case $\mathbb{S}_1 = \emptyset$, $\mathbb{S}_2 = \emptyset$ and the case $\mathbb{S}_1 = \{\theta, \tau\}$, $\mathbb{S}_2 = \emptyset$ have the same rate-distortion function. This is consistent with the classic result of Berger [19].

REFERENCES

[1] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
[2] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Probl. Contr. and Inf. Theory*, no. 1, pp. 19–31, 1980.
[3] B. Chen and G. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.
[4] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, no. 7, pp. 1691–1706, Jul. 2003.
[5] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inf. Theory*, no. 6, pp. 1629–1638, Jun. 2002.
[6] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 460–472, Jul. 1972.
[7] S. Arimoto, "An algorithm for calculating the capacity of an arbitrary discrete memoryless channel," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 14–20, Jan. 1972.
[8] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statist. Decisions*, pp. 205–237, Suppl.1 1984.
[9] P. O. Vontobel, "A generalized Blahut-Arimoto algorithm," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 53.
[10] A. Kavčić, "On the capacity of Markov sources over noisy channels," in *Proc. IEEE GLOBECOM*, San Antonio, TX, Nov. 2001, pp. 2997–3001.
[11] F. Dupuis, W. Yu, and F. Willems, "Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 179.
[12] C. Heegard and A. A. E. Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 6, pp. 731–739, Sep. 1983.
[13] M. Rezaeian and A. Grant, "A generalization of Arimoto-Blahut algorithm," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 181.
[14] M. Chiang and S. Boyd, "Geometric programming duals of channel capacity and rate distortion," *IEEE Trans. Inf. Theory*, vol. 50, no. 2, pp. 245–259, Feb. 2004.
[15] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[16] R. W. Yeung, *A First Course in Information Theory*. New York: Kluwer Academic/Plenum, 2002.
[17] S. Cheng, V. Stanković, and Z. Xiong, "Computing the channel capacity and rate-distortion function with two-sided information," in *Proc. CISS'05*, Baltimore, MD, Mar. 2005.
[18] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, vol. 2, pp. 289–293, 1958.
[19] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

Generalization of Hinging Hyperplanes

Shuning Wang and Xusheng Sun

***Abstract*—The model of hinging hyperplanes (HH) can approximate a large class of nonlinear functions to arbitrary precision, but represent only a small part of continuous piecewise-linear (CPWL) functions in two or more dimensions. In this correspondence, the influence of this drawback for black-box modeling is first illustrated by a simple example. Then it is shown that the above shortcoming can be amended by adding a sufficient number of linear functions to current hinges. It is proven that any CPWL function of n variables can be represented by a sum of hinges containing at most $n+1$ linear functions. Hence the model of a sum of such expanded hinges is a general representation for all CPWL functions. The structure of the novel general representation is much simpler than the existing generalized canonical representation that consists of nested absolute-value functions. This characteristic is very useful for black-box modeling. Based on the new general representation, an upper bound on the number of nestings of nested absolute-value functions of a generalized canonical representation is established, which is much smaller than the known result.**

***Index Terms*—Black-box modeling, canonical representation, continuous piecewise-linear function (CPWL), function approximation, hinging hyperplanes (HHs).**

I. INTRODUCTION

The model of hinging hyperplanes (HH) is a sum of hinges like

$$\pm \max\left\{\ell(x,a), \ell(x,b)\right\} \tag{1}$$

where $\ell(x,a)$ denotes the linear (affine) function $[1\ x^T]a$ of $x \in R^n$ for any parameter vector $a \in R^{n+1}$. This model can approximate a large class of nonlinear functions to arbitrary precision as long as it contains sufficient hinges [1]. And, due to its particular configuration, the least-squares formula can be used to construct efficient algorithms for identifying a desired HH from sampled data [1]–[4]. Because of these reasons, the HH model can be a very practical tool for modeling nonlinear relations with black-box approach [5].

Geometrically, the approximation mechanism of the HH model can be well appreciated from the point of view of continuous piecewise-linear (CPWL) approximation. When a nonlinear function is approximated by an HH in a region, the region is partitioned into a number of polyhedrons by ridges of concerned hinges. In each polyhedron the HH becomes a linear function which is used to approximate the nonlinear function. As an HH is continuous, the partition of a region and the resulting linear functions have to satisfy certain constraints. It is just because of these constraints that the power of CPWL approximation can not be fully utilized by the HH model. This can be illustrated by a two-dimensional example.

Let the sampled data come from the nonlinear function

$$\hat{f}(x) = \frac{0.2}{(x_1-x_2)^2+0.01}\, e^{-4(x_1-1)^2-4(x_2-1)^2} \tag{2}$$

Manuscript received June 13, 2004; revised July 6, 2005. This work was supported in part by the National Natural Science Foundation of China (60374061) and the 973 program of China (2002CB312200).

The authors are with the Department of Automation, Tsinghua University, Beijing, 100084, China (e-mail: swang@mail.tsinghua.edu.cn; sunxs03@mails.tsinghua.edu.cn).

Communicated by P. L. Bartlett, Associate Editor for Pattern Recognition, Statistical Learning and Inference.

Digital Object Identifier 10.1109/TIT.2005.859246

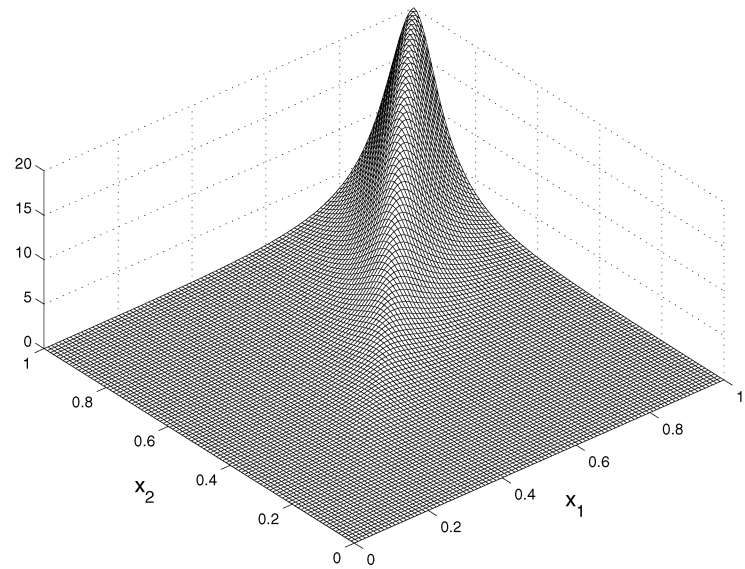

Fig. 1. Diagrams of $\hat{f}(x)$.

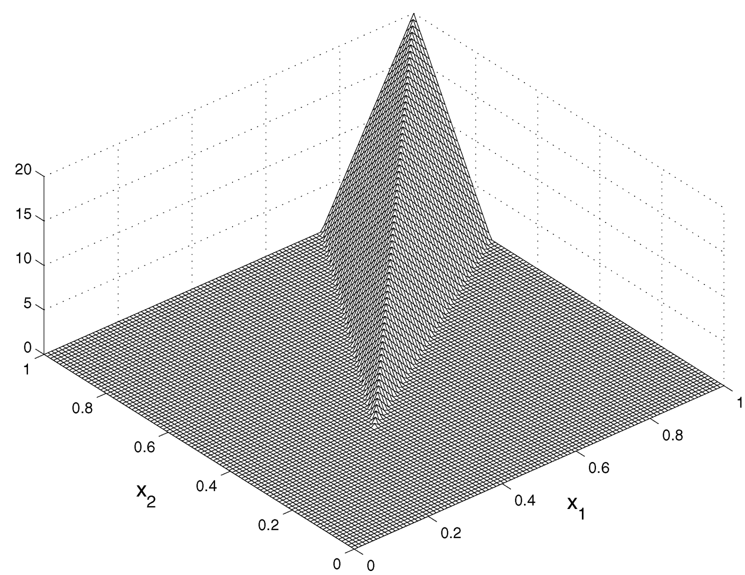

Fig. 2. Diagrams of $\hat{p}(x)$.

and be distributed in the region $0 \le x_1 \le 1$, $0 \le x_2 \le 1$. The diagrams of this nonlinear function and a CPWL function obtained to approximate it are drawn in Figs. 1 and 2. The CPWL function is

$$\hat{p}(x) = \begin{cases} 80x_1 - 50x_2 - 10, & \text{if } x \in \Omega_1 \\ -50x_1 + 80x_2 - 10, & \text{if } x \in \Omega_2 \\ 0, & \text{if } x \in \Omega_3 \end{cases} \tag{3}$$

whose definition domains of local linear functions are depicted in Fig. 3. Notice that the ridges $x_1 = x_2$, $8x_1 - 5x_2 = 1$, and $-5x_1 + 8x_2 = 1$ are cancelled in the region $3x_1 \le 1, 3x_2 \le 1$, which fits $\hat{p}(x)$ to the feature of $\hat{f}(x)$ perfectly. As each ridge of any HH is active in whole space, one can not achieve this effect using an HH.

The above fact on the HH model is actually implied by an early result on the so-called canonical representation, which is a linear function plus a sum of absolute-value functions like

$$\pm\left|\ell(x, a)\right|. \tag{4}$$

This model is first introduced in [6] for analyzing nonlinear circuits. It is shown in [7] that a canonical representation can not represent some CPWL functions even in two dimensions. Because the following identity holds for arbitrary functions $f(\cdot)$ and $g(\cdot)$ from R^n to R

$$\max\left\{f(x), g(x)\right\} = 0.5\left(f(x) + g(x) + \left|f(x) - g(x)\right|\right), \ \forall x \in R^n \tag{5}$$

the HH model and the canonical representation can be equivalently transformed from one to another. Therefore, neither is the HH model a general representation for all CPWL functions.

In order to remedy the shortcoming of the canonical representation, the two-level nested absolute-value functions like

$$\pm\left|\ell(x, b) + \left|\ell(x, c)\right|\right| \tag{6}$$

are introduced in [7], and it is proven that any CPWL function of $x \in R^2$ can be written as a canonical representation plus a sum of two-level nested absolute-value functions. Since this correspondence, it seems to have been a conventional approach to represent CPWL functions in high dimensions using more nestings of absolute-value functions. Such an expression is usually called a generalized canonical representation.

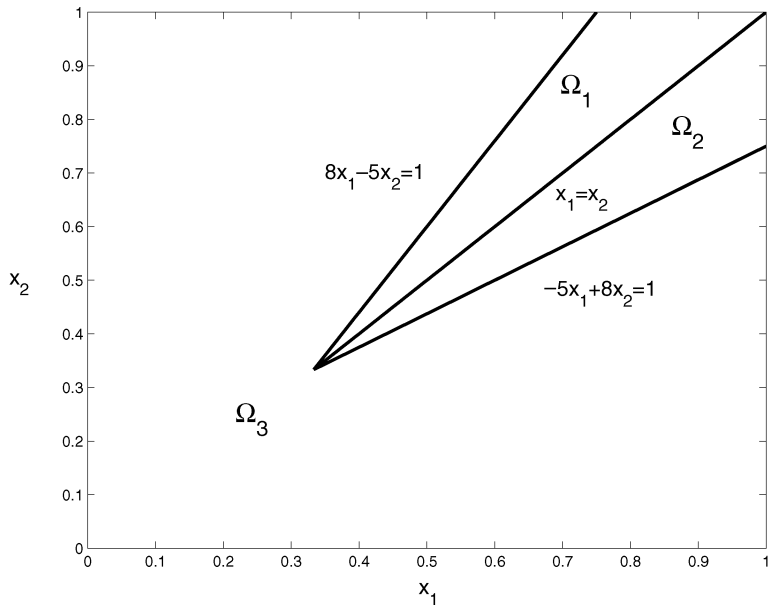

Fig. 3. Definition domains of $\hat{p}(x)$.

It is pointed out in [8] that using n-level nested absolute-value functions is sufficient for representing all CPWL function in n dimensions. Further in [9], a constructive method is given which can determine a generalized canonical representation for any CPWL function in arbitrary dimensions.

Though a generalized canonical representation can represent all CPWL functions, it is unpractical to use it in black-box modeling, for it is very difficult to identify a desired high-level nested absolute-value function from sampled data. In addition, the aforementioned papers provide only an upper bound on the number of nestings necessary for representing all CPWL functions. It is not known whether or not the same goal can be attained using less nestings of absolute-value functions.

The main contribution of this correspondence is to provide a general representation whose structure is much simpler than that of a generalized canonical representation. Hence it is likely to be used in black-box modeling. This general representation is constructed simply by adding a sufficient number of linear functions to current hinges. Specifically speaking, let functions like

$$\pm \max\left\{\ell\Big(x,\theta_1\Big),\, \ell\Big(x,\theta_2\Big),\ldots,\ell\Big(x,\theta_{k+1}\Big)\right\} \tag{7}$$

be called k-order hinges, and a sum of multiorder hinges like

$$\sum_i \sigma_i \max\left\{\ell\Big(x,\theta_1(i)\Big),\, \ell\Big(x,\theta_2(i)\Big),\ldots,\ell\Big(x,\theta_{k_i+1}(i)\Big)\right\} \tag{8}$$

with $\sigma_i \in \{1,-1\}$ and $k_i \leq n$ be called n-order hinging hyperplanes (n-HH). We will prove that any CPWL function of n variables can be written as an n-HH. For example, it is not hard to verify that the CPWL function in (3) can be written as

$$\begin{aligned}\max\{65(x_1-x_2),\, 65(x_2-x_1),\, 15(x_1+x_2)-10\}&\\ -\max\{65(x_1-x_2),\, 65(x_2-x_1)\}.&\end{aligned} \tag{9}$$

Therefore, it is possible to get the best CPWL approximation of $\hat{f}(x)$ in (2) using the 2-HH model.

Once the above result is proven, an interesting fact on the generalized canonical representation can be immediately derived. Consider the case $n = 3$. Let

$$\begin{aligned} f(x) &= \max\left\{\ell\Big(x,\theta_1\Big),\, \ell\Big(x,\theta_2\Big)\right\}\\ g(x) &= \max\left\{\ell\Big(x,\theta_3\Big),\, \ell\Big(x,\theta_4\Big)\right\}.\end{aligned}$$

Using (5) we can get

$$f(x) = 0.5\left(\ell\Big(x,\theta_1\Big)+\ell\Big(x,\theta_2\Big)+\left|\ell\Big(x,\theta_1\Big)-\ell\Big(x,\theta_2\Big)\right|\right) \tag{10}$$

$$g(x) = 0.5\left(\ell\Big(x,\theta_3\Big)+\ell\Big(x,\theta_4\Big)+\left|\ell\Big(x,\theta_3\Big)-\ell\Big(x,\theta_4\Big)\right|\right). \tag{11}$$

Therefore,

$$\begin{aligned}&\max\Big\{\ell\Big(x,\theta_1\Big),\ell\Big(x,\theta_2\Big),\, \ell\Big(x,\theta_3\Big),\, \ell\Big(x,\theta_4\Big)\Big\}\\ &\quad=\max\Big\{f(x),g(x)\Big\}=0.5\left(f(x)+g(x)+\Big|f(x)-g(x)\Big|\right).\end{aligned} \tag{12}$$

Note that the expression in the right-hand side (RHS) of (12) is a two-level nested absolute-value function. According to the result to be proven, it is sufficient to use two-level nestings of absolute-value functions to represent all CPWL functions in three dimensions. This is a bit surprising for it is exactly the same as in two dimensions. Therefore, the number of nestings of absolute-value functions is not crucial for representing CPWL functions.

In general, for any $n \geq 1$, let $\kappa(n)$ be the smallest integer that satisfies $\kappa(n) \geq \ln(n+1)/\ln(2)$. Using (5) repeatedly we can rewrite $\max\{\ell(x,\theta_1),\, \ell(x,\theta_2),\ldots,\ell(x,\theta_{n+1})\}$ as a $\kappa(n)$-level nested absolute-value function. Hence $\kappa(n)$ is at least an upper bound on the number of nestings necessary for representing all CPWL functions in n dimensions, which is much smaller than the known upper bound n.

The other parts of this correspondence are organized as follows. Some preliminary identities are proposed in Section II. The main result is proven in Section III. A brief conclusion and some comments are given in Section IV.

II. Preliminaries

We will prove the main result of this correspondence based on an existing general representation of CPWL functions, which is called a lattice representation. Let $p(\cdot)$ be an arbitrary CPWL function which has finite distinct local linear functions $\ell(\cdot, \theta_i)$, $1 \le i \le m$ from R^n to R, i.e., for any $x \in R^n$ there is at least a $1 \le k \le m$ such that $p(x) = \ell(x, \theta_k)$. It is shown in [10] that there exist finite nonempty subsets of $\{1, 2, \ldots, m\}$, say $s_j \subset \{1, 2, \ldots, m\}$, $1 \le j \le M$, such that

$$p(x) = \max_{1 \le j \le M} \left\{ \min_{i \in s_j} \ell(x, \theta_i) \right\}, \qquad \forall x \in R^n. \tag{13}$$

The expression in the RHS of this equation is just a lattice representation. As this result holds for any positive integer n and arbitrary CPWL function $p : R^n \mapsto R$, a lattice representation can really represent all CPWL functions in all dimensions. In what follows, we will show that any lattice representation in n dimensions can be represented by the n-HH model, hence the latter can represent all CPWL functions.

Two identities will play a fundamental role for deriving the main result of this correspondence. They are given in the following lemma.

Lemma 1: Let $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$ be arbitrary functions from R^n to R. Let α be an arbitrary real number which does not equal one. Let $\bar{\alpha} = 1/(1-\alpha)$. Then the following two identities are valid for all $x \in R^n$

$$\begin{aligned} &\max\Big\{f(x),\, \min\{g(x),\, h(x)\}\Big\} \\ &\quad = \max\{f(x),\, g(x)\} + \max\{f(x),\, h(x)\} \\ &\qquad - \max\{f(x),\, g(x),\, h(x)\}, \end{aligned} \tag{14}$$

$$\begin{aligned} &\max\Big\{f(x),\, g(x),\, \alpha g(x) + h(x)\Big\} \\ &\quad = \max\Big\{f(x),\, \max\{1, \alpha\} \max\Big\{g(x) - \bar{\alpha} h(x),\, 0\Big\} + \bar{\alpha} h(x)\Big\} \end{aligned} \tag{15}$$

$$+ \max\Big\{f(x),\, \min\{1, \alpha\} \min\Big\{g(x) - \bar{\alpha} h(x), 0\Big\} + \bar{\alpha} h(x)\Big\} \tag{16}$$

$$- \max\Big\{f(x),\, \bar{\alpha} h(x)\Big\}. \tag{17}$$

Proof: If $g(x) \le h(x)$, the left-hand side (LHS) of (14) becomes $\max\{f(x), g(x)\}$. As $\max\{f(x), g(x), h(x)\}$ equals $\max\{f(x), h(x)\}$ in this case, the RHS of (14) becomes $\max\{f(x), g(x)\}$, too. Otherwise, i.e., $g(x) \ge h(x)$, the LHS of (14) equals $\max\{f(x), h(x)\}$. As $\max\{f(x), g(x), h(x)\}$ equals $\max\{f(x), g(x)\}$ at this time, the identity in (14) still holds.

To prove the second identity, first consider the case

$$g(x) - (\alpha g(x) + h(x)) \ge 0.$$

The LHS of (15) equals $\max\Big\{f(x),\, g(x)\Big\}$. Note that

$$g(x) - (\alpha g(x) + h(x)) = (1 - \alpha)(g(x) - \bar{\alpha} h(x)).$$

If $\alpha < 1$, we have $g(x) - \bar{\alpha} h(x) \ge 0$. The expressions in the RHS of (15) and (16) become $\max\{f(x), g(x)\}$ and $\max\{f(x), \bar{\alpha} h(x)\}$, respectively. Thus the identity holds obviously. Otherwise, i.e., $\alpha > 1$, we have $g(x) - \bar{\alpha} h(x) \le 0$. The expressions in the RHS of (15) and (16) would be $\max\{f(x), \bar{\alpha} h(x)\}$ and $\max\{f(x), g(x)\}$. The identity holds, too.

Similarly, when

$$g(x) - (\alpha g(x) + h(x)) = (1 - \alpha)(g(x) - \bar{\alpha} h(x)) \le 0$$

the LHS of (15) becomes $\max\{f(x), \alpha g(x) + h(x)\}$. If $\alpha < 1$, we have $g(x) - \bar{\alpha} h(x) \le 0$. The expressions in the RHS of (15) and (16) become $\max\Big\{f(x),\, \bar{\alpha} h(x)\Big\}$ and

$$\begin{aligned} &\max\Big\{f(x),\, \alpha(g(x) - \bar{\alpha} h(x)) + \bar{\alpha} h(x)\Big\} \\ &\qquad = \max\Big\{f(x),\, \alpha g(x) + h(x)\Big\} \end{aligned}$$

respectively. The identity is valid. Otherwise, i.e., $\alpha > 1$, we have $g(x) - \bar{\alpha} h(x) \ge 0$. The expressions in the RHS of (15) and (16) would be $\max\{f(x), \alpha g(x) + h(x)\}$ and $\max\{f(x), \bar{\alpha} h(x)\}$. The identity is still valid.

Corollary 1: For any integer $L \ge 1$ and arbitrary functions $f(\cdot), g_1(\cdot), \ldots, g_L(\cdot)$ from R^n to R, there exist finite subsets of the index set $\{1, \ldots, L\}$, say $\bar{s}_1, \ldots, \bar{s}_K$, and constants $\sigma_1, \ldots, \sigma_K \in \{1, -1\}$ such that

$$\begin{aligned} &\max\Big\{f(x),\, \min_{1 \le i \le L} g_i(x)\Big\} \\ &\qquad = \sum_{k=1}^{K} \sigma_k \max\Big\{f(x),\, \max_{i \in \bar{s}_k} g_i(x)\Big\}, \qquad \forall x \in R^n. \end{aligned} \tag{18}$$

Proof: The corollary holds obviously when $L = 1$. Let $\bar{g}_2(x) = \min\Big\{g_2(x), \ldots, g_L(x)\Big\}$. Because

$$\max\Big\{f(x),\, \min_{1 \le i \le L} g_i(x)\Big\} = \max\Big\{f(x),\, \min\{g_1(x), \bar{g}_2(x)\}\Big\} \tag{19}$$

using the identity in (14) we can write the LHS of (18) as

$$\begin{aligned} &\max\Big\{f(x),\, g_1(x)\Big\} + \max\Big\{f(x),\, \bar{g}_2(x)\Big\} \\ &\qquad - \max\Big\{\max\Big\{f(x),\, g_1(x)\Big\},\, \bar{g}_2(x)\Big\}. \end{aligned} \tag{20}$$

Then we can complete the proof by mathematical induction, for $\bar{g}_2(x)$ consists of only $L - 1$ functions.

Corollary 2: For $f(\cdot), g(\cdot), h(\cdot), \alpha$ and $\bar{\alpha}$ defined in lemma 1, there exist constants $\sigma_1, \sigma_2, \sigma_3 \in \{1, -1\}$ such that for all $x \in R^n$ the function $\max\Big\{f(x),\, g(x),\, \alpha g(x) + h(x)\Big\}$ equals

$$\begin{aligned} &\sigma_1 \max\Big\{f(x),\, g(x),\, \bar{\alpha} h(x)\Big\} \\ &\quad + \sigma_2 \max\Big\{f(x), \alpha g(x) + h(x),\, \bar{\alpha} h(x)\Big\} + \sigma_3 \max\Big\{f(x), \bar{g}(x)\Big\} \end{aligned} \tag{21}$$

with $\bar{g}(x)$ being one of $g(x), \alpha g(x) + h(x)$ or $\bar{\alpha} h(x)$.

Proof: We prove the corollary for the cases $\alpha > 1$, $0 < \alpha < 1$ and $\alpha \le 0$, respectively. According to the second identity in lemma 1 we have the following equation with $F_1(x)$ and $F_2(x)$ being defined by (15) and (16)

$$\begin{aligned} &\max\Big\{f(x),\, g(x),\, \alpha g(x) + h(x)\Big\} \\ &\qquad = F_1(x) + F_2(x) - \max\Big\{f(x),\, \bar{\alpha} h(x)\Big\}. \end{aligned} \tag{22}$$

When $\alpha > 1$, we have

$$\begin{aligned} F_1(x) &= \max\Big\{f(x),\, \alpha \max\Big\{g(x) - \bar{\alpha} h(x),\, 0\Big\} + \bar{\alpha} h(x)\Big\} \\ &= \max\Big\{f(x),\, \alpha g(x) + h(x),\, \bar{\alpha} h(x)\Big\}, \\ F_2(x) &= \max\Big\{f(x),\, \min\Big\{g(x) - \bar{\alpha} h(x),\, 0\Big\} + \bar{\alpha} h(x)\Big\} \\ &= \max\Big\{f(x),\, \min\Big\{g(x),\, \bar{\alpha} h(x)\Big\}\Big\}. \end{aligned}$$

Due to the first identity in lemma 1 we can further get

$$F_2(x) = \max\{f(x),\, g(x)\} + \max\{f(x),\, \bar{\alpha}h(x)\} - \max\{f(x),\, g(x),\, \bar{\alpha}h(x)\}.$$

Hence, the corollary is valid for $\sigma_1 = -1, \sigma_2 = 1, \sigma_3 = 1$ and $\bar{g}(x) = g(x)$.

Similarly, when $0 < \alpha < 1$, we have

$$F_1(x) = \max\left\{f(x),\, g(x),\, \bar{\alpha}h(x)\right\}$$

and

$$F_2(x) = \max\left\{f(x),\, \min\left\{\alpha g(x) + h(x),\, \bar{\alpha}h(x)\right\}\right\}.$$

Using again the first identity in lemma 1 we can further obtain

$$F_2(x) = \max\{f(x),\, \alpha g(x) + h(x)\} + \max\{f(x),\, \bar{\alpha}h(x)\} - \max\{f(x),\, \alpha g(x) + h(x),\, \bar{\alpha}h(x)\}. \quad (23)$$

In this case, the corollary is valid for $\sigma_1 = 1, \sigma_2 = -1, \sigma_3 = 1$ and $\bar{g}(x) = \alpha g(x) + h(x)$.

Finally, when $\alpha \le 0$, we have $F_1(x) = \max\left\{f(x),\, g(x),\, \bar{\alpha}h(x)\right\}$ and

$$\begin{aligned} F_2(x) &= \max\left\{f(x), -\alpha \max\left\{-g(x) + \bar{\alpha}h(x),\, 0\right\} + \bar{\alpha}h(x)\right\} \\ &= \max\left\{f(x),\, \max\left\{(-\alpha)(-g(x) + \bar{\alpha}h(x)), 0\right\} + \bar{\alpha}h(x)\right\} \\ &= \max\left\{f(x),\, \alpha g(x) + h(x),\, \bar{\alpha}h(x)\right\}. \end{aligned} \quad (24)$$

In this case the corollary is valid for $\sigma_1 = 1, \sigma_2 = 1, \sigma_3 = -1$ and $\bar{g}(x) = \bar{\alpha}h(x)$.

Corollary 3: For any integer $L > n \ge 1$, $c_0 \in R$ and arbitrary linear functions $\ell(x, a_1), \ldots, \ell(x, a_L)$ of $x \in R^n$, there exist finite groups of $L - 1$ linear functions, say $\ell(x, b_1(k)), \ldots, \ell(x, b_{L-1}(k))$, $1 \le k \le K$, and corresponding $c_k \in R$, $\sigma_k \in \{1, -1\}$ such that

$$\max\left\{c_0,\, \ell(x, a_1), \ldots, \ell(x, a_L)\right\} = \sum_{k=1}^{K} \sigma_k \max\left\{c_k,\, \ell(x, b_1(k)), \ldots, \ell(x, b_{L-1}(k))\right\}, \quad \forall x \in R^n. \quad (25)$$

Proof: Let $\ell(x, a_i) = a_{i0} + \bar{a}_i^T x$ with $a_{i0} \in R$ and $\bar{a}_i \in R^n$ for all $1 \le i \le L$. Assume that there are at most $\bar{n}$ linearly independent vectors among $\bar{a}_1, \ldots, \bar{a}_L$. Assume further that $\bar{a}_1, \ldots, \bar{a}_{\bar{n}}$ are linearly independent. Let $A = [\bar{a}_1 \cdots \bar{a}_{\bar{n}}]$. Clearly, $\bar{n} \le n$ and for any $\bar{n} + 1 \le i \le L$ there is a $\beta_i \in R^{\bar{n}}$ such that $\bar{a}_i = A\beta_i$. Define a linear transformation from R^n to $R^{\bar{n}}$ by $x' = [x'_1 \cdots x'_{\bar{n}}]^T = A^T x + \gamma$ with $\gamma = [a_{10} \cdots a_{\bar{n}0}]^T$. It is not hard to see that

$$\max\left\{c_0,\, \ell(x, a_1), \ldots, \ell(x, a_L)\right\} = \max\left\{c_0,\, x'_1, \ldots, x'_{\bar{n}},\, \ell(x', a'_{\bar{n}+1}), \ldots, \ell(x', a'_L)\right\}, \quad (26)$$

where $\ell(x', a'_i) = \beta_i^T(x' - \gamma) + a_{i0}, \forall \bar{n} + 1 \le i \le L$. Let

$$\mu(x') = \max\left\{\ell(x', a'_{\bar{n}+1}), \ldots, \ell(x', a'_{L-1})\right\}.$$

Suppose we conjecture that there are finite groups of $\bar{n}$ linear functions of x', say $\ell(x', \bar{b}_1(k)), \ldots, \ell(x', \bar{b}_{\bar{n}}(k)), 1 \le k \le \bar{K}$, and corresponding constants $\bar{c}_k \in R, \bar{\sigma}_k \in \{1, -1\}$ such that

$$\max\left\{\mu(x'),\, c_0,\, x'_1, \ldots, x'_{\bar{n}},\, \ell(x', a'_L)\right\} = \sum_{k=1}^{\bar{K}} \bar{\sigma}_k \max\left\{\mu(x'),\, \bar{c}_k,\, \ell(x', \bar{b}_1(k)), \ldots, \ell(x', \bar{b}_{\bar{n}}(k))\right\}, \quad \forall x' \in R^{\bar{n}}. \quad (27)$$

Then we can complete the proof by simply putting $x' = A^T x + \gamma$ into (27). In what follows we will show that this conjecture is really correct.

We prove the above conjecture by mathematical induction. For the convenience of statement, we denote by $\alpha_0, \alpha_1, \ldots, \alpha_{\bar{n}}$ the components of a'_L, i.e., $a'_L = [\alpha_0\ \alpha_1 \cdots \alpha_{\bar{n}}]^T$. If each of α_i, $1 \le i \le \bar{n}$ equals zero, (27) is obviously valid for $\bar{K} = 1, \bar{\sigma}_1 = 1, \bar{c}_1 = \max\{c_0, \alpha_0\}$ and $\ell(x', \bar{b}_i(k)) = x'_i$, $1 \le i \le \bar{n}$. Assume that the conjecture is correct when $\bar{n} - \eta$ $(0 \le \eta \le \bar{n} - 1)$ numbers among α_i, $1 \le i \le \bar{n}$ equal zero. If we can prove that under this assumption the conjecture is also correct when $\bar{n} - (\eta + 1)$ numbers among α_i, $1 \le i \le \bar{n}$ equal zero, then the conjecture is proven by mathematical induction. Without loss of generality, we assume that $\alpha_i \ne 0$, $1 \le i \le \eta + 1$. Hence $\ell(x', a'_L) = \alpha_0 + \alpha_1 x'_1 + \cdots + \alpha_{\eta+1} x'_{\eta+1}$.

First consider the case that at least one of α_i, $1 \le i \le \eta + 1$, say $\alpha_{\eta+1}$, is not equal to one. If we regard

$$\max\{\mu(x'),\quad c_0,\, x'_1, \ldots, x'_\eta,\, x'_{\eta+2}, \ldots, x'_{\bar{n}}\}, \quad x'_{\eta+1} \quad \text{and} \quad \alpha_0 + \alpha_1 x'_1 + \cdots + \alpha_\eta x'_\eta$$

as $f(x)$, $g(x)$ and $h(x)$, respectively, in Corollary 2, according to this corollary there are $\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3 \in \{1, -1\}$ such that the function in the LHS of (27) can be written as shown in (28)–(30) at the bottom of the page, with $\ell(x', \tilde{\alpha})$ being one of the linear functions $x'_{\eta+1}$, $\ell(x', a'_L)$, or $(\alpha_0 + \alpha_1 x'_1 + \cdots + \alpha_\eta x'_\eta)/(1 - \alpha_{\eta+1})$.

According to the assumption for mathematical induction, the function in (28) can be written as

$$\hat{\sigma}_1 \sum_{k=1}^{K_1} \sigma_k^{(1)} \max\left\{\mu(x'),\, c_k^{(1)},\, \ell(x', b_1^{(1)}(k)), \ldots, \ell(x', b_{\bar{n}}^{(1)}(k))\right\} \quad (31)$$

with some constants $c_k^{(1)} \in R$, $\sigma_k^{(1)} \in \{1, -1\}$ and linear functions $\ell(x', b_1^{(1)}(k)), \ldots, \ell(x', b_{\bar{n}}^{(1)}(k))$ for $1 \le k \le K_1$, which meets the requirement of the inductive step being proven as $\hat{\sigma}_1 \sigma_k^{(1)} \in \{1, -1\}$. Note that the expression in (30) is already in the required form. Hence, we need to consider only how to transform the function in (29) into the required form.

Let a linear transformation $x'' = Bx'$ from $R^{\bar{n}}$ to $R^{\bar{n}}$ be defined by the following equations

$$x''_i = \begin{cases} x'_i, & \text{if } i \ne \eta + 1 \\ \alpha_0 + \alpha_1 x'_1 + \cdots + \alpha_{\eta+1} x'_{\eta+1}, & \text{if } i = \eta + 1 \end{cases} \quad (32)$$

where x''_i denotes the ith component of x''. As $\alpha_{\eta+1} \ne 0$, the above transformation is clearly invertible. Thus, B^{-1} exists. Recall that $\ell(x', a'_L) = \alpha_0 + \alpha_1 x'_1 + \cdots + \alpha_{\eta+1} x'_{\eta+1}$. The function in (29) can

$$\hat{\sigma}_1 \max\left\{\mu(x'),\, c_0,\, x'_1, \ldots, x'_{\bar{n}},\, (\alpha_0 + \alpha_1 x'_1 + \cdots + \alpha_\eta x'_\eta)/(1 - \alpha_{\eta+1})\right\} \quad (28)$$

$$+\, \hat{\sigma}_2 \max\left\{\mu(x'),\, c_0,\, x'_1, \ldots, x'_\eta,\, x'_{\eta+2}, \ldots, x'_{\bar{n}},\, \ell(x', a'_L),\, (\alpha_0 + \alpha_1 x'_1 + \cdots + \alpha_\eta x'_\eta)/(1 - \alpha_{\eta+1})\right\} \quad (29)$$

$$+\, \hat{\sigma}_3 \max\left\{\mu(x'),\, c_0,\, x'_1, \ldots, x'_\eta,\, x'_{\eta+2}, \ldots, x'_{\bar{n}},\, \ell(x', \tilde{\alpha})\right\} \quad (30)$$

be transformed into (33), as shown at the bottom of the page. Using again the assumption for mathematical induction, it can be further written as

$$\hat{\sigma}_2\sum_{k=1}^{K_2}\sigma_k^{(2)}\max\Big\{\mu(B^{-1}x''), c_k^{(2)}, \ell(x'', b_1^{(2)}(k)),\ldots,\ell(x'', b_{\bar{n}}^{(2)}(k))\Big\} \tag{34}$$

with some constants $c_k^{(2)} \in R$, $\sigma_k^{(2)} \in \{1,-1\}$ and linear functions $\ell(x'', b_1^{(2)}(k)),\ldots,\ell(x'', b_{\bar{n}}^{(2)}(k))$ for $1 \le k \le K_2$. Putting back $x'' = Bx'$ in (34), we can see that the function in (29) can also be transformed into the required form. Thus the inductive step being proven holds whenever at least one of α_i, $1 \le i \le \eta+1$ is not equal to one.

When the other case occurs, i.e., $\alpha_i = 1$, $1 \le i \le \eta+1$, the LHS of (27) becomes

$$\max\Big\{\mu(x'),\ c_0,\ x_1',\ldots,x_{\bar{n}}',\ \alpha_0 + x_1' + \cdots + x_\eta' + x_{\eta+1}'\Big\}. \tag{35}$$

If $\eta = 0$, because $\max\{x_1',\ \alpha_0 + x_1'\} = x_1' + \max\{0,\ \alpha_0\}$, the inductive step being proven is obviously correct. Otherwise, using the same linear transformation defined by (32) we can get

$$\begin{aligned}&\max\Big\{\mu(x'),\ c_0,\ x_1',\ldots,x_{\bar{n}}',\ \ell(x', a_L')\Big\}\\&=\max\Big\{\mu(B^{-1}x''), c_0, x_1'',\ldots,x_{\bar{n}}'',-\alpha_0 - x_1'' - \cdots - x_\eta'' + x_{\eta+1}''\Big\}.\end{aligned} \tag{36}$$

Noticing that the parameters before x_i'', $1 \le i \le \eta$ of the last linear function in (36) do not equal one, according to the result we just proved, the function in the RHS of (36) can be transformed into

$$\sum_{k=1}^{K_3}\sigma_k^{(3)}\max\Big\{\mu(B^{-1}x''), c_k^{(3)}, \ell(x'', b_1^{(3)}(k)),\ldots,\ell(x'', b_{\bar{n}}^{(3)}(k))\Big\} \tag{37}$$

with some constants $c_k^{(3)} \in R$, $\sigma_k^{(3)} \in \{1,-1\}$ and linear functions $\ell(x'', b_1^{(3)}(k)),\ldots,\ell(x'', b_{\bar{n}}^{(3)}(k))$ for $1 \le k \le K_3$. Then we can complete the proof of the inductive step in this case by putting back $x'' = Bx'$ in (37), which completes the proof of the corollary, too.

III. Main Result

In this section, we first use corollary 1 to rewrite a lattice representation as a sum of simple functions like those in (25), then use corollary 3 to obtain the required CPWL representation. The detailed process is given below.

Lemma 2: For the lattice representation in (13), there exist finite subsets of $\{1, 2, \ldots, m\}$, say s_k', $1 \le k \le M'$, and corresponding $\sigma_k \in \{1,-1\}$ such that

$$\max_{1\le j\le M}\left\{\min_{i\in s_j}\ell(x,\theta_i)\right\} = \sum_{k=1}^{M'}\sigma_k \max_{i\in s_k'}\ell(x,\theta_i),\quad \forall x\in R^n. \tag{38}$$

Proof: For the convenience of statement, define

$$\Phi_j(x) = \max_{1\le t\le j}\left\{\min_{i\in s_t}\ell(x,\theta_i)\right\},\quad 1\le j\le M. \tag{39}$$

Clearly, the lattice function in (38) is simply $\Phi_M(x)$.

When $M = 1$, we have $\Phi_1(x) = \min_{i\in s_1}\ell(x,\theta_i) = -\max_{i\in s_1}\ell(x,-\theta_i)$, which is already an expression of the RHS of (38).

When $M > 1$, according to corollary 1 we can get

$$\begin{aligned}\Phi_M(x) &= \max\left\{\Phi_{M-1}(x),\ \min_{i\in s_M}\ell(x,\theta_i)\right\}\\&= \sum_{k=1}^{K_1}\sigma_k^{(1)}\ \max\left\{\Phi_{M-1}(x),\ \max_{i\in s_k^{(1)}}\ell(x,\theta_i)\right\}\end{aligned} \tag{40}$$

with some $\sigma_k^{(1)} \in \{1,-1\}$, $s_k^{(1)} \subseteq s_M$, $1 \le k \le K_1$. Using the same corollary for every $1 \le k \le K_1$ we can further get

$$\begin{aligned}&\max\left\{\Phi_{M-1}(x),\ \max_{i\in s_k^{(1)}}\ell(x,\theta_i)\right\}\\&=\sum_{t=1}^{K_{1k}}\sigma_{kt}^{(1)}\max\left\{\Phi_{M-2}(x),\ \max_{i\in s_k^{(1)}}\ell(x,\theta_i),\ \max_{i\in s_{kt}^{(1)}}\ell(x,\theta_i)\right\}\end{aligned} \tag{41}$$

with some $\sigma_{kt}^{(1)} \in \{1,-1\}$, $s_{kt}^{(1)} \subseteq s_{M-1}$, $1 \le t \le K_{1k}$. Note that

$$\begin{aligned}&\max\left\{\Phi_{M-2}(x),\ \max_{i\in s_k^{(1)}}\ell(x,\theta_i),\ \max_{i\in s_{kt}^{(1)}}\ell(x,\theta_i)\right\}\\&\qquad= \max\left\{\Phi_{M-2}(x),\ \max_{i\in s_k^{(1)}\cup s_{kt}^{(1)}}\ell(x,\theta_i)\right\}\end{aligned} \tag{42}$$

and $\left(s_k^{(1)} \cup s_{kt}^{(1)}\right) \subseteq (s_M \cup s_{M-1}) \subseteq \{1, 2, \ldots, m\}$. Combining (40) and (41) we can get

$$\Phi_M(x) = \sum_{k=1}^{K_2}\sigma_k^{(2)}\ \max\left\{\Phi_{M-2}(x),\ \max_{i\in s_k^{(2)}}\ell(x,\theta_i)\right\} \tag{43}$$

with some $\sigma_k^{(2)} \in \{1,-1\}$, $s_k^{(2)} \subseteq \{1, 2, \ldots, m\}$, $1 \le k \le K_2$. Repeating the above process for another $M-2$ times $\Phi_M(x)$ will be necessarily transformed into the required form.

Theorem 1: For any positive integer n and arbitrary CPWL function $p : R^n \mapsto R$, there exist finite, say N, positive integers $\eta(k) \le n+1$, $1 \le k \le N$ and corresponding $\sigma_k \in \{1,-1\}$, $\theta_1(k),\ldots,\theta_{\eta(k)}(k) \in R^{n+1}$ such that

$$p(x)=\sum_{k=1}^{N}\sigma_k\max\Big\{\ell\Big(x,\theta_1(k)\Big),\ell\Big(x,\theta_2(k)\Big),\ldots,\ell\Big(x,\theta_{\eta(k)}(k)\Big)\Big\},\quad \forall x\in R^n. \tag{44}$$

Proof: According to [10], $p(x)$ can be written as a lattice representation like that in (13), i.e.,

$$p(x) = \max_{1\le j\le M}\left\{\min_{i\in s_j}\ell(x,\theta_i)\right\},\ \forall x\in R^n \tag{45}$$

where $s_j \subseteq \{1, 2, \ldots, m\}$, $1 \le j \le M$, and m is the number of distinct local linear functions of $p(x)$. According to lemma 2, we have some $s_k' \subseteq \{1, 2, \ldots, m\}$, $\sigma_k' \in \{1,-1\}$, $1 \le k \le M'$ such that

$$p(x) = \sum_{k=1}^{M'}\sigma_k'\ \max_{i\in s_k'}\ell(x,\theta_i). \tag{46}$$

If we can prove that for any $K > n+1$ and arbitrary vectors $\alpha_i \in R^{n+1}$, $1 \le i \le K$ the function

$$\varphi_K(x) = \max\{\ell(x,\alpha_1),\ldots,\ell(x,\alpha_K)\}$$

$$\hat{\sigma}_2\max\Big\{\mu(B^{-1}x''),\ c_0,\ x_1'',\ldots,x_{\bar{n}}'',\ (\alpha_0 + \alpha_1 x_1'' + \cdots + \alpha_\eta x_\eta'')/(1-\alpha_{\eta+1})\Big\}. \tag{33}$$

can be transformed into an expression of the RHS of (44), the theorem is proven.

Let $\beta_i = \alpha_i - \alpha_K$, $1 \le i \le K-1$. We have

$$\varphi_K(x) = \ell(x, \alpha_K) + \max\{0, \ell(x, \beta_1), \ldots, \ell(x, \beta_{K-1})\}. \quad (47)$$

Using repeatedly corollary 3 for $K-(n+1)$ times we can transform (47) into

$$\varphi_K(x) = \ell(x, \alpha_K) + \sum_{k=1}^{M''} \sigma_k'' \max\Big\{c_k, \ell(x, b_1(k)), \ldots, \ell(x, b_n(k))\Big\} \quad (48)$$

with $c_k \in R$, $\sigma_k'' \in \{1, -1\}$, $\forall 1 \le k \le M''$, which can be further written as

$$\varphi_K(x) = \sum_{k=1}^{M''} \sigma_k'' \max\Big\{\ell(x, b_1'(k)), \ell(x, b_2'(k)), \ldots, \ell(x, b_{n+1}'(k))\Big\} \quad (49)$$

by simply letting $\ell(x, b_{n+1}'(1)) = c_1 + \ell(x, \alpha_K)/\sigma_1''$, $\ell(x, b_j'(1)) = \ell(x, b_j(1)) + \ell(x, \alpha_K)/\sigma_1''$, $1 \le j \le n$, and $\ell(x, b_{n+1}'(k)) = c_k$, $\ell(x, b_j'(k)) = \ell(x, b_j(k))$, $1 \le j \le n$, $2 \le k \le M''$. Hence, the proof is completed.

IV. Conclusion and Comments

It has been shown that via adding at most $n-1$ linear functions to some hinges of an HH, the consequent n-HH model can represent all CPWL functions of n variables. Based on this result a novel upper bound on the number of the nestings of absolute-value functions necessary for representing all CPWL functions is established, which is much smaller than the known result. According to the new upper bound, the number of nestings necessary for representing all CPWL functions of three variables is exactly the same as that of two variables. Hence the number of nestings of absolute-value functions is not crucial for representing all CPWL functions.

Note that $\max\{0, x_1, x_2, \ldots, x_n\}$ is a CPWL function of n variables. It seems impossible to write this function as a $\bar{n}$-HH with some $\bar{n} < n$. Therefore, in general the n-HH model can be the simplest model for representing all CPWL functions of n variables.

As the n-HH model covers more CPWL functions, it is more flexible than the HH for black-box modeling. The potential advantage of such flexibility can be appreciated from the example presented in the first section of this correspondence. To clarify its real effect in general cases, it is necessary to consider how many sampled data are required for identifying a reliable n-HH function. A satisfactory answer to this question can be derived from the main result of [11]. Briefly speaking, according to [11], if a feedforward neural network with a fixed piecewise polynomial activation function and a fixed number of layers is to be identified from sampled data, a reliable result can be obtained as long as $Mlog(M)log(N)/N$ is sufficiently small, where M is the number of the parameters in the network to be identified, N is the number of the sampled data. Let $\sigma(u) = \max\{0, u\}$, $\forall u \in R$. Obviously it is a piecewise polynomial function. Recalling the discussion after (9) concerning the upper bound on the number of nestings necessary for representing all CPWL functions in n dimensions, it is not hard to see that an n-HH model can be expressed as a feedforward neural network of a fixed number of layers with $\sigma(u)$ as its activation function. Therefore, the result of [11] just mentioned is applicable to the n-HH model. Notice that if $N = M^\alpha$ with an arbitrary real number $\alpha > 1$, the number $Mlog(M)log(N)/N$ will approach zero as M becomes very large. Thus it can be inferred that the number of sampled data required for identifying a reliable n-HH function grows almost linearly with respect to the number of the parameters in the model to be identified. This is feasible for general nonlinear regression problems.

While the earlier comment explains theoretically that better results may be obtained using the n-HH model instead of the HH in black-box modeling, practical effects depend on specific identification algorithms. Since the structure of the n-HH model is quite similar to that of the HH, it is possible to construct efficient algorithms via extending the existing ones for the latter. For example, when better fitting can not be achieved by adjusting model parameters, one can increase either the number of hinges, as current HH algorithms do, or the numbers of linear functions in some hinges, as the main result of this correspondence suggests. Considering both of these choices may achieve a satisfactory approximation by means of as less as possible model parameters. In fact, the CPWL function in (3) is just obtained in this way. Notice that the number of possible choices will grow rapidly as n increases. This idea seems realizable only for relatively smaller n. How to develop efficient algorithms for identification of high-dimensional n-HH model would be an interesting subject for future research.

Acknowledgment

The authors are grateful to P. L. Bartlett for his useful suggestion on the identifiability of the n-HH model.

References

[1] L. Breiman, “Hinging hyperplanes for regression, classification, and function approximation,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 999–1013, May 1993.

[2] P. Pucar and J. Sjöberg, “On the hinge-finding algorithm for hinge hyperplanes,” *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1310–1319, May 1998.

[3] X. Li, S. Wang, and W. Wang, “A novel nonlinear approximation approach: Enhanced hinge-finding algorithm,” *Acta Automatica Sinica*, vol. 29, no. 1, pp. 8–13, 2003.

[4] J. Roll, A. Bemporad, and L. Ljung, “Identification of piecewise affine systems via mixed-integer programming,” *Automatica*, vol. 40, no. 1, pp. 37–50, 2004.

[5] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Y. Glorennec, H. Hjalmarsson, and A. Juditsky, “Nonlinear black-box modeling in system identification: A unified overview,” *Automatica*, vol. 31, no. 12, pp. 1691–1724, 1995.

[6] L. O. Chua and S. M. Kang, “Section-wise piecewise-linear functions: Canonical representation, properties, and applications,” *IEEE Trans. Circuits Syst.*, vol. CAS-30, no. 3, pp. 125–140, Mar. 1977.

[7] L. O. Chua and A. C. Deng, “Canonical piecewise-linear representation,” *IEEE Trans. Circuits Syst.*, vol. 35, no. 1, pp. 101–111, Jan. 1988.

[8] J. Lin and R. Unbehauen, “Canonical piecewise-linear networks,” *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 43–50, Jan. 1995.

[9] S. Wang, “General constructive representations for continuous piecewise-linear functions,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 9, pp. 1889–1896, Sep. 2004.

[10] J. M. Tarela and M. V. Martínez, “Region configurations for realizability of lattice piecewise-linear models,” *Math. Comp. Modeling*, vol. 30, no. 11-12, pp. 17–27, 1999.

[11] P. L. Bartlett, V. Maiorov, and R. Meir, “Almost linear VC-dimension bounds for piecewise polynomial networks,” *Neural Computat.*, vol. 10, pp. 2159–2173, 1998.