

===

The Mean-Variance-Skewness-Kurtosis Problem (Draft)

===

Note type project
Tags [#NAG](#) [#finance](#) [#optimization](#)

Table of contents

- ☐ Abstract
- ☐ Introduction
- ☐ Preliminaries
 - ☒ POP
 - ☒ Δ^n – "grid" search
 - ☒ SAGE
 - ☐ Cardinality restrictions.
 - ☐ Cardinality restrictions on POP
 - ☐ Cardinality restrictions on Δ^n – grid search
 - ☐ Cardinality restrictions in SAGE
- ☐ Modeling
 - ☒ Moments
 - ☒ Multi-objective problem:
 - ☒ Approaches
 - ☒ Minkowski Distance objective function (MDOF)
 - ☒ Aspirational problems
 - ☒ MDOF opt. prob.
 - ☒ Bounding the the MDOF
 - ☒ Upper bound via POP
 - ☒ Lower bound via sampling
 - ☐ Looser bound that can be computed.
 - ☐ sampling upper bounds
 - ☐ SAGE lower bounds.
- ☐ Numerics
 - ☐

Abstract

Introduction

Problem

Mean-Varaince-Skewness-Kurtosis (MVS_K) optimization problem comes from portfolio optimization in finance. It is an extension of the better known Markowitz or Mean-Variance optimization problem. The task is to diversify a selection of stocks, thereby protecting against the possibility that a significant portion of the portfolio suffers losses at the same time. This extension promises to also consider rare events, i.e. fat tails.

(Other peoples work)

(Our contribution)

(Paper structure)

Preliminaries

In this section we review techniques from three different domains of optimization: Polynomial optimization (POP), Signomial optimization (SOP), and Sampling. These techniques are complimentary as will be seen in the [Modeling](#) section.

Polynomial Optimization

We start by recalling the definition of a POP:

$$\begin{aligned} f_* &:= \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \\ \text{s.t. } & g_i(\mathbf{x}) \geq 0 \quad \forall i \in [m] \\ & h_j(\mathbf{x}) = 0 \quad \forall j \in [k]. \end{aligned} \tag{POP}$$

where $f, g_1, g_2, \dots, g_m, h_1, \dots, h_k \in \mathbb{R}[\mathbf{x}]$ are polynomials in n variables.

Measure reformulation

One can reformulate (POP) into an optimization problems over measures.

$$\begin{aligned} f_{\text{mea}} &:= \inf \int f(x) d\mu \\ \text{s.t. } & \mu \in \mathcal{M}_+(K), \\ & \int d\mu = 1, \end{aligned} \tag{MEA}$$

Where, $K := \{x \in \mathbb{R}^n \mid g_i(x) \geq 0 \ i \in [m], \ h_j(x) = 0 \ j \in [k]\}$ and $\mathcal{M}_+(K)$ is the set of all positive measures supported on K .

Proof: [Result Polynomial Optimization is Equivalent to Optimization over Probability Measures](#)

Moment reformulation

We can further reformulate (MEA) into an optimization problem over moments:

$$\begin{aligned} f_{\text{mom}} &:= \inf L(f) \\ \text{s.t. } & L : \mathbb{R}[\mathbf{x}] \rightarrow \mathbb{R} \text{ linear} \\ & L(1) = 1 \\ & M(g_0 L) := L([\mathbf{x}][\mathbf{x}]^T) \succeq 0 \\ & M(g_i L) := L(g_i(\mathbf{x})[\mathbf{x}][\mathbf{x}]^T) \succeq 0 \quad \forall i \in [m] \\ & M(h_j L) := L(h_j(\mathbf{x})[\mathbf{x}][\mathbf{x}]^T) = 0 \quad \forall j \in [k] \end{aligned} \tag{MOM}$$

We use the standard proof [Result Optimization over Probability Measures is equivalent to Optimization over Moments](#)

Relaxation

All three of the above formulations of (POP) are computationally intractable so we resort to approximations. The reformulation that is most susceptible to relaxation is (MOM). In fact one can build a hierarchy of relaxations, \citation, the result is

$$\begin{aligned} f_{mom,t} &:= \inf L(f) \\ s.t. \quad &L : \mathbb{R}[\mathbf{x}]_{2t} \rightarrow \mathbb{R} \text{ linear} \\ &L(1) = 1 \\ &M(g_0 L) := L([\mathbf{x}]_t [\mathbf{x}]_t^T) \succeq 0 \\ &M(g_i L) := L(g_i(\mathbf{x}) [\mathbf{x}]_{t-dg_i} [\mathbf{x}]_{t-dg_i}^T) \succeq 0 \quad \forall i \in [m] \\ &M(h_j L) := L(h_j(\mathbf{x}) [\mathbf{x}]_{t-dh_j} [\mathbf{x}]_{t-dh_j}^T) = 0 \quad \forall j \in [k] \end{aligned} \quad (\text{reMOM})$$

where $dg_i := \lceil \frac{\deg g_i}{2} \rceil$. If $\deg g_i$ is odd then some modifications can be applied for a stronger relaxation.

The usual results form POP (I still don't know if I want to say much in this regard)

I just collect the usual things people do in POP. Most of these theorems are taken from [Utility - References > ^LV2018](#)

\todo Lookup the proper references:

Convergence (Theorem 13.1.1 from [Utility - References > ^LV2018](#))

Assume that $\mathcal{M}(g, h) := \{p(\mathbf{x}) := \sum_{i=0}^m c_i g_i(\mathbf{x}) : c \in \mathbb{R}_+^{m+1}\} + \langle h_1, \dots, h_k \rangle$ is Archimedean. Then, the bounds $f_{p_{mom,t}} \rightarrow f_*$ as $t \rightarrow \infty$.

The theorem and proof do not mention the ideal. I don't see anything changing if it does.

The Archimedean property holds for the simplex constraints because:

$$0 \leq x_i \leq 1 \text{ for all } i \text{'s so } n - \sum_{i=1}^n x_i^2 \in \mathcal{M}(g, h).$$

Flatness (Theorem 13.5.1 from [Utility - References > ^LV2018](#))

Assume L is an optimal solution to the program (reMOM) and that there exists an $s \in \mathbb{N}$ for which: $\max\{\deg(f), \deg(K)\} \leq s \leq t$ and $\text{rank} M_s(L) = \text{rank} M_{s-d_k}(L)$ then:

$$f_{mom,t} = f_*$$

The theorem and proof do not mention the ideal. I don't see anything changing if it does.

Recovery of the solution. (Algorithm 4.2 in [Utility - References > Las2010](#))

\todo

Signomial optimization (SOP)

We recall some definition from SOP and move on to the result we will be using.

Definition: A function $f(x)$ is called a [Signomial](#) if it is of the form

$$\sum_{j \in [\ell]} c_j \exp(\alpha^{(j)^T} x)$$

Where $c \in \mathbb{R}^\ell$ and $\alpha^{(j)} \in \mathbb{R}^n$ for all $j \in [n]$.

Definition: We define a constrained [signomial optimization](#) problem as follows:

$$\min f(x)$$

(SIG)

$$\text{s. t. } g_i(x) \geq 0 \forall i \in [m]$$

where f, g_1, g_2, \dots, g_m are all signomials having exponents in $\mathcal{A} := \{\alpha^{(j)}\}_{j=1}^\ell \subset \mathbb{Q}^n$.

Definition: The [SAGE relaxation](#) of the (SIG) is

$$\begin{aligned} f_{\text{SAGE}}^{(p,q)} &= \sup_{\gamma \in \mathbb{R}, s_h \in \text{SAGE}(E_p(\mathcal{A}))} \gamma \\ \text{s. t. } f(x) - \gamma - \sum_{h(x) \in R_q(C)} s_h(x)h(x) &\in \text{SAGE}(E_{p+q}(\mathcal{A})) \end{aligned}$$

Where

$$C := \{g_i(x)\}_{i=1}^m$$

$$R_q(C) := \left\{ \prod_{k=1}^q h_k \mid h_k \in \{1\} \cup C \right\}$$

$$E_{p+1}(\mathcal{A}) := \left\{ \sum_{j=1}^\ell \lambda_j \alpha^{(j)} \mid \sum_{j=1}^\ell \lambda_j \leq p+1, \lambda_j \in \mathbb{N} \right\}$$

$$\text{SAGE}(\mathcal{B}) := \left\{ f \mid f = \sum_{i=1}^n f_i \text{ where } f_i \text{ is an AM-GM exponentia} \right\} \text{ [Sums of AM-GM exponentials](#)}$$

Theorem (4.2) : [Utility - References > ^CS2016](#)

Fix a set of rational exponents $\mathcal{A} := \{\alpha^{(j)}\}_{j=1}^\ell \subset \mathbb{Q}^n$, and let $f(x)$ and $C := \{g_i(x)\}_{i=1}^m$ be signomials with respect to these exponents. Let $R_q(C)$ be as defined above, and let the constraint set be

$$K_C = \{x \mid g_i(x) \geq 0, i = 1, \dots, m\}. \text{ Suppose inequalities of the form } U \geq \exp(\alpha^{(j)^T} x) \geq L, j = 1, \dots, \ell \text{ for}$$

$U, L \in \mathbb{R}_{++}$ are formally specified in the list of constraints C (explicitly serving as witnesses of the compactness of K_C), and suppose $f(x)$ is strictly positive for all $x \in K_C$. Then there exist $p, q \in \mathbb{N}$ and SAGE

functions $s_h(x) \in \text{SAGE}(E_p(\mathcal{A}))$ indexed by $h \in R_q(C)$ such that $f(x) - \sum_{h(x) \in R_q(C)} s_h(x)h(x) \in \text{SAGE}(E_{p+q}(\mathcal{A}))$.

Proof (incomplete) I'll look this up for the talk on 24 Nov anyway \square

Other Optimization approaches

The topic of sampling based optimization is vast and this paper cannot hope to cover the topic in sufficient detail. We will focus only on a lattice search approach [\citation](#). The reader is free to skip the details of this chapter or substitute the techniques for any other sampling based optimization technique like random walk,..... [\citations](#) to name a few.

[Difference Between Grid Bound and Optima on a Simplex](#)

Theorem from [Utility - References > ^z2015](#)

For any homogeneous polynomial f in n -variables of degree d let $r \in \mathbb{N}$ then

$$f_{\Delta(n,r)} - f_{\min, \Delta^n} \leq \left(1 - \frac{r^d}{r^d}\right) \binom{2d-1}{d} d^d (f_{\max, \Delta^n} - f_{\min, \Delta^n})$$

Where

$$r^{\underline{d}} = r(r-1)(r-2) \cdots (r-d+1)$$

$f_{\Delta(n,r)}$ is the [Grid search bound on polynomial optimization over the simplex](#), i.e., the optima over

$$\Delta(n, r) := \{x \in \Delta^n : rx \in \mathbb{N}^n\}.$$

f_{\max, Δ^n} is the [Optima over a Simplex](#).

Cardinality restrictions (No need to look at this just yet.)

The above techniques all suffer from computational cost as their precision increases. In this section we propose a natural constraint that one can impose on the MVSK problem in order to reduce the memory and solve time for these various approaches.

Our technique is to restrict the number of nonzero variables. This cardinality constraint on the solution is innovative on our part as it has been studied in [\citations](#). However its application to POP in this setting is novel.

Practically speaking it would be better have fewer rather more stocks in ones portfolio. The reasons being that it reduces complexity and transaction costs. One way to do this is to say that we are only interested in portfolios that consist of $k < N$ stocks where N , recall is the number of candidate stocks to choose from. We hence, impose the constraint that at most $k < N$ for the weights we assign may be nonzero. This can be phrased as a polynomial constraint:

$$\mathbf{x}^\alpha = 0 \quad \forall \quad \alpha \in \mathbb{N}^n \text{ s.t. } |\text{supp}(\alpha)| = k + 1$$

This constraint is saying is that among $k + 1$ distinct variables at least one variable is zero. This will have consequences in the relaxations [Support Constraint on Variables in Polynomial Optimization](#)

Cardinality restriction in POP

We begin with the tools of polynomial optimization and show how the cardinality constraints reduce complexity in contradistinction to the case of mixed -integer programming [CITATIONS].

Cardinality restriction in SAGE

Cardinality restriction in simplex sampling

The cardinality constraint effectively reduces the simplex domain into several smaller simplexes as follows.

$$\{w \in \Delta_n : \text{supp}(w) \leq \ell < n\} = \bigcup_{s \in S} P_s[\Delta_\ell]$$

where:

- $\Delta_n \subset \mathbb{R}^n$ is the standard n -simplex
- $S := \{y \in \{0, 1\}^n : |y| = \ell\}$

$$\bullet \quad P_s : \Delta_\ell \ni y \mapsto x := \begin{cases} x_i = y_i & i \in s \\ 0 & \text{else} \end{cases} \in \Delta_n$$

We then use the result from Zhao [Utility - References > ^z2015](#)

Modeling

The Data

We are given the returns $r \in \mathbb{R}^{N \times M}$ of N risky assets where M is the number of data points, one can think of this as the time component. The data is daily measurements of a stock price over the span of several years. We will be using ≈ 100 days of data and ≈ 20 stocks.

The Variables

Portfolio weights can fall into any of the following domains:

- **The Standard Simplex** in this setting there is no shorting nor leveraged positions: $w \in \Delta^N$, where

$\Delta^N := \{x \in [0, 1]^N : \sum_{i=1}^N x_i = 1\}$.

- **The Box** In this setting the investor has shorted and leveraged positions in the market. Obviously there is a bound on how leveraged a position can be $w \in \mathbb{B}^N$, where $\mathbb{B}^N := \{x \in [-b, b]^N\}$ with $b < \infty$. \todo :

what changes?

- "Quadratic domain" \todo : look into what Sh. sent me.

The Moments (Centralized)

We define the k^{th} -moment with weights w to be:

$$\phi_k(r, w) := \mathbb{E}[(w^T r)^k] = \begin{cases} w^T \Phi^{(k)}(\underbrace{w \otimes w \otimes \dots \otimes w}_{(k-1)-many}) & k \text{ is odd} \\ (\underbrace{w \otimes w \otimes \dots \otimes w}_{(k/2)-many})^T \Phi^{(k)}(\underbrace{w \otimes w \otimes \dots \otimes w}_{(k/2)-many}) & k \text{ is even} \end{cases}$$

where

$$\Phi^{(k)} := \begin{cases} \mathbb{E}[\underbrace{r \otimes r \otimes \dots \otimes r}_{(k-1)-many}]^T & k \text{ is odd} \\ \mathbb{E}[(\underbrace{r \otimes r \otimes \dots \otimes r}_{(k/2)-many})(\underbrace{r \otimes r \otimes \dots \otimes r}_{(k/2)-many})^T] & k \text{ is even} \end{cases}$$

with the expectation taken over the data's time component. In order to centralize the moments around the mean we use $\bar{r} := r - \Phi^{(1)}$. This then gives the first four moments:

1. **Mean:** $\mathbb{E}[w^T r] = w^T \Phi^{(1)}$, $\Phi^{(1)}$ is called the expected returns.
2. **Variance:** $\mathbb{E}[w^T \bar{r} \bar{r}^T w] = w^T \Phi^{(2)} w$, where $\Phi^{(2)}$ is the covariance matrix.
3. **Skewness:** $\mathbb{E}[w^T \bar{r} \bar{r}^T w r^T w] = \mathbb{E}[w^T \bar{r} (\bar{r} \otimes \bar{r})^T (w \otimes w)] = w^T \Phi^{(3)} (w \otimes w)$, where $\Phi^{(3)}$ is the "Skewness matrix".
4. **Kurtosis:** $\mathbb{E}[w^T \bar{r} (\bar{r} \otimes \bar{r} \otimes \bar{r})^T (w \otimes w \otimes w)] = (w \otimes w)^T \Phi^{(4)} (w \otimes w)$, where $\Phi^{(4)}$ is the "Kurtosis matrix".

Some observation on the moments:

- Form the definition it is mathematically clear that Variance and Kurtosis are always non negative. Furthermore, it is reasonable to assume that they will also be nonzero in practice.
- The expected returns and skewness can be zero and negative however it is assumed that this can be circumvented by prudent selection of stocks prior to optimization.
- The above process can be further extended to even higher moments. We omit to do so as not to bog the reader down with technical notation.

The Multi Objective Optimization Problem

The goal is to "solve" the following **multi-objective polynomial optimization** problem:

$$\begin{aligned} \max \quad & f_1(w) := w^T \mu \\ \min \quad & f_2(w) := w^T \Phi^{(2)} w \\ \max \quad & f_3(w) := w^T \Phi^{(3)} (w \otimes w) \\ \min \quad & f_4(w) := w^T \Phi^{(4)} (w \otimes w \otimes w) \\ \text{s.t.} \quad & w \in \Delta^N \end{aligned} \tag{0}$$

The hope is that in solving this problem the investor obtains a stock selection w that is both lucrative in that it maximizes the expected returns and skewness while at the same time limiting exposure to risk as quantified by variance and kurtosis. The greatest impediment to solving multi objective optimization problems is weighing the different objectives against each other. This is especially true in the case of MVSK as the difference objective differ in scale and as such are not directly comparable. We briefly review some ways of reformulating (0) into a single objective function optimization problem.

Linear Objective Function

One way to balance the different objectives would be to extend the idea of the mean-variance version and simply consider a linear combination of the four objectives:

$$\begin{aligned} \max \quad & \lambda_1 f_1(w) - \lambda_2 f_2(w) + \lambda_3 f_3(w) - \lambda_4 f_4(w) \\ \text{s.t.} \quad & w \in \Delta^N \end{aligned} \quad (?)$$

Where $\lambda \in \mathbb{R}_+^4$ or (Δ^4) depending on modeling choices. The problem with this objective function is the difference in scale between the different moments. *TODO* show. The advantage of this formulation is its conceptual simplicity.

Polynomial Objective Function

We can extend the previous idea to arbitrary polynomial functions in $f_1(w), \dots, f_4(w)$ as follows:

$$\begin{aligned} \max \quad & P(f_1(w), f_2(w), f_3(w), f_4(w)) \\ \text{s.t.} \quad & w \in \Delta^N \end{aligned} \quad (?)$$

Where $P \in \mathbb{R}[x_1, x_2, x_3, x_4]$ is a polynomial. For example:

$$P(f_1(w), f_2(w), f_3(w), f_4(w)) := f_1(w) + -10f_2(w) + 40f_4(w) - 39f_3(w)f_4(w)^2.$$

However, great care must be taken to ensure that the chosen polynomial will have a meaning relating to the underlying problem. The advantage of this formulation is that we can lower bound it using tools from POP mentioned in the [Preliminaries](#) section.

Minkowski Distance Objective Function

An idea that is prevalent in portfolio optimization is to choose as objective the distance from an idealized solution \citations. To begin consider the individual optimization problems whose solutions are sometimes called the *aspirational* solutions \citations:

$$\begin{aligned} f_{1,\max} = \max \quad & f_1(w) := w^T \Phi^{(1)} \\ \text{s.t.} \quad & w \in \Delta^N \end{aligned} \quad (1)$$

$$\begin{aligned} f_{2,\min} = \min \quad & f_2(w) := w^T \Phi^{(2)} w \\ \text{s.t.} \quad & w \in \Delta^N \end{aligned} \quad (2)$$

$$\begin{aligned} f_{3,\max} = \max \quad & f_3(w) := w^T \Phi^{(3)}(w \otimes w) \\ \text{s.t.} \quad & w \in \Delta^N \end{aligned} \quad (3)$$

$$\begin{aligned} f_{4,\min} = \min \quad & f_4(w) := w^T \Phi^{(4)}(w \otimes w \otimes w) \\ \text{s.t.} \quad & w \in \Delta^N \end{aligned} \quad (4)$$

For a fixed choice (made by the investor) of $\lambda \in \Delta^4$ define the following optimization problem:

$$\begin{aligned} \ell_{\min} := \min \quad & \ell(w) := \left(1 - \frac{f_1(w)}{f_{1,\max}}\right)^{\lambda_1} + \left(\frac{f_2(w)}{f_{2,\min}} - 1\right)^{\lambda_2} \\ & + \left(1 - \frac{f_3(w)}{f_{3,\max}}\right)^{\lambda_3} + \left(\frac{f_4(w)}{f_{4,\min}} - 1\right)^{\lambda_4} \\ \text{s.t.} \quad & w \in \Delta^N \end{aligned} \quad (5)$$

Problem (5) can in practice not be written down explicitly on account of the parameters $\psi_+^{(3)}$ and $\psi_-^{(4)}$ which require the solution of NP-hard problems. In contradistinction the parameters $\psi_+^{(1)}$ and $\psi_-^{(2)}$ can be obtained efficiently from theory and [convex quadratic optimization](#) respectively. First, $\psi_+^{(1)} = \max_{i \in [n]} \mathbb{E}(r_i)$ because the max of a linear objective $c^T x$ over the simplex is simply the maximal entry of c . Second, $\psi_-^{(2)}$ is obtained via quadratic programming exploiting the convexity...(\citation).

We hence introduce two approximate objectives that can be written down explicitly:

$$\begin{aligned}\tilde{\mathcal{F}}_{\min,k,l} := \min \quad \tilde{\mathcal{F}}_{k,l}(w) := & \left(1 - \frac{f_1(w)}{\psi_{\max}^{(1)}}\right)^{\lambda_1} + \left(\frac{f_2(w)}{\psi_{\min}^{(2)}} - 1\right)^{\lambda_2} \\ & + \left(1 - \frac{f_3(w)}{\tilde{\psi}_{\max,k}^{(3)}}\right)^{\lambda_3} + \left(\frac{f_4(w)}{\tilde{\psi}_{\min,l}^{(4)}} - 1\right)^{\lambda_4} \\ s. t. \quad & w \in \Delta^N\end{aligned}\tag{tilde 5}$$

where $\tilde{\psi}_{+,k}^{(3)}$ and $\tilde{\psi}_{-,l}^{(4)}$ are the k -level upper bound on $\psi_+^{(3)}$ and l -level lower bound on $\psi_-^{(4)}$ respectively.

Similarly

$$\begin{aligned}\hat{\mathcal{F}}_{\min} := \min \quad \hat{\mathcal{F}}(w) := & \left(1 - \frac{f_1(w)}{\psi_{\max}^{(1)}}\right)^{\lambda_1} + \left(\frac{f_2(w)}{\psi_{\min}^{(2)}} - 1\right)^{\lambda_2} \\ & + \left(1 - \frac{f_3(w)}{\hat{\psi}^{(3)}}\right)^{\lambda_3} + \left(\frac{f_4(w)}{\hat{\psi}^{(4)}} - 1\right)^{\lambda_4} \\ s. t. \quad & w \in \Delta^N \\ & \hat{\psi}^{(3)} \geq f_3(w) \\ & \hat{\psi}^{(4)} \leq f_4(w)\end{aligned}\tag{hat 5}$$

Where $\hat{\psi}^{(3)}$ and $\hat{\psi}^{(4)}$ can be obtained via sampling techniques like those mentioned in [Sampling based approaches](#).

Lemma (bounding \mathcal{F}_{\min}) If $f_3(w) \geq 0$ for all w then $\hat{\mathcal{F}}_{\min} \leq \mathcal{F}_{\min} \leq \tilde{\mathcal{F}}_{\min,k,l}$ for all $k, l \in \mathbb{N}$.

Proof

$$(\mathcal{F}_{\min} \leq \tilde{\mathcal{F}}_{\min,k,l})$$

$$0 < \psi_{\max}^{(3)} \leq \tilde{\psi}_{\max,k}^{(3)} \iff -\frac{1}{\psi_{\max}^{(3)}} \leq -\frac{1}{\tilde{\psi}_{\max,k}^{(3)}} \iff \left(1 - \frac{f_3(w)}{\psi_{\max}^{(3)}}\right)^{\lambda_3} \leq \left(1 - \frac{f_3(w)}{\tilde{\psi}_{\max,k}^{(3)}}\right)^{\lambda_3}$$

$$\psi_{\min}^{(4)} \geq \tilde{\psi}_{\min,l}^{(4)} > 0 \iff \frac{1}{\psi_{\min}^{(4)}} \leq \frac{1}{\tilde{\psi}_{\min,l}^{(4)}} \iff \left(\frac{f_4(w)}{\psi_{\min}^{(4)}} - 1\right)^{\lambda_4} \leq \left(\frac{f_4(w)}{\tilde{\psi}_{\min,l}^{(4)}} - 1\right)^{\lambda_4}$$

$$(\hat{\mathcal{F}}_{\min} \leq \mathcal{F}_{\min})$$

Only consider w for which $\hat{\psi}^{(3)} \geq f_3(w)$ and $\hat{\psi}^{(4)} \leq f_4(w)$, else we can set the respective terms to zero, then

$$\psi_{\max}^{(3)} \geq \hat{\psi}^{(3)} > 0 \iff -\frac{1}{\psi_{\max}^{(3)}} \geq -\frac{1}{\hat{\psi}^{(3)}} \iff \left(1 - \frac{f_3(w)}{\psi_{\max}^{(3)}}\right)^{\lambda_3} \geq \left(1 - \frac{f_3(w)}{\hat{\psi}^{(3)}}\right)^{\lambda_3}$$

$$0 < \psi_{\min}^{(4)} \leq \hat{\psi}^{(4)} \iff \frac{f_4(w)}{\psi_{\min}^{(4)}} \geq \frac{f_4(w)}{\hat{\psi}^{(4)}} \iff \left(\frac{f_4(w)}{\psi_{\min}^{(4)}} - 1\right)^{\lambda_4} \geq \left(\frac{f_4(w)}{\hat{\psi}^{(4)}} - 1\right)^{\lambda_4}$$

□

Lemma (Recasting (5) as a SOP)

Problem (5) is equivalent to the following signomial optimization problem:

$$\begin{aligned}
f_{\min} &= \min f(x, y) := \sum_{j=1}^4 \exp(\lambda_j x_j) \\
s. t. \quad &\sum_{j=1}^n \exp(y_j) = 1 \\
&\exp(x_1) = \left(1 - \sum_{i=1}^n \frac{\Phi_i^{(1)}}{\psi_{\max}^{(1)}} \exp(y_i)\right) \\
&\exp(x_2) = \left(\sum_{i,j=1}^n \frac{\Phi_{i,j}^{(2)}}{\psi_{\min}^{(2)}} \exp(y_i + y_j) - 1\right) \\
&\exp(x_3) = \left(1 - \sum_{i=1}^n \frac{\Phi_{i,jk}^{(3)}}{\psi_{\max}^{(3)}} \exp(y_i + y_j + y_k)\right) \\
&\exp(x_4) = \left(\sum_{i,j,k,l=1}^n \frac{\Phi_{i,jkl}^{(4)}}{\psi_{\min}^{(4)}} \exp(y_i + y_j + y_k + y_l) - 1\right)
\end{aligned} \tag{5-sig}$$

Proof

Formulation (5-sig) follows directly from (5) after the following changes of variables:

$$\begin{aligned}
\exp(y_i) &:= w_i \text{ if } w_i \neq 0 \ \forall i \in [n]; \ \exp(x_1) := \left(1 - \frac{f_1(w)}{\psi_{\max}^{(1)}}\right); \ \exp(x_2) := \left(\frac{f_2(w)}{\psi_{\min}^{(2)}} - 1\right) \\
\exp(x_3) &:= \left(1 - \frac{f_3(w)}{\psi_{\max}^{(3)}}\right); \ \exp(x_4) := \left(\frac{f_4(w)}{\psi_{\min}^{(4)}} - 1\right)
\end{aligned}$$

□

Similarly we have a signomial formulation for (hat 5) :

$$\begin{aligned}
\hat{f}_{\min} &= \min \hat{f}(x, y) := \sum_{j=1}^4 \exp(\lambda_j x_j) \\
s. t. \quad &\sum_{j=1}^n \exp(y_j) = 1 \\
&\exp(x_1) = \left(1 - \sum_{i=1}^n \frac{\Phi_i^{(1)}}{\psi_{\max}^{(1)}} \exp(y_i)\right) \\
&\exp(x_2) = \left(\sum_{i,j=1}^n \frac{\Phi_{i,j}^{(2)}}{\psi_{\min}^{(2)}} \exp(y_i + y_j) - 1\right) \\
&\exp(x_3) = \left(1 - \sum_{i=1}^n \frac{\Phi_{i,jk}^{(3)}}{\hat{\psi}^{(3)}} \exp(y_i + y_j + y_k)\right) \\
&\exp(x_4) = \left(\sum_{i,j,k,l=1}^n \frac{\Phi_{i,jkl}^{(4)}}{\hat{\psi}^{(4)}} \exp(y_i + y_j + y_k + y_l) - 1\right) \\
&\sum_{i=1}^n \Phi_{i,jk}^{(3)} \exp(y_i + y_j + y_k) \leq \hat{\psi}^{(3)} \\
&\sum_{i,j,k,l=1}^n \Phi_{i,jkl}^{(4)} \exp(y_i + y_j + y_k + y_l) \geq \hat{\psi}^{(4)}
\end{aligned} \tag{hat 5-sig}$$

We further "reformulate" (HOW) (hat 5-sig) in order to show how Theorem 4.2 from [Signomial optimization SOP](#) in [Preliminaries](#) applies.

$$\begin{aligned}
\widehat{f}_{\min, \approx} &= \min \widehat{f}_{\approx}(x, y) := \sum_{j=1}^4 \exp(\lambda_j x_j) \\
s. t. \quad &\sum_{j=1}^n \exp(y_j) = 1 \\
U_1 &\geq \exp(x_1) - \left(1 - \sum_{i=1}^n \frac{\Phi_i^{(1)}}{\psi_{\max}^{(1)}} \exp(y_i)\right) \geq L \\
U_2 &\geq \exp(x_2) - \left(\sum_{i,j=1}^n \frac{\Phi_{i,j}^{(2)}}{\psi_{\min}^{(2)}} \exp(y_i + y_j) - 1\right) \geq L \\
U_3 &\geq \exp(x_3) - \left(1 - \sum_{i=1}^n \frac{\Phi_{i,jk}^{(3)}}{\widehat{\psi}^{(3)}} \exp(y_i + y_j + y_k)\right) \geq L \quad (\text{mod-hat 5-sig}) \\
U_4 &\geq \exp(x_4) - \left(\sum_{i,j,k,l=1}^n \frac{\Phi_{i,jkl}^{(4)}}{\widehat{\psi}^{(4)}} \exp(y_i + y_j + y_k + y_l) - 1\right) \geq L \\
\widehat{\psi}^{(3)} &\geq \sum_{i=1}^n \Phi_{i,jk}^{(3)} \exp(y_i + y_j + y_k) \geq L \\
? &\geq \sum_{i,j,k,l=1}^n \Phi_{i,jkl}^{(4)} \exp(y_i + y_j + y_k + y_l) \geq \widehat{\psi}^{(4)} \geq L
\end{aligned}$$

Where one can take $L > 0$ (assumption) and $U < 2$ (I have to show this.)

"Proof" (very incomplete)

We have a minimization problem over nonnegative variables that so we can interchanged the equality for a lower bound. Hence, $\exp(x_1) = \left(1 - \sum_{i=1}^n \frac{\Phi_i^{(1)}}{\psi_{\max}^{(1)}} \exp(y_i)\right)$ becomes $\exp(x_1) - \left(1 - \sum_{i=1}^n \frac{\Phi_i^{(1)}}{\psi_{\max}^{(1)}} \exp(y_i)\right) \geq 0$ etc... We can further impose the upper bound of $2 > U$ as ...

Ok, let me level with you. Theoretically, these two programs differ but up to machine precision I think they will give the same result. However, I probably will not be able to falsify this claim as I know of now way to

□

Lemma (lower bounding \widehat{f}_{\min} with SAGE)

$$\widehat{f}_{\min}^{SAGE_{p,q}} \leq \widehat{f}_{\min}$$

where

$$\begin{aligned}
\widehat{f}_{\min}^{SAGE_{p,q}} &:= \sup_{\gamma \in \mathbb{R}, s_h \in \text{SAGE}(E_p(\mathcal{A}))} \gamma \\
s. t. \quad &\widehat{f}(w) - \gamma - \sum_{h(w) \in R_q(C)} s_h(w) h(w) \in \text{SAGE}(E_{p+q}(\mathcal{A}))
\end{aligned}$$

"Proof" (incomplete) This just something to look up □

Lemma (upper bounding $\widetilde{f}_{\min, k, l}$) via sampling

$$\widetilde{f}_{\min, k, l} \leq \widetilde{f}_{k, l}(w^{(r)})$$

where $w^{(r)}$ is the minimizer of $\widetilde{f}_{k, l}$ over $\Delta(n, r)$.

"Proof" Follows from the definition of the minimum. □

Numerical results

- POP

- [PMO](#)

- This is just for parsing/saving/loading POPs, SDPs and MomPs

- [SumOfSquares.jl](#)
- [TSSOS](#)
- [Ipopt](#)
 - I cant the damn thing to run on generic higher order polynomials
- [YALMIP](#) Is just for MATLAB
- [Polyopt.jl](#) Outdated
- [GloptiPoly_3](#) MATLAB and OCTAVE
- [MathProgComplex.jl](#) For complex polynomial optimization.
- **SAGE**
 - CVXPY
 - ECOS,
- **Grid search**
 - This I may have to code my self.

Without cardinality constraints

Skewness

level	#stocks	primal_status	dual_status	obj_val	comp_time(s)
2	5	FEASIBLE_POINT	FEASIBLE_POINT	-1.020754455691659e-6	0.031
2	6	FEASIBLE_POINT	FEASIBLE_POINT	-2.920655586956662e-6	0.047
2	7	FEASIBLE_POINT	FEASIBLE_POINT	-2.915511951729846e-6	0.125
2	8	FEASIBLE_POINT	FEASIBLE_POINT	-2.9075592284715473e-6	0.219
2	9	FEASIBLE_POINT	FEASIBLE_POINT	-2.9286345295789347e-6	0.515
2	10	FEASIBLE_POINT	FEASIBLE_POINT	-2.9259693152874718e-6	1.079
2	11	FEASIBLE_POINT	FEASIBLE_POINT	-2.9228832066662866e-6	2.578
2	12	FEASIBLE_POINT	FEASIBLE_POINT	-5.3433302709466476e-5	4.609
2	13	FEASIBLE_POINT	FEASIBLE_POINT	-5.34352809246628e-5	9.235
2	14	FEASIBLE_POINT	FEASIBLE_POINT	-5.343514597411791e-5	17.531
2	15	FEASIBLE_POINT	FEASIBLE_POINT	-5.343466851953268e-5	33.234
mem.	---	---	---	---	---
3	5	FEASIBLE_POINT	FEASIBLE_POINT	-1.002897385505069e-6	0.782
3	6	FEASIBLE_POINT	FEASIBLE_POINT	-2.9289651669618774e-6	4.593
3	7	FEASIBLE_POINT	FEASIBLE_POINT	-2.849054542182358e-6	24.485
3	8	FEASIBLE_POINT	FEASIBLE_POINT	-2.929254429232411e-6	142.0
mem.	---	---	---	---	---

Kurtosis (currently under revision)

The "UNKNOWN_RESULT_STATUS" seems to stem from the objective value. That is to say, using 1 instead of $f_4(w)$ makes the program feasible again which is weird.

level	#stocks	primal_status	dual_status	obj_val	
2	5	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-5.462410379852476e6	(
2	6	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-6.003107515609454e6	(

level	#stocks	primal_status	dual_status	obj_val	
2	7	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-1.1609449508052422e7	0
2	8	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-883351.2287677532	0
2	9	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-210134.43819556796	0
2	10	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-7.149278898484302e6	2
2	11	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-21849.10743635129	3
2	12	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-30742.43665861959	3
2	13	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-1072.1779065118176	1
2	14	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-728.9859449572551	2
mem.	---	---	---	---	-
2	15	UNKNOWN_RESULT_STATUS	UNKNOWN_RESULT_STATUS	-413.9730106071591	0
3	5	FEASIBLE_POINT	FEASIBLE_POINT	-9.814954929710079e-7	0
3	6	FEASIBLE_POINT	FEASIBLE_POINT	-1.0110424132951022e-6	4
3	7	FEASIBLE_POINT	FEASIBLE_POINT	-0.0001658148457561855	2
3	8	FEASIBLE_POINT	FEASIBLE_POINT	-0.0001658149202996121	1
mem.	---	---	---	---	-

With cardinality constraints

Skewness

Kurtosis

Grid search upper bounds

Signomial optimization.

Extensions and related topics

References

- ☐ <https://github.com/jump-dev/SumOfSquares.jl>
- ☐ 2005,De Klerk, Den Hertog, Elabwabi,on the Complexity of Optimization over The Simplex.pdf
- ☒ **2005,Jurczenko,Maillet,Merlin,Hedge Funds Portfolio Selection with Higher-order Moments A Non-parametric Mean-Variance-Skewness-Kurtosis Efficient Frontier.pdf**
- ☐ 2006,Cornuejols,Optimization Methods in Finance.pdf
- ☒ (quick intro to problem)**2006,Lai,Yu,Wang,Mean-Variance-Skewness-Kurtosis-based Portfolio Optimization.pdf**
- ☒ (does not give a clear forumuation)**2007,Maringer,Parpas,Global optimization of higher order moments in portfolio selection.pdf**
- ☐ 2010,Lasserre,Moments, Positive Polynomials and Their Applications.pdf
- ☐ 2014,de Klerk,Laurent,Sun,An alternative proof of a PTAS for fixed-degree polynomial optimization over the simplex.pdf

- ☐ 2015,Zhao,Polynomial optimization.pdf
- ☐ 2016,de Klerk,Laurent,A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis.pdf
- ☐ 2016,de Klerk,Laurent,Sun,Vera,On the convergence rate of grid search for polynomial optimization over the simplex.pdf
- ☒ (very little details on how the POP is solved and the entropy is just slapped on)2017,Aksarayli,Pala,a Polynomial Goal Programming Model for Portfolio Optimization Based on Entropy and Higher Moments.pdf
- ☒ (other methods) 2017,Naipas,Simar,Vanhems,Portfolio Selection in a Multi-Moment Setting A Simple Monte-Carlo-FDH Algorithm.pdf
- ☒ (Fin. math. app. overview but not math (so skip)) 2019,Andriosopoulos,Doumpos,Pardalos,Zopounidis,Computational approaches and data analytics in financial services A literature review.pdf
- ☒ (Not core) 2020, Pauwels, Putinar, Lasserre, Data analysis from empirical moments and the Christoffel function.pdf
- ☒ (NOT APPLICABLE) 2020,Li,Dong,Qian,Higher-Order Analysis of Probabilistic Long-Term Loss under Nonstationary Hazards.pdf
- ☒ 2020,Zhou,Palomar,Solving High-Order Portfolios via Successive Convex Approximation Algorithms.pdf
- ☒ (other methods) Ang,Majorization Minimization – the Technique of Surrogate.pdf
- ☐ NazarathySSAJuly2020Julia.pdf
- ☐ 2010, Prigent,Mhiri, International Portfolio Optimization with Higher Moments.pdf
- ☐ 2011,Kemalbay,Özkut,Franco Portfolio Selection with Higher Moments A Polynomial Goal Programming Approach to ISE-30 Index.pdf
- ☐ 2011,Peng,Wang,Semidenite Programming Relaxation for Portfolio Selection With Higher Order Moments .pdf
- ☐ 2013,Škrinjarić,Portfolio Selection with Higher Moments and Application on Zagreb Stock Exchange.pdf
- ☐ 2014, Saranya, Prasanna,Portfolio Selection and Optimization with Higher Moments Evidence from the Indian Stock Mark.pdf
- ☐ 2017,Naqvi,Mirza,Naqvi,Rizvi,Portfolio optimisation with higher moments of risk.pdf
- ☐ 2019,Gepp ,Harris, Vanstone,Financial applications of semidefinite programming a review and call for interdisciplinary research.pdf
- ☒ 2019,Li,Zhang,High-Order Portfolio Optimization Problem.pdf
- ☐ Jasour,Moment-Sum-Of-Squares based Semidefinite Programming.pdf

Cut content

Observations:

- (1,2,3,4) is an instance of polynomial optimization over the [simplex].
 - We can attack (3,4) with Polynomial optimization techniques
- (2,3,4) Contain STABLE SET *TODO* show
- (2,3) add non-convexity *TODO* show.
- (1) is just an LP more over the simplex. The solution is the stock with highest expected returns.
- Problem (5) can in fact be reformulated as a signomial optimization problem.

Bounds Upper

We saw in the lower bounds section that we can recover exact solutions in the case where the solution is *flat*. This is however not the norm and as such we need upper bounds to gauge the quality of our approximations. One way to obtain upper bound is to samples the feasible some feasible point and compute their objective values. We then take optimal values among those tested. However, sampling the feasible region could be a costly endeavor and we are not guaranteed to obtain good bounds. Fortunately for the simplex there exists results guaranteeing

((from Zhou thesis))

[Difference Between Grid Bound and Optima on a Simplex](#)

- ☒ Look into "resource allocation in networks" [Geometric Programming for Communication Systems](#)
 - ☐ section 2.2.5. (god that equation reminds me of the first paper)
- ☐ Dynamic data model, talk to <https://homepages.laas.fr/henrion/>
- ☐ [Didier HENRION](#)
- ☐ [SumsOfSquare.jl]