IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 29, 2018, accepted May 4, 2018, date of publication May 23, 2018, date of current version June 26, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2839079

A 3-Step Process to Estimate Phenylalanine in Commercial Foods for PKU Management

JIEUN KIM[1], AMRUTHAVARSHINI TALIKOTI[2], AND MIREILLE BOUTIN[2], (Senior Member, IEEE)

[1]Bloomberg L.P., New York, NY 10022, USA
[2]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA

Corresponding author: Mireille Boutin (mboutin@purdue.edu)

ABSTRACT Phenylalanine (Phe) is a key nutrient in the dietary management of the metabolic disease phenylketonuria (PKU). To give more freedom to PKU patients, we propose a numerical process to estimate the Phe content of a commercial food using the information printed on the label (Nutritional Fact Label and ordered ingredient list). The process is amenable to implementation as a smart phone/Web app and could be integrated into a personalized dietary management software. Our tests show that the results are very accurate (± 13 mg per serving) in the vast majority of foods considered. Our approach can be modified to handle any other amino acid for the management of other inborn errors of metabolism. A similar approach can also be used for improving the precision (i.e., increasing the quantization resolution) of the nutrients already listed (e.g., protein).

INDEX TERMS Food analysis, medical diet, PKU, mobile health.

I. INTRODUCTION

Some medical diets require keeping track of one's intake of certain nutrients. In order to do this, patients need to have access to the nutritional information for the food they consume. While many nutrients are listed on the Nutrition Facts Label of commercial foods, the information provided is not complete: not all nutrients are listed on the label, and the content for the ones that are listed is rounded. Being able to automatically determine the amount of a nutrient contained in the food, or being able to increase the precision of an amount listed in the Nutrition Facts Label, would thus be helpful for these patients. As pointed out in [1], this is an area of application of mathematics and numerical computing that could greatly benefit from interest from the related communities.

In the language of signal processing, the problems at hand are to *increase the quantization resolution* and *estimate latent quantities*. This work proposes numerical approaches for effectively doing so: we are proposing a food analysis tool based on mathematical optimization and thus amenable to implementation on smart phones or web apps.

Our test target disease is phenylketonuria (PKU), an inherited metabolic disorder affecting approximately 1 in 15,000 people in the United States, which is typically managed with diet restricted in the amino acid phenylalanine (Phe) [2]. We present an approach to numerically estimate the Phe content of a commercial food based on the information given on the food label (Nutrition Facts Label and ordered ingredient list). However, the approach we propose can be modified in order to estimate other amino acids (e.g. lysine) or nutrients, and therefore could be useful for the dietary management of other diseases.

The use of mobile technology such as smart phones in the dietary management of diseases (e.g. kidney disease [3]) or general health is an important component of mobile health [4]. For example, there are many apps for weight-loss [5], [6] as well as app-based food recommendation systems for various purposes [7]–[9]. Since diet recording is an essential part of the assessment and personalization, many efforts are being deployed to try to decrease the burden on the user in order to increase accuracy and compliance. For example, methods have been developed to automatically identify and estimate foods consumed using smart phones based on before/after pictures [10]–[13]. Other wearable systems for dietary monitoring are also being investigated [14].

Our work contributes to the array of mobile health technology by increasing the number of food choices available to metabolic disorder patients and with the hope of improving dietary compliance. Indeed, accurate food recording necessitates access to precise Phe data. While many commercial foods may be technically within the acceptable range of

Phe content, they cannot be safely consumed without that information.

Our proposed three-step process is described in Section II. The first step relies on conclusions of a previous analysis of the phe:protein ratios of common foods [15] to give a first estimate of the Phe content using the rounded protein content printed on the food label. Foods that contain aspartame are also identified and rejected in this step.

The second step uses a previously constructed database of phe:protein ratio [16] along with the food's ingredient list to refine the first estimate. More specifically, the second step also produces a range of possible Phe values whose middle point is used as an estimate with maximum error given by half the size of the possible Phe range. The third step involves a numerical optimization based on the (ordered) ingredient list and the nutrition facts. This optimization is possible when the nutrition information for at least some ingredients can be obtained from an existing food database. The result of the optimization is a range of possible Phe values that is valid when no part of any ingredient is removed during the preparation process. Note that a summary version of this optimization procedure for the general case of nutrient estimation was previously presented at a conference [17], [18]. The final estimate combines the results from all three steps into one single Phe value and maximum error. Some numerical examples are presented in Section III and discussed in Section IV. We conclude in Section V.

II. METHODS

Each step of the process involves a different Phe estimation method. Each of these methods yields a Phe estimate expressed either as a range of possible Phe values (described using a minimum and maximum Phe value in milligrams) or as a Phe estimate in milligrams $\pm$ some error. For example, if the minimum Phe value is $20mg$ and the maximum Phe value is $40mg$, then the possible range of Phe values $[20, 40]$ can be expressed as $30 \pm 10\ mg$.

The range of possible Phe values becomes narrower with each consecutive method, and the intersection of all the ranges is used to produce the final estimate. The amount of information required as input also increases with each method. For convenience, a user satisfied with the precision of the first or second method might elect to stop the process at that point and use the middle point of the Phe range in their diet records.

A. STEP 1: PHE FROM PROTEIN ESTIMATION

Step 1 takes as input the rounded protein content of the food label and considers whether the food is made only of fruit ingredients and whether the food contains aspartame. As nearly half the weight of aspartame is Phe, this sweetener is generally avoided in the PKU diet. The Phe estimate at this step is obtained by multiplying the minimum and maximum protein content by the appropriate minimum and maximum Phe:protein ratio, respectively, in order to obtain a minimum and maximum Phe amount.

For general foods not containing aspartame, we use the minimum and maximum Phe:protein ratios suggested in [15], namely 20mg and 64.5 mg of Phe per gram protein. For fruit-based foods, we use the minimum and maximum Phe:protein ratios suggested in [16], namely 20mg and 39mg of Phe per gram protein. The Phe:protein ratio for aspartame, namely 547 mg Phe per gram protein [16], is also used.

More specifically, let p be the rounded protein value and let Δ be the maximum rounding error. For example, if the label of a food sold in the US states that it contains 1g of protein, then $p = 1$ and $\Delta = 0.5$. Let $minprotein = p - \Delta$ and $maxprotein = p + \Delta$. Let $minphetoprotein = 20$ and $maxphetoprotein = 64.5$. If the food is known to be only made of fruit ingredients and Phe-free ingredients, then replace the value of $maxphetoprotein$ by 39. If the food contains aspartame (or if it is not know whether it does), replace the value of $maxphetoprotein$ by 547. The minimum and maximum Phe values for the first step are then set to

$$minphe_1 = (minphetoprotein) \times (minprotein),$$
$$maxphe_1 = (maxphetoprotein) \times (maxprotein).$$

If $minphe_1$ is high considering the individual's personal Phe tolerance, the user is advised not to consume the food and the process is terminated. For example, for classical PKU patients whose daily Phe allowance is below $400mg$, a minimum Phe value of 100mg should be ground for dismissing the food. Likewise, if the food contains aspartame (or if it is not know whether it does), the user is advised not to consume the food and the process is terminated.

The Phe estimate for Step 1 is taken to be the middle point of the interval $[minphe_1, maxphe_1]$, and the error of that estimate is set to $\frac{maxphe_1 - minphe_1}{2}$. If the size of the error is considered to be small enough, the user may choose to terminate the process and use the estimate of Step 1 in their diet records. For example, considering the precision of the Phe values obtained by laboratory measurements and the many possible causes of individual food variations, an error value below about $10 - 15mg$ may be considered acceptable.

Observe that the more precise the protein value, the smaller the error of the Phe estimate. When the protein content is rounded to the nearest 0.1g (e.g., for some imported foods sold in the US), the estimate provided is quite accurate. However, Nutrition Facts Labels in the US give the protein content rounded to the nearest 1g. For general foods without aspartame, the smallest maximal error one can obtain is for foods with 0g of protein ($\pm 16.13mg$ Phe). For foods made of fruit-based ingredients, that maximal error decreases to a mere $\pm 9.8mg$ Phe. However, the size of the maximal error grows with the protein content. Thus for US foods whose protein content is 1g or more, this initial step only provides a rough range of possible Phe values and thus is mostly used to quickly screen for foods that are obviously too high in Phe for the patient based on their individual tolerance.

A web implementation of this Phe estimation method, along with a free smart phone app version for Android, can be accessed at https://engineering.purdue.edu/brl/PKU/.

B. STEP 2: PHE FROM PROTEIN AND INGREDIENT ESTIMATION

The second step takes as input the previously mentioned protein content p and maximum rounding error Δ as well as the ingredient list. Let n be the number of ingredients in the list, and let $phetoprot_i$ be the Phe:protein ratio for ingredient i (from this Phe:protein database [16] or some other database). We consider the maximum and minimum Phe:protein ratio for all the ingredients:

$$minphetoprotein = \min\{phetoprot_i\}_{i=1}^{n}$$
$$maxphetoprotein = \max\{phetoprot_i\}_{i=1}^{n}$$

If more than one possibility for an ingredient is found in the Phe:protein database, and thus the phe:protein value is unclear, all values are added to the set before picking the maximum and the minimum. If an ingredient does not contain protein (or only traces of it), or if a minuscule amount of the ingredient is used in the food, then it may be discarded from the list.

Again, we let $minprotein = p-\Delta$ and $maxprotein = p+\Delta$, and the minimum and maximum Phe values for the second step are set to

$$minphe_2 = (minphetoprotein) \times (minprotein),$$
$$maxphe_2 = (maxphetoprotein) \times (maxprotein).$$

The Phe estimate for Step 2 is taken to be the middle point of the interval $[minphe_2, maxphe_2]$, and the error of that estimate is set to $\frac{maxphe_2-minphe_2}{2}$. If the size of the error is considered to be small enough, the user may choose to terminate the process and use the estimate of Step 2 in their diet records. Note that the estimate of Step 2 should be more accurate (smaller error) than the estimate of Step 1. A web implementation of this step can be accessed at https://engineering.purdue.edu/brl/PKU/.

C. STEP 3: NUMERICAL OPTIMIZATION AND INTERVAL INTERSECTION

The third step uses the ingredient list and the Nutrition Fact Label. This information is used to set up a set of inequalities which are then solved in order to find the values of $minphe_3$ and $maxphe_3$ using a third method for Phe estimation. The corresponding Phe interval is then intersected with that of Step 2 in order to produce the final estimate.

To apply the third method of Phe estimation, the ingredients must be listed in decreasing order of weight. This gives us a set of inequality constraints. The method also assumes that there is no loss during the preparation process (e.g., nothing is discarded). This gives us two equality constraints: the sum of each ingredient content equals to a serving size and the weighted sum of a nutrient content for one gram of each ingredient equals to the nutrient content for a serving size. We further consider inequality constraints obtained from the Nutrition Facts Label. The proposed method is applicable even if the nutrient content of some of the ingredients is not fully known. But, in general, the more nutrient information is known, the better the accuracy of the final estimate.

This Phe estimation method proceeds in two phases, which we describe after setting up the problem in mathematical terms below. Note that a web implementation of the first phase of Phe estimation method, which includes the use of up to 5 nutrition facts, can be accessed at https://engineering.purdue.edu/brl/PKU/ (Method 5). The link also contains a free smart phone app version for Android.

Denote by PH the Phe in one serving (in mg). Let n be the number of ingredients and x be the serving size (in grams) listed on the Nutrition Facts Label. Denote by A_i the weight (in grams) of ingredient i, starting from the weight of the first ingredient A_1 (largest) to the weight of the last ingredient A_n (smallest). Thus, $A_i \geq A_{i+1}$, for all $i = 1, \ldots, n-1$. If the preparation process is lossless, we have

$$x \geq A_1 \geq A_2 \geq \ldots \geq A_n > 0, \tag{1}$$

$$A_1 + A_2 + \ldots + A_n = x. \tag{2}$$

Let y^{nut} be the (rounded) nutrient content of nutrient nut listed on the Nutrition Facts Label, let Δ^{nut} be the rounding error for the content of nutrient nut, and let y_i^{nut} be the amount of nutrient nut in one gram of ingredient i. Assuming no loss in the preparation process, we have

$$y^{nut} - \Delta^{nut} \leq \sum_{i=1}^{n} y_i^{nut} A_i \leq y^{nut} + \Delta^{nut}. \tag{3}$$

Bounds for A_i can be found using linear programming methods for the optimization problem defined by (1)-(3). But when some of the nutrient data y_i^{nut} is unknown (as is the case for many commercially used ingredients), the problem becomes non-linear. However, we can find a relaxed rectangular region satisfying the constraints (1)-(3) *as much as possible* by finding bounds, $A_{i_{min}}$ and $A_{i_{max}}$, for each ingredient amount A_i, in a first phase we call the "Inverse Recipe."

1) PHASE 1: NUTRIENT CONTENT ESTIMATION USING APPROXIMATE INGREDIENT AMOUNTS (INVERSE RECIPE)

In this phase, we first obtain an initial range for the amount A_i of each ingredient i and then proceed iteratively to refine (i.e., shrink) that range. The final range $(A_{i_{min}}, A_{i_{max}})$ will give us a first set of bounds for the Phe content of the food:

$$\Sigma_{i=1}^{n} p_i A_{i_{min}} \leq PH \leq \Sigma_{i=1}^{n} p_i A_{i_{max}}, \tag{4}$$

where p_i is the Phe amount per gram of ingredient i.

Our initial range is based on the following two lemmas.

Lemma 1: If the ingredient amounts A_i, $i = 1, \ldots, n$, satisfy Equations (1)-(2), then we have the following bounds

for each ingredient:

$$\frac{x}{n} \leq A_1 \leq x, \tag{5}$$

$$0 < A_i \leq \frac{x}{i}, \quad \text{for } i = 2, 3, \ldots, n. \tag{6}$$

Proof: We have $A_i \leq A_{i-1} \leq \ldots \leq A_1$ and so $iA_i \leq \sum_{k=1}^{i} A_k \leq \sum_{k=1}^{n} A_k = x$, therefore $A_i \leq \frac{x}{i}$. Now $x = \sum_{k=1}^{n} A_k \leq \sum_{k=1}^{n} A_1 = nA_1$, and so $\frac{x}{n} \leq A_1$. □

Lemma 2: If the ingredient amounts A_i, $i = 1, \ldots, n$, satisfy Inequality (3) and if $y_i^{nut} \neq 0$, then

$$A_i \leq \frac{y^{nut}}{y_i^{nut}}, \quad \text{for } i = 1, 2, \ldots, n. \tag{7}$$

with equality for some ingredient i if and only if it is the only ingredient containing the nutrient.

Proof: Assume the contrary: $A_i > \frac{y^{nut}}{y_i^{nut}}$, for some i. Then $y_i^{nut} A_i > y^{nut}$. But $y_i^{nut} A_i \leq \sum y_j^{nut} A_j = y^{nut}$, for all i, which is a contradiction. Now if $A_i \leq \frac{y^{nut}}{y_i^{nut}}$, suppose that there is another ingredient i_0 containing that nutrient. If $A_i = \frac{y^{nut}}{y_i^{nut}}$, then $y_i^{nut} A_i = y^{nut}$. Therefore, $y^{nut} = \sum_{k=1}^{n} y_k^{nut} A_k = y_i^{nut} A_i + \sum_{k=1, i \neq i}^{n} y_k^{nut} A_k = y^{nut} + \sum_{k=1, k \neq i}^{n} y_k^{nut} A_k$. Thus we have $0 = \sum_{k=1, k \neq i}^{n} y_k^{nut} A_k$. However, $\sum_{k=1, k \neq i}^{n} y_k^{nut} A_k > 0$ since there exists another ingredient i_0 such that $y_{i_0}^{nut} > 0$, which is a contradiction. □

The initial bounds for each A_i are obtained by combining Equations (5), (6) and (7), as described in Procedure 1. Note that the Procedure takes into account the rounding error (Δ^{nut}) in the nutrient contents listed on the food label and the rounding errors (Δ_i^{nut}) in the USDA database.

Procedure 1 Bound (Initial)

$A_{1_{min}} \leftarrow \frac{x}{n}$, $A_{1_{max}} \leftarrow x$
for $k = 2$ to n **do**
  $A_{k_{min}} \leftarrow 0$, $A_{k_{max}} \leftarrow \frac{x}{k}$
for nutrient *nut* with content y^{nut} **do**
  if $y_1^{nut} \neq 0$ **then**
    $A_{1_{max}} \leftarrow min(A_{1_{max}}, \frac{y^{nut} + \Delta^{nut}}{y_1^{nut} - \Delta_1^{nut}})$
  for $k = 2$ to n **do**
    if $y_k \neq 0$ **then**
      $A_{k_{max}} \leftarrow min(A_{k_{max}}, A_{k-1_{max}}, \frac{y^{nut} + \Delta^{nut}}{y_k^{nut} - \Delta_k^{nut}})$
    else
      $A_{k_{max}} \leftarrow min(A_{k_{max}}, A_{k-1_{max}})$

The initial bounds $A_{i_{min}} \leq A_i \leq A_{i_{max}}$ can be refined using the equation $x = \Sigma_{i=1}^{n} A_i$. Specifically, we have

$$x - \Sigma_{j=1, j \neq i}^{n} A_{j_{max}} \leq A_i \leq x - \Sigma_{j=1, j \neq i}^{n} A_{j_{min}}$$

and so Procedure 2 can be used to narrow the range of each A_i. Lemma 3 yields further refinement.

Lemma 3: If $A_{i_{min}} \leq A_i \leq A_{i_{max}}$ for $i = 1, \ldots, n$, and if $y_{i_0}^{nut} \neq 0$ for some i_0, then $A_{i_0} \leq \frac{y^{nut} - y_i^{nut} A_{i_{min}}}{y_{i_0}^{nut}}$ for all $i \neq i_0$,

Procedure 2 Refining Bound

for $k = 1$ to n **do**
  $A_{k_{min}} \leftarrow max(A_{k_{min}}, x - \Sigma_{m=1, m \neq k}^{n} A_{m_{max}})$
  $A_{k_{max}} \leftarrow min(A_{k_{max}}, x - \Sigma_{m=1, m \neq k}^{n} A_{m_{min}})$

and $A_{i_0} \leq \frac{y^{nut} - \Sigma_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_{min}}}{y_{i_0}^{nut}}$. If y_i^{nut} is known for all i, then we have $A_{i_0} \geq \frac{y^{nut} - \Sigma_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_{max}}}{y_{i_0}^{nut}}$.

Proof: We have $\sum_{i=1}^{n} y_i^{nut} A_i = y^{nut}$ and so

$$y_{i_0}^{nut} A_{i_0} = y^{nut} - \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_i. \tag{8}$$

Multiplying y_i^{nut} by the (initial) bounds for A_i, we get the new bounds for $y_i^{nut} A_i$: $y_i^{nut} A_{i_{min}} \leq y_i^{nut} A_i \leq y_i^{nut} A_{i_{max}}$ for all i. By Equation (8),

$$\begin{aligned} y_{i_0}^{nut} A_{i_0} &= y^{nut} - \sum_{j=1, j \neq i_0}^{n} y_j^{nut} A_j \\ &\leq y^{nut} - y_i^{nut} A_i, \quad \text{for all } i \neq i_0, \\ &\leq y^{nut} - y_i^{nut} A_{i_{min}}, \quad \text{for all } i \neq i_0. \end{aligned}$$

Dividing each side by $y_{i_0}^{nut}$, we get $A_{i_0} \leq \frac{y^{nut} - y_i^{nut} A_{i_{min}}}{y_{i_0}^{nut}}$. In addition, if y_i^{nut} is known for all $i = 1, \ldots, n$, we have

$$\sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_{min}} \leq \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_i \leq \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_{max}}.$$

Therefore, $y^{nut} - \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_{max}} \leq y^{nut} - \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_i$, and $y^{nut} - \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_i \leq y^{nut} - \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_{min}}$. Combining these two inequalities with Equation (8), we get

$$y^{nut} - \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_{max}} \leq y_{i_0}^{nut} A_{i_0} \leq y^{nut} - \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_{min}}.$$

By dividing each side by $y_{i_0}^{nut}$, we obtain

$$\frac{y^{nut} - \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_m ax}}{y_{i_0}^{nut}} \leq A_{i_0}$$

and

$$A_{i_0} \leq \frac{y^{nut} - \sum_{i=1, i \neq i_0}^{n} y_i^{nut} A_{i_{min}}}{y_{i_0}^{nut}}. \tag{9}$$

Note that if y_i^{nut} is unknown, the maximal bound for A_{i_0} given by Equation (9) is still valid with $y_i^{nut} = 0$. □

Lemma 3 yields methods to increase the minimal bound (Procedure 3) and to decrease the upper bound (Procedure 4). Note that the minimal bound can only be refined if y_i^{nut} is known for all i; otherwise, it remains as is.

Let us summarize our proposed Phase 1. To estimate the A_i's, we first select a set of nutrients that are listed on the Nutrition Facts Label (e.g., carbohydrates, sodium,

Procedure 3 For Minimal Bound Increase

for nutrient *nut* with content y^{nut} such that y_k^{nut} exists $\forall k$ **do**
  if $y_n^{nut} \neq 0$ **then**
    $A_{n_{min}} \leftarrow max(A_{n_{min}}, \frac{(y^{nut}-\Delta^{nut})-\sum_{k=1}^{n-1}(y_k^{nut}+\Delta_k^{nut})A_{k_{max}}}{y_n^{nut}+\Delta_n^{nut}})$
  for $j = n-1$ to 1 **do**
    if $y_j^{nut} \neq 0$ **then**
      $A_{j_{min}} \leftarrow max(A_{j_{min}}, A_{j+1_{min}}, \frac{(y^{nut}-\Delta^{nut})-\sum_{k=1,k\neq j}^{n}(y_k^{nut}+\Delta_k^{nut})A_{k_{max}}}{y_j^{nut}+\Delta_j^{nut}})$

Procedure 4 For Maximal Bound Decrease

for nutrient *nut* with content y^{nut} **do**
  for $k = 1$ to n **do**
    if y_k^{nut} does not exist **then**
      $y_k^{nut} \leftarrow 0, \quad \Delta_k^{nut} \leftarrow 0$
  if $y_1^{nut} \neq 0$ **then**
    $A_{1_{max}} \leftarrow min(A_{1_{max}}, \frac{(y^{nut}+\Delta^{nut})-\sum_{k=2}^{n}(y_k^{nut}-\Delta_k^{nut})A_{k_{min}}}{y_1^{nut}-\Delta_1^{nut}})$
  for $j = 2$ to n **do**
    if $y_j^{nut} \neq 0$ **then**
      $A_{j_{max}} \leftarrow min(A_{j_{max}}, A_{j-1_{max}}, \frac{(y^{nut}+\Delta^{nut})-\sum_{k=1,k\neq j}^{n}(y_k^{nut}-\Delta_k^{nut})A_{k_{min}}}{y_j^{nut}-\Delta_j^{nut}})$

protein, etc.). The Nutrition facts, ingredients and nutrient database serve as inputs. We then apply Procedure 1 to estimate the initial bound for all the nutrients selected, and we subsequently follow with Procedure 2 for bound refinement. Afterwards, Procedure 3 and Procedure 4 (for all the nutrients selected) are repeated, followed by Procedure 2, until the total amount of change in the bounds is less than 10^{-5}. Once the required accuracy is reached the obtained min and max bounds are taken as final values from Step 1. (Figure 1).

2) PHASE 2: NUTRIENT CONTENT ESTIMATE REFINEMENT USING SIMPLEX ALGORITHM

The initial bounds on *PH* obtained using Equation (4) in Step 1 are very loose. This is because neither $\sum_{i=1}^{n} A_{i_{min}}$ nor $\sum_{i=1}^{n} A_{i_{max}}$ equal to a serving size x in general. We tighten these initial bounds using the Simplex algorithm [19]. This algorithm is a well known linear programming tool that optimizes a cost function by first finding an initial feasible solution and then refining it. In our case, the cost function is the sum of the amount of Phe contributed by each ingredient: cost $= \sum_{i=1}^{n} p_i A_i (= \text{PH})$.

There are different ways to write the linear constraints of the problem. We start from the constraints of Phase 1: $A_{i_{min}} \leq A_i \leq A_{i_{max}}$, and introduce nonnegative variables

$$\begin{aligned} a_n &= A_n, \\ a_i &= A_i - A_{i+1}, \quad \text{for } i = 1, \ldots, n-1, \\ d_i &\leq A_{i_{max}} - A_{i_{min}}, \quad \text{for } i = 1, 2, \ldots, n, \end{aligned}$$

and slack variables $s_i \geq 0$ for $i = 1, 2, \ldots, n$ such that

$$\begin{aligned} A_i + d_i &= A_{i_{max}}, \\ d_i + s_i &= A_{i_{max}} - A_{i_{min}}. \end{aligned}$$

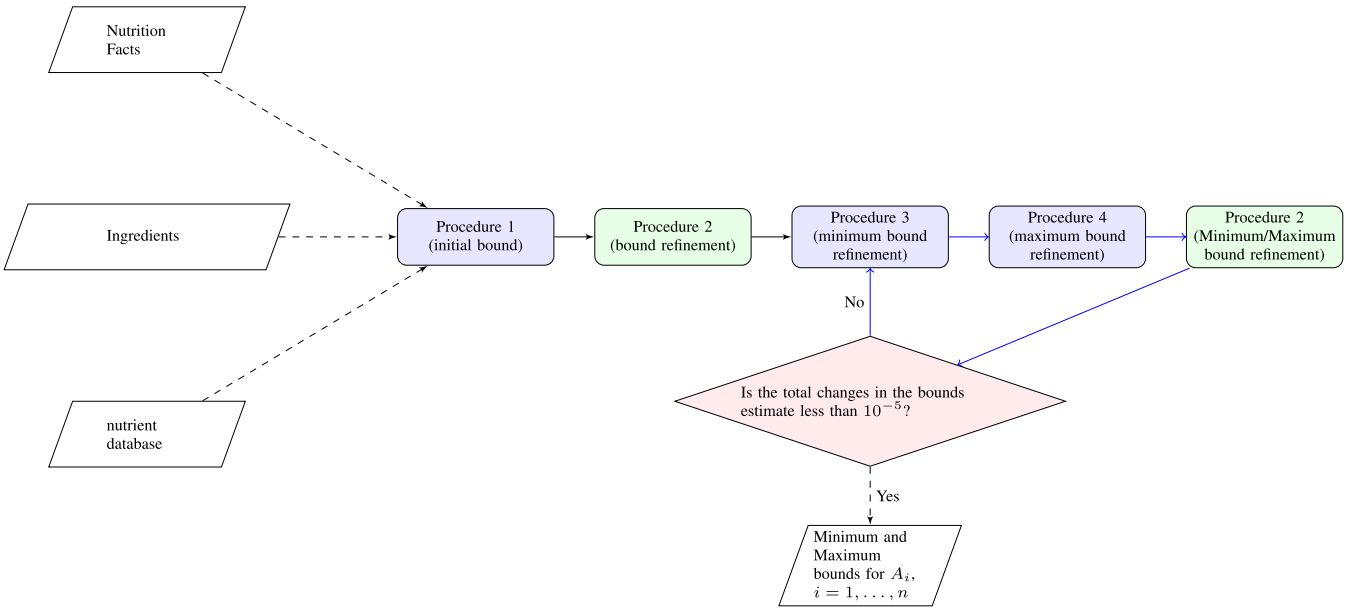

FIGURE 1. Schematic diagram of proposed method to estimate the ingredient amounts (inverse recipe).

TABLE 1. Results of Method 1 for the phenylalanine content estimation (step 1).

Food Number #	Description (serving size)	Protein Content (in g)	Min Phe (in mg)	Max Phe (in mg)	Phe estimate (in mg)	Error (in mg)
1	Carr's Whole Wheat Crackers (17 g)	1	10	96.75	53.38	43.38
2	Heinz Tomato Ketchup (17 g)	0	0	32.25	16.13	16.13
3	KIT KAT Milk Chocolate (42 g)	3	50	225.75	137.88	87.88
4	Campbell's Tomato soup (122 g)	2	30	161.25	95.63	65.63
5	Cheerios Cereal (28 g)	3	50	225.75	137.88	87.88
6	Rice Krispies Cereal (33 g)	2	30	161.25	95.63	65.63
7	Enchilada Sauce (60 g)	1	10	96.75	53.38	43.38
8	Eggo waffle (70 g)	4	70	290.25	180.13	110.13
9	Garlic chili pepper sauce (9 g)	0	0	32.25	16.13	16.13
10	Salsa sauce (30 g)	0	0	32.25	16.13	16.13
11	Simply potatoes Garlic mashed potatoes (124 g)	3	50	225.75	137.88	87.88
12	Butter with Canola Oil (14 g)	0	0	32.25	16.13	16.13
13	Go-Gurt (64 g)	2	30	161.25	95.63	65.63
14	Jell-O Gelatin Snacks-Strawberry (98 g)	1	10	96.75	53.38	43.38
15	Ore-Ida French fries (84 g)	2	30	161.25	95.63	65.63
16	Spicy Brown Mustard (5 g)	0	0	32.25	16.13	16.13
17	Starburst Fruit Chews (40 g)	0	0	32.25	16.13	16.13
18	Vinaigrette Balsamic Dressing (31 g)	0	0	32.25	16.13	16.13
19	Yoplait Original Strawberry (170 g)	6	110	419.25	264.63	154.63
20	ALTOIDS peppermint (2 g)	0	0	32.25	16.13	16.13

Then the amount of ingredient i is given by

$$A_i = A_n + \sum_{k=i}^{n-1}(A_k - A_{k+1}) = \sum_{k=i}^{n} a_k.$$

We can also rewrite Eq. (2) in terms of these new variables,

$$x = \sum_{i=1}^{n} A_i = \sum_{i=1}^{n}\sum_{k=i}^{n} a_k = \sum_{i=1}^{n} i a_i.$$

Secondly, PH can be obtained by

$$PH = \sum_{i=1}^{n} p_i A_i = \sum_{i=1}^{n} p_i \sum_{k=i}^{n} a_k = \sum_{i=1}^{n}(\sum_{m=1}^{i} p_m) a_i.$$

Lastly, by subtracting $\sum_{k=i+1}^{n} a_k + d_{i+1} = A_{i+1_{max}}$ from $\sum_{k=i}^{n} a_k + d_i = A_{i_{max}}$, we have the constraints $a_i + d_i - d_{i+1} = A_{i_{max}} - A_{i+1_{max}}$, for $i = 1, 2, \ldots, n-1$. Therefore, we define the PH estimation problem by Definition 1.

Definition 1 Nutrient Content Estimate Problem Statement for the Simplex Algorithm

minimize, maximize $\sum_{i=1}^{n}(\sum_{k=1}^{i} p_k) a_i$ *where*

$$\begin{cases} \sum_{i=1}^{n} i a_i = x, \\ a_i + d_i - d_{i+1} = A_{i_{max}} - A_{i+1_{max}}, & i = 1, \ldots, n-1, \\ a_n + d_n = A_{n_{max}}, \\ d_i + s_i = A_{i_{max}} - A_{i_{min}}, & i = 1, \ldots, n, \\ a_i, d_i, s_i \geq 0, & i = 1, \ldots, n. \end{cases}$$

Since all constraints are equalities, any feasible solutions satisfying the constraints are points on the edges of a (n-1)-dimensional polytope. Hence, we first find an initial feasible point in the same manner as for the Simplex algorithm. Subsequently, we look through the extreme points of the polytope until the cost at any adjacent points of an extreme point does not decrease anymore. The cost at the point becomes the minimum of the nutrient content for a serving size x gram of a food. Similarly, once the cost function does not increase anymore, we set the maximum bound for the nutrient content to the value of the cost function.

The web app takes into account the values for Phe:Protein ratios of ingredients of various food products from standard databases [16]. These databases are kept updated and can be used as a look-up table to search for the values required in Step 2. Currently, the app requires the user to manually enter the ingredients, answer questions about the presence of aspertame and indicate if the food is solely fruit-based. However, its implementation can be further enhanced to include OCR (Optical Character Recognition) techniques to read the values of protein content and the ingredient composition directly from the scanned images of the food label taken by a smart-phone to make it more user-friendly. This would extend the scope of the app for estimation of other amino acids and could thereby serve as a useful dietary management tool for other inborn metabolic disorders. Hence, it can be tested using a wider set of individuals. Further, each of the three steps discussed above can be similarly applied for

TABLE 2. Results of Method 2 and 3 for the phenylalanine content estimation and final phe estimates.

#	Method 2				Method 3				Final Intersection	
	Min Phe (in mg)	Max Phe (in mg)	Phe estimate (in mg)	Error (in mg)	Min Phe (in mg)	Max Phe (in mg)	Phe estimate (in mg)	Error (in mg)	Phe estimate (in mg)	Error (in mg)
1	20.55	79.69	50.12	29.57	53.61	85.11	69.36	15.75	66.65	13.04
2	0.00	32.14	16.07	16.07	1.20	6.57	3.89	2.69	3.89	2.69
3	87.72	185.94	136.83	49.11	144.27	191.53	167.90	23.63	165.11	20.84
4	30.91	132.81	81.86	50.95	40.69	95.45	68.07	27.38	68.07	27.38
5	120.72	199.23	159.97	39.26	179.86	180.51	180.19	0.32	180.19	0.32
6	78.87	134.62	106.74	27.87	91.54	94.80	93.17	1.63	93.17	1.63
7	12.20	96.43	54.31	42.12	0.41	34.14	17.28	16.87	23.17	10.97
8	143.82	297.25	220.53	76.71	196.26	216.35	206.31	10.05	206.31	10.05
9	0.00	16.58	8.29	8.29	2.65	5.27	3.96	1.31	3.96	1.31
10	0.00	26.73	13.37	13.37	7.90	18.23	13.07	5.17	13.07	5.17
11	70.31	183.33	126.82	56.51	139.51	162.23	150.87	11.36	150.87	11.36
12	0.00	26.19	13.10	13.10	12.06	17.66	14.86	2.80	14.86	2.80
13	31.07	129.55	80.31	49.24	116.38	120.95	118.67	2.29	118.67	2.29
14	10.00	51.00	30.50	20.50	10.01	30.44	20.23	10.22	20.23	10.22
15	40.35	160.72	100.53	60.19	77.64	78.76	78.20	0.56	78.20	0.56
16	0.00	32.14	16.07	16.07	10.11	10.16	10.14	0.03	10.14	0.03
17	0.00	18.00	9.00	9.00	0.00	4.48	2.24	2.24	2.24	2.24
18	0.00	32.14	16.07	16.07	0.00	5.53	2.77	2.77	2.77	2.77
19	113.92	336.82	225.37	111.45	287.11	291.08	289.10	1.98	289.10	1.98
20	0.00	10.36	5.18	5.18	0.43	4.22	2.33	1.90	2.33	1.90

protein values with a precision of 0.1g or 0.01g to incorporate for the food labels of products from non-US countries.

III. RESULTS

We estimated the Phe of 20 commercial foods using our process. None of the food chosen contains aspartame, and none of them is made solely of fruit-based ingredients. Details of our data are available at [20]. We used the protein content rounded to the nearest gram in order to show the accuracy one would expect when using US food labels; the accuracy obtained using a more precise protein estimate or in the case of fruit-based products would be even better. Table 1 shows the Phe estimation result obtained with the method of Step 1. Table 2 shows the Phe estimation results obtained with the methods of Step 2 and Step 3, along with the final estimates. Step 3 was performed using six nutrients (protein, sodium, calories, carbohydrates, fat, and cholesterol).

IV. DISCUSSION

The Phe estimation results after Step 1 (Table 1) show that the Phe of many foods can be accurately estimated from the protein content. Indeed, the error in the Phe estimate is only 16.13mg for 8 of the 20 foods considered, namely all foods with a 0g protein content. The results would have been even more accurate had we considered food made solely of fruit-based and Phe-free ingredients. For example, if such food had a listed 0g protein content, then the error would have been only 9.8mg. Note that Food #19 would likely be ruled out at this point, as the minimum Phe value (110mg) is over 100mg.

The Phe estimation results using the methods of Step 2 and Step 3 are even more accurate. The largest error in final estimates is only 27.38mg, the next largest error is 20.84, and all other errors are no more than 13.04. Note that using more nutrients than six for Step 3 might have further improved the accuracy of our estimates. It is interesting to note that the Phe intervals obtained at Step 2 and Step 3 are not necessarily nested into one another (see Food #1,3,7). Thus, the estimate obtained by combining these two steps (intersecting their Phe ranges) is more accurate than the estimate obtained by performing Step 3 only.

V. CONCLUSION

The information provided by the ingredient list and Nutritional Facts Label of commercial foods is important but incomplete as the precision of the nutrients listed is larger than required for managing certain medical diets, and some key nutrients (e.g., Phe) are simply not listed. This is problematic for individuals with inherited metabolic disorders such as PKU who must carefully monitor their Phe intake throughout their entire life [21]. Thus, we proposed a three-step process for estimating the Phe content of a food automatically using the information provided on the label. As a bonus, our process also returns an estimate for the amount of each ingredient used to prepare the food, and thus provides an approximate solution for the ''inverse recipe problem.''

Because of assumptions made in the third step of the process, our estimates are only valid when no part of any ingredient is discarded during the preparation process. No restricting assumption is made for the first two steps, and the overall method is applicable even if the nutrient content of some of the ingredients is not fully known. In our numerical tests, our process yield very accurate estimate for Phe (less than about 13mg error) in 90% of the foods considered and thus could be a useful tool for PKU management. Furthermore, our proposed process could be modified slightly in order to estimate other amino acids, or any other nutrient of interest that is not listed on the Nutritional Facts Label. One could also use a similar process for improving the precision (i.e., increase the quantization resolution) of the nutrients already listed (e.g., protein).

ACKNOWLEDGMENT

J. Kim was with the Department of Mathematics, Purdue University, West Lafayette, IN, USA.

REFERENCES

[1] J. Kim and M. Boutin, "Deriving nutrition information using mathematical estimation: The example of phenylalanine in sweets with gelatin," *J. Acad. Nutrition Dietetics*, vol. 115, no. 9, pp. 1384–1386, 2015.
[2] P. R. Huttenlocher, "The neuropathology of phenylketonuria: Human and animal studies," *Eur. J. Pediatrics*, vol. 159, no. 2, pp. S102–S106, 2000.
[3] J. Campbell and J. Porter, "Dietary mobile apps and their effect on nutritional indicators in chronic renal disease: A systematic review," *Nephrology*, vol. 20, no. 10, pp. 744–751, 2015.
[4] M. Hingle and H. Patrick, "There are thousands of apps for that: Navigating mobile technology for nutrition education and behavior," *J. Nutrition Edu. Behav.*, vol. 48, no. 3, pp. 213–218, 2016.
[5] K. M. Azar *et al.*, "Mobile applications for weight management: Theory-based content analysis," *Amer. J. Preventive Med.*, vol. 45, no. 5, pp. 583–589, 2013.
[6] J. Chen, J. E. Cade, and M. Allman-Farinelli, "The most popular smartphone apps for weight loss: A quality assessment," *JMIR mHealth uHealth*, vol. 3, no. 4, p. e104, 2015.
[7] M. A. El-Dosuky, M. Z. Rashad, T. T. Hamza, and A. H. El-Bassiouny, "Food recommendation using ontology and heuristics," in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl.* Berlin, Germany: Springer, 2012, pp. 423–429.
[8] G. Agapito *et al.*, "Profiling basic health information of tourists: Towards a recommendation system for the adaptive delivery of medical certified nutrition contents," in *Proc. IEEE Int. Conf. High Perform. Comput. Simulation (HPCS)*, Jul. 2014, pp. 616–620.
[9] G. Agapito *et al.*, "DIETOS: A recommender system for adaptive diet monitoring and personalized food suggestion," in *Proc. IEEE 12th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2016, pp. 1–8.
[10] F. Zhu *et al.*, "The use of mobile devices in aiding dietary assessment and evaluation," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 4, pp. 756–766, Aug. 2010.
[11] M. M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem, and S. G. Mougiakakou, "A food recognition system for diabetic patients based on an optimized bag-of-features model," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1261–1271, Jul. 2014.
[12] F. Zhu, M. Bosch, N. Khanna, C. J. Boushey, and E. J. Delp, "Multiple hypotheses image segmentation and classification with application to dietary assessment," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 377–388, Jan. 2015.
[13] G. Ciocca, P. Napoletano, and R. Schettini, "Food recognition: A new dataset, experiments, and results," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 588–598, May 2017.
[14] T. Prioleau, E. Moore, II, and M. Ghovanloo, "Unobtrusive and wearable systems for automatic dietary monitoring," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2075–2089, Sep. 2017.
[15] J. Kim and M. Boutin, "New multipliers for estimating the phenylalanine content of foods from the protein content," *J. Food Composition Anal.*, vol. 42, pp. 117–119, Sep. 2015.
[16] J. Kim and M. Boutin, "A list of phenylalanine to protein ratios for common foods," School Elect. Comput. Eng., Purdue Univ., Lafayette, IN, USA, ECE Tech. Rep. 456, 2014. [Online]. Available: http://docs.lib.purdue.edu/ecetr/456
[17] J. Kim and M. Boutin, "An approximate inverse recipe method with application to automatic food analysis," in *Proc. IEEE Symp. Comput. Intell. Healthcare e-Health (CICARE)*, Dec. 2014, pp. 32–39.
[18] J. Kim and M. Boutin, "Estimating the nutrient content of commercial foods from their label using numerical optimization," in *Proc. New Trends Image Anal. Process. ICIAP Workshops*. Cham, Switzerland: Springer, 2015, pp. 309–316.
[19] M. Benhamadou, "On the simplex algorithm 'revised form,'" *Adv. Eng. Softw.*, vol. 33, nos. 11–12, pp. 769–777, 2002.
[20] A. Talikoti and M. Boutin. (Feb. 2018). *20 Commercial Foods to Test 3-Step Phenylalanine Estimation Process*. [Online]. Available: https://purr.purdue.edu/publications/2940/1
[21] National Institutes of Health Consensus Development Panel, "National institutes of health consensus development conference statement: Phenylketonuria: Screening and management, October 16–18, 2000," *Pediatrics*, vol. 108, no. 4, pp. 972–982, Oct. 2001. [Online]. Available: http://www.nichd.nih.gov/publications/pubs/pku/sub3.cfm

JIEUN KIM was born in Seoul, South Korea, in 1984. She received the B.S. degree in mathematics from Hanyang University, Seoul, South Korea, in 2008, the M.S. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2015, and the Ph.D. degree in mathematics. While studying at Purdue University, she developed novel optimization methods to estimate phenylalanine (Phe) content in foods, using food labels and USDA database. The optimization methods help to monitor intake of certain nutrients (e.g., Phe), thus can assist planning medical diets of metabolic disease patients, such as phenylketonuria (PKU). She is currently a Software Engineer with Bloomberg L.P., New York.

AMRUTHAVARSHINI TALIKOTI received the B.E. degree in electronics and communication engineering from the R.V. College of Engineering, Bengaluru, India, in 2017. She is currently pursuing the M.S. degree in electrical and computer engineering with Purdue University, West Lafayette, USA. From 2015 to 2016, she was a part-time Intern with Siemens Technologies Pvt., Ltd. In 2016, she was also a summer Intern with ARM Embedded Technologies Pvt., Ltd. Her research interests include signal processing, communication systems, and data science.

MIREILLE (MIMI) BOUTIN (SM'17) received the Ph.D. degree in mathematics from the University of Minnesota in 2001 under the supervision of P. J. Olver. She held a post-doctoral position at the Max Plank Institute for Mathematics in the Sciences, Leipzig, Germany. Then, she joined Purdue University. She is currently an Associate Professor with the School of Electrical and Computer Engineering, with a courtesy appointment in the Department of Mathematics. Her research is in the area of signal processing, machine learning, and applied mathematics. Dr. Boutin is a member of the Image, Video, and Multidimensional Signal Processing Technical Committee of the IEEE Signal Processing Society. She was a three-time recipient of the Purdue's Seed for Success Award. She was also a recipient of the Eta Kappa Nu Outstanding Faculty Award, the Eta Kappa Nu Outstanding Teaching Award, and the Wilfred Duke Hesselberth Award for Teaching Excellence. She is currently an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.