Tractable Approximate Robust Geometric Programming

Kan-Lin Hsiung   Seung-Jean Kim   Stephen Boyd

Electrical Engineering Department
Stanford University
Stanford, CA 94305

{klhsiung, sjkim, boyd}@stanford.edu

Abstract

The optimal solution of a geometric program (GP) can be sensitive to variations in the problem data. Robust geometric programming can systematically alleviate the sensitivity problem by explicitly incorporating a model of data uncertainty in a GP and optimizing for the worst-case scenario under this model. However, it is not known whether a general robust GP can be reformulated as a tractable optimization problem that interior-point or other algorithms can efficiently solve. In this paper we propose an approximation method that seeks a compromise between solution accuracy and computational efficiency.

The method is based on approximating the robust GP as a robust linear program (LP), by replacing each nonlinear constraint function with a piecewise-linear (PWL) convex approximation. With a polyhedral or ellipsoidal description of the uncertain data, the resulting robust LP can be formulated as a standard convex optimization problem that interior-point methods can solve. The drawback of this basic method is that the number of terms in the PWL approximations required to obtain an acceptable approximation error can be very large. To overcome the "curse of dimensionality" that arises in directly approximating the nonlinear constraint functions in the original robust GP, we form a conservative approximation of the original robust GP, which contains only bivariate constraint functions. We show how to find globally optimal PWL approximations of these bivariate constraint functions.

Key words: Geometric programming, linear programming, piecewise-linear function, robust geometric programming, robust linear programming, robust optimization.

1 Introduction

1.1 Geometric programming

The convex function $\mathbf{lse} : \mathbf{R}^k \to \mathbf{R}$, defined as

$$\mathbf{lse}(z_1, \ldots, z_k) = \log(e^{z_1} + \cdots + e^{z_k}), \tag{1}$$

is called the (k-term) *log-sum-exp* function. (We use the same notation, no matter what k is; the context will always unambiguously determine the number of exponential terms.) When $k = 1$, the log-sum-exp function reduces to the identity.

A *geometric program* (in convex form) has the form

$$\begin{array}{ll} \text{minimize} & c^T y \\ \text{subject to} & \mathbf{lse}(A_i y + b_i) \le 0, \quad i = 1, \ldots, m, \\ & Gy + h = 0, \end{array} \tag{2}$$

where the optimization variable is $y \in \mathbf{R}^n$ and the problem data are $A_i \in \mathbf{R}^{K_i \times n}$, $b_i \in \mathbf{R}^{K_i}$, $c \in \mathbf{R}^n$, $G \in \mathbf{R}^{l \times n}$, and $h \in \mathbf{R}^l$. We call the inequality constraints in the GP (2) *log-sum-exp (inequality) constraints*. In many applications, GPs arise in *posynomial form*, and are then transformed by a standard change of coordinates and constraint functions to the convex form (2); see Appendix A. This transformation does not in any way change the problem data, which are the same for the posynomial form and convex form problems.

Geometric programming has been used in various fields since the late 1960s; early applications of geometric programming can be found in the books [Avr80, DPZ67, Zen71] and the survey papers [Eck80, Pet76, BKVH05]. More recent applications can be found in various fields including circuit design [BKPH05, CHP00, DBHL01, DGS03, Her02, HBL01, MHBL00, Sap96, SNLS05, SRVK93, YCLW01], chemical process control [WGW86], environment quality control [Gre95], resource allocation in communication systems [DR92], information theory [CB04, KC97], power control of wireless communication networks [KB02, OJB03], and statistics [MJ83].

Algorithms for solving geometric programs appeared in the late 1960s, and research on this topic continued until the early 1990s; see, *e.g.*, [ADP75, RB90]. A huge improvement in computational efficiency was achieved in 1994, when Nesterov and Nemirovsky developed provably efficient interior-point methods for many nonlinear convex optimization problems, including GPs [NN94]. A bit later, Kortanek, Xu, and Ye developed a primal-dual interior-point method for geometric programming, with efficiency approaching that of interior-point linear programming solvers [KXY97].

1.2 Robust geometric programming

In *robust geometric programming* (RGP), we include an explicit model of uncertainty or variation in the data that defines the GP. We assume that the problem data (A_i, b_i) depend affinely on a vector of uncertain parameters u, that belongs to a set $\mathcal{U} \subseteq \mathbf{R}^L$:

$$\left(\tilde{A}_i(u), \tilde{b}_i(u)\right) = \left(A_i^0 + \sum_{j=1}^{L} u_j A_i^j,\ b_i^0 + \sum_{j=1}^{L} u_j b_i^j\right), \qquad u \in \mathcal{U} \subseteq \mathbf{R}^L. \tag{3}$$

The data variation is described by $A_i^j \in \mathbf{R}^{K_i \times n}$, $b_i^j \in \mathbf{R}^{K_i}$, and the uncertainty set $\mathcal{U}$. We assume that all of these are known.

We consider two types of uncertainty sets. One is *polyhedral uncertainty*, in which $\mathcal{U}$ is a polyhedron, *i.e.*, the intersection of a finite number of halfspaces:

$$\mathcal{U} = \left\{ u \in \mathbf{R}^L \;\middle|\; Du \preceq d \right\}, \tag{4}$$

where $d \in \mathbf{R}^K$, $D \in \mathbf{R}^{K \times L}$, and the symbol $\preceq$ denotes the componentwise inequality between two vectors: $w \preceq v$ means $w_i \leq v_i$ for all i. The other is *ellipsoidal uncertainty*, in which $\mathcal{U}$ is an ellipsoid:

$$\mathcal{U} = \left\{ \bar{u} + P\rho \;\middle|\; \|\rho\|_2 \leq 1,\ \rho \in \mathbf{R}^L \right\}, \tag{5}$$

where $\bar{u} \in \mathbf{R}^L$ and $P \in \mathbf{R}^{L \times L}$. Here, the matrix P describes the variation in u and can be singular, in order to model the situation when the variation in u is restricted to a subspace. Note that due to the affine structure in (3), the ellipsoid uncertainty set $\mathcal{U}$ can be transformed to a unit ball (*i.e.*, P can be assumed to be an identity matrix) without loss of generality.

A (worst-case) robust GP (RGP) has the form

$$\begin{array}{ll} \text{minimize} & c^T y \\ \text{subject to} & \sup_{u \in \mathcal{U}} \mathbf{lse}\left(\tilde{A}_i(u) y + \tilde{b}_i(u)\right) \leq 0, \quad i = 1, \ldots, m, \\ & Gy + h = 0. \end{array} \tag{6}$$

The inequality constraints in the RGP (6) are called *robust log-sum-exp (inequality) constraints.*

The RGP (6) is a special type of robust convex optimization problem; see, *e.g.*, [BTN98] for more on robust convex optimization. Unlike the various types of robust convex optimization problems that have been studied in the literature [BTN99, BTNR02, GL97, GL98, GI03], the computational tractability of the RGP (6) is not clear; it is not yet known whether one can reformulate a general RGP as a tractable optimization problem that interior-point or other algorithms can efficiently solve.

1.3 Brief overview and outline

We first observe that a log-sum-exp function can be approximated arbitrarily well by a piecewise-linear (PWL) convex function. Using these approximations, the RGP can be approximated arbitrarily well as a robust LP, with polyhedral or ellipsoidal data uncertainty. Since robust LPs, with polyhedral or ellipsoidal uncertainty, can be tractably solved (see Appendix B), this gives us an approximation method for the RGP. In fact, this general approach can be used for any robust convex optimization problem with polyhderal or ellipsoidal uncertainty. Piecewise-linear approximation has been used in prior work on approximation methods for nonlinear convex optimization problems, since it allows us to approximately solve a nonlinear convex problem by solving a linear program; see, *e.g.*, [BTN01, FM88, Gli00, Tha78].

The problem with the basic PWL approach is that the number of terms needed in a PWL approximation of the log-sum-exp function (1), to obtain a given level of accuracy, grows rapidly with the dimension k. Thus, the size of the resulting robust LP is prohibitively large, unless all K_i are small. To overcome this "curse of dimensionality", we propose the following approach. We first replace the RGP with a new RGP, in which each log-sum-exp function has only one or two terms. This transformation to a two-term GP is exact for a nonrobust GP, and conservative for a RGP. We then use the PWL approximation method on the reduced RGP.

In §2, we show how PWL approximation of the constraint functions in the RGP (6) leads to a robust LP. We also describe how to approximate a general RGP with a more tractable RGP which contains only bivariate constraint functions.

In §3, we develop a constructive algorithm to solve the best PWL convex lower and upper approximation problems for the bivariate log-sum-exp function. Some numerical examples are presented in §4. Our conclusions are given in §5. Supplementary material is collected in the appendices.

2 Solving robust GPs via PWL approximation

2.1 Robust LP approximation

Suppose we have PWL lower and upper bounds on the log-sum-exp function in the ith constraint of the RGP (6),

$$\max_{j=1,\dots,I_i} \left\{ \underline{f}_{ij}^T y + \underline{g}_{ij} \right\} \leq \mathbf{lse}(y) \leq \max_{j=1,\dots,J_i} \left\{ \overline{f}_{ij}^T y + \overline{g}_{ij} \right\}, \quad \forall\, y \in \mathbf{R}^{K_i},$$

where $\underline{f}_{ij}, \overline{f}_{ij} \in \mathbf{R}^{K_i}$ and $\underline{g}_{ij}, \overline{g}_{ij} \in \mathbf{R}$. Replacing the log-sum-exp functions in the RGP (6) with the PWL bounds above, we obtain the two problems

$$\begin{array}{ll} \text{minimize} & c^T y \\ \text{subject to} & \sup_{u \in \mathcal{U}} \max_{j=1,\dots,J_i} \left\{ \overline{f}_{ij}^T \tilde{A}_i(u) y + \overline{f}_{ij}^T \tilde{b}_i(u) + \overline{g}_{ij} \right\} \leq 0, \quad i = 1, \dots, m, \\ & Gy + h = 0, \end{array} \tag{7}$$

and

$$\begin{array}{ll} \text{minimize} & c^T y \\ \text{subject to} & \sup_{u \in \mathcal{U}} \max_{j=1,\dots,I_i} \left\{ \underline{f}_{ij}^T \tilde{A}_i(u) y + \underline{f}_{ij}^T \tilde{b}_i(u) + \underline{g}_{ij} \right\} \leq 0, \quad i = 1, \dots, m, \\ & Gy + h = 0. \end{array} \tag{8}$$

These problems can be reformulated as the robust LPs

$$\begin{array}{ll}\text{minimize} & c^T y \\ \text{subject to} & \sup_{u \in \mathcal{U}} \left\{ \overline{f}_{ij}^T \tilde{A}_i(u) y + \overline{f}_{ij}^T \tilde{b}_i(u) + \overline{g}_{ij} \right\} \leq 0, \quad i = 1, \ldots, m, \quad j = 1, \ldots, J_i, \\ & Gy + h = 0, \end{array} \tag{9}$$

and

$$\begin{array}{ll}\text{minimize} & c^T y \\ \text{subject to} & \sup_{u \in \mathcal{U}} \left\{ \underline{f}_{ij}^T \tilde{A}_i(u) y + \underline{f}_{ij}^T \tilde{b}_i(u) + \underline{g}_{ij} \right\} \leq 0, \quad i = 1, \ldots, m, \quad j = 1, \ldots, I_i, \\ & Gy + h = 0. \end{array} \tag{10}$$

With a polyhedral uncertainty set, these can be cast as (larger) LPs, and for ellipsoidal uncertainty sets, they can be cast as SOCPs; see Appendix B.

Note that an optimal solution, say $\overline{y}$, of the robust LP (9) is also a feasible solution to the RGP (6). In other words, the robust LP (9) gives a conservative approximation of the RGP (6). The robust LP (10) has the opposite property: its feasible set covers the feasible set of the RGP (6). Therefore, the optimal value of the robust LP (10), say, $c^T \underline{y}$, gives a lower bound on the optimal value of the original RGP (6), and in particular, allows us to bound the error in the feasible, suboptimal point $\overline{y}$, for the RGP. In other words, we have

$$0 \leq c^T(\overline{y} - y^\star) \leq c^T(\overline{y} - \underline{y}), \tag{11}$$

where $y^\star$ is an optimal solution of the RGP. Finally, it is not difficult to see that as the PWL convex approximations of the log-sum-exp functions are made finer, the optimal values of the robust LPs (9) and (10) get closer to that of the RGP (6).

2.2 Tractable robust GP approximation

The RGP (6) can be reformulated as another RGP

$$\begin{array}{ll}\text{minimize} & \bar{c}^T \eta \\ \text{subject to} & \sup_{u \in \mathcal{U}} \mathbf{lse}\left((\bar{a}_{i1} + \bar{B}_{i1} u)^T \eta, \ldots, (\bar{a}_{iK_i} + \bar{B}_{iK_i} u)^T \eta \right) \leq 0, \quad i = 1, \ldots, m, \\ & \bar{G}\eta + \bar{h} = 0 \end{array} \tag{12}$$

with the optimization variables $\eta = (y, t) \in \mathbf{R}^n \times \mathbf{R}$. Here the problem data

$$\bar{c} \in \mathbf{R}^{n+1}, \quad \bar{G} \in \mathbf{R}^{(l+1)\times(n+1)}, \quad \bar{h} \in \mathbf{R}^{l+1}, \quad \bar{a}_{is} \in \mathbf{R}^{n+1}, \quad \bar{B}_{is} \in \mathbf{R}^{(n+1)\times L}$$

can be readily obtained from the problem data of the RGP (6); see Appendix C for the details. The RGPs (6) and (12) are equivalent: $\bar{y} \in \mathbf{R}^n$ is feasible to (6) if and only if $(\bar{y}, \bar{t}) \in \mathbf{R}^{n+1}$ is feasible to (12) for some $\bar{t} \in \mathbf{R}$. In the following we form a conservative approximation of the RGP (12), in which all the nonlinear constraint functions are bivariate.

Consider a k-term robust log-sum-exp constraint in the following generic form:

$$\sup_{u \in \mathcal{U}} \mathbf{lse}\left((a_1 + B_1 u)^T \eta, \ldots, (a_k + B_k u)^T \eta \right) \leq 0, \tag{13}$$

where $a_i \in \mathbf{R}^{n+1}$, $B_i \in \mathbf{R}^{(n+1)\times L}$. An approximate reduction procedure, described in Appendix D, shows that $\eta \in \mathbf{R}^{n+1}$ satisfies (13) if there exists $z = (z_1, \ldots, z_{k-2}) \in \mathbf{R}^{k-2}$ such that (η, z) satisfies the following system of $k-1$ *two-term* robust log-sum-exp constraints:

$$
\begin{aligned}
\sup_{u\in\mathcal{U}} \mathbf{lse}\left((a_1 + B_1 u)^T \eta, z_1\right) &\le 0, \\
\sup_{u\in\mathcal{U}} \mathbf{lse}\left((a_{s+1} + B_{s+1} u)^T \eta - z_s, z_{s+1} - z_s\right) &\le 0, \quad s = 1, \ldots, k-3, \\
\sup_{u\in\mathcal{U}} \mathbf{lse}\left((a_{k-1} + B_{k-1} u)^T \eta - z_{k-2}, (a_k + B_k u)^T \eta - z_{k-2}\right) &\le 0,
\end{aligned}
\tag{14}
$$

in which all the constraint functions are bivariate. We will call (14) a "two-term (conservative) approximation" of the k-term robust log-sum-exp constraint (13).

The idea of tractable RGP approximation is simple: we replace every robust log-sum-exp constraint(with more than two terms) by its two-term conservative approximation to obtain a "two-term RGP", which gives a conservative approximation of the original RGP. Although with more variables and constraints, the two-term RGP is much more tractable, in the sense that we can approximate the bivariate log-sum-exp function well with a small number of hyperplanes, as described in §3. Then the two-term RGP can be further solved via robust LP approximation, as shown in §2.1.

Now we give an exact expression of the two-term RGP approximation. First note that a one-term robust log-sum-exp constraint is simply a robust linear inequality. Since no PWL approximation for a one-term constraint is necessary, we can simply keep all the one-term constraints of a RGP in its two-term RGP approximation (and the consequent robust LP approximation). Therefore for simplicity, in the following we assume all the robust log-sum-exp constraints in RGP (12) have at least two terms, *i.e.*, $K_i \ge 2$, $i = 1, \ldots, m$. The two-term RGP has the form

$$
\begin{array}{ll}
\text{minimize} & \hat{c}^T x \\
\text{subject to} & \sup_{u\in\mathcal{U}} \mathbf{lse}\left((\hat{a}_i^1 + \hat{B}_i^1 u)^T x, (\hat{a}_i^2 + \hat{B}_i^2 u)^T x\right) \le 0, \quad i = 1, \ldots, K_c, \\
& \hat{G}x + \hat{h} = 0,
\end{array}
\tag{15}
$$

where the optimization variables are $x = (y, t, z) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R}^{K_v}$, and the problem data are

$$
\begin{array}{l}
\hat{a}_i^j \in \mathbf{R}^{n+K_v+1}, \quad \hat{B}_i^j \in \mathbf{R}^{(n+K_v+1)\times L}, \quad i = 1, \ldots, K_c, \quad j = 1, 2, \\
\hat{c} = (\bar{c}, 0) \in \mathbf{R}^{n+K_v+1}, \quad \hat{G} = [\bar{G} \;\; 0] \in \mathbf{R}^{(l+1)\times(n+K_v+1)}, \quad \hat{h} = \bar{h} \in \mathbf{R}^{l+1}.
\end{array}
$$

Here $K_v = \sum_{i=1}^m (K_i - 2)$ is the number of additional variables and $K_c = \sum_{i=1}^m (K_i - 1)$ is the number of two-term log-sum-exp constraints.

With general uncertainty structures, the RGP (15) is a conservative approximation of the original RGP (6). In other words, if $\hat{x} = (\hat{y}, \hat{t}, \hat{z}) \in \mathbf{R}^n \times \mathbf{R} \times \mathbf{R}^{K_v}$ is feasible to (15), $\hat{y}$ is feasible to (6). Hence the optimal value of the two-term RGP (15), if feasible, is an upper bound on that of the RGP (6).

3 PWL approximation of two-term log-sum-exp function

There has been growing interest in approximation and interpolation with convexity constraints [Bea81, Bea82, GNP95, Hu91, MR78]. However, relatively little attention has been paid to the best PWL convex approximation problem for multivariate, or even bivariate, convex functions. (A heuristic method, based on the K-means clustering algorithm, is developed in [MB05].) In this section, the problem of finding the best PWL convex approximation of the two-term (*i.e.*, bivariate) log-sum-exp function is solved and a constructive algorithm is provided.

3.1 Definitions

Let $\mathbf{int}\, X$ denote the interior of $X \subseteq \mathbf{R}^m$. A function $h : \mathbf{R}^m \to \mathbf{R}$ is called *(r-term) piecewise-linear* if there exists a partition of $\mathbf{R}^m$ as

$$\mathbf{R}^m = X_1 \cup X_2 \cup \cdots \cup X_r,$$

where $\mathbf{int}\, X_i \neq \emptyset$ and $\mathbf{int}\, X_i \cap \mathbf{int}\, X_j = \emptyset$ for $i \neq j$, and a family of affine functions $a_1^T x + b_1, \ldots, a_r^T x + b_r$ such that $h(x) = a_i^T x + b_i$ for $x \in X_i$. If an r-term PWL function h is convex, it can be expressed as the maximum of r affine functions: $h(x) = \max\{a_1^T x + b_1, \ldots, a_r^T x + b_r\}$. (See, *e.g.*, [BV04].) Let $\mathcal{P}_r^m$ denote the set of r-term PWL convex functions from $\mathbf{R}^m$ into $\mathbf{R}$. Note that $h \in \mathcal{P}_r^1$ if and only if there exist $x_i,\ i = 1, \ldots, r-1$ and $a_i, b_i,\ i = 1, \ldots, r$ with $x_1 < \cdots < x_{r-1}$ and $a_1 < \cdots < a_r$ such that h can be expressed as

$$h(x) = \begin{cases} a_1 x + b_1, & x \in (-\infty,\ x_1], \\ a_i x + b_i, & x \in [x_{i-1},\ x_i], \quad i = 2, \ldots, r-1, \\ a_r x + b_r, & x \in [x_{r-1},\ \infty). \end{cases}$$

The points $x_1, \ldots, x_{r-1}$ are called the *break points* of h, and the affine functions $a_i x + b_i,\ i = 1, \ldots, r$ are called the *segments*.

Let f be a continuous function from $\mathbf{R}^m$ into $\mathbf{R}$. A function $h : \mathbf{R}^m \to \mathbf{R}$ is called an *r-term PWL convex lower (respectively, upper) approximation* to f if $h \in \mathcal{P}_r^m$ and $h(x) \leq f(x)$ (respectively, $h(x) \geq f(x)$) for all $x \in \mathbf{R}^m$. An r-term PWL convex lower (respectively, upper) approximation $\underline{f}_r \in \mathcal{P}_r^m$ (respectively, $\overline{f}_r \in \mathcal{P}_r^m$) to f is called a *best r-term PWL convex lower (respectively, upper) approximation* if it has the minimum approximation error in the uniform norm among all r-term PWL convex lower (respectively, upper) approximations to f, which is denoted by $\underline{\epsilon}_f(r)$ (respectively, $\overline{\epsilon}_f(r)$):

$$\begin{aligned} \underline{\epsilon}_f(r) &= \sup_{x \in \mathbf{R}^m} \left(f(x) - \underline{f}_r(x)\right) = \inf_{h \in \mathcal{P}_r^m} \left\{ \sup_{x \in \mathbf{R}^m} (f(x) - h(x)) \;\middle|\; h(x) \leq f(x),\ \forall\, x \in \mathbf{R}^m \right\}, \\ \overline{\epsilon}_f(r) &= \sup_{x \in \mathbf{R}^m} \left(\overline{f}_r(x) - f(x)\right) = \inf_{h \in \mathcal{P}_r^m} \left\{ \sup_{x \in \mathbf{R}^m} (h(x) - f(x)) \;\middle|\; h(x) \geq f(x),\ \forall\, x \in \mathbf{R}^m \right\}. \end{aligned}$$

3.2 Best PWL approximation of two-term log-sum-exp function

3.2.1 Equivalent univariate best approximation problem

Finding the best r-term PWL convex approximation to the two-term log-sum-exp function is a "bivariate" best approximation problem over $\mathcal{P}_r^2$. In the following we show that this bivariate best approximation problem can be simplified as an equivalent "univariate" best approximation problem over $\mathcal{P}_r^1$.

We define the function $\phi : \mathbf{R} \to \mathbf{R}$ as

$$\phi(x) = \log(1 + e^x). \tag{16}$$

Note that ϕ satisfies

$$\lim_{x\to-\infty} \phi(x) = \lim_{x\to\infty} (\phi(x) - x) = 0. \tag{17}$$

Thus,

$$\underline{\epsilon}_\phi(1) = \inf_{(a,b)\in\mathbf{R}^2} \sup_{x\in\mathbf{R}} (\phi(x) - ax - b) = \infty, \tag{18}$$

$$\underline{\epsilon}_\phi(2) = \sup_{x\in\mathbf{R}} (\phi(x) - \max\{0, x\}) = \log 2. \tag{19}$$

Now, note that the two-term log-sum-exp function can be expressed as

$$\mathbf{lse}(y_1, y_2) = y_1 + \phi(y_2 - y_1) = y_2 + \phi(y_1 - y_2), \quad \forall\ (y_1, y_2) \in \mathbf{R}^2. \tag{20}$$

Therefore we see from (18–20) that the two-term log-sum-exp function cannot be approximated by a single affine function with a finite approximation error over $\mathbf{R}^2$, but has the unique best two-term PWL convex lower approximation $\underline{h}_2 : \mathbf{R}^2 \to \mathbf{R}$ and upper approximation $\overline{h}_2 : \mathbf{R}^2 \to \mathbf{R}$ defined as $\underline{h}_2(y_1, y_2) = \max\{y_1, y_2\}$ and $\overline{h}_2(y_1, y_2) = \max\{y_1 + \log 2, y_2 + \log 2\}$ respectively.

From now on, we restrict our discussion to the case $r \geq 3$. The following proposition establishes the uniqueness and some useful properties of the best r-term PWL convex lower approximation $\underline{\phi}_r$ to ϕ for $r \geq 3$.

Proposition 1. *For $r \geq 3$, there exist $x_1, \ldots, x_{r-1}$ and $(\underline{a}_i^\star, \underline{b}_i^\star) \in \mathbf{R}^2$, $i = 1, \ldots, r-2$ with*

$$x_1 < \cdots < x_{r-1}, \qquad 0 < \underline{a}_1^\star < \underline{a}_2^\star < \cdots < \underline{a}_{r-2}^\star < 1, \tag{21}$$

$$\underline{a}_i^\star + \underline{a}_{r-i-1}^\star = 1, \qquad \underline{b}_i^\star = \underline{b}_{r-i-1}^\star, \qquad i = 1, \ldots, r-2, \tag{22}$$

such that the function ϕ has the unique best r-term PWL convex lower approximation $\underline{\phi}_r$ defined as

$$\underline{\phi}_r(x) = \begin{cases} 0, & x \in (-\infty,\ x_1], \\ \underline{a}_i^\star x + \underline{b}_i^\star, & x \in [x_i,\ x_{i+1}], \quad i = 1, \ldots, r-2, \\ x, & x \in [x_{r-1},\ \infty). \end{cases} \tag{23}$$

Moreover, there exist $\tilde{x}_1, \ldots, \tilde{x}_{r-2} \in \mathbf{R}$ which satisfy

$$x_1 < \tilde{x}_1 < x_2 < \tilde{x}_2 < \cdots < x_{r-2} < \tilde{x}_{r-2} < x_{r-1}$$

such that the segments $\underline{a}_1^\star x + \underline{b}_1^\star, \ldots, \underline{a}_{r-2}^\star x + \underline{b}_{r-2}^\star$ *are tangent to* ϕ *at the points* $\tilde{x}_1, \ldots, \tilde{x}_{r-2}$ *respectively. Finally, the maximum approximation error occurs only at the break points of* $\underline{\phi}_r$*:*

$$\begin{array}{rcll} \phi(x) - \underline{\phi}_r(x) & < & \underline{\epsilon}_\phi(r), & x \notin \{x_1, \ldots, x_{r-1}\}, \\ \phi(x_i) - \underline{\phi}_r(x_i) & = & \underline{\epsilon}_\phi(r), & i = 1, \ldots, r-1. \end{array}$$

The proof is given in Appendix E.1.

As a consequence of Proposition 1 and (20), we have the following corollary.

Corollary 1. *For* $r \geq 3$*, the unique best* r*-term PWL convex lower approximation* $\underline{h}_r : \mathbf{R}^2 \to \mathbf{R}$ *of the two-term log-sum-exp function is*

$$\underline{h}_r(y_1, y_2) = \max\left\{y_1, \underline{a}_{r-2}^\star y_1 + \underline{a}_1^\star y_2 + \underline{b}_1^\star, \underline{a}_{r-3}^\star y_1 + \underline{a}_2^\star y_2 + \underline{b}_2^\star, \ldots, \underline{a}_1^\star y_1 + \underline{a}_{r-2}^\star y_2 + \underline{b}_{r-2}^\star, y_2\right\} \tag{24}$$

and the unique best r*-term PWL convex upper approximation* $\overline{h}_r : \mathbf{R}^2 \to \mathbf{R}$ *is*

$$\overline{h}_r(y_1, y_2) = \underline{h}_r(y_1, y_2) + \underline{\epsilon}_\phi(r), \tag{25}$$

where $a_i^\star, b_i^\star$*,* $i = 1, \ldots, r-2$ *are the coefficients of the segments of* $\underline{\phi}_r$ *defined in (23).*

The proof is given in Appendix E.2.

This corollary shows that both the best r-term PWL convex upper and lower approximations to the two-term log-sum-exp function can be readily obtained, provided that $\underline{\phi}_r$ is given. Hence we can restrict our attention on solving the best PWL convex lower approximation problem for the univariate function ϕ.

3.2.2 Constructive algorithm

Proposition 1 implies that a function $h \in \mathcal{P}_r^1$ $(r \geq 3)$ with $r-1$ break points $x_1 < \cdots < x_{r-1}$ solves the best PWL convex lower approximation problem for ϕ with approximation error $\epsilon \in (0,\ \log 2)$ (*i.e.*, $h \equiv \underline{\phi}_r$ and $\epsilon = \underline{\epsilon}_\phi(r)$) if and only if

$$h(x) \leq \phi(x), \quad \forall\, x \in \mathbf{R}, \tag{26}$$

$$\lim_{x \to -\infty} h(x) - \phi(x) = 0, \quad \lim_{x \to \infty} h(x) - \phi(x) = 0, \tag{27}$$

$$h(x_i) - \phi(x_i) = \epsilon, \quad i = 2, \ldots, r-2, \tag{28}$$

$$x_1 = \log(e^\epsilon - 1), \tag{29}$$

$$x_{r-1} = -\log(e^\epsilon - 1), \tag{30}$$

and there exist $\tilde{x}_1, \ldots, \tilde{x}_{r-2} \in \mathbf{R}$ such that

$$h(\tilde{x}_i) - \phi(\tilde{x}_i) = 0, \quad i = 1, \ldots, r-2, \tag{31}$$

$$x_1 < \tilde{x}_1 < x_2 < \tilde{x}_2 < \cdots < x_{r-2} < \tilde{x}_{r-2} < x_{r-1}. \tag{32}$$

Using these properties of the best r-term best PWL convex lower approximation, for any given $\epsilon \in (0,\ \log 2)$ and $r \geq 3$, the following algorithm can verify if $\epsilon = \underline{\epsilon}_\phi(r)$ holds.

given $\epsilon \in (0,\ \log 2)$, $r \geq 3$
define $\underline{x}_\epsilon = \log(e^\epsilon - 1)$ and $\overline{x}_\epsilon = -\log(e^\epsilon - 1)$
$k := 1$, $x_1^\epsilon := \underline{x}_\epsilon$
repeat

1. find the line $y = a_k^\epsilon x + b_k^\epsilon$ passing through the point $(x_k^\epsilon, \phi(x_k^\epsilon) - \epsilon)$ and tangent to the curve $y = \phi(x)$ at a point $(\tilde{x}_k^\epsilon, \phi(\tilde{x}_k^\epsilon))$ with $\tilde{x}_k^\epsilon > x_k^\epsilon$
2. find $x_{k+1}^\epsilon > \tilde{x}_k^\epsilon$ such that $a_k^\epsilon x_{k+1}^\epsilon + b_k^\epsilon = \phi(x_{k+1}^\epsilon) - \epsilon$
3. $k := k + 1$

until $k \geq r - 1$

This algorithm is illustrated in Figure 1.

Now, define an r-term PWL convex function $h^\epsilon : \mathbf{R} \to \mathbf{R}$ as

$$h^\epsilon(x) = \max\{0, a_1^\epsilon x + b_1^\epsilon, \ldots, a_{r-2}^\epsilon x + b_{r-2}^\epsilon, x\}.$$

Note that $x_1^\epsilon < \cdots < x_{r-1}^\epsilon$ and h^ϵ satisfy (26–29), and $\tilde{x}_1^\epsilon < \cdots < \tilde{x}_{r-2}^\epsilon$ satisfies (31–32). Thus $h^\epsilon \equiv \underline{\phi}_r$ if and only if (30) holds, which further implies

$$\epsilon = \underline{\epsilon}_\phi(r) \qquad \Longleftrightarrow \qquad x_{r-1}^\epsilon = \overline{x}_\epsilon. \tag{33}$$

Moreover, (30) implies

$$\epsilon < \underline{\epsilon}_\phi(r) \qquad \Longleftrightarrow \qquad x_{r-1}^\epsilon < \overline{x}_\epsilon \tag{34}$$

$$\epsilon > \underline{\epsilon}_\phi(r) \qquad \Longleftrightarrow \qquad x_{r-1}^\epsilon > \overline{x}_\epsilon. \tag{35}$$

Observing (33–35), we can see that the following simple bisection algorithm finds $\underline{\epsilon}_\phi(r)$ and $\underline{\phi}_r$ for any given $r \geq 3$.

given $r \geq 3$ and $\delta > 0$
$\underline{\epsilon} := 0$ and $\overline{\epsilon} := \log 2$
repeat

1. $\epsilon := (\underline{\epsilon} + \overline{\epsilon})/2$
2. find the points $\underline{x}_\epsilon$, $\overline{x}_\epsilon$, the segments $a_k^\epsilon x + b_k^\epsilon$, $k = 1, \ldots, r-1$, and the break points x_k^ϵ, $k = 1, \ldots, r-1$ by the algorithm described above
3. if $x_{r-1}^\epsilon > \overline{x}_\epsilon$, $\overline{\epsilon} := \epsilon$; otherwise, $\underline{\epsilon} := \epsilon$

until $|x_{r-1}^\epsilon - \overline{x}_\epsilon| \leq \delta$
let $\epsilon^\delta = \epsilon$ and define an r-term PWL convex function $\underline{\phi}_r^\delta : \mathbf{R} \to \mathbf{R}$ as

$$\underline{\phi}_r^\delta(x) = \max\{0, a_1^\epsilon x + b_1^\epsilon, \ldots, a_{r-2}^\epsilon x + b_{r-2}^\epsilon, x\}$$

Here, we have

$$\lim_{\delta \to 0} \sup_{x \in \mathbf{R}} |\underline{\phi}_r^\delta(x) - \underline{\phi}_r(x)| = 0, \quad \lim_{\delta \to 0} \epsilon^\delta = \underline{\epsilon}_\phi(r),$$

i.e., $\delta > 0$ controls the tolerance.

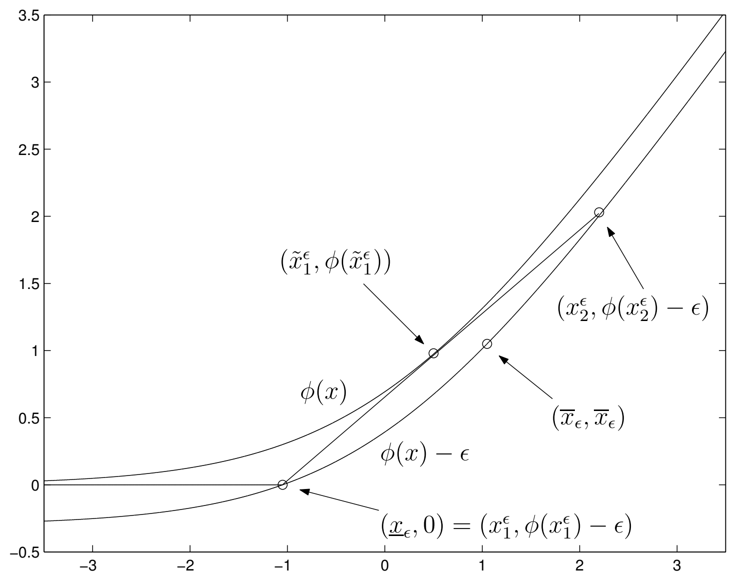

Figure 1: An illustration of the algorithm which checks if $\epsilon = \underline{\epsilon}_\phi(r)$ holds for given $\epsilon \in (0, \ \log 2)$ and $r \geq 3$. In this example we let $\epsilon = 0.3$ and $r = 3$. Since $x_2^\epsilon > \overline{x}_\epsilon$, we can conclude that $\underline{\epsilon}_\phi(3) < 0.3$.

3.2.3 Some approximation results

Table 1 shows the best r-term PWL convex lower approximation to the two-term log-sum-exp function for $r = 2, \ldots, 5$ and the corresponding approximation error $\underline{\epsilon}_\phi(r)$. As will be shown in §4, the approximation method described in §2.2 with the five-term PWL convex lower approximation provides a quite accurate approximate solution for the RGP (6). In practical applications we are usually interested in r in the range $5 \leq r \leq 10$, but we can estimate the error decay rate for large r. Figure 2 shows the optimal error $\underline{\epsilon}_\phi(r)$ for $2 \leq r \leq 1000$. We observe that the curve is almost linear in log-log scale, and using a least-squares fit to the data points $(\log r, \log \underline{\epsilon}_\phi(r))$, $r = 2, \ldots, 1000$, we obtain

$$\log \underline{\epsilon}_\phi(r) \approx -2.0215 \log r + 0.3457.$$

In normal scale,

$$\underline{\epsilon}_\phi(r) \approx \frac{1.4130}{r^{2.0215}} \leq \frac{\sqrt{2}}{r^2}.$$

4 Numerical examples

In the following we use some simple RGP numerical examples to demonstrate the robust LP approximation method described in §2.1. Practical engineering applications, such as power control in lognormal wireless communication channel [HKB05] and robust analog/RF circuit

r	Approximation Error $\underline{\epsilon}_\phi(r)$	Best r-Term PWL Convex Lower Approximation $\underline{\phi}_r$
2	0.693	$\max\{\ y_1,\ \ y_2\ \}$
3	0.223	$\max\{\ y_1,$ $0.500y_1 + 0.500y_2 + 0.693,$ $y_2\ \}$
4	0.109	$\max\{\ y_1,$ $0.271y_1 + 0.729y_2 + 0.584,$ $0.729y_1 + 0.271y_2 + 0.584,$ $y_2\ \}$
5	0.065	$\max\{\ y_1,$ $0.167y_1 + 0.833y_2 + 0.450,$ $0.500y_1 + 0.500y_2 + 0.693,$ $0.833y_1 + 0.167y_2 + 0.450,$ $y_2\ \}$

Table 1: Some best PWL convex lower approximations to the two-term log-sum-exp function.

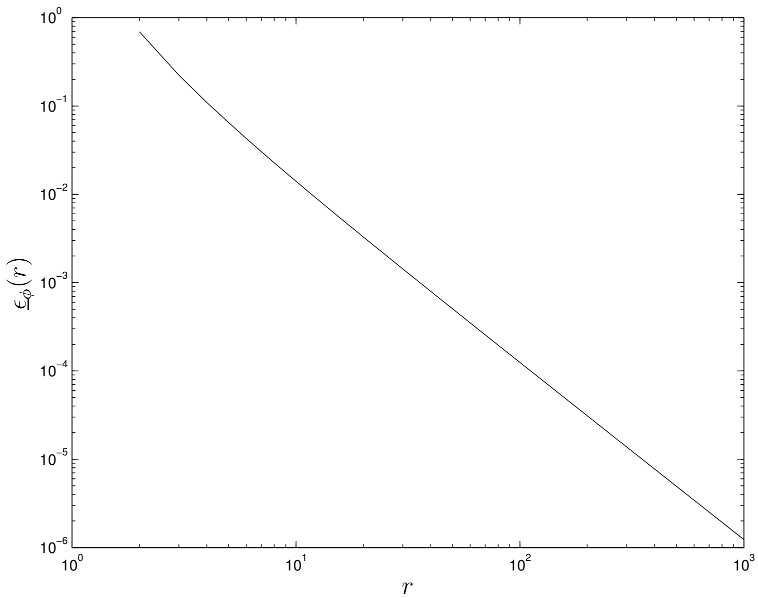

Figure 2: Approximation error $\underline{\epsilon}_\phi(r)$ vs. the degree of PWL approximation r in log-log scale: $r = 2, \ldots, 1000$.

design [YHL+05], have been reported to reveal the effectiveness of the tractable robust GP approximation method proposed in §2.2.

4.1 Two random families

We consider the following RGP, with 500 optimization variables, 500 two-term log-sum-exp inequality constraints, and no equality constraints:

$$R_L: \quad \begin{array}{ll} \text{minimize} & c^T y \\ \text{subject to} & \sup_{u \in \mathcal{U}} \mathbf{lse}\left((a_i^1 + B_i^1 u)^T y, (a_i^2 + B_i^2 u)^T y\right) \le 0, \quad i = 1, \ldots, 500. \end{array} \tag{36}$$

The optimization variable is $y \in \mathbf{R}^{500}$, $u \in \mathbf{R}^L$ represents the uncertain problem data, B_i^1 and B_i^2 are sparse matrices in $\mathbf{R}^{500 \times L}$, and

$$c = \mathbf{1} \in \mathbf{R}^{500}, \quad a_i^1 = a_i^2 = -\mathbf{1} \in \mathbf{R}^{500}.$$

Here, $\mathbf{1}$ is the vector with all entries one. The uncertainty set $\mathcal{U} \subseteq \mathbf{R}^L$ is given by the box in $\mathbf{R}^L$:

$$\mathcal{U} = \left\{ u \in \mathbf{R}^L \middle| \, \|u\|_\infty \le 1 \right\}, \tag{37}$$

where $\|u\|_\infty$ denotes the ℓ_∞-norm of u.

We generated 20 feasible instances, $R_5^1, \ldots, R_5^{20}$, of the RGP (36) with $L = 5$, by randomly generating the sparse matrices $B_i^1, B_i^2 \in \mathbf{R}^{500 \times 5}$, $i = 1, \ldots, 500$ with sparsity density 0.1 and nonzero entries independently uniformly distributed on the interval $[-1, 1]$. The family $\{R_5^1, \ldots, R_5^{20}\}$ is denoted by $\mathcal{F}_5$. With $L = 20$, we also generated a family $\mathcal{F}_{20}$ of 20 feasible instances, $R_{20}^1, \ldots, R_{20}^{20}$, in a similar way.

4.2 Approximation results

Before presenting the approximation results for the two random families $\mathcal{F}_5$ and $\mathcal{F}_{20}$, we describe the error measure associated with the approximation method described in this paper.

Suppose the r-term PWL approximation of the two-term log-sum-exp function is used to obtain approximate solutions of the RGP (36). We call r the *degree of PWL approximation*, and call the solution $\overline{y}_r$ of the robust LP (7) corresponding to the RGP (36) the *r-term upper approximate solution* and the solution $\underline{y}_r$ of the robust LP (8) the *r-term lower approximate solution*. Let $\overline{y}_r$ and $y^\star$ be an r-term upper approximate solution and an exact optimal solution of the RGP (36) respectively. Then, $e^{c^T y^\star}$ is the optimal value of the corresponding RGP in *posynomial* form. To express the difference between $e^{c^T y^\star}$ and $e^{c^T \overline{y}_r}$, we use the fractional difference in percentage α, given by

$$\alpha = 100 \left(\frac{e^{c^T \overline{y}_r}}{e^{c^T y^\star}} - 1 \right) = 100 \left(e^{c^T (\overline{y}_r - y^\star)} - 1 \right).$$

We call the value α the *r-term PWL approximation error (in percentage)* of the RGP (36).

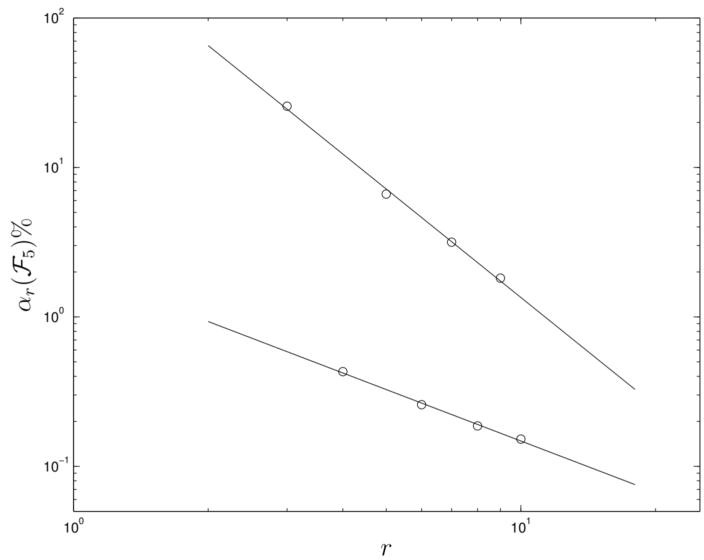

Figure 3: Approximation results for the random family $\mathcal{F}_5$: the degree of PWL approximation r vs. the mean $\alpha_r(\mathcal{F}_5)$ of the r-term PWL approximation errors in log-log scale. The upper solid line is obtained from linear least-squares fitting of the data points $(\log r, \log \alpha_r(\mathcal{F}_5))$, $r = 3, 5, 7, 9$ (shown as circles), while the lower one is obtained from linear least-squares fitting of the data points $(\log r, \log \alpha_r(\mathcal{F}_5))$, $r = 4, 6, 8, 10$.

We first describe the approximation results for $\mathcal{F}_5$. For each $r = 3, \ldots, 10$, we found the r-term upper approximate solutions $\overline{y}_r(1), \ldots, \overline{y}_r(20)$ of the randomly generated instances $R_5^1, \ldots, R_5^{20}$. We also found the exact optimal solutions $y^\star(1), \ldots, y^\star(20)$ of the instances, by solving the equivalent GPs with 16,000 inequality constraints obtained by replicating the inequality constraints for all vertices of the uncertainty box $\mathcal{U}$ in (37).

Figure 3 shows the degree of PWL approximation r vs. the mean $\alpha_r(\mathcal{F}_5)$ of the r-term PWL approximation errors $100(e^{c^T(\overline{y}_r(i) - y^\star(i))} - 1)$, $i = 1, \ldots, 20$, where

$$\alpha_r(\mathcal{F}_5) = \frac{1}{20} \sum_{i=1}^{20} 100 \left(e^{c^T(\overline{y}_r(i) - y^\star(i))} - 1 \right).$$

This figure shows that, in the region of interest, $\alpha_r(\mathcal{F}_5)$ decreases faster than quadratically with increasing r, since $\alpha_r(\mathcal{F}_5)$, $r = 3, 5, 7, 9$ decrease faster than quadratically. The variance of the r-term PWL approximation errors $100(e^{c^T(\overline{y}_r(i) - y^\star(i))} - 1)$, $i = 1, \ldots, 20$ was found to be less than 10^{-6}, regardless of r. The four-term PWL convex upper approximation therefore provides an approximate solution with less than 1% approximation error *quite consistently* for each of the randomly generated instances $R_5^1, \ldots, R_5^{20}$.

Note that $\alpha_r(\mathcal{F}_5)$ does not decrease monotonically with increasing r. This is mainly because it does not necessarily hold that

$$r_1 \geq r_2 \quad \Longrightarrow \quad \overline{h}_{r_2}(y_1, y_2) \geq \overline{h}_{r_1}(y_1, y_2) \geq \mathbf{lse}(y_1, y_2), \quad \forall\, (y_1, y_2) \in \mathbf{R}^2,$$

although

$$r_1 \geq r_2 \quad \Longrightarrow \quad \sup_{(y_1,y_2)\in\mathbf{R}^2} \left(\overline{h}_{r_1}(y_1,y_2) - \mathbf{lse}(y_1,y_2)\right) < \sup_{(y_1,y_2)\in\mathbf{R}^2} \left(\overline{h}_{r_2}(y_1,y_2) - \mathbf{lse}(y_1,y_2)\right),$$

where $\overline{h}_r$ denotes the best r-term PWL convex upper approximation to the two-term log-sum-exp function.

We next describe the approximation results for $\mathcal{F}_{20}$. For each $r = 3,\ldots,10$, we found the r-term upper approximate solutions $\overline{y}_r(1),\ldots,\overline{y}_r(20)$ of the randomly generated instances $R_{20}^1,\ldots,R_{20}^{20}$. Replicating the inequality constraints for all the vertices was not possible for the random family $\mathcal{F}_{20}$, since the corresponding uncertainty box $\mathcal{U}$ has approximately 10^6 vertices. Thus, it is too expensive to find the optimal solutions $y^\star(1),\ldots,y^\star(20)$ of the instances $R_{20}^1,\ldots,R_{20}^{20}$. Instead, we found the r-term lower approximate solutions $\underline{y}_r(1),\ldots,\underline{y}_r(20)$ of the instances $R_{20}^1,\ldots,R_{20}^{20}$ for each $r = 3,\ldots,10$.

Note from (11) that

$$0 \leq e^{c^T(\overline{y}_r(i)-y^\star(i))} - 1 \leq e^{c^T(\overline{y}_r(i)-\underline{y}_r(i))} - 1, \quad i = 1,\ldots,20.$$

The mean $\overline{\alpha}_r(\mathcal{F}_{20})$ of the r-term approximation errors $e^{c^T(\overline{y}_r(i)-\underline{y}_r(i))}-1,\ i = 1,\ldots,20$ is therefore an upper bound on the mean $\alpha_r(\mathcal{F}_{20})$ of the r-term approximation errors $e^{c^T(\overline{y}_r(i)-y^\star(i))} - 1,\ i = 1,\ldots,20$:

$$\overline{\alpha}_r(\mathcal{F}_{20}) = \frac{1}{20}\sum_{i=1}^{20} 100\left(e^{c^T(\overline{y}_r(i)-\underline{y}_r(i))} - 1\right) \geq \alpha_r(\mathcal{F}_{20}) = \frac{1}{20}\sum_{i=1}^{20} 100\left(e^{c^T(\overline{y}_r(i)-y^\star(i))} - 1\right).$$

Figure 4 shows the degree of PWL approximation r vs. $\overline{\alpha}_r(\mathcal{F}_{20})$. This figure shows that, in the region of interest, $\overline{\alpha}_r(\mathcal{F}_{20})$ decreases faster than quadratically with increasing r. The variance of the upper bounds $100(e^{c^T(\overline{y}_r(i)-\underline{y}_r(i))} - 1),\ i = 1,\ldots,20$ was found to be less than 10^{-4}, regardless of r. The seven-term PWL convex upper approximation therefore provides an approximate solution with less than 5% approximation error consistently for each of the instances $R_{20}^1,\ldots,R_{20}^{20}$.

5 Conclusions

We have described an approximation method for a RGP with polyhedral or ellipsoidal uncertainty. The approximation method is based on conservatively approximating the original RGP (6) with a more tractable robust two-term GP in which every nonlinear function in the constraints is bivariate. The idea can be extended to a (small) k-term RGP approximation in which every nonlinear function in the constraints has at most k exponential terms. The extension relies on accurate PWL approximations of k-term log-sum-exp functions. We are currently working on the extension using the heuristic for PWL approximation of convex functions developed in [MB05].

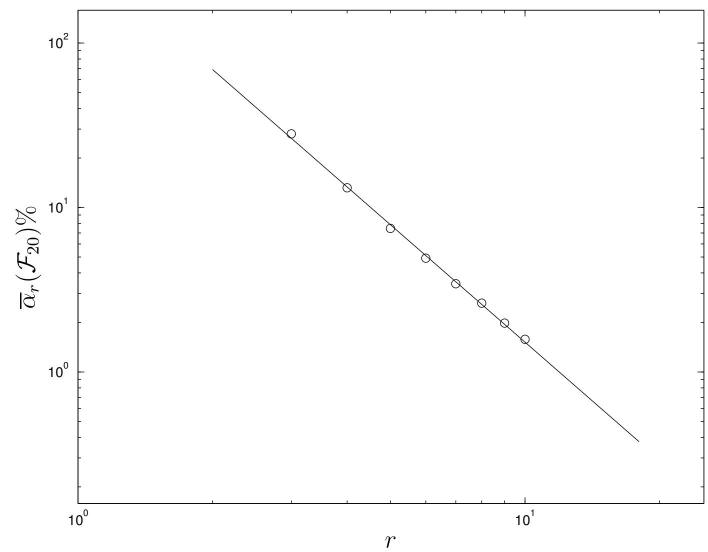

Figure 4: Approximation results for the random family $\mathcal{F}_{20}$: the degree of PWL approximation r vs. the upper bound $\overline{\alpha}_r(\mathcal{F}_{20})$ on the mean $\alpha_r(\mathcal{F}_{20})$ of the r-term PWL approximation errors in log-log scale. The solid line is obtained from linear least-squares fitting of the data points $(\log r, \log \overline{\alpha}_r(\mathcal{F}_{20}))$, $r = 3, 4, \ldots, 10$, shown as circles.

Acknowledgments

This material is based upon work supported by the National Science Foundation under grants #0423905 and (through October 2005) #0140700, by the Air Force Office of Scientific Research under grant #F49620-01-1-0365, by MARCO Focus center for Circuit & System Solutions contract #2003-CT-888, by MIT DARPA contract #N00014-05-1-0700, and by the Post-doctoral Fellowship Program of Korea Science and Engineering Foundation (KOSEF). The authors thank Alessandro Magnani for helpful comments and suggestions.

References

[ADP75] M. Avriel, R. Dembo, and U. Passy. Solution of generalized geometric programs. *International Journal for Numerical Methods in Engineering*, 9:149–168, 1975.

[Avr80] M. Avriel, editor. *Advances in Geometric Programming*, volume 21 of *Mathematical Concept and Methods in Science and Engineering*. Plenum Press, New York, 1980.

[Bea81] R. Beatson. Convex approximation by splines. *SIAM Journal on Mathematical Analysis*, 12:549–559, 1981.

[Bea82] R. Beatson. Monotone and convex approximation by splines: Error estimates and a curve fitting algorithm. *SIAM Journal on Numerical Analysis*, 19:1278–1285, 1982.

[BKPH05] S. Boyd, S.-J. Kim, D. Patil, and M. Horowitz. Digital circuit optimization via geometric programming. To appear in *Operations Research*, 2005. Available at `http://www.stanford.edu/~boyd/gp_digital_ckt.html`.

[BKVH05] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi. A tutorial on geometric programming. To appear in *Optimization and Engineering*, 2005. Available at `http://www.stanford.edu/~boyd/gp_tutorial.html`.

[BTN98] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[BTN99] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25:1–13, 1999.

[BTN01] A. Ben-Tal and A. Nemirovski. On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205, 2001.

[BTNR02] A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560, 2002.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[CB04] M. Chiang and S. Boyd. Geometric programming duals of channel capacity and rate distortion. *IEEE Transactions on Information Theorey*, 50(2):245–258, 2004.

[CHP00] W. Chen, C.-T. Hsieh, and M. Pedram. Simultaneous gate sizing and placement. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 19(2):206–214, 2000.

[DBHL01] J. Dawson, S. Boyd, M. Hershenson, and T. Lee. Optimal allocation of local feedback in multistage amplifiers via geometric programming. *IEEE Transactions on Circuits and Systems I*, 48(1):1–11, January 2001.

[DGS03] W. Daems, G. Gielen, and W. Sansen. Simulation-based generation of posynomial performance models for the sizing of analog integrated circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 22(5):517–534, May 2003.

[DPZ67] R. Duffin, E. Peterson, and C. Zener. *Geometric Programming–Theory and Application.* Wiley, New York, 1967.

[DR92] A. Dutta and D. Rama. An optimization model of communications satellite planning. *IEEE Transactions on Communications*, 40(9):1463–1473, 1992.

[Eck80] J. Ecker. Geometric programming: Methods, computations and applications. *SIAM Review*, 22(3):338–362, 1980.

[FM88] B. Feijoo and P. Meyer. Piecewise-linear approximation methods for nonseparable convex optimization. *Management Science*, 34(3):411–419, 1988.

[GI03] D. Goldfarb and G. Iyengar. Robust convex quadratically constrained programming. *Mathematical Programming*, 97(3):495–515, 2003.

[GL97] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

[GL98] L. El Ghaoui and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1):33–52, 1998.

[Gli00] F. Glineur. Polyhedral approximation of the second-order cone: Computational experiments. IMAGE Technical Report 0001, Faculté Polytechnique de Mons, Mons, Belgium, 2000.

[GNP95] B. Gao, D. Newman, and V. Popov. Convex approximation by rational functions. *SIAM Journal on Mathematical Analysis*, 26(2):488–499, 1995.

[Gre95] H. Greenberg. Mathematical programming models for environmental quality control. *Operations Research*, 43(4):578–622, 1995.

[HBL01] M. Hershenson, S. Boyd, and T. H. Lee. Optimal design of a CMOS op-amp via geometric programming. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 20(1):1–21, January 2001.

[Her02] M. Hershenson. Design of pipeline analog-to-digital converters via geometric programming. In *Proceedings of the IEEE/ACM International Conference on Computer Aided Design*, pages 317–324, San Jose, CA, November 2002.

[HKB05] K.-L. Hsiung, S.-J. Kim, and S. Boyd. Power control in lognormal fading wireless channels with uptime probability specifications via robust geometric programming. In *Proc. 2005 IEEE American Control Conference*, volume 6, pages 3955–3959, Portland, Oregon, June 2005.

[Hu91] H. Hu. Convexity preserving approximation by free knot splines. *SIAM Journal on Mathematical Analysis*, 22:1183–1191, 1991.

[KB02] S. Kandukuri and S. Boyd. Optimal power control in interference-limited fading wireless channels with outage-probability specifications. *IEEE Transactions on Wireless Communications*, 1(1):46–55, January 2002.

[KC97] J. Karlof and Y. Chang. Optimal permutation codes for the Gaussian channel. *IEEE Transactions on Information Theory*, 43(1):356–358, 1997.

[KXY97] K. Kortanek, X. Xu, and Y. Ye. An infeasible interior-point algorithm for solving primal and dual geometric progams. *Mathematical Programming*, 76(1):155–181, January 1997.

[LVBL98] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and Applications*, 284(1-3):193–228, 1998.

[MB05] A. Magnani and S. Boyd. Convex piecewise linear fitting. Manuscript, 2005. Available at `http://www.stanford.edu/~boyd/cvx_pwl_fitting.html`.

[MHBL00] S. Mohan, M. Hershenson, S. Boyd, and T. Lee. Bandwidth extension in CMOS with optimized on-chip inductors. *IEEE Journal of Solid-State Circuits*, 35(3):346–355, March 2000.

[MJ83] M. Mazumdar and T. Jefferson. Maximum likelihood estimates for multinomial probabilities via geometric programming. *Biometrika*, 70(1):257–261, 1983.

[MR78] D. McAllister and J. Roullier. Interpolation by convex quadratic splines. *Mathematics of Computation*, 32:1154–1162, 1978.

[NN94] Y. Nesterov and A. Nemirovsky. *Interior-Point Polynomial Methods in Convex Programming*, volume 13 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.

[OJB03] D. O'Neill, D. Julian, and S. Boyd. Seeking Foschini's genie: Optimal rates and powers in wireless networks. To appear in *IEEE Transactions on Vehicular Technology*, April 2003. Available at `http://www.stanford.edu/~boyd/foschini_genie.html`.

[Pet76] E. Peterson. Geometric programming. *SIAM Review*, 18(1):1–51, 1976.

[RB90] J. Rajpogal and D. Bricker. Posynomial geometric programming as a special case of semi-infinite linear programming. *Journal of Optimization Theory and Applications*, 66:444–475, 1990.

[Sap96] S. Sapatnekar. Wire sizing as a convex optimization problem: Exploring the area-delay tradeoff. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 15:1001–1011, 1996.

[SNLS05] J. Singh, V. Nookala, Z.-Q. Luo, and S. S. Sapatnekar. Robust gate sizing by geometric programming. In *Proc. 42nd ACM/IEEE Design Automation Conference*, pages 315–320, Anaheim, CA, June 2005.

[SRVK93] S. Sapatnekar, V. Rao, P. Vaidya, and S. Kang. An exact solution to the transistor sizing problem for CMOS circuits using convex optimization. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 12(11):1621–1634, 1993.

[Tha78] L. Thakur. Error analysis for convex separable programs: The piecewise linear approximation and the bounds on the optimal objective value. *SIAM Journal on Applied Mathematics*, 34(4):704–714, 1978.

[WGW86] T. Wall, D. Greening, and R. Woolsey. Solving complex chemical equilibria using a geometric-programming based technique. *Operations Research*, 34(3):345–355, 1986.

[YCLW01] F. Young, C. Chu, W. Luk, and Y. Wong. Handling soft modules in general nonslicing floorplan using lagrangian relaxation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 20(5):687–629, 2001.

[YHL+05] X. Yang, K.-L. Hsiung, X. Li, I. Nausieda, S. Boyd, and L. Pileggi. OPERA: Optimization with elipsoidal uncertainty for robust analog IC design. In *Proc. 42nd IEEE/ACM Design Automation Conference*, pages 632–637, Anaheim, CA, June 2005.

[Zen71] C. Zener. *Engineering Design by Geometric Programming*. Wiley, New York, 1971.

A Convex formulation of GP

Let $\mathbf{R}^n_{++}$ denote the set of real n-vectors whose components are positive. Let $x_1, \ldots, x_n$ be n real positive variables. A function $f : \mathbf{R}^n_{++} \to \mathbf{R}$, defined as

$$f(x) = d \prod_{j=1}^{n} x_j^{a_j}, \tag{38}$$

where $d \geq 0$ and $a_j \in \mathbf{R}$, is called a *monomial*. A sum of monomials, *i.e.*, a function of the form

$$f(x) = \sum_{k=1}^{K} d_k \prod_{j=1}^{n} x_j^{a_{jk}}, \tag{39}$$

where $d_k \geq 0$ and $a_{jk} \in \mathbf{R}$, is called a *posynomial* (with K terms).

An optimization problem of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 1, \quad i = 1, \ldots, m, \\ & h_i(x) = 1, \quad i = 1, \ldots, l, \end{array} \tag{40}$$

where $f_0, \ldots, f_m$ are posynomials and $h_1, \ldots, h_p$ are monomials, is called a *geometric program in posynomial form*. Here, the constraints $x_i > 0$, $i = 1, \ldots, n$ are implicit. The corresponding robust convex optimization problem is called a *RGP in posynomial form*.

We assume without loss of generality that the objective function f_0 is a monomial whose coefficient is one:

$$f_0(x) = \prod_{j=1}^{n} x_j^{c_j}.$$

If f_0 is not a monomial, we can equivalently reformulate the GP (40) as the following GP whose objective function is a monomial:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_0(x) t^{-1} \leq 1, \\ & f_i(x) \leq 1, \quad i = 1, \ldots, m, \\ & h_i(x) = 1, \quad i = 1, \ldots, l, \end{array}$$

where $(x, t) \in \mathbf{R}^n_{++} \times \mathbf{R}_{++}$ are the optimization variables.

GPs in posynomial form are not (in general) convex optimization problems, but they can be reformulated as convex problems by a change of variables and a transformation of the objective and constraint functions. To show this, we define new variables $y_i = \log x_i$, and take the logarithm of the posynomial f of x given in (39) to get

$$\tilde{f}(y) = \log(f(e^{y_1}, \ldots, e^{y_n})) = \log\left(\sum_{i=1}^{K} e^{a_k^T y + b_k}\right) = \mathbf{lse}(a_1^T y + b_1, \ldots, a_K^T y + b_K),$$

where $a_k = (a_{1k}, \ldots, a_{nk}) \in \mathbf{R}^n$ and $b_k = \log d_k$, *i.e.*, a posynomial becomes a sum of exponentials of affine functions after the change of variables. (Note that if the posynomial f

is a monomial, then the transformed function $\tilde{f}$ is an affine function.) This converts the original GP (40) into a GP:

$$\begin{array}{ll} \text{minimize} & c^T y \\ \text{subject to} & \mathbf{lse}(a_{i1}^T y + b_{i1}, \ldots, a_{iK_i}^T y + b_{iK_i}) \le 0, \quad i = 1, \ldots, m, \\ & g_i^T y + h_i = 0, \quad i = 1, \ldots, l, \end{array} \tag{41}$$

where $a_{ij} \in \mathbf{R}^n, \ i = 1, \ldots, m, \ j = 1, \ldots, K_i$ contain the exponents of the posynomial inequality constraints, $c \in \mathbf{R}^n$ contains the exponents of the monomial objective function of the original GP, and $g_i \in \mathbf{R}^n, \ i = 1, \ldots, l$ contain the exponents of the monomial equality constraints of the original GP.

B Robust linear programming

Consider the robust LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \sup_{u \in \mathcal{U}} (\bar{a}_i + B_i u)^T x + b_i \le 0, \quad i = 1, \ldots, m, \end{array} \tag{42}$$

where the optimization variable is $x \in \mathbf{R}^n$, $u \in \mathbf{R}^L$ represents the uncertain problem data, the set $\mathcal{U} \subseteq \mathbf{R}^L$ describes the uncertainty in u, and $c \in \mathbf{R}^n$, $\bar{a}_i \in \mathbf{R}^n$, $B_i \in \mathbf{R}^{n \times L}$, $b \in \mathbf{R}^m$. When the uncertainty set $\mathcal{U}$ is given by a bounded polyhedron or an ellipsoid, the robust LP (42) can be cast as a standard convex optimization problem, as shown below.

B.1 Polyhedral uncertainty

Let the uncertainty set $\mathcal{U}$ be a polyhedron:

$$\mathcal{U} = \left\{ u \in \mathbf{R}^L \middle| \, Du \preceq d \right\},$$

where $D \in \mathbf{R}^{K \times L}$ and $d \in \mathbf{R}^K$. We assume that $\mathcal{U}$ is non-empty and bounded. Using the duality theorem for linear programming, we can equivalently reformulate the robust LP (42) as the following LP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & D^T z_i = B_i^T x, \quad i = 1, \ldots, m, \\ & \bar{a}_i^T x + d^T z_i + b_i \le 0, \quad i = 1, \ldots, m, \\ & z_i \ge 0, \quad i = 1, \ldots, m, \end{array} \tag{43}$$

where the optimization variables are $(x, z_1, \ldots, z_m) \in \mathbf{R}^n \times \mathbf{R}^K \times \cdots \times \mathbf{R}^K$.

B.2 Ellipsoidal uncertainty

Without loss of generality, we assume that the uncertainty set $\mathcal{U}$ is a unit ball:

$$\mathcal{U} = \left\{ u \in \mathbf{R}^L \middle| \|u\|_2 \leq 1 \right\}.$$

Then, the robust LP (42) can be cast as the second-order cone program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \overline{a}_i^T x + \|B_i^T x\|_2 + b_i \leq 0, \quad i = 1, \ldots, m. \end{array}$$

See, *e.g.*, [LVBL98] for details.

C Reformulation of the robust GP

We start with reformulating the RGP (6) as the equivalent RGP

$$\begin{array}{ll} \text{minimize} & \bar{c}^T \begin{bmatrix} y \\ t \end{bmatrix} \\ \text{subject to} & \sup_{u \in \mathcal{U}} \mathbf{lse} \left(\left(\tilde{A}_i^0 + \sum_{j=1}^L u_j \tilde{A}_i^j \right) \begin{bmatrix} y \\ t \end{bmatrix} \right) \leq 0, \quad i = 1, \ldots, m, \\ & \bar{G} \begin{bmatrix} y \\ t \end{bmatrix} + \bar{h} = 0, \end{array} \tag{44}$$

where $(y, t) \in \mathbf{R}^n \times \mathbf{R}$ are the optimization variables, and the problem data are

$$\bar{c} = (c, 0) \in \mathbf{R}^{n+1}, \quad \bar{G} = \begin{bmatrix} G & 0 \\ 0 & 1 \end{bmatrix} \in \mathbf{R}^{(l+1)\times(n+1)}, \quad \bar{h} = \begin{bmatrix} h \\ -1 \end{bmatrix} \in \mathbf{R}^{l+1},$$
$$\tilde{A}_i^j = \begin{bmatrix} A_i^j & b_i^j \end{bmatrix} \in \mathbf{R}^{K_i \times (n+1)}, \quad i = 1, \ldots, m, \quad j = 0, 1, \ldots, L.$$

Denote the sth row of $\tilde{A}_i^j$ as $\tilde{a}_{is}^{jT}$, $s = 1, \ldots, K_i$, *i.e.*,

$$\tilde{A}_i^j = \begin{bmatrix} \tilde{a}_{i1}^{jT} \\ \vdots \\ \tilde{a}_{iK_i}^{jT} \end{bmatrix} \in \mathbf{R}^{K_i \times (n+1)}, \quad i = 1, \ldots, m, \quad j = 0, 1, \ldots, L.$$

Then the RGP (44) can be readily rewritten as the equivalent RGP (12) with the optimization variables $\eta = (y, t) \in \mathbf{R}^n \times \mathbf{R}$ and the problem data

$$\bar{a}_{is} = \tilde{a}_{is}^0 \in \mathbf{R}^{n+1}, \quad \bar{B}_{is} = \begin{bmatrix} \tilde{a}_{is}^1 & \cdots & \tilde{a}_{is}^L \end{bmatrix} \in \mathbf{R}^{(n+1)\times L}, \quad s = 1, \ldots, K_i, \quad i = 1, \ldots, m.$$

D Details of the two-term robust GP approximation

Consider a k-term log-sum-exp constraint:

$$\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_1+B_1u)^T\eta,\ldots,(a_k+B_ku)^T\eta\Big) \le 0,$$

where $a_i \in \mathbf{R}^{n+1}$, $B_i \in \mathbf{R}^{(n+1)\times L}$. It is easy to see that

$$\begin{aligned}
&\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_1+B_1u)^T\eta,\ldots,(a_k+B_ku)^T\eta\Big)\\
&\quad= \sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_1+B_1u)^T\eta, \mathbf{lse}\Big((a_2+B_2u)^T\eta,\ldots,(a_k+B_ku)^T\Big)\Big)\\
&\quad\le \sup_{u\in\mathcal{U}} \mathbf{lse}\left((a_1+B_1u)^T\eta, \sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_2+B_2u)^T\eta,\ldots,(a_k+B_ku)^T\eta\Big)\right).
\end{aligned}$$

Therefore a sufficient condition for the k-term robust log-sum-exp constraint (13) is that there exists $z_1 \in \mathbf{R}$ such that

$$\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_1+B_1u)^T\eta, z_1\Big) \le 0, \quad \sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_2+B_2u)^T\eta,\ldots,(a_k+B_ku)^T\eta\Big) \le z_1. \tag{45}$$

Similarly, since

$$\begin{aligned}
&\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_2+B_2u)^T\eta,\ldots,(a_k+B_ku)^T\eta\Big)\\
&\quad\le \sup_{u\in\mathcal{U}} \mathbf{lse}\left((a_2+B_2u)^T\eta, \sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_3+B_3u)^T\eta,\ldots,(a_k+B_ku)^T\eta\Big)\right),
\end{aligned}$$

a sufficient condition for (45) is that there exist z_1, $z_2 \in \mathbf{R}$ such that

$$\begin{aligned}
\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_1+B_1u)^T\eta, z_1\Big) &\le 0,\\
\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_2+B_2u)^T\eta, z_2\Big) &\le z_1,\\
\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_3+B_3u)^T\eta,\ldots,(a_k+B_ku)^T\eta\Big) &\le z_2.
\end{aligned}$$

Now it is clear that η satisfies (13) if there exists $z = (z_1,\ldots,z_{k-2}) \in \mathbf{R}^{k-2}$ such that (η, z) satisfies the system of $k-1$ two-term robust log-sum-exp constraints:

$$\begin{aligned}
\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_1+B_1u)^T\eta, z_1\Big) &\le 0,\\
\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_{s+1}+B_{s+1}u)^T\eta, z_{s+1}\Big) &\le z_s, \quad s=1,\ldots,k-3,\\
\sup_{u\in\mathcal{U}} \mathbf{lse}\Big((a_{k-1}+B_{k-1}u)^T\eta, (a_k+B_ku)^T\eta\Big) &\le z_{k-2},
\end{aligned}$$

which is obviously equivalent to (14).

E Proofs

E.1 Proof of Proposition 1

We first establish the existence of an optimal solution for the optimization problem

$$\begin{array}{ll} \text{minimize} & \sup_{x\in\mathbf{R}} \left(\phi(x) - \max\{0, a_1x + b_1, \dots, a_{r-2}x + b_{r-2}, x\}\right) \\ \text{subject to} & \phi(x) \geq \max\{0, a_1x + b_1, \dots, a_{r-2}x + b_{r-2}, x\}, \quad \forall\, x \in \mathbf{R} \end{array} \tag{46}$$

with the optimization variable $(a, b) \in \mathbf{R}^{r-2} \times \mathbf{R}^{r-2}$.

Lemma 1. *The optimization problem (46) has a solution, say* $(a^\star, b^\star) \in \mathbf{R}^{r-2} \times \mathbf{R}^{r-2}$*, which satisfies*

$$0 < a_1^\star < \cdots < a_{r-2}^\star < 1, \tag{47}$$

such that the function $h^\star : \mathbf{R} \to \mathbf{R}$*, defined by*

$$h^\star(x) = \max\{0, a_1^\star x + b_1^\star, \dots, a_{r-2}^\star x + b_{r-2}^\star, x\},$$

is r*-term PWL convex,* i.e.*, it can be written as*

$$h^\star(x) = \begin{cases} 0, & x \in (-\infty,\ x_1], \\ a_i^\star x + b_i^\star, & x \in [x_i,\ x_{i+1}], \quad i \in \{1, \dots, r-2\}, \\ x, & x \in [x_{r-1},\ \infty) \end{cases} \tag{48}$$

for some $x_1 < \cdots < x_{r-1}$*.*

Proof. Obviously the feasible set of (46) is nonempty. Let $(a, b) \in \mathbf{R}^{r-2} \times \mathbf{R}^{r-2}$ be a feasible solution and define the continuous function $h : \mathbf{R} \to \mathbf{R}$ as

$$h(x) = \max\{0, a_1x + b_1, \dots, a_{r-2}x + b_{r-2}, x\},$$

which satisfies $h(x) \leq \phi(x)$ for all $x \in \mathbf{R}$. The derivative of ϕ, $\phi'(x) = e^x/(1 + e^x)$, satisfies $0 < \phi'(x) < 1$ for all $x \in \mathbf{R}$. Note from (17) that if $a_j > 1$ for some j, $\lim_{x\to\infty}(\phi(x) - h(x)) = -\infty$. Then it follows from the continuity of $\phi(x) - h(x)$ over $\mathbf{R}$ that there exists $\hat{x} \in \mathbf{R}$ such that $h(\hat{x}) > \phi(\hat{x})$. Similarly, if $a_j < 0$ for some j, $\lim_{x\to-\infty}(\phi(x) - h(x)) = -\infty$, which also implies $h(\bar{x}) > \phi(\bar{x})$ for some $\bar{x} \in \mathbf{R}$. Hence,

$$0 \leq a_i \leq 1, \quad i = 1, \dots, r-2. \tag{49}$$

Since $\phi(0) = \log 2$, we also have $b_i \leq \log 2,\ i = 1, \dots, r-2$. It is also obvious from (17) and (49) that if $b_i < 0$, then $a_ix + b_i < \max\{0, x\}$, for all $x \in \mathbf{R}$ and hence

$$h(x) = \max\{0, a_1x + b_1, \dots, a_{i-1}x + b_{i-1}, a_{i+1}x + b_{i+1}, \dots, a_{r-2}x + b_{r-2}, x\}, \quad \forall\, x \in \mathbf{R}.$$

Thus far, we have seen that (46) is equivalent to

$$\begin{array}{ll} \text{minimize} & \sup_{x\in\mathbf{R}} \left(\phi(x) - \max\{0, a_1x + b_1, \dots, a_{r-2}x + b_{r-2}, x\}\right) \\ \text{subject to} & \phi(x) \geq \max\{0, a_1x + b_1, \dots, a_{r-2}x + b_{r-2}, x\}, \quad \forall\, x \in \mathbf{R}, \\ & 0 \leq a_i \leq 1, \quad i = 1, \dots, r-2, \\ & 0 \leq b_i \leq \log 2, \quad i = 1, \dots, r-2. \end{array} \tag{50}$$

Denote the feasible set of the above optimization problem as $\mathcal{F}$. Notice that the objective function of (50) is continuous over $\mathcal{F}$, which is nonempty and compact. Thus (50) has at least one optimal solution, say $(a^\star, b^\star) \in \mathbf{R}^{r-2} \times \mathbf{R}^{r-2}$.

We can assume without loss of generality that $a_i^\star$ are in increasing order:

$$0 < a_1^\star \leq \cdots \leq a_{r-2}^\star < 1.$$

Suppose that $a_i^\star = \cdots = a_{i+s}^\star < a_{i+s+1}^\star$. Then we can always replace the segment $a_{i+s}^\star x + b_{i+s}^\star$ with a new affine function $\bar{a}_{i+1}x + \bar{b}_{i+1}$ such that

$$a_i^\star = \cdots = a_{i+s-1}^\star < \bar{a}_{i+s} < a_{i+2}^\star, \qquad \sup_{x \in \mathbf{R}} \left(\phi(x) - \bar{h}(x) \right) = \underline{\epsilon}_\phi(r),$$

and $\bar{h}(x) = \bar{a}_{i+s}x + \bar{b}_{i+s}$ on some interval, where $\bar{h}$ is the PWL function

$$\bar{h}(x) = \max\{0, a_1^\star x + b_1^\star, \ldots, a_{i+s-1}^\star x + b_{i+s-1}^\star, \bar{a}_{i+s}x + \bar{b}_{i+s}, a_{i+s+1}^\star x + b_{i+s+1}^\star, \ldots, x\}.$$

Repeating the arguments above, we can see that if an optimal solution does not satisfy (47), then we can always find a new optimal solution which satisfies (47). $\square$

To proceed, we need the following technical lemma which implies that the maximum error between the function ϕ and an affine function $cx + d$ on an interval can arise only at its endpoints.

Lemma 2. *Suppose that, on an interval $(x_1,\ x_2)$,*

$$cx + d \leq \phi(x), \quad \forall\, x \in [x_1,\ x_2].$$

Then,

$$\max_{x \in (x_1,\ x_2)} (\phi(x) - cx - d) <= \max\{\phi(x_1) - cx_1 - d, \phi(x_2) - cx_2 - d\}.$$

Proof. The function

$$\psi(x) = \phi(x) - cx - d$$

is convex and positive on $[x_1,\ x_2]$. The claim of this lemma directly follows from the convexity and positivity of this function. $\square$

We also need the following three technical lemmas.

Lemma 3. *Suppose that, for some $2 \leq j \leq r-1$, the function $h^\star$ defined in (48) satisfies*

$$\phi(x_j) - h^\star(x_j) < \max\{\phi(x_{j-1}) - h^\star(x_{j-1}), \phi(x_{j+1}) - h^\star(x_{j+1})\}. \tag{51}$$

Then, there exist $\bar{c}_{j-1},\ \bar{c}_j \in \mathbf{R}$ with $c_{j-2}^\star < \bar{c}_{j-1} < c_{j-1}^\star < c_j^\star < \bar{c}_j < c_{j+1}^\star$ and $\bar{d}_{j-1}, \bar{d}_j \in \mathbf{R}$ such that

$$\begin{aligned}
\bar{c}_j x + \bar{d}_j \leq \bar{c}_{j-1}x + \bar{d}_{j-1} \leq \phi(x), &\quad \forall\, x \in [x_{j-1},\ x_j], \\
\bar{c}_{j-1}x + \bar{d}_{j-1} \leq \bar{c}_j x + \bar{d}_j \leq \phi(x), &\quad \forall\, x \in [x_j,\ x_{j-1}], \\
c_{j-1}^\star x_{j-1} + d_{j-1}^\star < \bar{c}_{j-1}x_{j-1} + \bar{d}_{j-1} < \phi(x_{j-1}),& \\
c_j^\star x_{j+1} + d_j^\star < \bar{c}_j x_{j+1} + \bar{d}_j < \phi(x_{j+1}),&
\end{aligned}$$

and

$$\max_{x\in[x_{j-1},\ x_{j+1}]}(\phi(x)-\max\{\bar{c}_{j-1}x+\bar{d}_{j-1},\bar{c}_jx+\bar{d}_j\}) < \max_{x\in[x_{j-1},\ x_{j+1}]}(\phi(x)-\max\{c^\star_{j-1}x+d^\star_{j-1},c^\star_jx+d^\star_j\}).$$

Proof. First, note from Lemma 2 that

$$\begin{aligned}\max_{x\in[x_{j-1},\ x_j]}(\phi(x)-h^\star(x)) &= \max_{x\in[x_{j-1},\ x_j]}(\phi(x)-c^\star_{j-1}x-d^\star_{j-1})\\ &= \max\{\phi(x_{j-1})-c^\star_{j-1}x_{j-1}-d^\star_{j-1},\phi(x_j)-c^\star_jx_j-d^\star_{j-1}\}\end{aligned}$$

and

$$\begin{aligned}\max_{x\in[x_j,\ x_{j+1}]}(\phi(x)-h^\star(x)) &= \max_{x\in[x_j,\ x_{j+1}]}(\phi(x)-c^\star_jx-d^\star_j)\\ &= \max\{\phi(x_j)-c^\star_jx_j-d^\star_j,\phi(x_{j+1})-c^\star_jx_{j+1}-d^\star_j\}.\end{aligned}$$

Then, we can see from (51) that

$$\begin{aligned}&\max_{x\in[x_{j-1},\ x_{j+1}]}(\phi(x)-h^\star(x))\\ &= \max_{x\in[x_{j-1},\ x_{j+1}]}(\phi(x)-\max\{c^\star_{j-1}x+d^\star_{j-1},c^\star_jx+d^\star_j\})\\ &= \max\{\phi(x_{j-1})-c^\star_{j-1}x_{j-1}-d^\star_{j-1},\phi(x_{j+1})-c^\star_jx_{j+1}-d^\star_j\}.\end{aligned} \tag{52}$$

Define $\hat{y}_k(\eta),\ k=j-1,j,j+1$ as

$$\begin{aligned}\hat{y}_{j-1}(\eta) &= c^\star_{j-1}x_{j-1}+d^\star_{j-1}+\eta,\\ \hat{y}_j(\eta) &= c^\star_{j-1}x_j+d^\star_{j-1}-\eta,\\ \hat{y}_3(\eta) &= c^\star_jx_{j+1}+d^\star_j+\eta.\end{aligned}$$

Let $\hat{c}_{j-1}(\eta)x+\hat{d}_{j-1}(\eta)$ (respectively, $\hat{c}_j(\eta)x+\hat{d}_j(\eta)$) denote the line passing the two points $(x_1,\hat{y}_1(\eta))$ and $(x_2,\hat{y}_2(\eta))$ (respectively, $(x_2,\hat{y}_2(\eta))$ and $(x_3,\hat{y}_3(\eta))$). Then, for sufficiently small $\eta>0$,

$$c^\star_{j-2}<\hat{c}_{j-1}(\eta)<c^\star_{j-1}<c^\star_j<\hat{c}_j(\eta)<c^\star_{j+1}$$

such that

$$c^\star_{j-1}x_{j-1}+d^\star_{j-1}<\hat{y}_{j-1}(\eta)=\hat{c}_{j-1}(\eta)x_{j-1}+\hat{d}_{j-1}(\eta)<\phi(x_{j-1}), \tag{53}$$

$$c^\star_jx_{j+1}+d^\star_j<\hat{y}_{j+1}(\eta)=\hat{c}_j(\eta)x_{j+1}+\hat{d}_j(\eta)<\phi(x_{j+1}), \tag{54}$$

$$\phi(x_j)-\hat{y}_j(\eta)<\max\{\phi(x_{j-1})-\hat{y}_{j-1}(\eta),\phi(x_{j+1})-\hat{y}_{j+1}(\eta)\} \tag{55}$$

and

$$\begin{aligned}\hat{c}_j(\eta)x+\hat{d}_j(\eta)\le\hat{c}_{j-1}(\eta)x+\hat{d}_{j-1}(\eta)\le\phi(x), &\quad \forall\, x\in[x_{j-1},\ x_j],\\ \hat{c}_{j-1}(\eta)x+\hat{d}_{j-1}(\eta)\le\hat{c}_j(\eta)x+\hat{d}_j(\eta)\le\phi(x), &\quad \forall\, x\in[x_j,\ x_{j-1}].\end{aligned}$$

Now, note from Lemma 2 that

$$\begin{aligned}&\max_{x\in[x_{j-1},\ x_{j+1}]}(\phi(x)-\max\{\hat{c}_{j-1}(\eta)x+\hat{d}_{j-1}(\eta),\hat{c}_j(\eta)x+\hat{d}_j(\eta)\})\\ &\quad=\ \max\{\phi(x_{j-1})-\hat{y}_{j-1}(\eta),\phi(x_{j+1})-\hat{y}_{j+1}(\eta)\}.\end{aligned}$$

This along with (52), (53), and (54) implies that

$$\begin{aligned}&\max_{x\in[x_{j-1},\ x_{j+1}]}(\phi(x)-\max\{\hat{c}_{j-1}(\eta)x+\hat{d}_{j-1}(\eta),\hat{c}_j(\eta)x+\hat{d}_j(\eta)\})\\ &\quad<\ \max_{x\in[x_{j-1},\ x_{j+1}]}(\phi(x)-\max\{c^\star_{j-1}x+d^\star_{j-1},c^\star_jx+d^\star_j\}).\end{aligned}$$

Finally letting $\bar{c}_j=\hat{c}_j(\eta)$, $\bar{c}_{j-1}=\hat{c}_{j-1}(\eta)$, $\bar{d}_j=\hat{d}_j(\eta)$, and $\bar{d}_{j-1}=\hat{d}_{j-1}(\eta)$ proves this lemma. □

Lemma 4. *Suppose that, for some $2\le j\le r-2$, there exist $\bar{c}_j\in(c^\star_j,\ c^\star_{j+1})$ and $\bar{d}_j\in\mathbf{R}$ such that*

$$c^\star_jx_{j+1}+d^\star_j<\bar{c}_jx_{j+1}+\bar{d}_j<\phi(x_{j+1}),\qquad \bar{c}_jx+d_j\le\phi(x),\quad\forall\ x\in[x_j,\ x_{j+1}].$$

Then, there exist $\bar{c}_{j+1}\in(\bar{c}_j,\ c^\star_{j+2})$ and $\bar{d}_{j+1}\in\mathbf{R}$ such that

$$\begin{aligned}c^\star_{j+1}x_{j+1}+d^\star_{j+1}\ &<\ \bar{c}_{j+1}x_{j+2}+\bar{d}_{i+1}<\phi(x_{j+2})\\ \bar{c}_{j+1}x+\bar{d}_{i+1}\ &\le\ \phi(x),\quad\forall\ x\in[x_{j+1},\ x_{j+2}]\\ \max_{x\in[x_{j+1},\ x_{j+2}]}(\phi(x)-\max\{\bar{c}_jx+\bar{d}_j,\bar{c}_{j+1}x+\bar{d}_{j+1}\})\ &<\ \max_{x\in[x_{j+1},\ x_{j+2}]}(\phi(x)-c^\star_{j+1}x-d^\star_{j+1}).\end{aligned}$$

Proof. First, note that there exists a point $\hat{x}_{j+1}\in(x_{j+1},\ x_{j+2})$ such that

$$\bar{c}_j\hat{x}_{j+1}+\bar{d}_j=c^\star_{j+1}\hat{x}_{j+1}+d^\star_{j+1}.$$

Thus,

$$\phi(\hat{x}_{j+1})-\bar{c}_j\hat{x}_{j+1}-\bar{d}_j=\phi(\hat{x}_{j+1})-c^\star_{j+1}\hat{x}_{j+1}-d^\star_{j+1}.$$

Here, it is obvious from Lemma 2 that

$$\phi(\hat{x}_{j+1})-c^\star_{j+1}\hat{x}_{j+1}-d^\star_{j+1}<\max\{\phi(x_{j+1})-c^\star_{j+1}x_{j+1}-d^\star_{j+1},\phi(x_{j+2})-c^\star_{j+1}x_{j+2}-d^\star_{j+1}\}.\quad(56)$$

Define $\hat{x}_{j+1}(\eta)$ and $\hat{y}_{j+2}(\eta)$ as

$$\hat{x}_{j+1}(\eta)=\hat{x}_{j+1}+\eta,\quad \hat{y}_{j+2}(\eta)=c^\star_{j+1}x_{j+2}+d^\star_{j+1}+\eta.$$

Let $\hat{c}_{j+1}(\eta)x+\hat{d}_{j+1}(\eta)$ denote the line passing the two points $(\hat{x}_{j+1}(\eta),\bar{c}_{j+1}\hat{x}_{j+1}(\eta)+\bar{d}_{j+1})$ and $(x_{j+2},\hat{y}_{j+2}(\eta))$. Then, it is clear from (56)) that, for sufficiently small $\eta>0$,

$$\bar{c}_j<\hat{c}_{j+1}(\eta)<c^\star_{j+2}\quad(57)$$

$$\hat{c}_{j+1}(\eta)x+\hat{d}_{j+1}(\eta)\le\phi(x),\quad x\in[x_{j+1},\ x_{j+2}]\quad(58)$$

$$\max_{x\in[x_{j+1},\ \hat{x}_{j+1}(\eta)]}(\phi(x)-\bar{c}_jx-\bar{d}_j)<\max_{x\in[x_{j+1},\ \hat{x}_{j+1}(\eta)]}(\phi(x)-c^\star_{j+1}x-d^\star_{j+1})\quad(59)$$

and

$$\max\{\phi(\hat{x}_{j+1}(\eta)) - \hat{c}_{j+1}(\eta)\hat{x}_{j+1}(\eta) - \hat{d}_{j+1}(\eta), \phi(x_{j+2}) - \hat{c}_{j+1}(\eta)x_{j+2} - \hat{d}_{j+1}(\eta)\}$$
$$< \max\{\phi(x_{j+1}) - c^\star_{j+1}x_{j+1} - d^\star_{j+1}, \phi(x_{j+2}) - c^\star_{j+1}x_{j+2} - d^\star_{j+1}\}. \quad (60)$$

Now, note from Lemma 2 that

$$\max_{x\in[\hat{x}_{j+1}(\eta),\ x_{j+2}]}(\phi(x) - \hat{c}_{j+1}(\eta)x - \hat{d}_{j+1}(\eta))$$
$$= \max\{\phi(\hat{x}_{j+1}(\eta)) - \hat{c}_{j+1}(\eta)\hat{x}_{j+1}(\eta) - \hat{d}_{j+1(\eta)}, \phi(x_{j+2}) - \hat{c}_{j+1}(\eta)x_{j+2} - \hat{d}_{j+1}(\eta)\}.$$

and, hence, that

$$\max_{x\in[x_{j+1},\ x_{j+2}]}(\phi(x) - \max\{\bar{c}_jx + \bar{d}_j, \hat{c}_{j+1}(\eta)x + \hat{d}_{j+1}(\eta)\})$$
$$= \max\{\max_{x\in[x_{j+1},\ \hat{x}_{j+1}(\eta)]}(\phi(x) - \bar{c}_jx - \bar{d}_j), \max_{x\in[\hat{x}_{j+1}(\eta),\ x_{j+2}]}(\phi(x) - \hat{c}_{j+1}(\eta)x - \hat{d}_{j+1}(\eta))\}. (61)$$

Also, note from Lemma 2 that

$$\max\{\phi(\hat{x}_{j+1}(\eta)) - \hat{c}_{j+1}(\eta)\hat{x}_{j+1}(\eta) - \hat{d}_{j+1}(\eta), \phi(x_{j+2}) - \hat{c}_{j+1}(\eta)x_{j+2} - \hat{d}_{j+1}(\eta)\}$$
$$= \max_{x\in[\hat{x}_{j+1}(\eta),\ x_{j+2}]}(\phi(x) - \hat{c}_{j+1}(\eta)x - \hat{d}_{j+1}(\eta))$$

and

$$\max\{\phi(x_{j+1}) - c^\star_{j+1}x_{j+1} - d^\star_{j+1}, \phi(x_{j+2}) - c^\star_{j+1}x_{j+2} - d^\star_{j+1}\}$$
$$= \max_{x\in[x_{j+1},\ x_{j+2}]}(\phi(x) - c^\star_{j+1}x - d^\star_{j+1}).$$

This along with (59)–(61) implies that

$$\max_{x\in[x_{j+1},\ x_{j+2}]}(\phi(x) - \max\{\bar{c}_jx + \bar{d}_j, \hat{c}_{j+1}(\eta)x + \hat{d}_{j+1}(\eta)\})$$
$$< \max\{\max_{x\in[x_{j+1},\ \hat{x}_{j+1}(\eta)]}(\phi(x) - c^\star_{j+1}x - d^\star_{j+1}), \max_{x\in[x_{j+1},\ x_{j+2}]}(\phi(x) - c^\star_{j+1}x - d^\star_{j+1})\}$$
$$= \max_{x\in[x_{j+1},\ x_{j+2}]}(\phi(x) - c^\star_{j+1}x - d^\star_{j+1}).$$

Letting $\bar{c}_{j+1} = \hat{c}_{j+1}(\eta)$ and $\bar{d}_{j+1} = \hat{d}_{j+1}(\eta)$ therefore proves this lemma. □

Lemma 5. *Suppose that, for some* $3 \le j \le r-2$*, there exist* $\bar{c}_{j-1} \in (c^\star_{j-1},\ c^\star_j)$ *and* $\bar{d}_{j-1} \in \mathbf{R}$ *such that*

$$c^\star_{j-1}x_{j-1} + d^\star_{j-1} < \bar{c}_{j-1}x_{j-1} + \bar{d}_{j-1} < \phi(x_{i-1}),$$
$$\bar{c}_{j-1}x + d_{j-1} \le \phi(x), \quad \forall\ x \in [x_{j-1},\ x_j].$$

Then, there exist $\bar{c}_{j-2} \in (c^\star_{j-2},\ \bar{c}_{j-1})$ *and* $\bar{d}_{j-2} \in \mathbf{R}$ *such that*

$$c^\star_{j-2}x_{j-2} + d^\star_{j-2} < \bar{c}_{j-2}x_{j-2} + \bar{d}_{i-2} < \phi(x_{j-2}),$$
$$\bar{c}_{j-2}x + \bar{d}_{j-2} \le \phi(x), \quad \forall\ x \in [x_{j-2},\ x_{j-1}],$$
$$\max_{x\in[x_{j-2},\ x_{j-1}]}(\phi(x) - c^\star_{j-2}x - d^\star_{j-2}) < \max_{x\in[x_{j-2},\ x_{j-1}]}(\phi(x) - \max\{\bar{c}_{j-2}x + \bar{d}_{j-2}, \bar{c}_{j-1}x + \bar{d}_{j-1}\}).$$

Through some arguments similar to those used to prove Lemma 4, we can show that the claim of this lemma holds.

We are ready to show that the best r-term PWL convex lower approximation $h^\star$ defined in (48) to ϕ has the equal approximation errors at its break points.

Lemma 6. *Let $(\underline{a}, \underline{b}) \in \mathbf{R}^{r-2} \times \mathbf{R}^{r-2}$ be an optimal solution to the optimization problem (46). Then,*

$$\phi(x_i) - h^\star(x_i) = \phi(x_j) - h^\star(x_j), \quad i, j \in \{1, \ldots, r-1\}. \tag{62}$$

Proof. Suppose that (62) is not true. Define $x_j = \operatorname{argmin}_{x_1,\ldots,x_{r-1}}(\phi(x_i) - h^\star(x_i))$. We only consider the case $2 \leq j \leq r-2$, since the arguments for the other cases are similar to those for this case.

Note from Lemma 3 that there exist $\bar{c}_{j-1},\ \bar{c}_j$ with $c^\star_{j-2} < \bar{c}_{j-1} < c^\star_{j-1} < c^\star_j < \bar{c}_j < c^\star_{j+1}$ and $\bar{d}_{j-1}, \bar{d}_j \in \mathbf{R}$ such that

$$c^\star_j x_{j+1} + d^\star_j < \bar{c}_j x_{j+1} + \bar{d}_j < \phi(x_{j+1}), \tag{63}$$

and

$$\max_{x \in [x_{j-1},\ x_{j+1}]} (\phi(x) - \max\{\bar{c}_{j-1}x + \bar{d}_{j-1}, \bar{c}_j x + \bar{d}_j\}) \ < \max_{x \in [x_{j-1},\ x_{j+1}]} (\phi(x) - h^\star(x)). \tag{64}$$

It follows from Lemma 4 and (63) that there exist $\bar{c}_{j+1} \in (\bar{c}_j,\ c^\star_{j+2})$ and $\bar{d}_{j+1} \in \mathbf{R}$ such that

$$\max_{x \in [x_{j+1},\ x_{j+2}]} (\phi(x) - \max\{\bar{c}_j x + \bar{d}_j, \bar{c}_{j+1}x + \bar{d}_{j+1}\}) \ < \max_{x \in [x_{j+1},\ x_{j+2}]} (\phi(x) - h^\star(x)). \tag{65}$$

By recursive application of Lemma 4, we can see that there exist $\bar{c}_{j+2}, \ldots, \bar{c}_{r-2}, \bar{d}_{j+1}, \ldots, \bar{d}_{r-2} \in \mathbf{R}$ such that

$$\bar{c}_{j+1} < c^\star_{j+1} < \bar{c}_{j+1} < c^\star_{j+2} < \cdots < \bar{c}_{r-3} < c^\star_{r-2} < \bar{c}_{r-2} < 1$$

and, for each $k \in \{j+2, \ldots, r-2\}$,

$$\max_{x \in [x_k,\ x_{k+1}]} (\phi(x) - \max\{\bar{c}_k x + \bar{d}_k, \bar{c}_k x + \bar{d}_k\}) \ < \max_{x \in [x_k,\ x_{k+1}]} (\phi(x) - h^\star(x)). \tag{66}$$

Now, it is clear from (64) and (66) that

$$\max_{x \in [x_{j-1},\ x_{r-1}]} (\phi(x) - \max\{\bar{c}_{j-2}x + \bar{d}_{j-2}, \ldots, \bar{c}_{r-1}x + \bar{d}_{r-1}\}) < \max_{x \in [x_{j-1},\ x_{r-1}]} (\phi(x) - h^\star(x)). \tag{67}$$

Through some arguments based on Lemmas 5, similar used to those in the preceding paragraph, we can show that there exist $\bar{c}_1, \ldots, \bar{c}_{j-1}, \bar{d}_1, \ldots, \bar{d}_{j-1} \in \mathbf{R}$ such that

$$0 < \bar{c}_1 < c^\star_1 < \bar{c}_2 < c^\star_2 < \cdots < \bar{c}_{j-1} < c^\star_{j-1}$$

and

$$\max_{x \in [x_1,\ x_{j+1}]} (\phi(x) - \max\{\bar{c}_1 x + \bar{d}_1, \ldots, \bar{c}_{j+1}x + \bar{d}_{j+1}\}) < \max_{x \in [x_1,\ x_{j+1}]} (\phi(x) - h^\star(x)\}). \tag{68}$$

Now, define a function $\bar{h} : \mathbf{R} \to \mathbf{R}$ by

$$\bar{h}(x) = \max\{0, \bar{c}_1 x + \bar{d}_1, \ldots, \bar{c}_{r-1} x + \bar{d}_{r-1}, x\},$$

which is r-term PWL. Then, it follows from (67) and (68) that

$$\max_{x \in [x_1,\ x_{r-1}]} (\phi(x) - \bar{h}(x)\}) < \max_{x \in [x_1,\ x_{r-1}]} (\phi(x) - h^\star(x)\}).$$

Moreover, it is clear from $\bar{c}_1 < c_1^\star$ and $\bar{c}_{r-2} < c_{r-2}^\star$ that

$$\begin{aligned} \max_{x \in (-\infty,\ x_1]} (\phi(x) - \bar{h}(x)\}) &< \max_{x \in (-\infty,\ x_1]} (\phi(x) - h^\star(x)\}), \\ \max_{x \in [x_{r-1},\ \infty)} (\phi(x) - \bar{h}(x)\}) &< \max_{x \in [x_{r-1},\ \infty)} (\phi(x) - h^\star(x)\}). \end{aligned}$$

Thus far, we have seen that the function g has a smaller uniform approximation error than $h^\star$. This is contradictory to the assumption that $h^\star$ is a best r-term PWL approximation to ϕ. Thus the claim of this lemma holds. □

As a consequence of Lemma 2 and Lemma 6, the maximum error between ϕ and $h^\star$ can occur only at the break points of $h^\star$. Thus,

$$\phi(x_i) - h^\star(x_i) = \underline{\epsilon}_\phi(r), \quad i = 1, \ldots, r-1$$

and

$$\phi(x) - h^\star(x) < \underline{\epsilon}_\phi(r), \quad x \notin \{x_1, \ldots, x_{r-1}\}.$$

The following lemma further implies that the segments $a_i^\star x + b_i^\star,\ i = 1, \ldots, r-2$ of the function $h^\star$ given in (48) are tangent to the function ϕ at a point $\tilde{x}_i \in (x_i,\ x_{i+1}),\ i = 1, \ldots, r-2$ respectively.

Lemma 7. *For each $i = 1, \ldots, r-2$, the segment $a_i^\star x + b_i^\star$ of the best r-term PWL convex lower approximation $h^\star$ in (48) is tangent to ϕ at a point $\tilde{x}_i \in (x_i,\ x_{i+1})$.*

Proof. Suppose that for some $j \in \{1, \ldots, r-2\}$, the segment $a_j^\star x + b_j^\star$ is not tangent to ϕ at any point on the interval $(x_i,\ x_{i+1})$, *i.e.*, $h(x) < \phi(x)$ for all $x \in (x_j,\ x_{j+1})$. Then there exists $\delta > 0$ such that the function $\tilde{h} : \mathbf{R} \to \mathbf{R}$, defined as

$$\tilde{h}(x) = \max\{0, a_1^\star x + b_1^\star, \ldots, a_{j-1}^\star x + b_{j-1}^\star, a_j^\star x + b_j^\star + \delta, a_{j+1}^\star x + b_{j+1}^\star, \ldots, a_{r-2}^\star x + b_{r-2}^\star, x\},$$

satisfies

$$\sup_{x \in (-\infty,\ \infty)} (\phi(x) - \tilde{h}(x)) \le \sup_{x \in (-\infty,\ \infty)} (\phi(x) - h^\star(x)) = \underline{\epsilon}_\phi(r).$$

Through some arguments similar to those to prove Lemma 6, we can show that there exists a r-term PWL function $\hat{h} : \mathbf{R} \to \mathbf{R}$ which satisfies

$$\sup_{x \in (-\infty,\ \infty)} (\phi(x) - \hat{h}(x)) < \sup_{x \in (-\infty,\ \infty)} (\phi(x) - \tilde{h}(x)) \le \sup_{x \in (-\infty,\ \infty)} (\phi(x) - h^\star(x)) = \underline{\epsilon}_\phi(r).$$

This is contradictory to the assumption that $h^\star$ is a best r-term PWL approximation to ϕ. □

We are now ready to establish the uniqueness of the best r-term PWL lower convex approximation to ϕ.

Lemma 8. *The optimization problem (46) has a unique solution.*

Proof. The arguments used to prove Lemma 6 and Lemma 7 show that if $\bar{h}$ is a best r-term PWL convex lower approximation to ϕ with break points $\bar{x}_1, \ldots, \bar{x}_{r-1}$, then it satisfies

$$\phi(\bar{x}_i) - \bar{h}(\bar{x}_i) = \underline{\epsilon}_\phi(r), \ i = 1, \ldots, r-1, \tag{69}$$

and for each $i = 1, \ldots, r-2$, the segment $\underline{a}_i x + \underline{b}_i$ must be tangent to ϕ at a point, say $\hat{x}_i$ on the interval $(\bar{x}_i, \ \bar{x}_{i+1})$: $\phi(\hat{x}_i) = \bar{h}(\hat{x}_i)$.

Now, note that the equation $\phi(x) = \underline{\epsilon}_\phi(r)$ has the unique solution, say, z_1. We can uniquely define $\tilde{z}_2$ and z_2 from the equations

$$\begin{aligned} \phi(z_1) - \phi'(\tilde{z}_2)(z_1 - \tilde{z}_2) - \phi(\tilde{z}_2) &= \underline{\epsilon}_\phi(r), \\ \phi(z_2) - \phi'(\tilde{z}_2)(z_2 - \tilde{z}_2) - \phi(\tilde{z}_2) &= \underline{\epsilon}_\phi(r). \end{aligned}$$

We can also uniquely define $z_i, \tilde{z}_i, \ i = 3, \ldots, r-1$ from the recursive equations

$$\begin{aligned} \phi(z_{i-1}) - \phi'(\tilde{z}_i)(z_{i-1} - \tilde{z}_i) - \phi(\tilde{z}_i) &= \underline{\epsilon}_\phi(r) \\ \phi(z_i) - \phi'(\tilde{z}_i)(z_i - \tilde{z}_i) - \phi(\tilde{z}_i) &= \underline{\epsilon}_\phi(r). \end{aligned}$$

Finally, it is obvious that $x_i = z_i, \ i = 1, \ldots, r-1$ and $\hat{x}_i = \tilde{z}_i, \ i = 1, \ldots, r-2$. The assertion of this lemma is an easy consequence of the fact that the points $z_1, \ldots, z_{r-1}, \tilde{z}_1, \ldots, \tilde{z}_{r-2}$ uniquely determine a PWL function, which is the unique best r-term PWL lower convex approximation to ϕ. □

So far, we have proved all the claims in Proposition 1 except for (22). To show this, we note that the function ϕ satisfies $\phi(x) = \phi(-x) + x, \ \forall \ x \in \mathbf{R}$. Then, the best r-term PWL convex lower approximation $h^\star$ given in (48) can also be written as $h^\star(x) = \max\{0, (1-a^\star_{r-2})x + b^\star_{r-2}, \ldots, (1-a^\star_1)x + b^\star_1, x\}$. Here note that $0 < 1-a^\star_{r-2} < \cdots < 1-a^\star_1 < 1$. By the uniqueness of the best r-term PWL convex lower approximation to ϕ, we finally have (22).

E.2 Proof of Corollary 1

The best PWL convex lower approximation problem for the two-term log-sum-exp function can be formulated as

$$\begin{array}{ll} \text{minimize} & \sup_{(y_1,y_2)\in\mathbf{R}^2} \left(\mathbf{lse}(y_1, y_2) - \max\limits_{i=1,\ldots,r} \{f_{i1}y_1 + f_{i2}y_2 + g_i\} \right) \\ \text{subject to} & \mathbf{lse}(y_1, y_2) \geq \max\limits_{i=1,\ldots,r} \{f_{i1}y_1 + f_{i2}y_2 + g_i\}, \quad \forall \ (y_1, y_2) \in \mathbf{R}^2, \end{array} \tag{70}$$

where $f_{i1}, f_{i2}, g_i \in \mathbf{R},\ i = 1, \ldots, r$ are the optimization variables. Here, note from (20) that

$$\begin{aligned}
&\mathbf{lse}(y_1, y_2) - \max_{i=1,\ldots,r} \{f_{i1}y_1 + f_{i2}y_2 + g_i\} \\
&= \ y_1 + \phi(y_2 - y_1) - \max_{i=1,\ldots,r} \{f_{i1}y_1 + f_{i2}y_2 + g_i\} \\
&= \ \phi(y_2 - y_1) - \max_{i=1,\ldots,r} \{(f_{i1} + f_{i2} - 1)y_1 + f_{i2}(y_2 - y_1) + g_i\}.
\end{aligned}$$

Obviously, if $\sup_{(y_1,y_2)\in\mathbf{R}^2} \left(\mathbf{lse}(y_1, y_2) - \max_{i=1,\ldots,r} \{f_{i1}y_1 + f_{i2}y_2 + g_i\}\right) < \infty$, then $f_{i1} + f_{i2} = 1$, $i = 1, \ldots, r$. Hence (70) is equivalent to

$$\begin{array}{ll}
\text{minimize} & \sup_{(y_1,y_2)\in\mathbf{R}^2} \left(y_1 + \phi(y_2 - y_1) - \max_{i=1,\ldots,r} \{f_{i1}y_1 + f_{i2}y_2 + g_i\}\right) \\
\text{subject to} & y_1 + \phi(y_2 - y_1) \geq \max_{i=1,\ldots,r} \{f_{i1}y_1 + f_{i2}y_2 + g_i\}, \quad \forall\ (y_1, y_2) \in \mathbf{R}^2, \\
& f_{i1} + f_{i2} = 1, \quad i = 1, \ldots, r.
\end{array} \tag{71}$$

This optimization problem is further equivalent to

$$\begin{array}{ll}
\text{minimize} & \sup_{x\in\mathbf{R}} \left(\phi(x) - \max_{i=1,\ldots,r}\{c_i x + d_i\}\right) \\
\text{subject to} & \phi(x) \geq \max_{i=1,\ldots,r}\{c_i x + d_i\}, \quad \forall\ x \in \mathbf{R}
\end{array} \tag{72}$$

in which $c_i, d_i \in \mathbf{R},\ i = 1, \ldots, r$ are the optimization variables. If $c_i^\star, d_i^\star \in \mathbf{R},\ i = 1, \ldots, r$ solve (72), then $f_{i1}^\star = 1 - c_i^\star,\ f_{i2}^\star = c_i^\star,\ g_i^\star = d_i^\star,\ i = 1, \ldots, r$ solve (71). Conversely, if $f_{i1}^\star, f_{i2}^\star, g_i^\star \in \mathbf{R},\ i = 1, \ldots, r$ solve (71), then $c_i^\star = 1 - f_{i1}^\star = f_{i2}^\star,\ d_i^\star = g_i^\star,\ i = 1, \ldots, r$ solve (72). Moreover, (71) and (72) have the same optimal value. Hence it is obvious from Proposition 1 that the two-term log-sum-exp function has the unique best r-term PWL convex lower approximation $\underline{h}_r$, given by (24).

We next show that the best r-term PWL convex upper approximation $\overline{h}_r$ to the two-term log-sum-exp function can be obtained from (25). To see this, we cast the optimization problem (70) as

$$\begin{array}{ll}
\text{minimize} & \epsilon \\
\text{subject to} & \max_{i=1,\ldots,r} \{f_{i1}y_1 + f_{i2}y_2 + g_i\} \leq \mathbf{lse}(y_1, y_2), \quad \forall\ (y_1, y_2) \in \mathbf{R}^2, \\
& \mathbf{lse}(y_1, y_2) \leq \max_{i=1,\ldots,r} \{f_{i1}y_1 + f_{i2}y_2 + g_i\} + \epsilon, \quad \forall\ (y_1, y_2) \in \mathbf{R}^2,
\end{array} \tag{73}$$

which is obviously equivalent to

$$\begin{array}{ll}
\text{minimize} & \epsilon \\
\text{subject to} & \max_{i=1,\ldots,r} \{\tilde{f}_{i1}y_1 + \tilde{f}_{i2}y_2 + \tilde{g}_i\} - \epsilon \leq \mathbf{lse}(y_1, y_2), \quad \forall\ (y_1, y_2) \in \mathbf{R}^2, \\
& \mathbf{lse}(y_1, y_2) \leq \max_{i=1,\ldots,r} \{\tilde{f}_{i1}y_1 + \tilde{f}_{i2}y_2 + \tilde{g}_i\}, \quad \forall\ (y_1, y_2) \in \mathbf{R}^2.
\end{array} \tag{74}$$

If $\underline{\epsilon},\ \underline{f}_{i1},\ \underline{f}_{i2},\ \underline{g}_i,\ i = 1, \ldots, r$ solve (73) and $\overline{\epsilon},\ \overline{f}_{i1},\ \overline{f}_{i2},\ \overline{g}_i,\ i = 1, \ldots, r$ solve (74) respectively, then $\underline{\epsilon} = \overline{\epsilon} = \underline{\epsilon}_\phi(r),\ \underline{f}_{i1} = \overline{f}_{i1},\ \underline{f}_{i2} = \overline{f}_{i2},\ \overline{g}_i = \underline{g}_i + \underline{\epsilon},\ i = 1, \ldots, r$. Here, note that the best PWL convex upper approximation problem for the two-term log-sum-exp function can be formulated exactly as (74). Finally, note from the uniqueness of the best r-term PWL convex lower approximation to ϕ that the two-term log-sum-exp function has the unique best r-term PWL convex upper approximation $\overline{h}_r$, given by (25).