J Optim Theory Appl (2008) 138: 407–427
DOI 10.1007/s10957-008-9402-6

Convexity of Products of Univariate Functions and Convexification Transformations for Geometric Programming

C.E. Gounaris · C.A. Floudas

Published online: 18 April 2008

Abstract We investigate the characteristics that have to be possessed by a functional mapping $f : \mathbb{R} \mapsto \mathbb{R}$ so that it is suitable to be employed in a variable transformation of the type $x \to f(y)$ in the convexification of posynomials. We study first the bilinear product of univariate functions $f_1(y_1)$, $f_2(y_2)$ and, based on convexity analysis, we derive sufficient conditions for these two functions so that $\mathcal{F}_2(y_1, y_2) = f_1(y_1) f_2(y_2)$ is convex for all (y_1, y_2) in some box domain. We then prove that these conditions suffice for the general case of products of univariate functions; that is, they are sufficient conditions for every $f_i(y_i)$, $i = 1, 2, \ldots, n$, so as $\mathcal{F}_n(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} f_i(y_i)$ to be convex. In order to address the transformation of variables that are exponentiated to some power $\kappa \neq 1$, we investigate under which further conditions would the function $(f)^\kappa$ be also suitable. The results provide rigorous reasoning on why transformations that have already appeared in the literature, like the exponential or reciprocal, work properly in convexifying posynomial programs. Furthermore, a useful contribution is in devising other transformation schemes that have the potential to work better with a particular formulation. Finally, the results can be used to infer the convexity of multivariate functions that can be expressed as products of univariate factors, through conditions on these factors on an individual basis.

Keywords Global optimization · Geometric programming · Multilinear products · Posynomial functions · Signomial functions · Convexification transformations

Communicated by P. M. Pardalos.

The authors gratefully acknowledge support from the National Science Foundation.

C.E. Gounaris · C.A. Floudas (✉)
Department of Chemical Engineering, Princeton University, Princeton, NJ 08544-5263, USA
e-mail: floudas@titan.princeton.edu

1 Introduction

Generalized geometric programming (GGP) refers to optimization problems that involve signomial functions, that is, sums of products of independent variables, each of which is exponentiated to some nonzero rational number. Models that involve such functions appear across a wide range of science and engineering. Illustratively, these involve applications in process synthesis [1–5] and process design [6–9], parameter estimation [10, 11], phase and chemical equilibrium [12, 13], molecular conformation [14], and metabolic pathways [15, 16].

The formulation (1) corresponds to a GGP problem. Note that, if a variable is allowed to assume negative values, the corresponding exponents should, of course, be restricted to the set of integers.

$$\min_x \quad s_0(x) = \sum_{t \in T_0} c_{0t} \prod_{i \in I_{0t}} x_i^{\kappa_{0ti}}, \tag{1a}$$

$$\text{s.t.} \quad s_q(x) = \sum_{t \in T_q} c_{qt} \prod_{i \in I_{qt}} x_i^{\kappa_{qti}} \leq 0, \quad \forall q = 1, 2, \ldots, Q, \tag{1b}$$

$$x \in X, \tag{1c}$$

where x is a vector of n variables, X is a polyhedral domain in dimension n that corresponds to bound and other linear constraints, Q is the number of signomial inequality constraints, T_q is the set of the indices of terms ($t = 1, 2, \ldots, |T_q|$) in the objective function ($q = 0$) and problem constraints ($q > 0$), c_{qt} are the nonzero coefficients of each term, κ_{qti} are the appropriate exponents in the various terms, and I_{qt} is the set of variable indices i for which the corresponding exponent is nonzero, that is, when the variable is indeed a factor in the term. Signomial equality constraints are treated as two signomial inequalities.

Note that the above formulation corresponds to a very broad class of problems in mathematical programming. For example, multilinear problems constitute the special case where every exponent is equal to one, that is, $\kappa_{qti} = 1, \ \forall (q, t, i)$. A more general class, polynomial problems, are also included in formulation (1). They correspond to the case where every exponent is a positive integer number, that is, $\kappa_{qti} \in \mathbb{N}^*, \ \forall (q, t, i)$.

A signomial function, $s(x)$, with positive coefficients ($c_t > 0, \ \forall t$) is called a posynomial $p(x)$. In the field of generalized geometric programming, it is typical to separate the terms into two sets, T^+ and T^-, according to the sign of their coefficient, thus consider every signomial function as the difference between two posynomials. The following formulation is equivalent to the formulation (1):

$$\min_x \quad p_0^+(x) - p_0^-(x) = \sum_{t \in T_0^+} c_{0t} \prod_{i \in I_{0t}} x_i^{\kappa_{0ti}} - \sum_{t \in T_0^-} (-c_{0t}) \prod_{i \in I_{0t}} x_i^{\kappa_{0ti}}, \tag{2a}$$

$$\text{s.t.} \quad p_q^+(x) - p_q^-(x) = \sum_{t \in T_q^+} c_{qt} \prod_{i \in I_{qt}} x_i^{\kappa_{qti}} - \sum_{t \in T_q^-} (-c_{qt}) \prod_{i \in I_{qt}} x_i^{\kappa_{qti}} \leq 0,$$
$$\forall q = 1, 2, \ldots, Q, \tag{2b}$$

$$x \in X. \tag{2c}$$

Zener [17, 18] was the first to notice the frequent occurence of signomial terms in various engineering models, including many empirical fits that involve power laws. He derived initially the global solution for an unconstrained problem ($Q = 0$) with a posynomial objective function ($|T_0^-| = 0$) in the special case where $|T_0^+| = n + 1$, that is, when the number of monomial terms is one greater than the number of problem variables. The intimate connection of his solution with geometric means led to the term "geometric programming", which has been used widely ever since. Unconstrained geometric programs, involving only posynomial functions but with no restrictions on the number of terms $|T_0^+|$, were later studied by Duffin and Peterson [19], based on duality theory. Duffin [20] addressed constrained posynomial problems based on a technique called polynomial condensation, that involves the classical inequality between the arithmetic and geometric means. Peterson [21] offered a review of these results.

Passy and Wilde [22], Blau and Wilde [23] and Avriel and Williams [24] addressed general signomial problems, that is, problems that include monomials with both positive and negative coefficients. The name "generalized geometric programming" was put forward for this class of problems. Duffin and Peterson [25] used harmonic means to address reverse geometric problems; that is, problems that involve functions that are either posynomial or negative posynomial. They also extended the applicability of their results by showing how a general signomial problem can be transformed into a reverse geometric one [26]. An extensive review of all these local approaches can be found in Sarma *et al.* [27].

Given the multiplicity of solutions of the generally nonconvex formulations (1) and (2), a global optimization approach was necessary and Falk [28] was the first to introduce one such. It was based on the exponential variable transformation, its convex relaxation, and branch and bounding on the space of exponents of the negative monomials. Maranas and Floudas [29, 30] proposed a deterministic global optimization method that was also based on the exponential transformation, but applied partitioning in the typically smaller space of the problem variables. Extensive reviews appear also in the textbooks by Floudas [31] and Floudas and Pardalos [32–34].

The exponential transformation $x \to e^y$ has also been utilized in other approaches [35, 36]. With the use of such a transformation, the positive coefficient monomials are convexified. The negative monomials do not in general and therefore have to be transformed differently or relaxed. The relaxation procedure is the main difference of the above methods, with a linear relaxation usually being proposed.

Since the exponential function assumes only strictly positive variables, the above methods postulate that this is the case. In their series of papers, Li and Tsai [37], Tsai and Lin [38], and Tsai *et al.* [39] presented reformulations that convert a signomial problem involving free variables to a problem that involves strictly positive ones. This occurs at the expense of introducing additional continuous and discrete variables. They presented also convexity rules that can be used to identify trivariate monomials that are already convex and therefore do not require some variable transformation in order to be convexified.

More recently, Li *et al.* [40] proposed the reciprocal transformation $x \to 1/y$ for the convexification of the positive coefficient monomials. The y-defining equalities involve xy bilinear terms and therefore can be relaxed with the use of their convex and concave envelopes. It is apparent that different convexification techniques, based on different transformations, are possible for a given problem. As Björk *et al.* [41] pointed out, the choice of transformation functions is clearly essential; they showed that it has an influence on the efficiency of the optimization approach used in each case. Westerlund [42] reviewed their work, which is based on power transformations.

The current paper deals with the identification of functional forms that would be suitable for monomial-convexifying variable transformations and therefore could be employed in a convexification/relaxation scheme for signomial programs. Section 2 presents a few important remarks for such functions, while Sect. 3 presents the convexity analysis for a transformed bilinear term. This analysis leads to sufficient conditions for a suitable function, which are summarized in Sect. 4. Section 5 generalizes the conditions for general products of univariate functions, thus addressing general multilinear terms; Sect. 6 further generalizes for generic monomials, where the exponents are arbitrary real numbers.

2 Remarks on Variable Transformations $x \to f(y)$

Let variable x participate in a signomial formulation and let $D_x = [x^L, x^U] \subset \mathbb{R}$ be its domain of interest. Let, also, variable $y \in D_y = [y^L, y^U] \subset \mathbb{R}$ and a mapping $f{:}D_y \mapsto R_f(D_y) \subset \mathbb{R}$, where $R_f(D_y)$ is the range of function f for the given domain. In an effort to eliminate the nonconvexities from the formulation, or just alter them in a favorable way, one typically introduces the variable y in the formulation by performing a transformation $x \to f(y)$. If this transformation is performed on every occurence of the original variable x, then this variable is completely replaced and the total number of variables remains unchanged. The desired value of the variable x at optimality, x^{opt}, can then be obtained from the optimal solution y^{opt} of the variable y after the optimization procedure has occured and through the relation $x^{\text{opt}} = f(y^{\text{opt}})$. If the replacement is partial, then both variables x and y participate in the transformed formulation, resulting in an increased number of variables. Furthermore in this case, the transformed formulation has to be augmented with the equality constraint $x = f(y)$ so that it is equivalent to the original one.

A first question that arises is whether the function $f(y)$ should correspond to an one-to-one mapping. As long as the variable y can be transformed uniquely back to a value for the variable x, something that is guaranteed by a valid functional mapping $f : y \mapsto f(y)$, there should be no restriction. However, the possibility for $f(y^\alpha) = f(y^\beta)$, with $y^\alpha \neq y^\beta$, allows for a multiplicity of solutions of the transformed problem. This can be viewed as an unnecessary degree of freedom; therefore, we can restrict the domain D_y so as to select only some strictly monotone portion of the function $f(y)$. We have to keep in mind that, by restricting the domain of the variable y as much as possible, we increase our chances of finding a transformation that would satisfy all the required sufficient conditions that have to be satisfied for every point $y^* \in D_y$ and that would guarantee desirable properties, such as the convexity property discussed in the later sections. Therefore we conclude that strict monotonicity is not a necessity, but it is however a desirable property.

A second question that is posed is what would be an appropriate range for function $f(y)$, $y \in D_y$. Clearly, $R_f(D_y) \supseteq D_x$ so as to allow the transformed problem to capture all the variation of the original. Furthermore, if for some y^* we have $f(y^*) \notin D_x$, then such an instance of the transformed problem corresponds to an infeasible instance of the original problem, which should not have been considered; that is, the transformed problem should also be infeasible. Therefore, we conclude that $R_f(D_y) = D_x = [x^L, x^U]$.

A final remark is on the selection of an appropriate domain D_y for the introduced variable y. It should be clear that, if strictly monotone functions are selected, the domain of variable y is uniquely defined as the range of the inverse function $f^{-1}(x)$, given the domain of variable x, that is, $[y^L, y^U] = D_y = R_{f^{-1}}(D_x)$. Such a selection not only spans over the full desired variation, but also results into a domain D_y that is as small as possible, leading to tighter relaxations (if necessary) of the transformed problem.

3 Convexity Analysis for the Bilinear Term

Let us consider the transformations $x_1 \to f_1(y_1)$ and $x_2 \to f_2(y_2)$. The bilinear term, x_1x_2, would then be replaced by $\mathcal{F}_2(y_1, y_2) = f_1(y_1)f_2(y_2)$, where $(y_1, y_2) \in [y_1^L, y_1^U] \times [y_2^L, y_2^U]$. This box domain corresponds to the required ranges of f_1^{-1} and f_2^{-1}, as described in the previous section.

The Hessian matrix of $\mathcal{F}_2$ is

$$\nabla^2\mathcal{F}_2(y_1, y_2) = H_2(y_1, y_2) = \begin{bmatrix} f_1''(y_1)f_2(y_1) & f_1'(y_1)f_2'(y_2) \\ f_1'(y_1)f_2'(y_1) & f_1(y_1)f_2''(y_2) \end{bmatrix},$$

where f_i' denotes the first and f_i'' denotes the second derivative of the univariate function f_i with respect to the corresponding variable y_i.

The two eigenvalues of this matrix are

$$\lambda_{1,2}(y_1, y_2) = \frac{1}{2}\left\{(f_1f_2'' + f_2f_1'') \pm \sqrt{(f_1f_2'' - f_2f_1'')^2 + 4(f_1')^2(f_2')^2}\right\}.$$

For the function $\mathcal{F}_2(y_1, y_2)$ to be convex, $H_2(y_1, y_2)$ needs to be positive semidefinite at every point (y_1, y_2), that is,

$$\lambda_{1,2}(y_1, y_2) \geq 0, \quad \forall (y_1, y_2) \in [y_1^L, y_1^U] \times [y_2^L, y_2^U].$$

This condition results equivalently into the following two conditions, both of which need to be satisfied $\forall (y_1, y_2)$:

$$f_1f_2'' + f_2f_1'' \geq 0, \tag{3}$$

$$f_1f_1''f_2f_2'' \geq (f_1')^2(f_2')^2. \tag{4}$$

These two inequalities constitute the necessary and sufficient conditions for the convexity of the transformed bilinear term. In the analysis that follows, we will investigate which functional forms are suitable to be selected for the individual functions $f_1(y_1)$ and $f_2(y_2)$.

Let us start by first observing that the individual functions cannot contain linear parts. The following lemma applies.

Lemma 3.1 *Neither $f_1(y_1)$ nor $f_2(y_2)$ can be a linear function at any given interval of interest.*

Proof We start the proof by noting that neither f_1 nor f_2 can be a constant function, since they should correspond to one-to-one mappings.

Let $f_1(y_1)=\alpha y_1+\beta$ at some interval $[y_1^l, y_1^u]\subseteq[y_1^L, y_1^U]$, where $y_1^l\neq y_1^u$. Therefore, we also have $f_1'(y_1)=\alpha$ and $f_1''(y_1)=0$ for every $y_1\in(y_1^l, y_1^u)$. For the condition (4) to hold at some $y_1^*\in(y_1^l, y_1^u)$, we need

$$0\geq\alpha^2(f_2'(y_2))^2,\quad \forall y_2\in[y_2^L, y_2^U].$$

However, since f_2 is not a constant function, there must be some $y_2^*\in[y_2^L, y_2^U]$ for which $f_2'(y_2^*)\neq 0$. Therefore, for the condition (4) to hold at the point (y_1^*, y_2^*), there must be $\alpha=0$, which means that f_1 needs to be constant at $[y_1^l, y_1^u]$. This is restricted; therefore, $f_1(y_1)$ cannot be linear at any given subinterval of $[y_1^L, y_1^U]$.

Similarly, $f_2(y_2)$ cannot be linear at any given interval $[y_2^l, y_2^u]\subseteq[y_2^L, y_2^U]$. □

An immediate corollary of this lemma is that a function of the form $y_1 f(y_2)$ (with f not being a constant), such as $y_1e^{y_2}$ or even the bilinear term itself (y_1y_2), cannot be convex at any given domain $D\subseteq\mathbb{R}^2$.

Lemma 3.2 *At every point (y_1, y_2), we need have: $f_1(y_1)f_2''(y_2)\geq 0$ and $f_2(y_2)\times f_1''(y_1)\geq 0$.*

Proof Suppose that there is some point $(y_1^*, y_2^*)\in[y_1^L, y_1^U]\times[y_2^L, y_2^U]$ for which $f_1(y_1^*)f_2''(y_2^*)<0$. Then, for the condition (3) to hold at this point, we need $f_2(y_2^*)f_1''(y_1^*)>0$. This leads to the conclusion that $f_1(y_1^*)f_1''(y_1^*)f_2(y_2^*)f_2''(y_2^*)<0$, which violates condition (4). Therefore

$$f_1(y_1)f_2''(y_2)\geq 0,\quad \forall(y_1, y_2)\in[y_1^L, y_1^U]\times[y_2^L, y_2^U].$$

Similarly, $f_2(y_2)f_1''(y_1)\geq 0$ at every point (y_1, y_2). □

The next lemma imposes restrictions on the sign of functions f_1 and f_2.

Lemma 3.3 *Each of the functions $f_1(y_1)$ and $f_2(y_2)$ need be either nonnegative or nonpositive.*

Proof Let us prove it for f_1. According to Lemma 3.1, f_2 is not linear. Therefore, there exists some $y_2^*\in[y_2^L, y_2^U]$ for which $f_2''(y_2^*)\neq 0$. From Lemma 3.2, we have

$$f_1(y_1)f_2''(y_2^*)\geq 0\ \forall y_1\Rightarrow\begin{cases} f_1(y_1)\geq 0\ \forall y_1, & \text{if } f_2''(y_2^*)>0,\\ \text{or} & \\ f_1(y_1)\leq 0\ \forall y_1, & \text{if } f_2''(y_2^*)<0.\end{cases}$$

Similarly, $f_2(y_2)\geq 0,\ \forall y_2$ or $f_2(y_2)\leq 0,\ \forall y_2$. □

Since f_i cannot change sign, $D_x = R_f(D_y)$ cannot span over both negative and positive values, that is: $x_i^L \geq 0$ or $x_i^U \leq 0$. The immediate corollary is that we cannot transform directly a variable $x : 0 \in (x^L, x^U)$. Some shifting transformation of the form $x \to \tilde{x} + c$, $c \leq x^L$, would be necessary. Alternatively and at the expense of incorporating additional variables in the formulation, the problematic variable can be replaced by the sum of a nonnegative and a nonpositive variable, or the difference of two nonnegative variables.

From now on, we will consider only the case where $x_i^L \geq 0, \forall i$. This poses no loss of generality, since if a nonpositive variable is present in the formulation, a transformation of the form $x \to -\tilde{x}$ could always be employed.

Under this restriction, that f_1 and f_2 are nonnegative, Lemma 3.2 yields that $f_1''(y_1) \geq 0$, $\forall y_1$, and $f_2''(y_2) \geq 0$, $\forall y_2$. Therefore, f_1 and f_2 need to be convex, as well. Furthermore, since they cannot contain linear parts (according to Lemma 3.1), their convexity needs to be strict.

We will now investigate the potential for the function f to actually attain the value of zero. If this is to happen, there can only be a single point $y^* \in D_y$ such that $f(y^*) = 0$, because of the strict monotonicity requirement. Furthermore, since f has to be nonnegative, this has to be one of the domain edges. In particular, $y^* = y^L$, if f is an increasing function, and $y^* = y^U$, if f is a decreasing function. The following lemma applies.

Lemma 3.4 *$f_1(y_1)$ and $f_2(y_2)$ cannot both have a root.*

Proof Consider the function f_i, where $i = 1$ or 2, and suppose that it attains the value zero at $y_i = y_i^*$. This point ought to be unique and also ought to be one of the two domain edges. Let us also consider the point $y_i = y_i^{**}$ to be the other domain edge, that is, the domain edge that is not a root of function f_i. If the function $\mathcal{F}_2(y_1, y_2)$ were to be convex, Jensen's inequality would have to hold for any pair of points in $D_{y_1} \times D_{y_2}$. In particular, let us select the points (y_1^*, y_2^{**}) and (y_1^{**}, y_2^*). According to the Jensen inequality, we need to have

$$\mathcal{F}_2\left(\frac{y_1^* + y_1^{**}}{2}, \frac{y_2^{**} + y_2^*}{2}\right) \leq \frac{\mathcal{F}_2(y_1^*, y_2^{**}) + \mathcal{F}_2(y_1^{**}, y_2^*)}{2}$$

$$\Rightarrow f_1\left(\frac{y_1^* + y_1^{**}}{2}\right) f_2\left(\frac{y_2^{**} + y_2^*}{2}\right) \leq \frac{f_1(y_1^*) f_2(y_2^{**}) + f_1(y_1^{**}) f_2(y_2^*)}{2}$$

$$\overset{\substack{f_1(y_1^*)=0 \\ f_2(y_2^*)=0}}{\Longrightarrow} f_1\left(\frac{y_1^* + y_1^{**}}{2}\right) f_2\left(\frac{y_2^{**} + y_2^*}{2}\right) \leq 0.$$

Clearly, the above cannot hold, since the left-hand side is a product of two strictly positive quantities. This results from the strict monotonicity and nonnegativity of both f_1,f_2, and the fact that $y_1^* \neq y_1^{**}$, $y_2^* \neq y_2^{**}$ (since the intervals need to have some nonzero length). We conclude that the Jensen inequality does not hold for the selected pair of points and, therefore, function $\mathcal{F}_2(y_1, y_2)$ cannot be convex when both f_1 and f_2 have a root. □

Since we aim at deriving conditions for each function that would be independent of the other one in the product, so that the results can be generalized to a higher number of factors, the above lemma imposes the restriction that f_1 and f_2 have to be strictly positive. The resulting corollary is that a variable x with variation $D_x = [0, x^U]$ (or equivalently $[x^L, 0]$) cannot be addressed. Again, some shifting transformation or a more rigorous treatment that involves additional binary variables might be applied. For practical purposes, one usually imposes an artificial ϵ-small lower bound, instead of a strict zero, and is willing to tolerate some loss of accuracy in this region.

Let us now refocus our attention on conditions (3) and (4). Because of the strict positivity and convexity requirements, (3) will always hold. As far as condition (4) is concerned, we rearrange it as follows:

$$\left\{f_1 f_1'' - (f_1')^2\right\}(f_2')^2 + \left\{f_1 f_1'' - (f_1')^2\right\}\left\{f_2 f_2'' - (f_2')^2\right\} + (f_1')^2\left\{f_2 f_2'' - (f_2')^2\right\} \geq 0. \tag{5}$$

The equivalence of the above inequality with condition (4) can be trivially shown by performing the multiplications and canceling out the excess terms.

It is clear that condition (5), and thus condition (4), will be satisfied (for every point) if both $f_1 f_1'' - (f_1')^2$ and $f_2 f_2'' - (f_2')^2$ are nonnegative throughout their respective domains $[y_1^L, y_1^U]$ and $[y_2^L, y_2^U]$. Of course, $f_i f_i'' - (f_i')^2 \geq 0$, $i = 1, 2$ is a conservative consideration and only constitutes a sufficient condition. There could still exist some combination of functions f_1, f_2 for which it does not hold, while condition (5) does. However, for generalization purposes, it is practical to impose independent conditions on the individual functions and not conditions that are based on their combined effect within a product.

Remark 3.1 In the special case where the two original variables, x_1 and x_2, vary over the same domain, $D_{x_1} = D_{x_2}$, and we restrict ourselves to using the same functional form for both transformations, the condition $f_i f_i'' - (f_i')^2 \geq 0$ is not only sufficient, but also necessary.

Before we conclude this section, let us present the following lemma and its important corollary.

Lemma 3.5 $\mathcal{F}_2(y_1, y_2) = f_1(y_1) f_2(y_2)$ *cannot be a linear function.*

Proof For $\mathcal{F}_2$ to be linear, both eigenvalues of H_2 need to be zero at every point $(y_1, y_2) \in [y_1^L, y_1^U] \times [y_2^L, y_2^U]$. For this special case, condition (3) becomes $f_1 f_2'' + f_2 f_1'' = 0,\ \ \forall (y_1, y_2)$, which in turn, after application of Lemma 3.2, becomes $f_1 f_2'' = f_2 f_1'' = 0,\ \ \forall (y_1, y_2)$.

This means that we need have $f_1(y_1) = 0$, $\forall y_1$ or $f_2''(y_2) = 0$, $\forall y_2$, as well as $f_2(y_2) = 0$, $\forall y_2$ or $f_1''(y_1) = 0$, $\forall y_1$. However, none of these conditions can be met, since f_1 and f_2 are not linear functions, according to Lemma 3.1. $\square$

Lemma 3.5 has an important corollary, which is presented as the following remark:

Remark 3.2 We cannot perform a set of variable transformations $x_i \to f_i(y_i)$, $i = 1, 2$, that will convexify simultaneously both a bilinear term $x_1 x_2$ and a negative bilinear term $-x_1 x_2$. Therefore, one cannot hope for a complete convexification of a general bilinear formulation, where at least one of the nonlinear variables participates in both the sets $T^+ = \bigcup_{q=0}^{Q} T_q^+$ and $T^- = \bigcup_{q=0}^{Q} T_q^-$; see the formulation (2).

Note that the above remark assumes that the formulation is irreducible; that is, the number of variables cannot be reduced by algebraic manipulations and reordering/restructuring of the terms.

4 Sufficient Conditions and Suitable Functions

Summarizing the analysis of Sect. 3, the following properties (when valid for the complete domain D_y) constitute sufficient conditions for a function $f(y)$ to be suitable to replace a strictly positive variable x during a bilinear term convexification transformation. Equivalently, such a function would be suitable to constitute one of the two factors in a convex bivariate product of two univariate functions. It is assumed that the function is properly defined over the required domain $D_y = R_{f^{-1}}(D_x)$.

Property (i): Strict positivity.
Property (ii): Strict monotonicity.
Property (iii): $f(y)f''(y) - (f'(y))^2 \geq 0$.

We introduce the following definition.

Definition 4.1 A continuous univariate function $f(y)$, $y \in D_y$, is said to be a "convexification transformation suitable function" (CTSF) in D_y if properties (i)–(iii) are satisfied $\forall y \in D_y$.

Note that the strict convexity property is not required explicitly, since it is implied through the combination of the other properties.[1] Therefore, a CTSF is always strictly convex.

Properties (i)–(iii) are in agreement with the transformation functions that have been proposed in the literature. In particular, both the exponential function [27–29], $f(y) = e^y$, $y \in \mathbb{R}$, and the reciprocal function [40], $f(y) = 1/y$, $y \in \mathbb{R}_+^*$, are functions that possess these properties and thus are suitable for the transformations under consideration. This translates to the fact that bivariate functions such as $e^{(y_1+y_2)}$, $1/(y_1 y_2)$, or even e^{y_1}/y_2, are indeed convex in $\mathbb{R}^2$,${\mathbb{R}_+^*}^2$, $\mathbb{R} \times \mathbb{R}_+^*$, respectively. Of course, proving the convexity of $\mathcal{F}_2$, given the functions f_1 and f_2, is typically not a difficult task, since the dimensionality of the Hessian matrix is small. The practical importance of the derived conditions will be more apparent when we address general multilinear functions. This will be the topic of Sect. 5.

In the case of the exponential and reciprocal functions, the sufficient conditions are satisfied for an infinite domain $\mathbb{R}$ and $\mathbb{R}_+^*$, respectively. In both cases, this results in

[1] (iii) $\Rightarrow ff'' \geq (f')^2 \geq 0 \overset{\text{(i)}}{\Rightarrow} f'' \geq 0$, $\forall y$, and also $f''(y^*) = 0 \overset{\text{(iii)}}{\Rightarrow} f'(y^*) = 0 \overset{\text{(ii)}}{\Rightarrow}$ isolated points y^*.

a range of $(0, +\infty)$; therefore, these functions can span over the full variation of any strictly positive variable. It should be noted, however, that one could use the above conditions to infer convexity of the bilinear product only within a bounded interval. The relevant functions would still be suitable, as long as their resulting range spans over the desired variation D_x. Let us present the following examples.

Example 4.1 Consider the function $f(y) = 1 - \log(1 - y)$ defined in the half-open interval $D_y = [0, 1)$.

The function $f(y)$ is CTSF in D_y, since it satisfies conditions (i)–(iii) as follows:

(i) $f(y) = 1 - \log(1 - y) \Rightarrow \inf_{y\in[0,1)} f(y) = f(0) = 1 > 0$,
(ii) $f'(y) = \frac{1}{1-y} \Rightarrow \inf_{y\in[0,1)} f'(y) = f'(0) = 1 > 0$,
(iii) $f(y)f''(y) - (f'(y))^2 = -\log(1 - y)(\frac{1}{1-y})^2 \overset{y\in[0,1)}{\geq} 0, \ \forall y$.

This example provides a range of $R_f(D_y) = [1, +\infty)$; thus, it is appropriate to replace the variables that attain values from within this range.

Example 4.2 Consider the function $f(y) = y^2 + 1$ defined in the closed interval $D_y = [0, 1]$.

The function $f(y)$ is CTSF in D_y, since it satisfies conditions (i)–(iii) as follows:

(i) $f(y) = y^2 + 1 \Rightarrow \inf_{y\in[0,1]} f(y) = f(0) = 1 > 0$,
(ii) $f'(y) = 2y \Rightarrow \inf_{y\in[0,1]} f'(y) = f'(0) = 0 \geq 0$.
Furthermore, $y = 0$ is a unique stationary point. Thus, the monotonicity is strict.
(iii) $f(y)f''(y) - (f'(y))^2 = 2 - 2y^2 \overset{y\in[0,1]}{\geq} 0, \ \forall y$.

This example provides a closed range of $R_f(D_y) = [1, 2]$.

Combining the above two examples, we can infer the convexity of the function

$$\mathcal{F}(y_1, y_2) = (1 - \log(1 - y_1))(y_2^2 + 1)$$

at any domain D such that $D \subset [0, 1) \times [0, 1]$.

When many choices are available, that is, when there are many CTSFs that provide a range that covers the desired variation D_x, the selection of the functional form to be used is important because it is related directly to the extent of underestimation that we will have to tolerate when the formulation is relaxed.

The following lemma can be used to adjust a CTSF (i.e., to shift and scale it), so as to meet the required range D_x of the variable x.

Lemma 4.1 *If $f(y)$ is CTSF in some interval D_y, then $h(y) = \alpha f(y) + \beta$, where $\alpha \neq 0$, is also CTSF, as long as $\alpha > 0$ and $\beta \geq \beta^L$, where β^L is a sufficiently large nonpositive constant* (*see proof for exact value*).

Proof We start the proof by noting that $h''(y) = \alpha f''(y)$. As has been noted before, the properties (i)–(iii) imply strict convexity; therefore, for h to be CTSF, it also needs to be strictly convex. For this, we need $h''(y) \geq 0,\ \forall y \in D_y$. Given that $f(y)$ is CTSF and thus is strictly convex in D_y, we need only $\alpha \geq 0 \overset{\alpha \neq 0}{\Rightarrow} \alpha > 0$.

This condition suffices for the strict monotonicity property to hold, since $h'(y) = \alpha f'(y)$ and the function f is itself strictly monotone. In fact, since $\alpha > 0$, the function h will exhibit the same type of monotonicity as the function f.

For property (iii) to hold, $\forall y \in D_y$ we need

$$\begin{aligned} hh'' - (h')^2 \geq 0 &\Leftrightarrow (\alpha f + \beta)(\alpha f'') - (\alpha f')^2 \geq 0 \\ &\overset{\alpha > 0}{\Leftrightarrow} \left(f + \frac{\beta}{\alpha}\right) f'' - (f')^2 \geq 0 \\ &\Leftrightarrow \frac{\beta}{\alpha} f'' \geq -\left\{ f f'' - (f')^2 \right\}. \end{aligned}$$

For those $y^* \in D_y$ for which $f''(y^*) = 0$, the above holds without any further restrictions, due to property (iii) of CTSF f. For the above to hold for the rest of the points in D_y, that is for points for which $f''(y) > 0$, we need to impose

$$\frac{\beta}{\alpha} \geq \sup_{\substack{y \in D_y \\ y: f''(y) > 0}} \frac{-\{f(y) f''(y) - (f'(y))^2\}}{f''(y)} = \frac{\beta^L}{\alpha}.$$

Since $f(y)$ cannot be linear, due to the strict convexity, there will always be points where the second derivative is strictly positive. Therefore, the bound β^L always constitutes a necessary restriction for the shifting parameter β.

For strict positivity to hold, $\forall y \in D_y$ we need

$$\begin{aligned} h(y) > 0 &\Leftrightarrow \alpha f(y) + \beta > 0 \\ &\overset{\alpha > 0}{\Leftrightarrow} \frac{\beta}{\alpha} > -f(y). \end{aligned}$$

This would hold, if we impose

$$\frac{\beta}{\alpha} > \sup_{y \in D_y} \{-f(y)\}.$$

The following relation reveals that the bound β^L, derived previously for property (iii), suffices for property (i) to hold as well:

$$\begin{aligned} \frac{\beta^L}{\alpha} &= \sup_{\substack{y \in D_y \\ y: f''(y) > 0}} \left\{ -f(y) + \frac{(f'(y))^2}{f''(y)} \right\} \\ &> \sup_{\substack{y \in D_y \\ y: f''(y) > 0}} \{-f(y)\} \overset{\text{note}^2}{=} \sup_{y \in D_y} \{-f(y)\}. \end{aligned}$$

Therefore, under the restrictions that $\alpha > 0$ and $\beta \geq \beta^L$, h would also be CTSF. □

[2] The two suprema are equal because the two domains differ only by the single inflection point, if one such exists, and function f is continuous.

Note that, because $\beta^L \leq 0$, the addition of a positive constant to a CTSF does not affect its suitability. Let us consider the following example:

Example 4.3 Show that $h(y) = 2y^2 + 3$ is CTSF in $D_y = [0, 1]$.

According to Example 4.2, the function $f(y) = y^2 + 1$ is CTSF in [0, 1]. Therefore, by application of Lemma 4.1, the function $h(y) = 2y^2 + 3 = 2f(y) + 1$ will also be CTSF, if $\alpha > 0$ and $\beta \geq \beta^L$. In this particular case, we have $\alpha = 2$, $\beta = 1$ and $\frac{\beta^L}{\alpha} = \max_{y\in[0,1]} \frac{-\{(y^2+1)2-(2y)^2\}}{2} = \max_{y\in[0,1]} y^2 - 1 = 0$. Therefore, both restrictions are satisfied and the function $h(y) = 2y_2^2 + 3$ is CTSF in [0, 1].

The following two lemmas consider sums and products of CTSFs and specify conditions for them to be CTSF as well. Note that these lemmas generalize trivially to a higher number of CTSFs as addends of the sum and factors of the product, respectively.

Lemma 4.2 *Let $f(y)$ and $g(y)$ both be CTSF in some interval D_y. Their sum $h(y) = f(y) + g(y)$ is also CTSF in D_y, as long as it is strictly monotone.*

Proof The strict positivity of the sum trivially derives from the strict positivity of its two CTSF addends, namely the functions f and g.

Since $h' = f' + g'$, with f and g satisfying property (ii), it is clear that the function h will be strictly monotone if f and g are both increasing or both decreasing. Note that, in the case when this is not true and functions f, g have slopes of opposite signs, there is still the possibility for strict monotonicity of the sum, depending on the actual functions and domain under consideration.

From property (iii) of the functions f and g, we have

$$\left.\begin{array}{r} ff'' \geq (f')^2 \geq 0 \\ gg'' \geq (g')^2 \geq 0 \end{array}\right\} \Rightarrow (f''g)(fg'') \geq (f'g')^2 \Rightarrow \sqrt{(f''g)(fg'')} \geq |f'g'|$$

$$\Rightarrow \sqrt{f''gfg''} \geq f'g' \Rightarrow -2f'g' \geq -2\sqrt{f''g}\sqrt{fg''}.$$

Regarding property (iii), and according to the above relation, we have

$$\begin{aligned} hh'' - (h')^2 &= (f+g)(f''+g'') - (f'+g')^2 \\ &= \left\{ff'' - (f')^2\right\} + \left\{gg'' - (g')^2\right\} + \left(f''g - 2f'g' + fg''\right) \\ &\geq \left\{ff'' - (f')^2\right\} + \left\{gg'' - (g')^2\right\} + \left(f''g - 2\sqrt{f''g}\sqrt{fg''} + fg''\right) \\ &= \left\{ff'' - (f')^2\right\} + \left\{gg'' - (g')^2\right\} + \left(\sqrt{f''g} - \sqrt{fg''}\right)^2, \end{aligned} \tag{6}$$

which is clearly a nonnegative quantity because functions f and g satisfy condition (iii) on an individual basis. □

Lemma 4.3 *Let $f(y)$ and $g(y)$ both be CTSF in some interval D_y. Their product $h(y) = f(y)g(y)$ is also CTSF in D_y, as long as it is strictly monotone.*

Proof The strict positivity property of the product derives trivially from the strict positivity of its two CTSF factors, namely the functions f and g.

Since $h' = f'g + fg'$, with f and g satisfying properties (i) and (ii), it is clear that the function h will be strictly monotone if f and g are both increasing or both decreasing. Note that, in the case when this is not true and functions f, g have slopes of opposite signs, there is still the possibility for strict monotonicity of the product, depending on the actual functions and domain under consideration.

Regarding property (iii), we have

$$\begin{aligned} hh'' - (h')^2 &= (fg)(f''g + 2f'g' + fg'') - (f'g + fg')^2 \\ &= g^2\{ff'' - (f')^2\} + f^2\{gg'' - (g')^2\}, \end{aligned} \tag{7}$$

which is clearly a nonnegative quantity because the functions f and g satisfy condition (iii) on an individual basis. □

An immediate corollary of Lemma 4.3 is that every CTSF can be used for the transformation of a variable that is exponentiated to a positive integer power. This results from the fact that, if we apply the transformation $x \to f(y)$ in a term x^n, where $n \in \mathbb{N}^*$, then the resulting transformed term, $(f(y))^n$, will be CTSF by sequential application of the above lemma $n - 1$ times. Section 6 will address the issue of exponentiated variables into greater depth.

Lemmas 4.1–4.3 provide us with tools to infer the suitability of a function through investigating the suitability of its various building components. Therefore, even when the treatment of the function as a whole cannot lead to the conclusion that the sufficient conditions (i)–(iii) are met, the suitability of the function can still be inferred if the preceding lemmas are applicable.

5 Convexity of $\mathcal{F}_n(y_1, y_2, \ldots, y_n)$

We will show that the conditions derived previously, through the analysis of the bilinear product, are indeed sufficient conditions for the convexity of general products of univariate functions. The following theorem holds:

Theorem 5.1 *If the univariate functions $f_i(y_i)$ are strictly positive and satisfy property* (iii), *for every $i = 1, 2, \ldots, n$, then the n-variate function $\mathcal{F}_n(y_1, y_2, \ldots, y_n) = \prod_{i=1}^n f_i(y_i)$ is convex.*

Proof Firstly note that, in addition, every function $f_i(y_i)$ would be also convex, since

$$\text{(iii)} \Rightarrow f_i f_i'' \geq (f_i')^2 \geq 0 \overset{f_i > 0}{\Rightarrow} f_i'' \geq 0.$$

We will prove the theorem by induction on the number of factors n. For $n = 1$, the result holds trivially, since $\mathcal{F}_1(y_1) = f_1(y_1)$, which is convex. In fact, the result also holds for $n = 2$, since it was the convexity analysis of the function $\mathcal{F}_2(y_1, y_2)$ that led to the derivation of properties (i), (iii).

As for the induction hypothesis, let us assume that the result holds for some $n = k$, that is,

$$\mathcal{F}_k(y_1, y_2, \ldots, y_k) = \prod_{i=1}^{k} f_i(y_i)$$

is convex. Let us also define H_k to be the Hessian matrix of $\mathcal{F}_k$. Obviously, H_k is positive semidefinite and we denote this by $H_k \succeq 0$.

We will show that, under the above hypothesis, the result holds for $n = k + 1$ as well, that is, $\mathcal{F}_{k+1}(y_1, y_2, \ldots, y_k, y_{k+1}) = \mathcal{F}_k(y_1, y_2, \ldots, y_k) f_{k+1}(y_{k+1})$ is convex or equivalently $H_{k+1} \succeq 0$.

In the analysis that follows, we will simplify the notation by dropping the parentheses with the list of variables of every function, since it is unambiguous what this list should be. Writing the Hessian matrix of $\mathcal{F}_{k+1}$ in block form, we have

$$H_{k+1} = \begin{bmatrix} f_{k+1} H_k & b \\ b^T & c \end{bmatrix}, \tag{8}$$

where b is a $n \times 1$ vector and c is a scalar. In particular, $b_i = f'_{k+1} f'_i \prod_{\substack{m=1 \\ m \neq i}}^{k} f_m$, $\forall i = 1, 2, \ldots, n$, and $c = \mathcal{F}_k f''_{k+1}$. Note that $c \geq 0$, due to the strict positivity and convexity of all the functions f_i.

We will first distinguish the special case where there is an inflection point of the function f_{k+1}, that is, the set of points $(y_1, y_2, \ldots, y_k, y_{k+1})$ such that $f''_{k+1} = 0$. In this case, we have $c = 0$. Furthermore, since f_{k+1} is CTSF, it satisfies property (iii) and a possible inflection point is also a stationary point; that is, $f'_{k+1} = 0$, as well. Therefore, $b_i = 0$, $\forall i = 1, 2, \ldots, k$, and the Hessian matrix becomes

$$H_{k+1} = \begin{bmatrix} f_{k+1} H_k & 0 \\ 0^T & 0 \end{bmatrix}. \tag{9}$$

This matrix is positive semidefinite if its n^{th} order leading principal minor $f_{k+1} H_k$ is positive semidefinite. This is indeed the case, because H_k is positive semidefinite (induction hypothesis) and f_{k+1} is nonnegative (strict positivity of CTSF f_{k+1}).

We now treat the general case, that is, for points where $f''_{k+1} > 0$. Since $\mathcal{F}_k > 0$ (strict positivity of CTSFs f_i, $\forall i = 1, 2, \ldots, k$), we have also $c > 0 \Rightarrow c \neq 0$ and there exists the following Schur factorization for the matrix H_{k+1}:

$$\begin{bmatrix} f_{k+1} H_k & b \\ b^T & c \end{bmatrix} = \begin{bmatrix} I & c^{-1} b \\ 0^T & 1 \end{bmatrix} \cdot \begin{bmatrix} S_c & 0 \\ 0^T & c \end{bmatrix} \cdot \begin{bmatrix} I & 0 \\ c^{-1} b^T & 1 \end{bmatrix}, \tag{10}$$

where $S_c = f_{k+1} H_k - \frac{1}{c} b \cdot b^T$ is the Schur complement of H_{k+1} with respect to its bottom-right diagonal element c.

The validity of (10) can be checked easily after the appropriate matrix multiplications in the right-hand side. Therefore, we have managed to express H_{k+1} as a matrix product $U^T D U$. Note that this does not happen always after a Schur factorization. It occurs here because of the special structure, mainly the symmetry of a Hessian matrix and the fact that the bottom-right block in the partitioning of H_{k+1} is just a scalar.

It is clear that, if $D \succeq 0$, then H_{k+1} is also positive semidefinite. However, since D is block diagonal, the problem reduces to proving the positive semidefiniteness of every one of its diagonal blocks. The second block is just the scalar $c > 0$ and therefore is positive semidefinite. It remains to show that S_c is also positive semidefinite.

S_c is a square matrix of size n. Consider its diagonal elements,

$$\begin{aligned}
[S_c]_{ii} &= f_{k+1}[H_k]_{ii} - \frac{b_i^2}{c} \\
&= f_{k+1}[H_k]_{ii} - \frac{(f'_{k+1})^2 (f'_i)^2}{f''_{k+1}} \frac{\prod_{\substack{m=1\\ m\neq i}}^k (f_m)^2}{\mathcal{F}_k} \\
&= f_{k+1}[H_k]_{ii} - \frac{(f'_{k+1})^2}{f''_{k+1}} \frac{(f'_i)^2}{f_i} \prod_{\substack{m=1\\ m\neq i}}^k f_m \\
&= f_{k+1}[H_k]_{ii} - \frac{(f'_{k+1})^2}{f''_{k+1}} \left\{ f''_i - f''_i + \frac{(f'_i)^2}{f_i} \right\} \prod_{\substack{m=1\\ m\neq i}}^k f_m \\
&= f_{k+1}[H_k]_{ii} - \frac{(f'_{k+1})^2}{f''_{k+1}} f''_i \prod_{\substack{m=1\\ m\neq i}}^k f_m + \frac{(f'_{k+1})^2}{f''_{k+1}} \left\{ f''_i - \frac{(f'_i)^2}{f_i} \right\} \prod_{\substack{m=1\\ m\neq i}}^k f_m \\
&= f_{k+1}[H_k]_{ii} - \frac{(f'_{k+1})^2}{f''_{k+1}} [H_k]_{ii} + \frac{(f'_{k+1})^2}{f''_{k+1}} \left\{ \frac{f_i f''_i - (f'_i)^2}{f_i} \right\} \prod_{\substack{m=1\\ m\neq i}}^k f_m \\
&= \left\{ \frac{f_{k+1} f''_{k+1} - (f'_{k+1})^2}{f''_{k+1}} \right\} [H_k]_{ii} + \frac{(f'_{k+1})^2 \mathcal{F}_k}{f''_{k+1}} \left\{ \frac{f_i f''_i - (f'_i)^2}{(f_i)^2} \right\}.
\end{aligned}$$

Similarly, for its offdiagonal elements, we have

$$\begin{aligned}
[S_c]_{ij} &= f_{k+1}[H_k]_{ij} - \frac{b_i b_j}{c} \\
&= f_{k+1}[H_k]_{ij} - \frac{(f'_{k+1})^2 f'_i f'_j}{f''_{k+1}} \frac{\prod_{\substack{m=1\\ m\neq i}}^k f_m \prod_{\substack{m=1\\ m\neq j}}^k f_m}{\mathcal{F}_k} \\
&= f_{k+1}[H_k]_{ij} - \frac{(f'_{k+1})^2}{f''_{k+1}} f'_i f'_j \prod_{\substack{m=1\\ m\neq i,j}}^k f_m \\
&= f_{k+1}[H_k]_{ij} - \frac{(f'_{k+1})^2}{f''_{k+1}} [H_k]_{ij} \\
&= \left\{ \frac{f_{k+1} f''_{k+1} - (f'_{k+1})^2}{f''_{k+1}} \right\} [H_k]_{ij}.
\end{aligned}$$

Therefore, we have been able to express the matrix S_c in the following form:

$$S_c = \nu_1 H_k + \nu_2 \text{diag}(d_i), \tag{11}$$

where ν_1, ν_2 are nonnegative scalars and d is a $n \times 1$ vector with nonnegative elements. In particular,

$$\nu_1 = \frac{f_{k+1} f''_{k+1} - (f'_{k+1})^2}{f''_{k+1}} \geq 0,$$

because of convexity and property (iii) of the function f_{k+1};

$$\nu_2 = \frac{(f'_{k+1})^2 \mathcal{F}_k}{f''_{k+1}} \geq 0,$$

because of convexity of f_{k+1} and positivity of $\mathcal{F}_k$; and

$$d_i = \frac{f_i f''_i - (f'_i)^2}{(f_i)^2} \geq 0, \quad \forall i = 1, 2, \ldots, n,$$

because of property (iii) of each corresponding function f_i.

Equation (11) decomposes the Schur complement into the sum of two matrices, namely $S_{c1} = \nu_1 H_k$ and $S_{c2} = \nu_2 \text{diag}(d_i)$. The first matrix S_{c1} is positive semidefinite because it is the product of the positive semidefinite matrix H_k (induction hypothesis) with the nonnegative scalar ν_1. The second matrix S_{c2} is also positive semidefinite, since it is a diagonal matrix with nonnegative diagonal elements $\nu_2 d_i$. The sum of two positive semidefinite matrices is also positive semidefinite; therefore, $S_c \succeq 0$ and the theorem has been proven. □

Theorem 5.1 (also with the potential application of Lemma 4.3) can be a powerful tool in inferring the convexity of the multivariate product function, since it involves considering each of the univariate factors individually; this can offer a significant advantage especially when the number of factors and thus variables is large, because the conventional Hessian eigenvalue analysis becomes nontractable and one is usually restricted to examining convexity only within a local context.

Let us present an example, where one can infer the convexity of a relatively complex multivariate function.

Example 5.1 Let

$$\mathcal{F}_4(y_1, y_2, y_3, y_4) = \left\{ \frac{1.8}{y_2} + 1.2 y_2 - \frac{3 \log(1 - y_1)}{y_2} - 2 y_2 \log(1 - y_1) \right\} \frac{e^{y_3 - y_4}}{y_4^{1.2}}.$$

Show that this function is convex in $D = [\frac{1}{3}, \frac{2}{3}]^4$.

First, we observe that the function $\mathcal{F}_4$ can be decomposed into a product of four univariate factors, as follows:

$$\mathcal{F}_4(y_1, y_2, y_3, y_4) = \left\{ \frac{1.8}{y_2} + 1.2 y_2 - \frac{3 \log(1 - y_1)}{y_2} - 2 y_2 \log(1 - y_1) \right\} \frac{e^{y_3 - y_4}}{y_4^{1.2}}$$

$$= \left\{ \frac{3}{y_2} \left[0.6 - \log(1-y_1) \right] + 2y_2 \left[0.6 - \log(1-y_1) \right] \right\} \frac{e^{y_3 - y_4}}{y_4^{1.2}}$$

$$= \{0.6 - \log(1-y_1)\} \left\{ \frac{3}{y_2} + 2y_2 \right\} \frac{e^{y_3} e^{-y_4}}{y_4^{1.2}}$$

$$= \{0.6 - \log(1-y_1)\} \left\{ \frac{2y_2^2 + 3}{y_2} \right\} \{e^{y_3}\} \left\{ \frac{e^{-y_4}}{y_4^{1.2}} \right\}.$$

Let us define

$$f_1(y_1) = 0.6 - \log(1-y_1), \tag{12}$$

$$f_2(y_2) = \frac{2y_2^2 + 3}{y_2}, \tag{13}$$

$$f_3(y_3) = e^{y_3}, \tag{14}$$

$$f_4(y_4) = \frac{e^{-y_4}}{y_4^{1.2}}. \tag{15}$$

According to Theorem 5.1, the function $\mathcal{F}_4$ is convex, if all four factors $f_i(y_i)$ are strictly positive and satisfy $f_i(y_i) f_i''(y_i) - (f_i'(y_i))^2 \geq 0$, $\forall y_i \in [\frac{1}{3}, \frac{2}{3}]$.

(a) For function (12), we recall Example 4.1. According to that, function $h(y_1) = 1 - \log(1-y_1) = f_1(y_1) + 0.4$ is CTSF in domain $[0, 1)$ and, since $[\frac{1}{3}, \frac{2}{3}] \subseteq [0, 1)$, it is CTSF in domain $[\frac{1}{3}, \frac{2}{3}]$ as well. Applying Lemma 4.1, we infer that the function h maintains its suitability if we add to it a constant, as long as the latter is greater or equal than β^L, where

$$\beta^L = \sup_{\substack{y_1 \in [\frac{1}{3}, \frac{2}{3}] \\ y_1 : h''(y_1) > 0}} \frac{-\{h(y_1) h''(y_1) - (h'(y_1))^2\}}{h''(y_1)}$$

$$\overset{\text{note}^3}{=} \max_{y_1 \in [\frac{1}{3}, \frac{2}{3}]} \log(1-y_1) = \log\left(\frac{2}{3}\right) < -0.4.$$

Therefore, the function $f(y_1) = h(y_1) - 0.4$ is also CTSF in $[\frac{1}{3}, \frac{2}{3}]$ and, thus, is strictly positive and satisfies property (iii).

(b) We consider function (13) as the product of two functions, namely $f_{21}(y_2) = \frac{1}{y_2}$ and $f_{22}(y_2) = 2y_2^2 + 3$. The function f_{21} is the reciprocal function, which is CTSF in every domain $D_y \subset \mathbb{R}_+^*$. The function f_{22} has been considered in Example 4.3, where it was proven that it is CTSF in $[0, 1] \supset [\frac{1}{3}, \frac{2}{3}]$. Thus, both are strictly positive and satisfy property (iii). According to (7) (with $\{h, f, g\} \equiv \{f_2, f_{21}, f_{22}\}$), property (iii) will hold also for their product, function $f_2(y_2)$.

(c) Function (14) is the exponential function, which is CTSF in every subset of $\mathbb{R}$; thus, it is strictly positive and satisfies property (iii) in $[\frac{1}{3}, \frac{2}{3}]$.

[3] The supremum has been replaced by maximum because the domain is closed and function h is continuous.

(d) Function (15) can be considered as the product of two functions, namely $f_{41}(y_4) = e^{-y_4}$ and $f_{42}(y_4) = \frac{1}{y_4^{1.2}}$. Both functions are strictly positive in $[\frac{1}{3}, \frac{2}{3}]$. Regarding property (iii), we have $f_{41} f_{41}'' - (f_{41}')^2 = 0 \geq 0 \ \ \forall y_4 \in \mathbb{R}$, and $f_{42} f_{42}'' - (f_{42}')^2 = \frac{1.2}{y_4^{4.4}} \geq 0 \ \ \forall y_4 \in \mathbb{R}_+^*$. According to (7) (with $\{h, f, g\} \equiv \{f_4, f_{41}, f_{42}\}$), property (iii) will hold also for their product, function $f_4(y_4)$.

Based on (a) to (d), we conclude that the function $\mathcal{F}_4$ is convex indeed.

Note that the convexity of the product function $\mathcal{F}_n$ is obviously preserved, if an even number of univariate functions f_i is instead strictly negative and, due to property (iii), concave. Also note that CTSFs are monotone functions that satisfy the requirements of Theorem 5.1. Therefore, they are appropriate for convexification transformation of monomials with more than just two factors, such as in the case of multilinear programming. We conclude this section by presenting the following generalization of Remark 3.2.

Remark 5.1 A general irreducible signomial formulation, where at least one of the nonlinear variables participates in both a positive and a negative coefficient monomial, is not convexifiable.

Note that the applicability of the above remark includes every signomial formulation with a nonlinear equality constraint.

6 From Multilinear to Posynomial Functions

The previous section generalized the convexity analysis of the transformed bilinear term, so as to cover products of univariate functions. However, in a general signomial problem, the monomials may include factors with arbitrary rational exponents. We want to investigate under which further restrictions a function that satisfies the properties of Sect. 4 is an appropriate choice for the transformation of a variable that participates in the formulation exponentiated to some arbitrary power $\kappa \in \mathbb{R}^*$.

The following lemma helps us to identify the required restrictions.

Lemma 6.1 *Let $f(y)$ be CTSF in some interval D_y. Then, $h(y) = (f(y))^\kappa$, where $\kappa \neq 0$, is CTSF only if $\kappa > 0$ or if $\{ff'' - (f')^2\} = 0, \ \forall y \in D_y$.*

Proof Since $h = (f)^\kappa$, we have:

$$h' = \kappa (f)^{\kappa-1} f'$$

and

$$h'' = \kappa (f)^{\kappa-2}\{ff'' - (f')^2\} + \kappa^2 (f)^{\kappa-2}(f')^2.$$

Note that the powers $(f)^{\kappa-1}$ and $(f)^{\kappa-2}$ are well-defined, since $f(y) > 0, \ \forall y \in D_y$.

The strict positivity property (i) of function h results trivially from the strict positivity of the function f for every possible exponent κ.

For the strict monotonicity property (ii), we examine the first derivative h'. Since $(f)^{\kappa-1} > 0$, $\forall y \in D_y$, and since the function f is strictly monotone, so will be the function h for every possible exponent κ. In particular, it will have the same type of monotonicity as f if $\kappa > 0$ and the opposite type if $\kappa < 0$.

Regarding property (iii), we have

$$\begin{aligned} hh'' - (h')^2 &= (f)^\kappa \left(\kappa(f)^{\kappa-2}\left\{ff'' - (f')^2\right\} + \kappa^2(f)^{\kappa-2}(f')^2\right) - \left(\kappa(f)^{\kappa-1}f'\right)^2 \\ &= \kappa(f)^{2\kappa-2}\left\{ff'' - (f')^2\right\}. \end{aligned} \tag{16}$$

Since f is a CTSF, it satisfies property (iii). Therefore, for this property to hold for the function h as well, and given that

$$(f)^{2\kappa-2} > 0, \ \forall y \in D_y,$$

we need

$$\text{either} \quad \kappa \ge 0 \overset{\kappa\neq 0}{\Rightarrow} \kappa > 0 \quad \text{or} \quad \{ff'' - (f')^2\} = 0, \quad \forall y \in D_y. \qquad \square$$

Under the restrictions imposed by Lemma 6.1, a variable transformation $x \to f(y)$ is appropriate for the transformation of a monomial factor that is exponentiated to some power κ, since the resulting factor $h(y) = (f(y))^\kappa$ is CTSF and may participate in a convex product of univariate functions, as dictated by Theorem 5.1. Therefore, we have established the sufficient conditions for transformation functions that can convexify a (positive coefficient) monomial and thus a complete posynomial formulation.

According to these results, every CTSF with an appropriate range is suitable to transform a monomial factor, as long as the latter is exponentiated to a positive constant. If we have the case where negative exponents are present, we need to restrict ourselves to CTSFs for which property (iii) holds as an equality. The following theorem investigates this latter case.

Theorem 6.1 *The exponential transformation*, $x \to ce^{\lambda y}$, *where* $c > 0$ *and* $\lambda \neq 0$, *is the only transformation that can convexify every possible term in a general posynomial program with strictly positive variables.*

Proof According to Lemma 6.1, for there to be no restriction on the exponent κ, the function $f(y)$ used has to be a solution of the following ordinary differential equation:

$$f(y)f''(y) - (f'(y))^2 = 0. \tag{17}$$

Of course, it should still satisfy properties (i) and (ii). We have

$$\begin{aligned} f(y)f''(y) - (f'(y))^2 = 0 \ &\overset{f\neq 0}{\Leftrightarrow} \ \frac{f(y)f''(y) - (f'(y))^2}{(f(y))^2} = 0 \\ &\Leftrightarrow \ \frac{d}{dy}\frac{f'(y)}{f(y)} = 0 \end{aligned}$$

$$
\begin{aligned}
&\Leftrightarrow \quad \frac{f'(y)}{f(y)} = \lambda \\
&\overset{f>0}{\Leftrightarrow} \quad \frac{d}{dy}\log(f(y)) = \lambda \\
&\Leftrightarrow \quad \log(f(y)) = \lambda y + c_1 \\
&\overset{e^{c_1}\to c}{\Leftrightarrow} \quad f(y) = ce^{\lambda y},
\end{aligned}
$$

where $c_1 \in \mathbb{R}$ is an auxiliary constant.

The uniqueness of this solution is thus established. Therefore, the exponential transformation is the only one that can address appropriately every variable that participates in a generic posynomial problem. The constants $c \in \mathbb{R}_+^*$ and $\lambda \in \mathbb{R}^*$ can be selected arbitrarily or can be used to properly scale the transformed problem. Note that $\lambda \neq 0$ because the function $f(y)$ cannot be constant. □

7 Conclusions

We have established a way to infer the convexity of multivariate functions that can be expressed as products of univariate functions. It is based on a set of sufficient conditions that have to be satisfied by every individual factor univariate function. The results can be used to devise transformation schemes that aim at alleviating the nonconvexities of a signomial problem. Furthermore, we have shown that the exponential transformation is the only one that can address variables that are exponentiated to a negative power, thus making it the only suitable functional form for the complete convexification of any given posynomial problem with strictly positive variables.

References

1. Floudas, C.A., Grossmann, I.E.: Synthesis of flexible heat-exchanger networks with uncertain flowrates and temperatures. Comput. Chem. Eng. **11**, 319–336 (1987)
2. Floudas, C.A., Anastasiades, S.H.: Synthesis of distillation sequences with several multicomponent feed and product streams. Chem. Eng. Sci. **43**, 2407–2419 (1988)
3. Aggarwal, A., Floudas, C.A.: Synthesis of general distillation sequences—nonsharp separations. Comput. Chem. Eng. **14**, 631–653 (1990)
4. Kokkosis, A.C., Floudas, C.A.: Optimization of complex reactor networks: 2. Nonisothermal operation. Chem. Eng. Sci. **49**, 1037–1051 (1994)
5. Papalexandri, K.P., Pistikopoulos, E.N., Floudas, C.A.: Mass-exchange networks for waste minimization—a simultaneous approach. Chem. Eng. Res. Des. **72**, 279–294 (1994)
6. Visweswaran, V., Floudas, C.A.: A global optimization algorithm (GOP) for certain classes of nonconvex NLPs: II. Application of theory and test problems. Comput. Chem. Eng. **14**, 1419–1434 (1990)
7. Adjiman, C.S., Androulakis, I.P., Maranas, C.D., Floudas, C.A.: A global optimization method αBB for process design. Comput. Chem. Eng. **20**, S419–S424 (1996)
8. Adjiman, C.S., Androulakis, I.P., Floudas, C.A.: Global optimization of MINLP problems in process synthesis and design. Comput. Chem. Eng. **21**, S445–S450 (1997)
9. Adjiman, C.S., Androulakis, I.P., Floudas, C.A.: Global optimization of mixed-integer nonlinear problems. AIChE J. **46**, 1769–1797 (2000)
10. Esposito, W.R., Floudas, C.A.: Global optimization for the parameter estimation of differential-algebraic systems. Ind. Eng. Chem. Res. **37**, 1841–1858 (1998)

11. Esposito, W.R., Floudas, C.A.: Global optimization in parameter estimation of nonlinear algebraic models via the error-in-variables approach. Ind. Eng. Chem. Res. **39**, 1291–1310 (2000)
12. McDonald, C.M., Floudas, C.A.: Global optimization and analysis of the Gibbs free-energy function using the UNIFAC, Wilson and ASOG equations. Comput. Chem. Eng. **19**, 1111–1141 (1995)
13. McDonald, C.M., Floudas, C.A.: Global optimization for the phase and chemical equilibrium problem: application to the NRTL equation. Ind. Eng. Chem. Res. **34**, 1674–1687 (1995)
14. Androulakis, I.P., Maranas, C.D., Floudas, C.A.: Prediction of oligopeptide conformations via deterministic global optimization. J. Glob. Optim. **11**, 1–34 (1997)
15. Hatzimanikatis, V., Floudas, C.A., Bailey, J.E.: Analysis and design of metabolic reaction networks via mixed-integer linear optimization. AIChE J. **42**, 1277–1292 (1996)
16. Hatzimanikatis, V., Floudas, C.A., Bailey, J.E.: Optimization of regulatory architectures in metabolic reaction networks. Biotechnol. Bioeng. **52**, 485–500 (1996)
17. Zener, C.: A mathematical aid in optimizing engineering designs. Proc. Natl. Acad. Sci. **47**, 537–539 (1961)
18. Zener, C.: A further mathematical aid in optimizing engineering designs. Proc. Natl. Acad. Sci. **48**, 512–522 (1962)
19. Duffin, R.J., Peterson, E.L.: A mathematical aid in optimizing engineering designs. SIAM J. Appl. Math. **14**, 1307–1349 (1966)
20. Duffin, R.J.: Linearizing geometric programs. SIAM Rev. **12**, 211–227 (1970)
21. Peterson, E.L.: Geometric programming. SIAM Rev. **18**, 1–51 (1976)
22. Passy, U., Wilde, D.J.: Generalized polynomial optimization. SIAM J. Appl. Math. **15**, 1344 (1967)
23. Blau, G.E., Wilde, D.J.: Generalized polynomial programming. Can. J. Chem. Eng. **47**, 317 (1969)
24. Avriel, M., Williams, A.C.: Complementary geometric programs. SIAM J. Appl. Math. **19**, 125–141 (1970)
25. Duffin, R.J., Peterson, E.L.: Reversed geometric programs treated by harmonic means. Indiana Univ. Math. J. **22**, 531–549 (1972)
26. Duffin, R.J., Peterson, E.L.: Geometric programming with signomials. J. Optim. Theory Appl. **11**, 3–35 (1973)
27. Sarma, P.V.N.L., Martens, X.M., Reklaitis, G.V., Rijckaert, M.J.: Comparison of computational strategies for geometric programs. J. Optim. Theory Appl. **26**, 185–203 (1978)
28. Falk, J.E.: Global solutions of signomial programs. Report, The George Washington University, Program in Logistics (1973)
29. Maranas, C.D., Floudas, C.A.: Finding all solutions of nonlinearly constrained systems of equations. J. Glob. Optim. **7**, 143–182 (1995)
30. Maranas, C.D., Floudas, C.A.: Global optimization in generalized geometric programming. Comput. Chem. Eng. **21**, 351–370 (1997)
31. Floudas, C.A.: Deterministic Global Optimization: Theory, Algorithms and Applications. Kluwer Academic, Dordrecht (2000)
32. Floudas, C.A., Pardalos, P.M.: State of the art in global optimization—computational methods and applications—preface. J. Glob. Optim. **7**, 113–113 (1995)
33. Floudas, C.A., Pardalos, P.M.: State of the Art in Global Optimization—Computational Methods and Applications. Kluwer Academic, Dordrecht (1996)
34. Floudas, C.A., Pardalos, P.M.: Frontiers in Global Optimization. Kluwer Academic, Dordrecht (2003)
35. Shen, P.: Linearization method of global optimization for generalized geometric programming. Appl. Math. Comput. **162**, 353–370 (2005)
36. Wang, Y., Liang, Z.: A deterministic global optimization algorithm for generalized geometric programming. SIAM J. Appl. Math. **168**, 722–737 (2005)
37. Li, H.L., Tsai, J.F.: Treating free variables in generalized geometric global optimization programs. J. Glob. Optim. **33**, 1–13 (2005)
38. Tsai, J.F., Lin, M.H.: An optimization approach for solving signomial discrete programming problems with free variables. Comput. Chem. Eng. **30**, 1256–1263 (2006)
39. Tsai, J.F., Lin, M.H., Hu, Y.C.: On generalized geometric programming problems with nonpositive variables. Eur. J. Oper. Res. **178**, 10–17 (2007)
40. Li, H.L., Tsai, J.F., Floudas, C.A.: Convex underestimation for posynomial functions of positive variables. Optim. Lett. **2** (2008, in press)
41. Björk, K.M., Lindberg, P.O., Westerlund, T.: Some convexifications in global optimization of problems containing signomial terms. Comput. Chem. Eng. **27**, 669–679 (2003)
42. Westerlund, T.: In: Liberti, L., Maculan, N. (eds.) Global Optimization: From Theory to Implementation, pp. 45–74. Springer, Berlin (2006)