A proof for multilinear error bounds

Alberto Costa

Future Resilient Systems, Singapore-ETH Centre, Singapore

Abstract

We derive the error bounds for multilinear terms in $[0,1]^n$ using a proof methodology based on the polyhedral representation of the convex hull. We extend the result for multilinear terms in $[\boldsymbol{L},\boldsymbol{0}]\times[\boldsymbol{0},\boldsymbol{U}]\subset\mathbb{R}^n$.

Keywords: convex envelope, convex hull, error analysis, multilinear term, relaxation

1. Introduction

A multilinear term is defined as follows:

$$w(\boldsymbol{x}) = \prod_{j=1}^{n} x_j, \tag{1}$$

where $n \in \mathbb{N}$, $n > 1$, and $\boldsymbol{x} \in [\boldsymbol{L},\boldsymbol{U}]$. As notation, bold symbols represent vectors.

Usually, nonconvex problems with multilinear terms are solved by means of spatial Branch-and-Bound methods, where each multilinear term is replaced by a convex relaxation [2]. The tightest of such relaxations is called convex hull. More precisely, let $W = \{(\boldsymbol{x},w)\,|\,w = \prod_{j=1}^{n} x_j \wedge \boldsymbol{x} \in [\boldsymbol{L},\boldsymbol{U}]\}$. For a single multilinear term, the convex hull $\tilde{W}$ of W is defined as follows:

$$\tilde{W} = \{(\boldsymbol{x},w)\,|\,w \geq \text{vex}[w](\boldsymbol{x}) \wedge w \leq \text{cav}[w](\boldsymbol{x}) \wedge \boldsymbol{x} \in [\boldsymbol{L},\boldsymbol{U}]\}, \tag{2}$$

where $\text{vex}[w](\boldsymbol{x})$ and $\text{cav}[w](\boldsymbol{x})$ are called convex and concave envelopes, respectively.

Email address: alberto.costa@sec.ethz.ch (Alberto Costa)

A meticulous analysis of the tightness of convex hulls for multilinear terms has been recently put forward in [1]. There, the authors provide bounds for the worst-case error between the real value of a monomial (i.e., $\prod_{j=1}^{n} x_j^{\alpha_j}$) and its convex hull for $\boldsymbol{x} \in S \subseteq [0,1]^n$. Moreover, worst-case errors are also provided for a multilinear term $\prod_{j=1}^{n} x_j$ when $\boldsymbol{x} \in [-1,1]^n$ and $\boldsymbol{x} \in [1,r]^n$, $r > 1$. In this paper, we first derive a new proof for the bound formula obtained in [1, Theorem 1.1] for the multilinear term $\prod_{j=1}^{n} x_j$ when $\boldsymbol{x} \in [0,1]^n$. This proof is based on the polyhedral representation of the convex hull. After that, we extend the results to the case where each variable x_i involved in the multilinear term has domain either $[l_i, 0]$ or $[0, u_i]$.

The rest of the paper is organized as follows. Section 2 describes the background on the polyhedral representation of convex hulls for multilinear terms in $[0,1]^n$. The worst-case error analysis is presented in Section 3. Finally, in Section 4 we extend our results to multilinear terms in $[\boldsymbol{L}, \mathbf{0}] \times [\mathbf{0}, \boldsymbol{U}] \subset \mathbb{R}^n$.

2. Preliminaries

In this section we introduce the polyhedral characterization of the convex hull $\tilde{W}$ of W (called dual in [4]). As the need arises, let V be the set of vertices of the hyperrectangle $[\boldsymbol{L}, \boldsymbol{U}]$, and V_W be the lifting of V in the space spanned by $(\boldsymbol{x}, w)$. Hence, $\forall \hat{\boldsymbol{x}} \in V$, $(\hat{\boldsymbol{x}}, w(\hat{\boldsymbol{x}})) \in V_W$. In [3] it was shown that envelopes of a multilinear term are vertex polyhedral. This means that $\tilde{W}$ is a polyhedron having V_W as vertex set. Hence, we can express a point in $\tilde{W}$ as a convex combinations of the points in V_W. Since we focus on multilinear terms in $[0,1]^n$, the components of each element of V_W are either 0 or 1. More formally, let $V_W = \{\boldsymbol{v_1}, \ldots, \boldsymbol{v_{2^n}}\} \subseteq \{0,1\}^{n+1}$, and $\boldsymbol{\lambda} \in [0,1]^{2^n}$. We can characterize $\tilde{W}$ as follows:

$$\forall j \in \{1, \ldots, n\} \quad x_j = \sum_{i=1}^{2^n} \lambda_i v_{i,j} \tag{3}$$

$$w = \sum_{i=1}^{2^n} \lambda_i v_{i,n+1} \tag{4}$$

$$\sum_{i=1}^{2^n} \lambda_i = 1. \tag{5}$$

We could derive the exact formula to express each $\boldsymbol{v_i}$ with respect to the lower and upper bounds of $\boldsymbol{x}$ (i.e, 0 and 1), as done in [4]. However, it is sufficient to understand the structure of V_W. More precisely, $\forall i \in \{1, \ldots, 2^n\}$ $(v_{i,1}, \ldots, v_{i,n})$ is the binary representation of the integer number $i-1$ (i.e., $i - 1 = \sum_{j=1}^{n} 2^{n-j} v_{i,j}$), and $\forall i \in \{1, \ldots, 2^n\}$ $v_{i,n+1} = \prod_{j=1}^{n} v_{i,j}$. For example, when $n = 3$ we obtain:

$$V_W = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \tag{6}$$

where the $i-th$ row represents $\boldsymbol{v_i} \in V_W$. At this point, it can be noticed from (4) that $w = \lambda_{2^n}$. This will be used in the proofs of the next section.

3. Error bounds for a multilinear term in $[0,1]^n$

In this section, the error introduced by the convex hull of a multilinear term is studied. More precisely, we want to quantify the upper and lower bounds of the difference between the real value of the multilinear term and that provided by the convex hull $\tilde{W}$. These two values, called $\mu_U(\tilde{W})$ and $\mu_L(\tilde{W})$, are defined as follows:

$$\mu_U(\tilde{W}) = \max_{(\boldsymbol{x},w)\in\tilde{W}} (w(\boldsymbol{x}) - w) \tag{7}$$

$$\mu_L(\tilde{W}) = \min_{(\boldsymbol{x},w)\in\tilde{W}} (w(\boldsymbol{x}) - w)\,. \tag{8}$$

Theorem 1 (Lower bound)**.** *Let $\boldsymbol{x} \in [0,1]^n$ and $(\boldsymbol{x}, w) \in \tilde{W}$. Moreover, let $w(\boldsymbol{x}) = \prod_{j=1}^{n} x_j$. Then, $w(\boldsymbol{x}) - w \geq \frac{1-n}{n} \sqrt[n-1]{\frac{1}{n}}$.*

Proof. Using the fact that $w = \lambda_{2^n}$, as explained in the previous section, we can rewrite the term $w(\boldsymbol{x}) - w$ in (8) as follows:

$$w(\boldsymbol{x}) - w = f(\lambda_2, \ldots, \lambda_{2^n}) + \lambda_{2^n}^n - \lambda_{2^n}, \tag{9}$$

where $f(\cdot)$ is a function involving sums of products between variables λ_i, except λ_1, and it can be obtained by replacing each x_j in $w(\boldsymbol{x})$ using Definition (3). The idea is to put outside $f(\cdot)$ all the terms (in this case only $\lambda_{2^n}^n$) where the variable λ_{2^n} is not multiplied by at least another variable λ_i, $i \in \{2, \ldots, 2^n-1\}$. Since $\forall i \in \{1, \ldots, 2^n\}$ $\lambda_i \geq 0$, then $f(\lambda_2, \ldots, \lambda_{2^n}) \geq 0$. Moreover, $\forall i \in \{1, \ldots, 2^n\}$ $\lambda_i \leq 1$, hence $\lambda_{2^n}^n - \lambda_{2^n} \leq 0$. If there exists a feasible solution $\tilde{\boldsymbol{\lambda}}$ such that $f(\tilde{\lambda}_2, \ldots, \tilde{\lambda}_{2^n}) = 0$ and $\tilde{\lambda}_{2^n} = \arg\min_{\lambda_{2^n} \in [0,1]}(\lambda_{2^n}^n - \lambda_{2^n})$, then $\tilde{\boldsymbol{\lambda}}$ is also optimal. Recall that $\tilde{\boldsymbol{\lambda}}$ is feasible if $||\tilde{\boldsymbol{\lambda}}||_1 = 1$ and $\tilde{\boldsymbol{\lambda}} \in [0,1]^n$.

It is easy to check that the optimal solution of $\min_{\lambda_{2^n} \in [0,1]}(\lambda_{2^n}^n - \lambda_{2^n})$ is $\lambda_{2^n}^* = \sqrt[n-1]{\frac{1}{n}}$. In addition, it is possible to construct a feasible solution $\boldsymbol{\lambda}^*$, starting from $\lambda_{2^n}^*$, such that $f(\lambda_2^*, \ldots, \lambda_{2^n}^*) = 0$. The solution is $\forall i \in \{2, \ldots, 2^n-1\}$ $\lambda_i^* = 0$, and $\lambda_1^* = 1 - \lambda_{2^n}^*$. According to (3)-(4), this means that the solution $\boldsymbol{x}^*$ for which the lower bound is attained is $\forall j \in \{1, \ldots, n\}$ $x_j^* = \sqrt[n-1]{\frac{1}{n}}$ and $\lambda_{2^n}^* = w^* = \sqrt[n-1]{\frac{1}{n}}$. Hence, $\mu_L(\tilde{W}) = \prod_{j=1}^n x_j^* - w^* = \lambda_{2^n}^{*n} - \lambda_{2^n}^* = \frac{1-n}{n} \sqrt[n-1]{\frac{1}{n}}$. $\square$

Theorem 2 (Upper bound)**.** *Let* $\boldsymbol{x} \in [0,1]^n$ *and* $(\boldsymbol{x}, w) \in \tilde{W}$. *Moreover, let* $w(\boldsymbol{x}) = \prod_{j=1}^n x_j$. *Then,* $w(\boldsymbol{x}) - w \leq \left(\frac{n-1}{n}\right)^n$.

Proof. we can rewrite the term $w(\boldsymbol{x}) - w$ in (7) as follows:

$$w(\boldsymbol{x}) - w = \prod_{j=1}^{n} p_j(\lambda_1, \ldots, \lambda_{2^n-1}) - 1 + \sum_{i=1}^{2^n-1} \lambda_i, \tag{10}$$

where $p_j(\cdot) = 1 - \lambda_1 - \sum_{i \in I(j)} \lambda_i$, and $I(j) \subset \{2, \ldots, 2^n - 1\}$. Each $p_j(\cdot)$ can be derived by expanding $w(\boldsymbol{x})$ using (3) and then replacing each term λ_{2^n}, which appears in each dual representation of x_j (see the last row of (6)), by $1 - \sum_{i=1}^{2^n-1} \lambda_i$ thanks to (5). Similarly, $-w = -\lambda_{2^n}$ has been replaced by $-1 + \sum_{i=1}^{2^n-1} \lambda_i$. Notice that each $p_j(\cdot)$ is nonnegative, since $||\boldsymbol{\lambda}||_1 = 1$, and it contains $-\lambda_1$. In order to maximize (10), the optimal solution has $\lambda_1^* = 0$. To prove it, suppose that there is an optimal solution $\bar{\boldsymbol{\lambda}}$ with $\bar{\lambda}_1 > 0$. One could construct another solution $\tilde{\boldsymbol{\lambda}}$ where $\tilde{\lambda}_1 = 0$, $\tilde{\lambda}_h = \bar{\lambda}_h + \bar{\lambda}_1$ for some $h \in \{2, \ldots, 2^n - 1\}$, and $\forall i \neq h \in \{2, \ldots, 2^n\}\, \tilde{\lambda}_i = \bar{\lambda}_i$. This solution would still be feasible, and by construction the last term in (10) would keep the same value (i.e., $\sum_{i=1}^{2^n-1} \tilde{\lambda}_i = \sum_{i=1}^{2^n-1} \bar{\lambda}_i$). Concerning the terms $p_j(\cdot)$, by construction if $h \in I(j)$ then $p_j(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{2^n-1}) = p_j(\bar{\lambda}_1, \ldots, \bar{\lambda}_{2^n-1})$.

Otherwise, $p_j(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{2^n-1}) > p_j(\bar{\lambda}_1, \ldots, \bar{\lambda}_{2^n-1})$. Hence $\tilde{\boldsymbol{\lambda}}$ is a solution at least as good as $\bar{\boldsymbol{\lambda}}$. However, we can further extend this reasoning as λ_1 is not the only variable equal to 0 in the optimal solution. It is tedious but not too hard to show that there are n variables λ_i which appear only once in the terms $p_j(\cdot)$, and each of these variables appear in a distinct term $p_j(\cdot)$. More specifically, they are the variables λ_i for which the coefficients of the binary decomposition of $i-1$ sum up to $n-1$. As the need arises, let us define the set of those indices as $U_n = \{i \in \{2, \ldots, 2^n - 1\} : i - 1 = \sum_{j=1}^{n} 2^{n-j} a_j \wedge \boldsymbol{a} \in \{0,1\}^n \wedge ||\boldsymbol{a}||_1 = n-1\}$. For the sake of the proof, it is not crucial to determine the elements of U_n, rather the cardinality of the set, and it can be checked that $|U_n| = n$. For example, $U_2 = \{2, 3\}$, $U_3 = \{4, 6, 7\}$, and $U_4 = \{8, 12, 14, 15\}$. The case U_3 is illustrated in (6), where the 4th, 6th, and 7th rows have exactly two elements equal to 1 (the last column is irrelevant). However, one could derive the structure of U_n in a recursive way, i.e., $U_2 = \{2, 3\}$ and $\forall n \in \mathbb{N} : n > 2$, $U_{n+1} = \{2i, \forall i \in U_n\} \cup \{2^{n+1} - 1\}$. With an argument similar to that used to prove that $\lambda_1^* = 0$, we can prove that $\forall i \in \{2, \ldots, 2^n - 1\} \setminus U_n$, $\lambda_i^* = 0$. In other words, it is better to set to 0 the variables appearing more frequently than others in the terms $p_j(\cdot)$: as they have negative sign, they decrease the value of $p_j(\cdot)$. At this point, we can rewrite (10) as follows:

$$\prod_{h \in U_n} (1 - \lambda_h) - 1 + \sum_{h \in U_n} \lambda_h. \tag{11}$$

The problem we have to solve can then be casted as:

$$\max \quad \left(\prod_{h \in U_n} (1 - \lambda_h) - 1 + \sum_{h \in U_n} \lambda_h \right) \tag{12}$$

$$\text{s.t.} \quad \sum_{h \in U_n} \lambda_h \leq 1 \tag{13}$$

$$\forall h \in U_n \quad \lambda_h \geq 0, \tag{14}$$

where (13) is not an equality constraint because in principle λ_{2^n} could be a nonzero slack variable. By applying the KKT conditions we obtain the following constraints:

$$\forall h \in U_n \prod_{k \neq h \in U_n} (1 - \lambda_k) = -\mu \tag{15}$$

$$\left(\sum_{h \in U_n} \lambda_h - 1 \right) \mu = 0. \tag{16}$$

From (15), we get:

$$(1 - \lambda^*_{i_1}) = \cdots = (1 - \lambda^*_{i_n}), \tag{17}$$

where $U_n = \{i_1, \dots, i_n\}$. Hence, $\lambda^*_{i_1} = \cdots = \lambda^*_{i_n}$. In addition, it is not possible that $\mu^* = 0$. As a matter of fact, from (15) and (17) this would mean that $\forall h \in U_n \quad \lambda_h = 1$, and that does not respect Constraint (13). Hence, $\mu^* \neq 0$ and Constraint (16) imposes that $\sum_{h \in U_n} \lambda_h = 1$. Therefore, we obtain:

$$\lambda^*_i = \begin{cases} \frac{1}{n}, \forall i \in U_n, \\ 0, \ \text{otherwise}. \end{cases} \tag{18}$$

This also means that $w^* = \lambda^*_{2^n} = 0$ (Constraint (13) is active at $\boldsymbol{\lambda}^*$). Hence, $\mu_U(\tilde{W})$ is equal to the optimal solution of (12)-(14), that is $\left(1 - \frac{1}{n}\right)^n = \left(\frac{n-1}{n}\right)^n$.

□

Notice that from Theorems 1 and 2, when $n \to \infty$ we can conclude that the difference between a multilinear term and its convex hull is always bounded below by -1 and bounded above by e^{-1}.

4. Error bounds for a multilinear term in $[\boldsymbol{L}, \boldsymbol{0}] \times [\boldsymbol{0}, \boldsymbol{U}] \subset \mathbb{R}^{\boldsymbol{n}}$

In this section, we extend the results of Theorems 1 and 2 from the domain $[0, 1]^n$ to $[\boldsymbol{L}, \boldsymbol{0}] \times [\boldsymbol{0}, \boldsymbol{U}] \subset \mathbb{R}^n$. Consider again the following multilinear term:

$$w(\boldsymbol{x}) = \prod_{j \in N} x_j, \tag{19}$$

where $N = \{1, \dots, n\}$. Let $N_L \subseteq N$, $N_U \subseteq N$, $\forall h \in N_L \ x_h \in [L_h, 0]$, and $\forall k \in N_U \ x_k \in [0, U_k]$, $U_k \neq 1$. We can now define the following auxiliary variables:

$$\forall h \in N_L, \quad \tilde{x}_h = \frac{x_h}{L_h} \in [0,1] \tag{20}$$

$$\forall k \in N_U, \quad \tilde{x}_k = \frac{x_k}{U_k} \in [0,1]. \tag{21}$$

As the need arises, let us also define $\tilde{x}_j = x_j$, $\forall j \in N \setminus \{N_L \cup N_U\}$ (i.e., $x_j \in [0,1]$). We can rewrite (19) as follows:

$$w(\boldsymbol{x}) = \prod_{j \in N \setminus \{N_L \cup N_U\}} \tilde{x}_j \prod_{h \in N_L} (\tilde{x}_h L_h) \prod_{k \in N_U} (\tilde{x}_k U_k) = \prod_{j \in N} \tilde{x}_j \prod_{h \in N_L} L_h \prod_{k \in N_U} U_k. \tag{22}$$

Since all the variables $\tilde{x}$ are in $[0,1]$, we can apply the results of Theorem 1 and 2 to obtain the following bounds:

$$\mu_L(\tilde{W}) = \begin{cases} \prod_{h \in N_L} L_h \prod_{k \in N_U} U_k \left(\frac{n-1}{n}\right)^n, & \text{if } \prod_{h \in N_L} L_h < 0, \\ \prod_{h \in N_L} L_h \prod_{k \in N_U} U_k \left(\frac{1-n}{n} \sqrt[n-1]{\frac{1}{n}}\right), & \text{otherwise.} \end{cases} \tag{23}$$

$$\mu_U(\tilde{W}) = \begin{cases} \prod_{h \in N_L} L_h \prod_{k \in N_U} U_k \left(\frac{1-n}{n} \sqrt[n-1]{\frac{1}{n}}\right), & \text{if } \prod_{h \in N_L} L_h < 0, \\ \prod_{h \in N_L} L_h \prod_{k \in N_U} U_k \left(\frac{n-1}{n}\right)^n, & \text{otherwise.} \end{cases} \tag{24}$$

By computing $\lim_{n \to \infty}$ it is easy to derive a result similar to that of the case $[0,1]^n$, where the scaling factors depend on $\boldsymbol{L}$ and $\boldsymbol{U}$ as summarized in (23) and (24).

Acknowledgements

The author would like to thank Leo Liberti for the useful comments. This research was conducted at the Future Resilient Systems and the Future Cities Laboratory at the Singapore-ETH Centre (SEC). The SEC was established as a collaboration between ETH Zurich and National Research Foundation (NRF) Singapore (FI 370074011 - FI 370074016) under the auspices of the NRF's Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

[1] W. Adams, A. Gupte, Y. Xu, Error bounds for monomial convexification in polynomial optimization, Mathematical Programming (2018).

[2] P. Belotti, J. Lee, L. Liberti, F. Margot, A. Wächter, Branching and bounds tightening techniques for non-convex MINLP, Optimization Methods and Software 24 (4-5) (2009) 597–634 (2009).

[3] A. Rikun, A convex envelope formula for multilinear functions, Journal of Global Optimization 10 (4) (1997) 425–437 (1997).

[4] A. Costa, L. Liberti, Relaxations of multilinear convex envelopes: Dual is better than primal, in: R. Klasing (Ed.), Experimental Algorithms, Vol. 7276 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2012, pp. 87–98 (2012).