

# An approximate inverse recipe method with application to automatic food analysis

Jieun Kim

Department of Mathematics  
Purdue University  
West Lafayette, Indiana 47907  
Email: kim793@purdue.edu

Mireille Boutin

School of Electrical and Computer Engineering  
Purdue University  
West Lafayette, Indiana 47907  
Email: mboutin@purdue.edu

**Abstract**—We propose a method for automatically determining the amount of each ingredient used to prepare a commercial food using the information provided on its label. The method applies when no part of any ingredient is removed in the preparation process and as long as we can collect the nutrition data (e.g., from the USDA Food Database) for at least some of the ingredients. Using this information, we first find a set of initial minimum and maximum bounds for each ingredient amount. Then we improve these maximum and minimum bounds using an iterative method. The resulting bounds on the ingredient amounts can then be used to estimate the nutrient content of the food. We tested this approach for estimating the phenylalanine content of various commercial foods. Phenylalanine is an amino acid that must be carefully monitored when treating patients with the metabolic disease phenylketonuria (PKU). Our numerical tests indicate that the accuracy of our method is within an acceptable range (10mg Phe) for most of the foods we considered. We implemented a web-based application of our proposed method for public use. Our method should be applicable to the estimation of nutrients involved in the management of other medical diets.

## I. INTRODUCTION

People on a medical diet must constantly be aware of the nutritional content of the food they consume. Unfortunately, the information printed on the label of a commercial food is nowhere near complete. In fact, it is not even sufficient to be able to re-create the food. In particular, while the ingredients are listed on the label, the exact amounts used are unknown. The Nutrition Facts Label provides additional information by listing the content of some nutrients. However, the numbers given are rounded up. For example, the protein content is usually rounded up to the nearest gram. For people who must follow a medical diet, the missing information can be problematic because following such diets can require knowing the nutrient content of the food beyond the precision of the food label. In some cases, the nutrients of interest are not even listed on the food label. Being able to automatically determine the ingredient amounts could solve this problem, as the nutrient content of most basic ingredients can be obtained from a food database. Unfortunately, as far as we know, there has been no previous work on this problem.

We propose to determine the amount of each ingredient used to prepare a commercial food using the food label, the ingredient list, along with the USDA Food Database [1] (USDA database). Our motivating application is the management of inherited metabolic diseases, the most common of which is phenylketonuria (PKU) with an incidence of about 350 in

million births in the United States [2]. PKU is characterized by an inability to metabolize the amino acid phenylalanine (Phe), which leads to an abnormal accumulation of Phe in the patient's blood [3]. The classical treatment for this disease involves following a strict Phe-restricted diet [4]. The Phe-restricted diet is required especially for newborn babies with PKU in order to avoid intellectual disorder [5][6]. In addition, this dietary treatment has positive effects even on previously untreated adults [7] as well as for pregnancy with PKU [8].

The USDA database is commonly used to determine the Phe content of foods consumed by PKU patients, as the Phe content of commercial foods is not typically listed in the food label. However, this database, which is considered very extensive, only lists the Phe content for a very limited number of food items (4843). There are other databases available, such as [9], [10] and [11], but the cost and time needed for testing foods for their Phe content limits the usability of such databases, considering the enormous variety of foods available on the market today. As the Phe content of commercial foods is not listed in the food label, being able to automatically estimate the Phe content of a food from the information found on its label would be a useful tool for managing PKU. In previous work [12], we have shown that the protein content can be used to roughly estimate the Phe content of the food (See Table I). The method we propose in this paper yields a lot more accurate results.

Given is a commercial food. Denote by  $x$  the serving size of the food, and let  $n$  be the number of ingredients used in the recipe. Both  $x$  and  $n$  can be obtained from the information printed on the packaging of the food, namely from the Nutrition Facts Label and the ingredient list. Let  $A_i$  denote the unknown weight (in grams) of ingredient  $i$ , for  $i = 1, \dots, n$ . Since the ingredients are listed in decreasing order, we have  $A_i \geq A_{i+1}$ . If no part of any ingredient is removed in the preparation process, we have

$$x \geq A_1 \geq A_2 \geq \dots \geq A_n > 0, \quad (1)$$

$$A_1 + A_2 + \dots + A_n = x. \quad (2)$$

The food label gives us the rounded up content  $y^{nut}$  of many nutrients (for example, the nutrient protein). We can look for the amount  $y_i^{nut}$  of nutrient “nut” in one gram of ingredient  $i$  in the USDA database. If no part of any ingredient is removed in the preparation process and  $\Delta^{nut}$  is the rounding error of

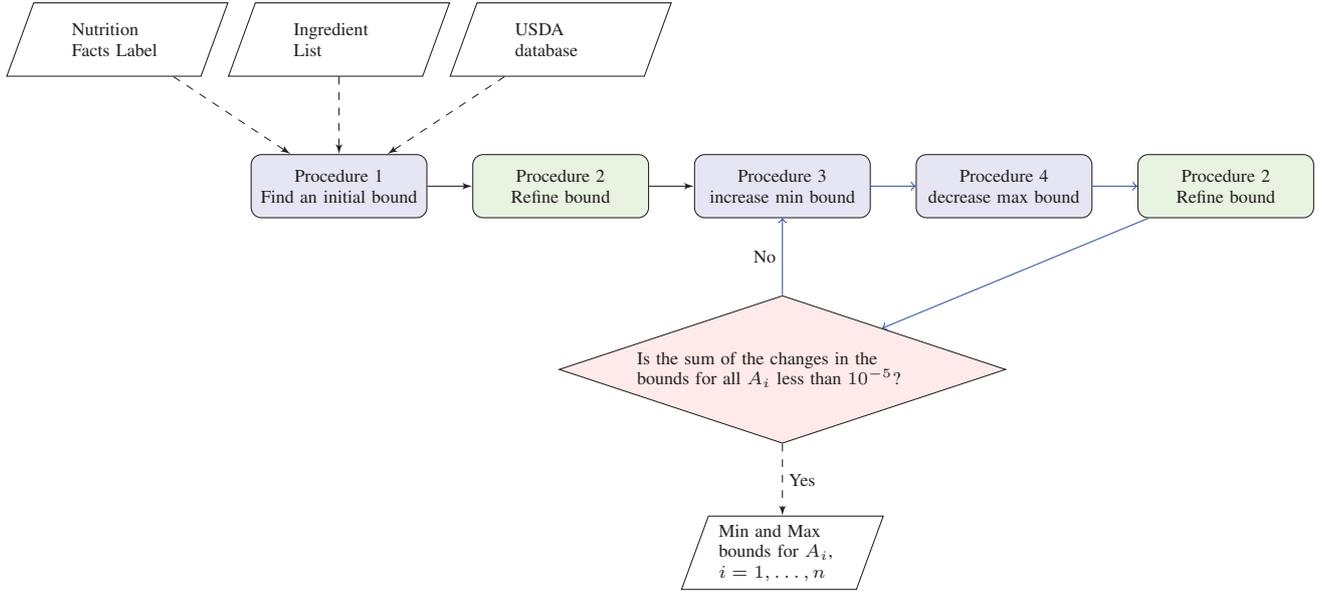


Fig. 1: Schematic Diagram of Proposed Method to Estimate the Ingredient Amounts

---

#### Procedure 1 Initial bound

---

$A_{1_{min}} \leftarrow \frac{x}{n}, \quad A_{1_{max}} \leftarrow x$   
**for**  $i = 2$  to  $n$  **do**  
 $A_{i_{min}} \leftarrow 0, \quad A_{i_{max}} \leftarrow \frac{x}{i}$   
**for** given nutrient 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**  $i = 2$  to  $n$  **do**  
**if**  $y_i \neq 0$  **then**  
 $A_{i_{max}} \leftarrow \min(A_{i_{max}}, A_{i-1_{max}}, \frac{y^{nut} + \Delta^{nut}}{y_i^{nut} - \Delta_i^{nut}})$   
**else**  
 $A_{i_{max}} \leftarrow \min(A_{i_{max}}, A_{i-1_{max}})$

---



---

#### Procedure 2 Refining bound

---

**for**  $i = 1$  to  $n$  **do**  
 $A_{i_{min}} \leftarrow \max(A_{i_{min}}, x - \sum_{j=1, j \neq i}^n A_{j_{max}})$   
 $A_{i_{max}} \leftarrow \min(A_{i_{max}}, x - \sum_{j=1, j \neq i}^n A_{j_{min}})$

---



---

#### Procedure 3 to increase the minimal bound

---

**for** given nutrient 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**  $i = n - 1$  to  $1$  **do**  
**if**  $y_i^{nut} \neq 0$  **then**  
 $A_{i_{min}} \leftarrow \max(A_{i_{min}}, A_{i+1_{min}}, \frac{(y^{nut} - \Delta^{nut}) - \sum_{k=1, k \neq i}^n (y_k^{nut} + \Delta_k^{nut}) A_{k_{max}}}{y_i^{nut} + \Delta_i^{nut}})$

---

---

**Procedure 4** to decrease the maximal bound

---

```

for given nutrient 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  $i = 2$  to  $n$  do
      if  $y_i^{nut} \neq 0$  then
         $A_{i_{max}} \leftarrow \min(A_{i_{max}}, A_{i-1_{max}}, \frac{(y^{nut} + \Delta^{nut}) - \sum_{k=1, k \neq i}^n (y_k^{nut} - \Delta_k^{nut}) A_{k_{min}}}{y_i^{nut} - \Delta_i^{nut}})$ 

```

---

**TABLE I: Rough estimate of Phe content based on protein content [12]** The Phe content can be estimated using the protein content rounded to the nearest gram.

Protein content	Phe content	
0 g	max	32.5 mg
	min	0 mg
1 g	max	97.5 mg
	min	10 mg
2 g	max	162.5 mg
	min	30 mg
3 g	max	227.5 mg
	min	50 mg

$y^{nut}$ , we have

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

One can look for bounds for the unknown nutrient content found by using linear programming methods for the optimization problem constrained by Equation (1)-(3). Unfortunately, many commercial foods include ingredients not listed in the USDA database:  $y_i^{nut}$  is unknown for these ingredients, making the problem non-linear. Furthermore, it is sometimes difficult to select the correct ingredients in the USDA database, as different variations of an ingredient can be listed. This may yield an inconsistent linear system for which linear programming methods would fail to find any possible solution.

To address these issues, we propose an iterative method for obtaining a maximum bound and a minimum bound for each ingredient amount  $A_i$ ,  $i = 1, \dots, n$ . This method, which can be viewed as an *approximate inverse recipe method*, is applicable even if the nutrient data  $y_i^{nut}$  of some ingredients is missing or unknown: only the nutrition information that is known is used in the optimization. We describe and test our proposed *approximate inverse recipe method* in Section II. In Section III, we apply our *approximate inverse recipe method* to the problem of estimating the Phe content of commercial foods. We conclude in Section IV.

## II. APPROXIMATE INVERSE RECIPE METHOD

We now describe our *approximate inverse recipe method* to obtain bounds for the amount  $A_i$  of each ingredient contained in a commercial food,  $i = 1, \dots, n$ .

### A. Initial range estimate

To obtain a set of initial values for the minimum bound and the maximum bound for each ingredient amount  $A_i$ , we use the following two lemmas.

**Lemma 1.** *If  $\{A_i\}_{i=1}^n$  satisfy Equation (1) and (2), then*

$$\frac{x}{n} \leq A_1 \leq x \quad (4)$$

$$0 < A_i \leq \frac{x}{i}, \quad \text{for } i = 2, 3, \dots, n \quad (5)$$

*Proof:* Since  $A_i \leq A_{i-1} \leq \dots \leq A_1$ ,

$$iA_i \leq \sum_{k=1}^i A_k \leq \sum_{k=1}^n A_k = x.$$

Dividing each side by  $i$ , we get

$$A_i \leq \frac{x}{i}.$$

In addition,

$$x = \sum_{k=1}^n A_k \leq \sum_{k=1}^n A_1 = nA_1.$$

Dividing each side by  $n$ , we have a minimum bound for  $A_1$ ,

$$\frac{x}{n} \leq A_1. \quad \blacksquare$$

**Lemma 2.** *If  $\{A_i\}_{i=1}^n$  satisfy Equation (3) and  $y_i^{nut} \neq 0$ , then*

$$A_i \leq \frac{y^{nut}}{y_i^{nut}} \quad \text{for } i = 1, 2, \dots, n. \quad (6)$$

*Equality holds for some  $i_0$  only if ingredient  $i_0$  is the sole ingredient containing the nutrient.*

*Proof:* Suppose that

$$A_i > \frac{y^{nut}}{y_i^{nut}}, \quad \text{for some } i.$$

This implies  $y_i^{nut} A_i > y^{nut}$ . However,

$$y_i^{nut} A_i \leq \sum_{j=1}^n y_j^{nut} A_j = y^{nut}, \quad \text{for all } i.$$

This is a contradiction, so

$$A_i \leq \frac{y^{nut}}{y_i^{nut}}$$

is true for all  $i$ . Now considering equality on (6), assume that there are more than one ingredient containing the nutrient. If  $A_{i_0} = \frac{y^{nut}}{y_{i_0}^{nut}}$ , then  $y_{i_0}^{nut} A_{i_0} = y^{nut}$ . Therefore,

$$\begin{aligned} y^{nut} &= \sum_{i=1}^n y_i^{nut} A_i \\ &= y_{i_0}^{nut} A_{i_0} + \sum_{i=1, i \neq i_0}^n y_i^{nut} A_i \\ &= y^{nut} + \sum_{i=1, i \neq i_0}^n y_i^{nut} A_i \end{aligned}$$

This gives

$$0 = \sum_{i=1, i \neq i_0}^n y_i^{nut} A_i.$$

But

$$\sum_{i=1, i \neq i_0}^n y_i^{nut} A_i > 0$$

since there exists another index  $t$  then  $i_0$  such that  $y_t^{nut} > 0$  by assumption. This is a contradiction.  $\blacksquare$

We obtain initial bounds for each  $A_i$  by combining Equations (4), (5) and (6), as described in Procedure 1. Note that the Procedure takes into account the rounding errors ( $\Delta^{nut}$  and  $\Delta_i^{nut}$ ) in the nutrient contents listed on the food label and in the USDA database.

### B. Iterative method to narrow the range estimate

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

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

and so Procedure 2 can be used to narrow the range of each  $A_i$ .

**Lemma 3.** Suppose  $A_{i_{min}} \leq A_i \leq A_{i_{max}}$  for  $i = 1, \dots, n$ . If  $y_k^{nut} \neq 0$  for some  $k$ , then

$$A_k \leq \frac{y^{nut} - y_i^{nut} A_{i_{min}}}{y_k^{nut}} \quad \text{for all } i \neq k.$$

Also,

$$A_k \leq \frac{y^{nut} - \sum_{i=1, i \neq k}^n y_i^{nut} A_{i_{min}}}{y_k^{nut}}.$$

Furthermore, if  $y_i^{nut}$  is known for all  $i$ ,

$$A_k \geq \frac{y^{nut} - \sum_{i=1, i \neq k}^n y_i^{nut} A_{i_{max}}}{y_k^{nut}}.$$

*Proof:* Since  $\sum_{i=1}^n y_i^{nut} A_i = y^{nut}$ , we have

$$y_k^{nut} A_k = y^{nut} - \sum_{i=1, i \neq k}^n y_i^{nut} A_i \quad (7)$$

Multiplying  $y_i^{nut}$  to the set of initial bounds for  $A_i$ , we get the set of bound for  $y_i^{nut} A_i$  such that

$$y_i^{nut} A_{i_{min}} \leq y_i^{nut} A_i \leq y_i^{nut} A_{i_{max}} \quad \text{for all } i.$$

Then by (7),

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

Since  $y_k^{nut} > 0$ , dividing each side by  $y_k^{nut}$  yields

$$A_k \leq \frac{y^{nut} - y_i^{nut} A_{i_{min}}}{y_k^{nut}}.$$

Furthermore,

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

Thus,

$$y^{nut} - \sum_{i=1, i \neq k}^n y_i^{nut} A_{i_{max}} \leq y^{nut} - \sum_{i=1, i \neq k}^n y_i^{nut} A_i,$$

and

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

Combining these inequalities with Equation (7),

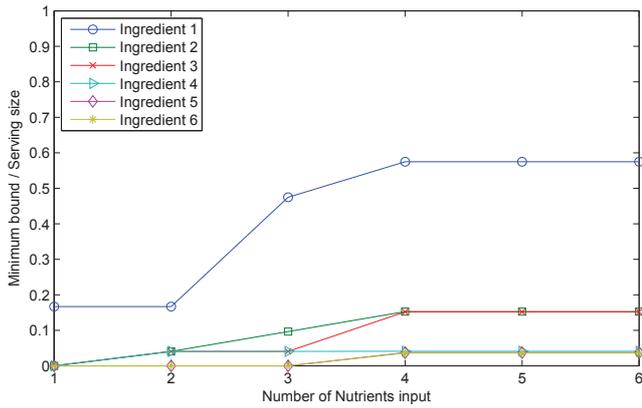
$$y^{nut} - \sum_{i=1, i \neq k}^n y_i^{nut} A_{i_{max}} \leq y_k^{nut} A_k \leq y^{nut} - \sum_{i=1, i \neq k}^n y_i^{nut} A_{i_{min}}.$$

Therefore, dividing each side by  $y_k^{nut}$ , we can conclude that

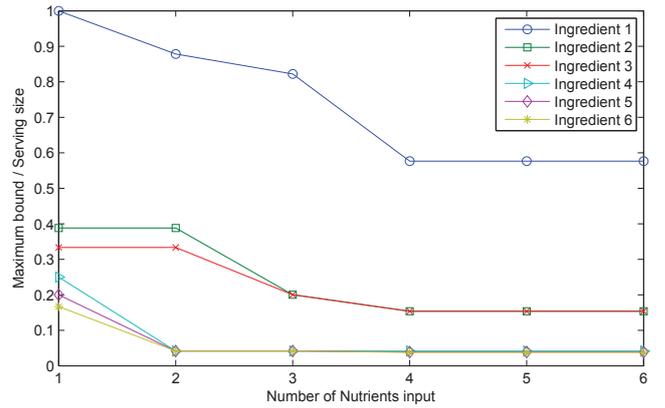
$$\begin{aligned} \frac{y^{nut} - \sum_{i=1, i \neq k}^n y_i^{nut} A_{i_{max}}}{y_k^{nut}} &\leq A_k \\ &\leq \frac{y^{nut} - \sum_{i=1, i \neq k}^n y_i^{nut} A_{i_{min}}}{y_k^{nut}}. \end{aligned}$$

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  $y_k^{nut}$  is known for all  $k$ . Otherwise, the bound remains as it is. This is not the case for the maximum bound.

To estimate the  $A_i$ 's, we first pick a set of nutrients of interest. We then apply Procedure 1 (running over all selected nutrients), followed by Procedure 2. After that, we keep repeating Procedure 3 and Procedure 4 (running over all selected nutrients), followed by Procedure 2, until our estimates change by a total of less than  $10^{-5}$  between consecutive repetitions. (In other words, we iterate until sum of the changes in all the minimum bounds and all the maximum bounds is less than  $10^{-5}$ .) Observe that, in this iterative method, the changes in the minimum and maximum bounds interact to help to reduce the gap between each other. The algorithm is described in Figure 1.



(a) Minimum bounds



(b) Maximum bounds

Fig. 2: Changes in Minimum and Maximum bounds for ingredient amounts depending on the number of nutrients input in case of Spicy brown mustard.

### C. Numerical Experiments

To directly test our method for estimating the amount of each ingredient in a commercial food, we would need to have the true ingredient amounts. For good reasons, manufacturers are unwilling to share this information. However, we can test the accuracy of our method by looking at the difference between the estimated maximum and the estimated minimum of each ingredient. If our method works well, these differences should become smaller as we consider more nutrients.

To test this, we estimated the ingredient amounts of foods using a subset (in order) of the following nutrients: protein, sodium, calories, carbohydrates, fat, and cholesterol. We first estimated the minimum and maximum ingredient amounts for a commercial food, called “Spicy Brown Mustard”. In Figure 2, we indicate the minimum and maximum bounds for each ingredient as the portions that the ingredient contributes to a serving size of the food. The numbers in the figure show that, as we took more nutrients into account, the minimum and maximum bounds for the ingredient amounts increased and decreased, respectively. Notice that an increment in minimum bounds occurs only when there exist some changes in maximum bounds. This shows that the minimum and maximum bounds for the ingredient amounts improve coherently and eventually approach each other.

This phenomenon is illustrated in Figure 3. As expected, the estimated maximum and minimum ingredient amounts tended to decrease and increase as we considered more nutrients. As a result, the range of the estimates (measured the ratio of the difference between the maximum and the minimum and the serving size in Figure 3) decreased. In some cases (e.g., Spicy Brown Mustard in Figure 3(a)), the range decreased to nearly zero ( $< 0.2\%$  of serving size) for all ingredients with only 4 nutrients. In other cases (e.g., Garlic mashed potatoes in Figure 3(b)), we failed to obtain a good estimates for some of the ingredients even though we obtained a near perfect estimate for the other ingredients with just two nutrients. Clearly, the accuracy of our method depends on the food considered and can vary for one ingredient to the next.

The experiment with one serving of a cracker named “Carr’s Whole Wheat Crackers” shows how a set of bounds converges toward each other as we iterate more. We approximated the minimum and maximum bounds for the amounts of ingredients used to prepare the cracker with six iterations described in section II-B using six nutrients (protein, sodium, calories, carbohydrates, fat, and cholesterol), presented in Table II. The initial minimum bounds of ingredient amounts were zero, except for Ingredient 1. However, as we iterate our method more times, the minimum bounds for Ingredient 2, 3, and 4 increased gradually toward their maximum bounds. At the same time, the maximum bound for each ingredient decreased.

### III. APPLICATION TO PHENYLALANINE (PHE) CONTENT ESTIMATION

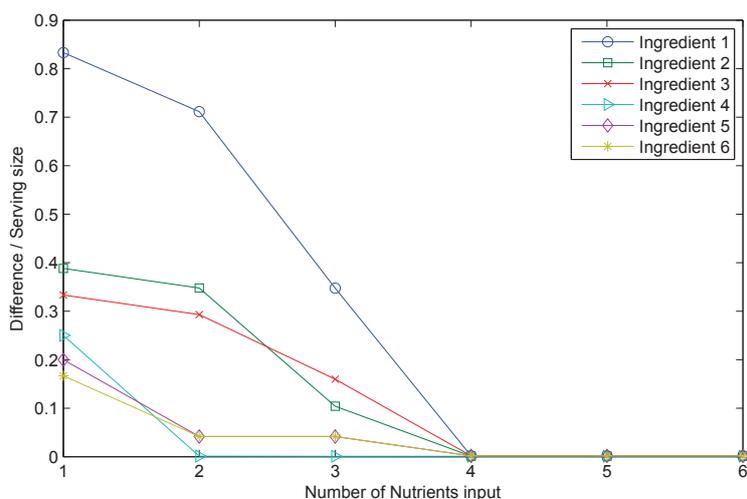
A nutrient that is not listed on the Nutrition Facts Label is the amino acid phenylalanine (Phe). If we knew  $A_i$ , the amount of ingredient  $i$ , along with  $p_i$ , the number of milligrams of Phe per gram of ingredient  $i$ , then  $p_i A_i$  would be the Phe contributed by ingredient  $i$ , and the total Phe in the food would be  $\sum_{i=1}^n p_i A_i$ . Therefore, we have the following bounds for the Phe content,

$$\sum_{i=1}^n p_i A_{i_{min}} \leq Phe \leq \sum_{i=1}^n p_i A_{i_{max}}, \quad (8)$$

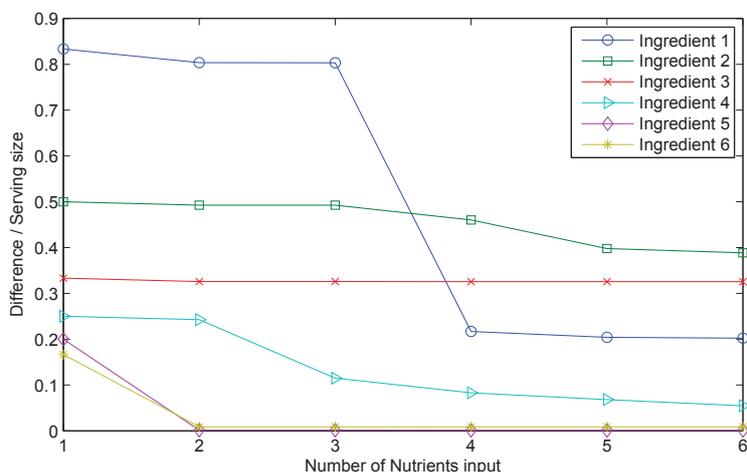
where  $A_{i_{min}}$  and  $A_{i_{max}}$  are the bounds for  $A_i$  obtained from our *approximate inverse recipe method*,  $i = 1, \dots, n$ .

#### A. Numerical Experiments

We estimated the Phe content of various commercial foods using our *approximate inverse recipe method* and six nutrients (protein, sodium, calories, carbohydrates, fat, and cholesterol). The minimum bound and the maximum bound obtained are written in parenthesis in the last column of Table III. A majority of the cases (16 out of 25), the range obtained is less than 10 mg Phe. Some of the ranges we obtained are much larger, but they could potentially be decreased by using more nutrients.



(a) Spicy brown mustard



(b) Simply potatoes, garlic mashed potatoes

Fig. 3: **Range of Estimates for Ingredient Amounts.** more nutrients are taken into account, the difference between the estimated maximum amount and the estimated minimum amount for each ingredient often decreases quickly.

For comparison, we also estimated the Phe content using the Simplex method to solve the linear problem defined by Equation (1), (2), and (3). Unfortunately, we were unable to get any result most of the time (18 out of 25). Due to an overdetermined linear system, we could not find an initial feasible solution for the foods. We indicated the case with superscript ‘c’ in Table III. However, when we did obtain a result, then the range was contained in the range we obtained with our method, as expected. This is shown in Figure 4. We can observe that the minimum and maximum bounds were below and above the equality line, respectively. Furthermore, the points in Figure 4 appear near the equality line except one maximum point and one minimum point. This implies that the results from our *approximate inverse recipe method* were fairly close to the results from the Simplex algorithm.

Table III also lists the Phe content listed in the USDA

database [1] and a low-protein food database [9] when available. We wrote ‘N/A’ when there was no data found for a food item in the database. We can observe that only a part of food items has Phe information from the databases, namely 24% (6 out of 25) of food items do not have any data from both databases.

One discrepancy between our result and ground truth observed is butter, for which the Phe content from the low-protein food database is almost half of our minimum bound. This is because part of the cream used to make butter is discarded in the preparation, and so our assumptions are not satisfied. Rice krispies cereal and waffles are also outside of the range we obtained, perhaps because of the significant drying of the ingredients in the cooking process. All other estimates (22 out of 25) are within no more than 3 mg from at least one of the database numbers.

TABLE II: Improvement of minimum and maximum bounds (in gram) for ingredient amounts contained in 17 g of Carr’s Whole Wheat Crackers As we iterate our algorithm six times, the minimum and maximum bounds respectively increase and decrease gradually.

		Ingredient 1	Ingredient 2	Ingredient 3	Ingredient 4	Ingredient 5
<b>Initial bound</b> (after Procedure 1)	max	12.47920133	4.500225011	4.500225011	0.259304651	0.259304651
	min	3.4	0	0	0	0
<b>After 1<sup>st</sup> iteration</b>	max	12.47920133	4.319626073	4.319626073	0.25901482	0.25901482
	min	7.480940676	3.17678213	0.31736349	0.254494463	0
<b>After 6<sup>th</sup> iteration</b>	max	12.47920133	4.310448532	4.310448532	0.259000667	0.259000667
	min	7.861101505	3.176791761	0.317609423	0.254510842	0

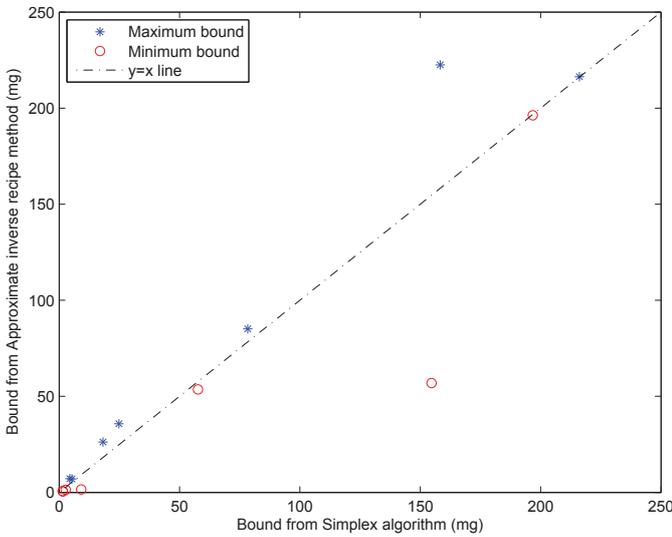


Fig. 4: Comparison with Simplex algorithm (possible only if all nutrients are known) As expected, the minimum bounds from our method were smaller than the minimum bounds from the Simplex algorithm, and the maximum bounds from our method were higher than the maximum bounds from the Simplex algorithm.

#### IV. SUMMARY AND CONCLUSIONS

We proposed a method for estimating the amount of each ingredient used to prepare a commercial food based on the Nutrition Facts Label and the Ingredient list. The method assumes that the ingredients are listed in decreasing order based on their weight and that no part of any ingredient was removed in the preparation process. The method uses a food nutrient database to look for the nutrient content of the ingredients. However, the method applies even if the nutrient contents of some ingredients are unknown or only partially known.

The output of this *approximate inverse recipe method* is a minimum bound and a maximum bound for the weight of each ingredient. The method is iterative, and takes into account a set of predetermined nutrients. As more nutrients are considered, the minimum bound and the maximum bound tend to get closer. In some cases, they become nearly identical with only

two nutrients. For a given food, the accuracy of the estimates for the different ingredients can vary. In particular, even if the amount of one ingredient can only be roughly estimated, the amount of the other ingredients may still be found accurately. Table II illustrates the effect of our iterative method: the zero minimum bounds used initially all increase to positive values after only one iteration, except for the ingredient with the smallest portion. As we iterate more, the estimates improve each other conjunctly. In other words, the increase in the minimum bounds decrease the maximum bounds, and vice versa. Furthermore, the estimates for an ingredient narrows the consecutive estimates for all other ingredients.

A direct application of our *approximate inverse recipe method* is the estimation of the nutrient content of a food. In the case where the rounded nutrient content is listed on the food label, our method can be used to obtain a more accurate estimate than the rounded number given. In the case where the nutrient content is not listed on the label, our method can be used to obtain an estimate for the nutrient content. The idea is to multiply the nutrient per weight factor for each of the ingredient (obtained from a food nutrient database) by the minimum and maximum bounds for the ingredient weight, respectively, to obtain minimum and maximum bounds for the nutrient contribution of each ingredient. (Note that this can be done even if the nutritional data for the ingredients is incomplete, as long as the nutrient per weight factor is known for that ingredient.) Adding up the contribution of all the ingredients gives an estimate for the total nutrient content of the food.

To illustrate this, we used our method to approximate the phenylalanine (Phe) content of various commercial foods. In a majority of cases, the range we obtained was less than 10 mg. When available, we compared our estimates with the USDA database [1] data and with a low-protein food database [9] data. Overall, we found a good agreements between the data found in these databases and our results (Table III).

When we were able to obtain the nutritional data for every ingredient, we also compared our estimate with that obtained by solving the linear optimization problem taking into account all the inequalities given by the nutritional data. Our implementation was based on the Simplex method. In many cases, numerical issues prevented us from getting any estimate at all with this method. In other cases, we were unable to obtain an answer with that method because we did not have the nutritional data for all the ingredients. However, when this

TABLE III: Comparison of phenylalanine content estimates obtained with our method, the full linear programming approach (Simplex Method) and two food databases.

Description ( serving size )	USDA database[1]	low-protein food database[9]	Simplex Method	Approximate inverse recipe method
Carr's Whole Wheat Crackers ( 17 g )	81.6 mg	75 mg	( 57.68 mg, 78.35 mg )	( 53.61 mg, 85.11 mg )
Ketchup ( 17 g )	4.42 mg	10.2 mg	( 1.44 mg, 4.42 mg )	( 0.70 mg, 7.09 mg )
KIT KAT Milk Chocolate ( 42 g )	113.4 mg	131.86 mg	DNE <sup>c</sup>	( 129.56 mg, 238.91 mg )
Campbell's Tomato soup ( 122 g )	68.32 mg <sup>a</sup>	66.90 mg	DNE <sup>c</sup>	( 33.21 mg, 102.91 mg )
Cheerios Cereal ( 28 g )	175.84 mg	165 mg	DNE <sup>c</sup>	( 176.42 mg, 177.00 mg )
Rice Krispies Cereal ( 33 g )	116.82 mg	107 mg	DNE <sup>c</sup>	( 91.54 mg, 94.80 mg )
Enchilada Sauce ( 60 g )	N/A	6 mg	( 1.53 mg, 24.83 mg )	( 0.41 mg, 35.69 mg )
Eggo waffle ( 70 g )	N/A	238 mg	( 196.73 mg, 216.09 mg )	( 196.26 mg, 216.35 mg )
Garlic chili pepper sauce ( 9 g )	N/A	1.93 mg	( 2.71 mg, 5.27 mg )	( 1.37 mg, 6.96 mg )
Salsa sauce ( 30 g )	N/A	11 mg	( 9.12 mg, 18.20 mg )	( 1.53 mg, 26.21 mg )
Simply potatoes Garlic mashed potatoes ( 124 g )	N/A	N/A <sup>b</sup>	( 154.71 mg, 158.29 mg )	( 56.89 mg, 222.50 mg )
Butter with Canola Oil ( 14 g )	N/A	6 mg	DNE <sup>c</sup>	( 11.88 mg, 17.66 mg )
Go-Gurt ( 64 g )	N/A	120 mg	DNE <sup>c</sup>	( 116.38 mg, 120.95 mg )
Jell-O Gelatin Snacks ( 98 g )	N/A	23.76 mg	DNE <sup>c</sup>	( 10.01 mg, 30.44 mg )
Marshmallow Peeps, Baby Chicks ( 42 g )	N/A	21 mg	DNE <sup>c</sup>	( 19.17 mg, 23.56 mg )
Ore-Ida French fries ( 84 g )	N/A	76 mg	DNE <sup>c</sup>	( 77.64 mg, 78.77 mg )
Spicy Brown Mustard ( 5 g )	N/A	8 mg	DNE <sup>c</sup>	( 8.95 mg, 9.01 mg )
Starburst Fruit Chews ( 40 g )	N/A	5.42 mg	DNE <sup>c</sup>	( 0.00 mg, 4.48 mg )
Vinaigrette Balsamic Dressing ( 31 g )	N/A	3 mg	DNE <sup>d</sup>	( 0.00 mg, 2.80 mg )
Yoplait Original Strawberry ( 170 g )	N/A	284.67 mg	DNE <sup>c</sup>	( 287.11 mg, 291.08 mg )
ALTOIDS peppermint ( 2 g )	N/A	N/A	DNE <sup>c</sup>	( 0.43 mg, 4.22 mg )
Jell-O Cheesecake Pudding Dessert ( 26 g )	N/A	N/A	DNE <sup>c</sup>	( 0.91 mg, 0.98 mg )
Sweet potato Tot ( 85 g )	N/A	N/A	DNE <sup>c</sup>	( 54.87 mg, 113.77 mg )
Taco Shells ( 32 g )	N/A	N/A	DNE <sup>c</sup>	( 36.69 mg, 38.31 mg )
Vanilla bean Ice cream ( 87 g )	N/A	N/A	DNE <sup>c</sup>	( 206.87 mg, 211.09 mg )

<sup>a</sup> Any brand Tomato soup, condensed. Not Campbell's product.

<sup>b</sup> Database has a value, but with a different protein content.

<sup>c</sup> Simplex algorithm could not find a solution

<sup>d</sup> Simplex algorithm is not applicable due to missing data.

full linear optimization approach did give us an estimate, the range of possible Phe values obtained was somewhat smaller than with our proposed approach, as expected.

One of the strength of our proposed method is that it provides a set of estimates for the nutrient content of foods for which neither the current nutrition databases nor linear programming methods would yield any information. Our work provides the first method for the automatic and accurate estimation of the Phe content of a food from its Nutrition Facts Label and Ingredient list. We have implemented our method in a web-based application and we are pleased to make this online tool available to the PKU community at <https://engineering.purdue.edu/brl/PKU/>.

#### REFERENCES

- [1] U.S. Department of Agriculture, Agricultural Research Service, "USDA national nutrient database for standard reference, release 25," Nutrient Data Laboratory Home Page, <http://www.ars.usda.gov/ba/bhnrc/ndl>, 2012.
- [2] G. L. Arnold and R. D. Steiner, "Phenylketonuria, Epidemiology," Medscape: Medscape Access. Available at: <http://emedicine.medscape.com/article/947781-overview#a0156>.
- [3] P. R. Huttenlocher, "The neuropathology of phenylketonuria: human and animal studies," *European Journal of Pediatrics*, vol. 159, no. 2, pp. S102–S106, 2000.
- [4] National Institute of Health, "Consensus development conference statement," Phenylketonuria: Screening and Management. Available at: <http://www.nichd.nih.gov/publications/pubs/pku/sub3.cfm>, October 16–18 2000.
- [5] M. Williamson, R. Koch, C. Azen, and C. Chang, "Correlates of intelligence test results in treated phenylketonuric children," *Pediatrics*, vol. 68, no. 2, pp. 161–167, 1981.
- [6] P. Burgard, "Development of intelligence in early treated phenylketonuria," *European journal of pediatrics*, vol. 159, no. 14, pp. 74–79, 2000.
- [7] S. Yannicelli and A. Ryan, "Improvements in behaviour and physical manifestations in previously untreated adults with phenylketonuria using a phenylalanine-restricted diet: a national survey," *Journal of inherited metabolic disease*, vol. 18, no. 2, pp. 131–134, 1995.
- [8] K. Matalon, P. Acosta, and C. Azen, "Role of nutrition in pregnancy with phenylketonuria and birth defects," *Pediatrics*, vol. 112, no. Supplement 4, pp. 1534–1536, 2003.
- [9] V. E. Schuett, *Low Protein Food List for PKU*, 3rd ed., 2010.
- [10] E. Saxholt, A. Christensen, A.T. and MÅller, H. Hartkopp, K. Hess Ygil, and O. Hels, "Danish food composition databank, revision 7," Department of Nutrition, National Food Institute, Technical University of Denmark. Available at: <http://www.foodcomp.dk/>, 2008.
- [11] J. Kim and M. Boutin, "A list of phenylalanine to protein ratios for common foods," ECE Technical Reports. Paper 456. Available at: <http://docs.lib.purdue.edu/ccetr/456>, 2014.
- [12] —, "New multipliers for estimating the phenylalanine content of foods from the protein content," 2014, submitted.