot e^{-3x} + 0.03 \cdot e^{-2x} + 0.41 \cdot e^x + 0.76 \cdot e^{2x} - e^{3x} \]
Applying Sums-of-Squares (SOS) to an AGE function

Consider the following function on \( \mathbb{R}^2 \)

\[
f(x, y) = 1 - 2e^{(2x+2y)} + \frac{1}{2} \left( e^{8x} + e^{8y} \right).
\]

Use sageopt, round solution, certify \( f \) is \( \mathbb{R}^2 \)-AGE with \( \nu^* = (1, -2, 1/2, 1/2) \).

We can express \( f \) as a sum-of-squares, but this requires new terms

\[
f(x, y) = \left( 1 - 2e^{(2x+2y)} \right)^2 + \frac{1}{2} \left( e^{4x} - e^{4y} \right)^2 \\
= \left( 1 - 2e^{(2x+2y)} + e^{(4x+4y)} \right) + \frac{1}{2} \left( e^{8x} + e^{8y} - 2e^{(4x+4y)} \right).
\]

SOS is the predominant way to certify polynomial nonnegativity.

SAGE can certify polynomial nonnegativity [46] with \( X \subsetneq \mathbb{R}^n \) [45].

Remark: In the special case \( X = \mathbb{R}^n \) with integer exponents, the sparsity result can also be deduced from Jie Wang’s work on Sums-of-Nonnegative-Circuit polynomials [47, 48].
Motivation.
Mathematical Preliminaries.
Sums-of-AM/GM-Exponentials.
Sparsity preservation.
A hierarchy.
Extreme rays.
Conclusion.
A hierarchy of stronger convex relaxations.

The earlier example with $X \subset \mathbb{R}^3$ –

$$f(x) = \exp(x_1 - x_2)/2 - \exp x_1 - 5 \exp(-x_2).$$

We found bounds

$$f_X^{\text{SAGE}} = -147.85713 \leq f^*_X \quad \text{and} \quad f^*_X \leq f(\tilde{x}) = -147.66666.$$

The *modulation trick* lets us construct a sequence of bounds

$$f_X^{(\ell)} := \sup\{\gamma : (\sum_{i=1}^m \exp(a_i \cdot x))^\ell (f(x) - \gamma) \text{ is } X\text{-SAGE}\}.$$

Using MOSEK + sageopt with this particular example,

<table>
<thead>
<tr>
<th>$\ell$</th>
<th>SAGE bound</th>
<th>solve time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-147.85713</td>
<td>0.01</td>
</tr>
<tr>
<td>1</td>
<td>-147.67225</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>-147.66680</td>
<td>0.08</td>
</tr>
<tr>
<td>3</td>
<td>-147.66666</td>
<td>0.26</td>
</tr>
</tbody>
</table>
When introducing $\mathbb{R}^n$-SAGE, Chandrasekaran and Shah proved two results for hierarchies.

No known convergence conditions for $f_X^{(\ell)}$, prior to March 2020.

**Theorem (Wang, Jaini, Yu, Poupart [49])**

*Let $A \in \mathbb{Q}^{m \times n}$ be a rank $n$ matrix with $a_1 = 0$, and consider $f(x) = \sum_{i=1}^{m} c_i \exp(a_i \cdot x)$. If $X$ is a compact convex set, then*

$$\lim_{\ell \to \infty} f_X^{(\ell)} = f_X^*.$$

*Assumes nothing about the representation of $X$.*

Compare to the canonical (non $X$-SAGE) approach, which uses a Lagrangian relaxation:

$$\inf_{x} \{ f(x) : g(x) \geq 0 \} \geq \sup_{\gamma, \lambda} \{ \gamma : f - \gamma - \sum_{i} \lambda_i \cdot g_i \in \Lambda, \lambda_i \in \Lambda' \}$$

$\Lambda, \Lambda'$ are sets of nonnegative functions.
<table>
<thead>
<tr>
<th>Motivation</th>
<th>Preliminaries</th>
<th>Sums-of-AM/GM-Exponentials</th>
<th>Sparsity preservation</th>
<th>A Hierarchy</th>
<th>Extreme rays</th>
<th>Conclusion</th>
<th>Appendices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Mathematical Preliminaries</td>
<td>Sums-of-AM/GM-Exponentials</td>
<td>Sparsity preservation</td>
<td>A hierarchy</td>
<td>Extreme rays</td>
<td>Conclusion</td>
<td>Appendices</td>
</tr>
<tr>
<td>Extreme rays</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Affine matroid-theoretic circuits

A **circuit** is a minimal affinely dependent \( \{a_i\}_{i \in I} \subset \mathbb{R}^n \).

A circuit is **simplicial** if \( \text{conv}\{a_i\}_{i \in I} \) has \(|I| - 1\) extreme points.

As a matter of notation, let \( \mathcal{L}^k = \{\lambda : \lambda^\top 1 = 0, \lambda_k = -1, \lambda \setminus k \geq 0\} \).

Simplicial circuits obtained from \( A \in \mathbb{R}^{m \times n} \) are 1-to-1 with certain \( \lambda \in \mathbb{R}^m \)

\[ \lambda \in \mathcal{L}^k \text{ for some } k \in [m], \quad A^\top \lambda = 0 \quad \text{and} \quad \{a_i : \lambda_i > 0\} \text{ is affinely independent.} \]
M., Chandrasekaran, and Wierman [46] determined extreme rays of

\[ C_{\text{AGE}}(A, i) = \{ c \in C_{\text{NNS}}(A, \mathbb{R}^n) : c_{\setminus i} \geq 0 \}. \]

The ordinary SAGE cone is a Minkowski sum

\[ C_{\text{SAGE}}(A) = \sum_{i=1}^{m} C_{\text{AGE}}(A, i). \]

Katthän, Naumann, and Theobald [50] completely determined \( \text{ext} C_{\text{SAGE}}(A) \).

Forsgård and de Wolff [51] studied \( \partial C_{\text{SAGE}}(A) \) in detail; defined

\[ \text{Rez}(A) = \text{co}\{ \lambda \in \mathbb{R}^m : \lambda \text{ is a simplicial circuit w.r.t. } A \}. \]

Combine [50, 51] to clearly link \( \text{ext} C_{\text{SAGE}}(A) \) and \( \text{ext} \text{Rez}(A) \).
The following is ongoing, joint work with Helen Naumann and Thorsten Theobald.

**Definition.** A simplicial $X$-circuit induced by $A \in \mathbb{R}^{m \times n}$ is a vector $\lambda^* \in \mathbb{R}^m$ where

1. $\lambda^* \in \mathcal{L}_k$ for some $k \in [m],$
2. $\sigma_X (-A^T \lambda^*) < +\infty$, and
3. if $\lambda \mapsto \sigma_X (-A^T \lambda)$ is linear on $[\lambda_1, \lambda_2] \subset \mathcal{L}^k$, then $\lambda^* \not\in \text{relint}[\lambda_1, \lambda_2].$

A clean generalization from $X = \mathbb{R}^n$.

Provides the basis for a “Reznick cone” with conditional SAGE certificates.

Particularly informative when $X$ is a polyhedron.

E.g., if $X$ is a polyhedron, then $C_{\text{SAGE}}(A, X)$ is power-cone representable.
Motivation.

Mathematical Preliminaries.

Sums-of-AM/GM-Exponentials.

Sparsity preservation.

A hierarchy.

Extreme rays.

Conclusion.
Some open problems

1. When do we have $C_{SAGE}(A, X) = C_{NNS}(A, X)$?

   For $X = \mathbb{R}^n$ see [46, 51], and also [46, 47] for polynomials.

2. If $f > 0$ on compact $X$, is there some $g > 0$ so $f \cdot g$ is $X$-SAGE?

   **Resolved in the affirmative for $A \in \mathbb{Q}^{m \times n}$ [49]!** Follow-up questions …

   If “$h =$standard multiplier,” how to bound least $\ell$ where $h^\ell \cdot f$ is $X$-SAGE?

   Irrational $A$? Perhaps leverage Hausdorff continuity.

3. Complexity of testing “$c \in C_{NNS}(\alpha, X)$” with two $c_i < 0$?

   Many possible algorithmic projects (ask me for details).

   More open problems to follow once “$X$-circuit” paper is put on arXiv.
Acknowledgements

Special thanks to my collaborators, on all things SAGE-related:

Venkat Chandrasekaran, Adam Wierman, Helen Naumann, Thorsten Theobald, Fang Xiao, and Berk Orztuk.

Thanks also to the many people who make working in this area exciting:

Sadik Iliman, Timo de Wolff, Marieke Dressler, Henning Seidler, Janin Heuer, Jie Wang, Jen Forsgård, and more!
*Interior-Point Polynomial Algorithms in Convex Programming.*

Digital circuit optimization via geometric programming. 

Application of geometric programming to transformer design. 

A new method for design of robust digital circuits. 
In *Sixth international symposium on quality electronic design (isqed’05)*, pages 676–681, 2005.

Structural optimization using single-term posynomial geometric programming. 
Geometric programming strategies in large-scale structural synthesis.  

A geometric programming model for the optimal design of wastewater treatment plants.  

Geometric programming for aircraft design optimization.  

Solar and gas powered long-endurance unmanned aircraft sizing via geometric programming.  

Optimal resource allocation for control of networked epidemic models.  
Optimal resource allocation for network protection against spreading processes.


Power control by geometric programming.

Optimal energy consumption model for smart grid households with energy storage.
Period optimization for hard real-time distributed automotive systems.


Optimal resource allocation for pervasive health monitoring systems with body sensor networks.

Optimization of biotechnological systems through geometric programming.
Theoretical Biology and Medical Modelling, 4(1), September 2007.
Estimating the firm value distribution function by entropy optimization and geometric programming.  

Fuzzy pricing, marketing and service planning in a fuzzy inventory model: A geometric programming approach.  

[21] Seyed Ahmad Yazdian, Kamran Shahanaghi, and Ahmad Makui.  
Joint optimisation of price, warranty and recovery planning in remanufacturing of used products under linear and non-linear demand, return and cost functions.  

[22] M. MAZUMDAR and T. R. JEFFERSON.  
Maximum likelihood estimates for multinomial probabilities via geometric programming.  
[23] Paul D. Feigin and Ury Passy.
The geometric programming dual to the extinction probability problem in simple branching processes.

Turbofan engine sizing and tradeoff analysis via signomial programming.

Efficient aircraft multidisciplinary design optimization and sensitivity analysis via signomial programming.

Optimal aircraft design decisions under uncertainty via robust signomial programming.
[27] Philippe G. Kirschen and Warren W. Hoburg.  
The power of log transformation: A comparison of geometric and signomial programming with general nonlinear programming techniques for aircraft design optimization.  

Application of signomial programming to aircraft design.  

Optimal design of pitched laminated wood beams.  

Efficient optimization of space trusses.  
References VIII

Optimum design of composite hybrid plate girders.

Second-order method of generalized geometric programming for spatial frame optimization.

Inductor design using signomial programming.

[34] Mung Chiang.
Nonconvex Optimization for Communication Networks, pages 137–196.
[35] Han Zhao, Pascal Poupart, and Geoffrey J Gordon. 
A unified approach for learning the parameters of sum-product networks. 

[36] Bruce Reznick. 
Extremal PSD forms with few terms. 

[37] Bruce Reznick. 
Forms derived from the arithmetic-geometric inequality. 

[38] Philippe P. Pébay, J. Maurice Rojas, and David C. Thompson. 
Optimization and NP-r-completeness of certain fewnomials. 
Lower bounds for polynomials using geometric programming.  

Lower bounds on the global minimum of a polynomial.  

[41] Casian Pantea, Heinz Koeppl, and Gheorghe Craciun.  
Global injectivity and multiple equilibria in uni- and bi-molecular reaction networks.  

Finding invariant sets for biological systems using monomial domination.  

[43] Sadik Iliman and Timo de Wolff.  
Amoebas, nonnegative polynomials and sums of squares supported on circuits.  
[44] Venkat Chandrasekaran and Parikshit Shah.  
Relative entropy relaxations for signomial optimization.  

[45] Riley Murray, Venkat Chandrasekaran, and Adam Wierman.  
Signomial and polynomial optimization via relative entropy and partial dualization, 2019.

[46] Riley Murray, Venkat Chandrasekaran, and Adam Wierman.  
Newton polytopes and relative entropy optimization, 2018.

Nonnegative polynomials and circuit polynomials, 2018.


The algebraic boundary of the sonc cone, 2019.
Appendices

- Exactness analysis.
- Software.
- Optimization with nonconvex constraints.
- Log-log convex ("geometrically convex") functions.
$\mathbb{R}^n$-SAGE Exactness

**Motivation**

**Preliminaries**

**Sums-of-AM/GM-Exponentials**

**Sparsity preservation**

**A Hierarchy**

**Extreme rays**

**Conclusion**

**Appendices**
Theorem (7)

If $\text{conv}(A)$ is simplicial, and $c_i \leq 0$ for all nonextremal $a_i$, then $c \in C_{\text{NNS}}(A, \mathbb{R}^n)$ if and only if $c \in C_{\text{SAGE}}(A, \mathbb{R}^n)$. 
Theorem (7)

If $\text{conv}(A)$ is simplicial, and $c_i \leq 0$ for all nonextremal $a_i$, then $c \in C_{\text{NNS}}(A, \mathbb{R}^n)$ if and only if $c \in C_{\text{SAGE}}(A, \mathbb{R}^n)$.

$$f(x) = (e^{x_1} - e^{x_2} - e^{x_3})^2$$
We say that $A$ can be **partitioned into $\ell$ faces** if we can permute its rows so that $A = [A^{(1)}; \ldots; A^{(\ell)}]$ where $\{\text{conv } A^{(i)}\}_{i=1}^{\ell}$ are mutually disjoint faces of $\text{conv}(A)$. 

![Partitioning a Newton polytope](image-url)
Partitioning a Newton polytope

Theorem (8)

If \( \{A^{(i)}\}_{i=1}^{\ell} \) are matrices partitioning \( A = [A^{(1)}; \ldots ; A^{(\ell)}] \), then

\[
C_{\text{NNS}}(A, \mathbb{R}^n) = \bigoplus_{i=1}^{\ell} C_{\text{NNS}}(A^{(i)}, \mathbb{R}^n)
\]

–and the same is true of \( C_{\text{SAGE}}(A, \mathbb{R}^n) \).

Sanity checks:

All matrices \( A \) admit a trivial partition with \( \ell = 1 \).

If all \( a_i \) are extremal, then \( C_{\text{NNS}}(A, \mathbb{R}^n) = \mathbb{R}_+^m \).
Theorem (9)

Suppose $A$ can be partitioned into faces where

1. each simplicial face has $\leq 2$ nonextremal exponents, and
2. all other faces contain at most one nonextremal exponent.

Then $C_{\text{SAGE}}(A, \mathbb{R}^n) = C_{\text{NNS}}(A, \mathbb{R}^n)$.

Violate the first hypothesis? Consider

$$f(x) = (e^{x_1} - e^{x_2} - e^{x_3})^2$$

not SAGE, per C&S’16.

Violate the second hypothesis? Consider

$$A^T = [e_1, e_2, 2e_1, 2e_2, 2(e_1 + e_2), 0],$$

for which $(-4, -2, 3, 2, 1, 1.8) \in C_{\text{NNS}}(A, \mathbb{R}^2) \setminus C_{\text{SAGE}}(A, \mathbb{R}^2)$. 
Optimization with nonconvex constraints

Q: What should we do when some constraints are nonconvex?

A: Combine $X$-SAGE certificates with Lagrangian relaxations.

Concretely, suppose we want to minimize $f$ over

$$\Omega = X \cap \{x : g(x) \leq 0\}$$

where $X$ is convex, but $g_1, \ldots, g_k$ are nonconvex signomials.

Then, if $\lambda_1, \ldots, \lambda_k$ are nonnegative dual variables, we have

$$\inf_{x \in \Omega} f(x) \geq \sup \left\{ \gamma : f + \sum_{i=1}^{k} \lambda_i g_i - \gamma \text{ is } X\text{-SAGE} \right\}.$$
The SimPleAC aircraft design problem

From Warren Hoburg’s PhD thesis.

Problem statistics:
- 140 variables.
- 89 inequality constraints (1 nonconvex).
- 67 equality constraints (15 nonconvex).

Performance of the most basic SAGE relaxation:
- bound "cost $\geq 2957"$ (roughly match a known solution).
- MOSEK solves in two seconds, on a six year old laptop.
- solution recovery fails (numerical issues).
import sageopt as so

y = so.standard_sig_monomials(3)
f = 0.5*y[0]/y[1] - y[0] - 5/y[1]
ineqs = [100 - y[1]/y[2] - y[1] - 0.05*y[0]*y[2],
         y[0] - 70, y[1] - 1, y[2] - 0.5,
         150 - y[0], 30 - y[1], 21 - y[2]]
X = so.infer_domain(f, gts, [])

prob = so.sig_relaxation(f, X, form='dual')
prob.solve()
solutions = so.sig_solrec(prob)
The `sageopt` python package

```python
...  # define f, X as before
from sageopt import coniclifts as cl

modulator = so.Signomial(f.alpha, np.ones(f.m)) ** 3
gamma = cl.Variable()

h = modulator * (f - gamma)
con = cl.PrimalSageCone(h.c, h.alpha, X, 'con_name')
prob = cl.Problem(cl.MAX, gamma, [con])

prob.solve()
age_vecs = [v.value for v in con.age_vectors.values()]
age_sigs = [so.Signomial(h.alpha, v) for v in age_vecs]
h_numeric = so.Signomial(h.alpha, h.c.value)
```
Log-log convexity: examples

With domains $D = \mathbb{R}^n_{++}$:

$$g(x) = \max\{x_1, \ldots, x_n\}$$

$$g(x) = x_1^{a_1} \cdots x_n^{a_n}$$

$$g(x) = \left( \int_x^\infty e^{-t^2} \, dt \right)^{-1}$$

With more restricted domains:

$$x \mapsto (-x \log x)^{-1} \quad D = (0, 1)$$

$$X \mapsto (I - X)^{-1} \quad D = \{ X \in \mathbb{R}^{n \times n}_{++} : \rho(X) < 1 \}$$

$$x \mapsto (\log x)^{-1} \quad D = (1, \infty)$$

Some tractable constraints for $X$-SAGE polynomials:

$$\|x\|_p \leq a \quad x_j^2 = a \quad a \leq \mathbb{P}\{\mathcal{N}(0, \sigma) \geq |x|\}$$

where $a > 0$. 