



A Utility Function to Solve Approximate Linear Equations for Decision Making

Kiyoshi Yoneda*, Walter Celaschi**

Abstract. Suppose there are a number of decision variables linearly related to a set of outcome variables. There are at least as many outcome variables as the number of decision variables since all decisions are outcomes by themselves. The quality of outcome is evaluated by a utility function. Given desired values for all outcome variables, decision making reduces to “solving” the system of linear equations with respect to the decision variables; the solution being defined as decision variable values such that maximize the utility function. This paper proposes a family of additively separable utility functions which can be defined by setting four intuitive parameters for each outcome variable: the desired value of the outcome, the lower and the upper limits of its admissible interval, and its importance weight. The utility function takes a nonnegative value within the admissible domain and negative outside; permits gradient methods for maximization; is designed to have a small dynamic range for numerical computation. Small examples are presented to illustrate the proposed method.

Keywords: individual behavior, inverse problems, simultaneous equations, optimization

Mathematics Subject Classification: 90B50

JEL Classification: C44

Revised: February 19, 2013

1. INTRODUCTION

We aim to devise a tool for decision making such that anyone with a rudimentary understanding of linear equations will be able to build and solve a model. The method is similar to the weighted least squares, except that instead of minimizing the usual quadratic loss function a slightly more elaborate utility function is maximized.

Consider approximate linear equations $\phi x \approx y$, where x is a vector of decision variables, y the desired outcomes, and ϕ a matrix describing the linear causality

* Department of Industrial Economics, Fukuoka University, Jounan-ku, Fukuoka, 814-0180 Japan, e-mail: yoneda@econ.fukuoka-u.ac.jp, corresponding author

** Departament of Production Engineering, Faculdades de Campinas (FACAMP), Estrada Municipal UNICAMP – Telebrás Km 1, s/n, Cidade Universitária – Campinas/SP, Brazil, e-mail: walter@flexys.com.br

relationship; $\dim x \leq \dim y$ is secured by including all equations of the form $x_j \approx y_j$. Its solution $\hat{x} := \arg \max_x U(x)$ is given by maximizing the utility function U defined as a weighted sum $U(x) := \sum w_i u(\hat{y}_i(x))$ of subutility functions u with $\hat{y}_i(x) := \sum \phi_{ij} x_j$, where w_i is a given importance weight.

Now define a subutility function $u(z) := u(z; y, a, b)$ with $y \in [a, b]$ satisfying the following requirements: u is continuously differentiable; u is quadratic in $[a, y]$ and also in $[y, b]$; u is linear for $z < a$ and $b < z$; and $u(y) = 1$, $\left. \frac{du(z)}{dz} \right|_{z=y} = 0$, with $u(a) = u(b) = 0$. Such u exists uniquely:

$$u(z; y, a, b) := \begin{cases} \frac{(c-z)(c-2y+z)}{(c-y)^2} & z \in [a, b] \\ \frac{2(c-z)}{c-y} & z \notin [a, b] \end{cases} \quad \text{where} \quad c := \begin{cases} a & z < y \\ b & y < z \end{cases} \quad (1)$$

A major advantage of this subutility function is that it has only three intuitive parameters: y , the ideal value of z ; a , the lowest acceptable value of z ; and b , the highest acceptable value of z . The method differs from the weighted least squares in that u is generally asymmetric around y , not necessarily $y - a = b - y$.

The optimization procedure permits gradient methods since u is continuously differentiable, which is important for scalability. The tails of u are chosen to be linear rather than quadratic aiming to prevent numerical overflow in a resource-constrained computation devoid of floating point arithmetics. This consideration is in the hope to facilitate implementations as embedded systems, such as for robots.

To build a model with this method one has to provide the causality matrix ϕ , and for each outcome the parameters y , a , b , and w , which have clear meanings, justifying the claim that anybody who understands linear equations can build and solve a model. Some application ideas and small numerical examples will be presented.

In a previous paper (Yoneda 2008) one of the authors proposed to solve similar problems using subutility functions defined only within the open interval $]a, b[$. After a few years of experience we concluded that the approach was impractical because:

- a feasible solution may be difficult to find when no *status quo* is available,
- no recommendation will be available when no feasible solution is found,
- evaluation of the subutility function was expensive, and
- it would cause numerical overflow near both extremes of the interval.

This paper corrects those inconveniences by eliminating the hard constraints. For further motivations of the research the readers are referred to (Yoneda 2008, Wichers 1996) and references therein.

The remainder of this paper is organized as follows. In Section 2 the linear inverse problem is briefly reviewed. Section 3 lists some attributes that the utility function should satisfy in order to formulate and solve it. The major contribution of this paper is in Section 4 which defines a subutility function to satisfy those desirable attributes. Sections 5 and 6 present small examples to illustrate the proposed methodology in detail. Section 7 discusses research directions including applications to embedded systems.

3. REQUIREMENTS

Of the subutility functions to be derived we require the following:

Asymmetry: a subutility function may be asymmetric around its mode.

Minimally specified: a subutility function may be specified with a minimal number of easily obtained parameters.

Easily optimizable: a utility function may easily be maximized with respect to the decision variables using resource-constrained computers, such as microcontrollers and digital signal processors.

An additively *separable* concave utility function is a function of the form:

$$U(\dots, z_i, \dots) := \sum_{z_i \in \mathbb{R}} w_i u_i(z_i), \quad 0 \leq w_i \leq 1, \quad \sum w_i = 1 \quad (3)$$

where the subutility functions $-\infty < u_i(z_i) \leq 1$ are *strictly concave* and \mathbb{R} is the set of real numbers. Such utility functions are known to be easily optimizable. In the subsequent part of this paper the suffix i will usually be dropped because we will be referring more to u than to U .

Consider a unimodal asymmetric strictly concave subutility function such that:

$$\begin{aligned} u(z) &= u(z; y, a, b) \\ a &< y & y < b \\ u(y) &= 1 & u(a) = u(b) = 0 \end{aligned} \quad (4)$$

permitting negative utilities beyond $[a, b]$. This specification is minimal in the sense that it would have to state at least the most desirable value y and the acceptable range $[a, b]$. A simplest implementation would be:

$$u(z; y, a, b) := \begin{cases} 1 - \frac{z - y}{a - y} & z < y \\ 1 - \frac{z - y}{b - y} & y \leq z \end{cases} \quad (5)$$

which is often found in control and artificial intelligence literatures. The maximization of U in this case would have to rely on subgradient methods which are inefficient compared to gradient methods.

4. SUBUTILITY FUNCTIONS

In order to make the utility function efficiently optimizable by permitting gradient methods we require that u be differentiable with

$$\left. \frac{du(z)}{dz} \right|_{z=y} = 0 \quad (6)$$

By requiring further that u be two-piece quadratic in $[a, y]$ and $[y, b]$ and linear outside the interval $[a, b]$, we have:

$$u(z; y, a, b) := \begin{cases} \frac{2(z - a)}{y - a} & z \leq a \\ \frac{(a - z)(a - 2y + z)}{(y - a)^2} & a < z < y \\ \frac{(b - z)(b - 2y + z)}{(y - b)^2} & y \leq z < b \\ \frac{2(z - b)}{y - b} & b \leq z \end{cases} \quad (7)$$

or, in a shorter form, (1).

The two central expressions¹ of (7), which are for the permissible range $z \in [a, b]$, follow from the conditions that:

$$u(a) = 0 \qquad u(y) = 1 \qquad u(b) = 0 \qquad \left. \frac{du(z)}{dz} \right|_{z=y} = 0.$$

The parts beyond the permissible range are linear extensions of the quadratic functions. These terminal parts are chosen to be linear rather than quadratic hoping to reduce numerical overflows: in resource-constrained computation floating points are usually unavailable, severely limiting the dynamic range of numbers.

The gradient of the utility function U may be calculated from the derivative of (1):

$$\frac{du}{dz}(z; y, a, b) = \begin{cases} \frac{2(y - z)}{(y - c)^2} & z \in [a, b] \\ \frac{2}{y - c} & z \notin [a, b] \end{cases} \quad \text{where} \quad c := \begin{cases} a & z < y \\ b & y < z \end{cases},$$

$$\nabla U = \left(\left[\dots w_i \frac{\partial u_i(z_i)}{\partial z_i} \dots \right] \phi \right)' \quad (8)$$

The procedures to follow are the same as in the weighted least squares, the only difference being that rather than minimizing the quadratic loss function, the utility function (7) is maximized. While (7) is easier than (5) to maximize permitting gradient methods, it excludes methods requiring continuous second derivatives.

An important remaining issue is how to set the weights w_i , which permits various ways. Methods such as AHP² could be used, for instance. Since there are various problems involved in setting up importance weights, an in-depth discussion would deserve a separate paper.

¹ This portion coincides with the *loss function* for the *asymmetric Gaussian distribution* (Kato et al. 2002) with the sign inverted and then shifted upwards by one: $u(z) = 1 - \text{loss}(z)$ for $z \in [a, b]$.

² Analytic hierarchy process. See for instance, http://en.wikipedia.org/wiki/Analytic_Hierarchy_Process

5. PEANUTS AND BEER EXAMPLE

5.1. PROBLEM FORMULATION

Suppose you wish to have some buttered peanuts and a beer under a budget and an energy intake specifications. Let:

$$\begin{aligned} x_1 &:= \text{peanuts amount [g = grams]}, \\ x_2 &:= \text{beer amount [ml = milliliters]}, \\ x_3 &:= \text{peanuts cost [cu = currency units]}, \\ x_4 &:= \text{beer cost [cu]}, \\ x_5 &:= \text{peanuts energy [kcal = kilocalories]}, \\ x_6 &:= \text{beer energy [kcal]}. \end{aligned}$$

Actually you wish to have a 50 g sack of peanuts with a 350 ml bottle of beer:

$$\begin{aligned} x_1 &= 50 \text{ [g]} \\ x_2 &= 350 \text{ [ml]} ; \end{aligned}$$

the bartender can adjust the quantities.

You know that both peanuts and beer require a fixed cost plus a cost proportional to the amount:

$$\begin{aligned} 1 + \frac{1}{50} x_1 & \quad \text{[cu]} \\ \frac{1}{2} & \quad + \frac{1}{100} x_2 \text{ [cu]} \end{aligned}$$

and that both have fixed calories per quantity:

$$\begin{aligned} \frac{592}{100} x_1 & \quad \text{[kcal]} \\ \frac{142}{350} x_2 & \quad \text{[kcal]} \end{aligned}$$

You are willing to pay the fixed costs of $1 + \frac{1}{2} = \frac{3}{2}$ cu to gain a total energy intake of 200 kcal:

$$\begin{aligned} \frac{3}{2} + \frac{1}{50} x_1 + \frac{1}{100} x_2 &= \frac{3}{2} \text{ [cu]} \\ \frac{592}{100} x_1 + \frac{142}{350} x_2 &= 200 \text{ [kcal]} \end{aligned}$$

But this is impossible since if the peanuts are $x_1 = 50$ g and the beer is $x_2 = 350$ ml the cost and the energy will be:

$$\begin{aligned} \frac{3}{2} + \frac{1}{50} \times 50 + \frac{1}{100} \times 350 &= 6 \text{ [cu]} \\ \frac{592}{100} \times 50 + \frac{142}{350} \times 350 &= 438 \text{ [kcal]} \end{aligned}$$

So you relax the requirement specifications to:

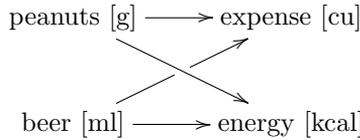
Peanuts should ideally be 50 g, hopefully between 30 and 100.

Beer should ideally be 350 ml, hopefully between 100 and 500.

Cost should ideally be $\frac{3}{2}$ cu, hopefully between 0 and 5.

Energy should ideally be 200 kcal, hopefully between 150 and 300.

Now you decide that you will try to fulfill your goals as much as possible by controlling the amounts to eat x_1 and drink x_2 rather than by controlling your expense or energy intake. This means that the causality relationship has been established as:



with the arrows indicating the direction of causality, so that the system of approximate equations has to be solved with respect to $x = [x_1, x_2]'$.

Furthermore, you give a clearer meaning to the near-equalities “ \approx ” by assigning importance weights to each outcome, reflecting your priorities. Suppose they are $w := \frac{1}{10} [3, 3, 1, 3]'$. Then, the entire situation may be restated as:

$$\begin{aligned}
 x_1 & \overset{.3}{\approx} 50 \in [30, 100] \text{ [g]} \\
 x_2 & \overset{.3}{\approx} 350 \in [100, 500] \text{ [ml]} \\
 \frac{3}{2} + \frac{1}{50} x_1 + \frac{1}{100} x_2 & \overset{.1}{\approx} \frac{3}{2} \in [0, 5] \text{ [cu]} \\
 \frac{592}{100} x_1 + \frac{142}{350} x_2 & \overset{.3}{\approx} 200 \in [150, 300] \text{ [kcal]}
 \end{aligned}$$

or, subtracting the constant term $\frac{3}{2}$ from both sides of the third equation:

$$\begin{array}{ccc}
 \phi x \overset{w}{\approx} y & & \text{[unit } a \quad b \quad w] \\
 \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ .02 & .01 \\ 5.92 & .406 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \overset{w}{\approx} \begin{bmatrix} 50 \\ 350 \\ 0 \\ 200 \end{bmatrix} & & \begin{bmatrix} \text{g} & 30 & 100 & .3 \\ \text{ml} & 100 & 500 & .3 \\ \text{cu} & -1.5 & 3.5 & .1 \\ \text{kcal} & 150 & 300 & .3 \end{bmatrix} \quad (9)
 \end{array}$$

5.2. UTILITY FUNCTION

The subutility function for beer looks like in Figure 1. Note that the subutility goes negative for beer less than 100 ml and over 500 ml. For comparison, the subutility function corresponding to (5) is illustrated in dashed lines. Similar subutility functions exist for all other equations, *viz.* for the peanuts, cost, and the energy intake.

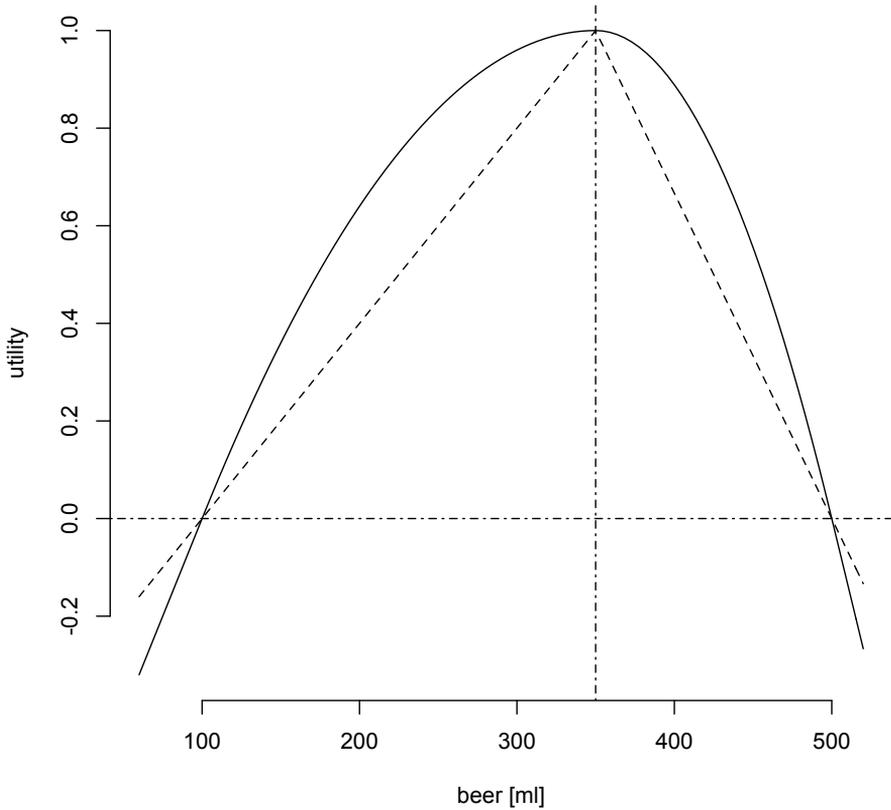


Fig. 1. *The subutility function for the beer.
The ideal amount of beer is 350 ml*

The utility function is a weighted sum of all those subutility functions:

$$\begin{aligned}
 U(x) &:= \sum_i w_i u(\phi_{i\bullet} x; y_i, a_i, b_i) \\
 &= .3 u(x_1; 50, 30, 100) + \\
 &\quad .3 u(x_2; 350, 100, 500) + \\
 &\quad .1 u(.02 x_1 + .01 x_2; 0, -1.5, 3.5) + \\
 &\quad .3 u(5.92 x_1 + .406 x_2; 200, 150, 300)
 \end{aligned}$$

A diagrammatical representation is in Figure 2.

Since there are only two decision variables, the quantities $x = [x_1 \ x_2]'$ of peanuts and beer, the utility function $U(x) = U(x_1, x_2)$ may be illustrated as in Figure 3. The portion below the flat surface of zero utility has negative utilities.

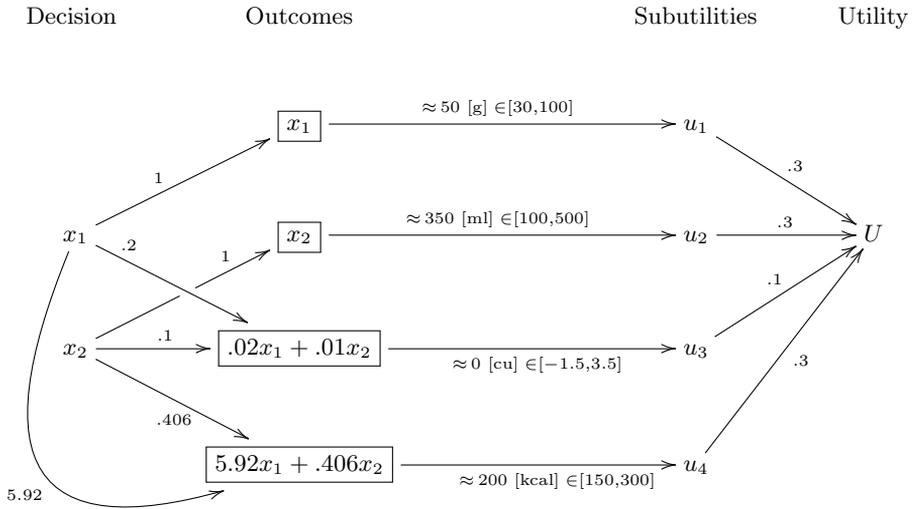


Fig. 2. Formulation

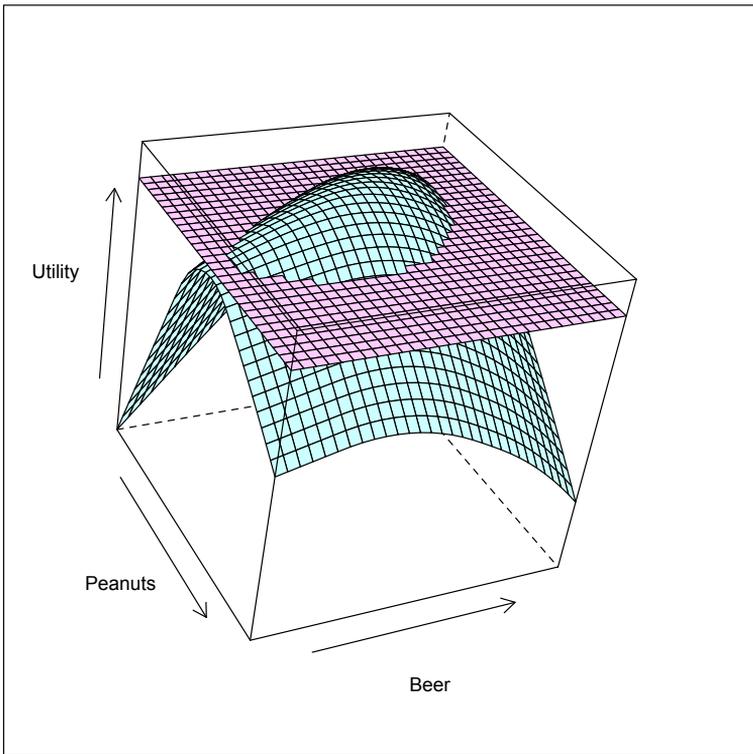


Fig. 3. The utility as a function of the amounts of peanuts and beer. Peanuts range from 0 to 80 g; beer ranges from 0 to 500 ml

5.3. RESULTS

The input items required for obtaining a result are all that are listed in (9) except the vector x of decision variables. The corresponding result is summarized in Table 1 for easy comparison of the recommendation against the original plan.

Table 1. *Beer specs against recommendation*

Item	Input				Output			
	y	a	b	w	\hat{y}	u_0	u	wu
Peanuts	50	30	100	.3	34	1	.386	.116
Beer	350	100	500	.3	150	1	.362	.109
Cost	0	-1.5	3.5	.1	2.19	0	.609	.061
Energy	200	150	300	.3	264	-2.76	.588	.176
Total								.462

The “Input” columns list the input except the technology coefficient matrix ϕ . The “Output” columns consist of:

- \hat{y} = recommended decision and corresponding outcome,
- u_0 = subutilities attained by the ideal x disregarding its outcome,
- u = subutilities attained by the recommended x ,
- wu = subutilities u above multiplied by the item’s importance weight w , with the utility $U = w'u = \sum w_i u_i$.

Items with unacceptable outcomes are easily identifiable by checking negative utilities in columns u_0 and u or wu . It is clear at a glance that the recommended decision $\hat{x} = [34 \ 150]'$ is superior to the initial decision $x_0 = [50 \ 350]'$ since u has no negative while u_0 has.

5.4. INTERPRETATION

Table 1 suggests that your utility will be maximized if you have peanuts of $x_1 = 34$ g and a beer of $x_2 = 150$ ml for a cost of $\frac{3}{2} + \frac{1}{50} x_1 + \frac{1}{100} x_2 = 3.7$ cu and an energy intake of $\frac{592}{100} x_1 + \frac{142}{350} x_2 = 264$ kcal at an utility value of $U(x) = .462$, which may be interpreted as “46% satisfied” against the 100% for the fictitious case in which all requirements are fully met. The largest source of frustration under this suggestion would be the amount of beer, which you get less than a half the desired amount. The next source of frustration is the amount of peanuts which you have only slightly more than the minimum. In compensation, you don’t have serious complaints regarding the cost and the energy intake.

It is worth noting that the information contained in the basic model (9) is enough to derive the consumer’s demand curve by changing the beer price, currently set at 0.01 cu/ml, through some range of interest. This is a great advantage in marketing if the consumer’s behavior is desired to be predicted. In this particular example the demand curve is found to be nearly flat, only slightly decreasing with respect to the price, indicating that the price is unimportant.

6. PRODUCTION PLANNING EXAMPLE

This section presents a problem with a flavor in production planning:

- You run a factory that produces plastics, hard and soft. You need to decide how much of them you will produce this coming week.
- The sales department says that they want 4 and 6 [t = tons] of hard and soft plastics. Since you have product stocks and some space in the warehouse, the quantities produced would be acceptable if they are between 4.5 and 5 t for hard and between 4 and 7 t for soft plastics.
- The purchase department says that they will have 10 t raw material available, which serves to produce both hard and soft plastics. Considering the relationship with the supplier, it is undesirable to order less than 8 t. Also, it will be difficult to prepare more than 13 t even considering purchase not only from the supplier but also from the market.
- The personnel department says that they have 8 [p = persons] available for the next week. It is possible, however, to adjust it between 7 and 9 p by reducing or extending work hours, without hiring or firing.
- The engineering department says that they have 15 [m = machines] leased for production, but can reduce to 12 m or increase to 17 m by adjusting operation time and machine speed.
- The production department says that in order to produce 1 t of hard plastics you need 2 t material, 1 p labor, and 3 m machines, while to produce 1 t of soft plastics you need 1 t material, 1 p labor, and 1 m machine.

This last paragraph means that if the decision variables are the amounts of hard and soft plastics produced, then the technical coefficient matrix is:

$$\phi := \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \\ 1 & 1 \\ 3 & 1 \end{bmatrix}$$

It is clear that the production as the sales department suggests is impossible, since the material, labor, and machines will all be short.

As the factory head you weigh the importance of each department's statement and come to conclude that their relative weights are $w := [.2 \ .2 \ .1 \ .3 \ .2]'$. By the department statements a specifications table may be prepared, which completes input necessary to run the software for recommendation. Table 2 shows the specifications against recommendation.

Table 2. *Plastics specs against recommendation*

Item	Input				Output			
	y	a	b	w	\hat{y}	u_0	u	wu
Hard	4	3.5	5	.2	3.8	1	.815	.163
Soft	6	4	7	.2	4.3	1	.290	.058
Material	10	8	13	.1	11.9	-.667	.605	.061
Labor	8	7	9	.3	8.1	-2	.990	.297
Machine	15	12	17	.2	15.7	-1	.888	.178
Total								.756

Following the recommendation, which suggests to produce 3.8 t hard and 4.3 t soft plastics, makes all requirements fall into the permissible range with utility .756. The utility function is as in Figure 4.

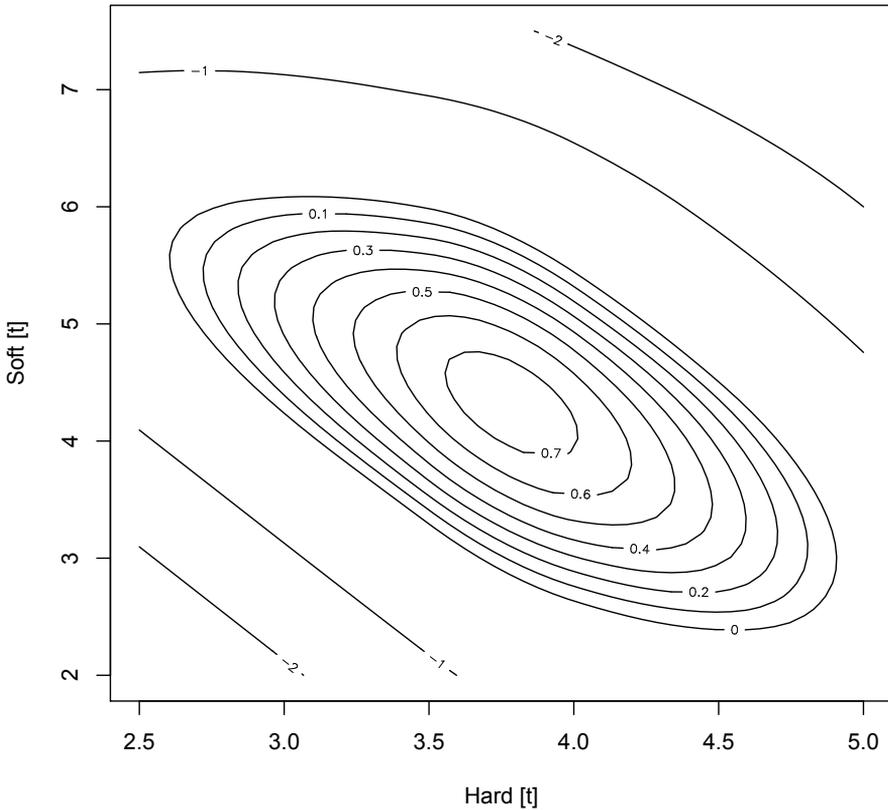


Fig. 4. *The utility function for plastics*

7. DISCUSSION

A practical advantage of the proposed method is that the model can be built, if necessary, entirely subjectively in the absence of objective data. The present implementation is in R (R Core Team 2012), in which all function graphs in this paper have been generated.

One of the authors has given a one-semester course on linear equations including the weighted least squares and the method proposed in this paper, with numerical computation using R. The class consisted of about 40 students, 3/4 of which being second-year undergraduates in economics with no previous exposure to linear algebra or computer programming. The rest of the students were mixed, ranging from graduate level to near-dropouts. Here is a part of a problem given in the final exam:

You wish to employ fifty skilled and three hundred unskilled workers. A skilled worker receives double the salary of an unskilled. A skilled worker's productivity is three times that of an unskilled. You are interested in the numbers of skilled and unskilled workers, the total salary to pay, and the number of products to deliver. Assuming that you adjust them by deciding upon the numbers of skilled and unskilled workers, describe a matrix relating the decision variables to the outcome variables, disregarding the fixed costs.

This requires at least an elementary understanding of matrix in order to answer. Among about 30 who took the exam only 1/3 was able to come up with the correct answer³. It seems to suggest that a serious stumbling block for a student to build a model of her own is in setting up a technical coefficient matrix ϕ . Incidentally, some understanding of optimization was helpful but not essential in learning.

Embedded systems, notably robotics among them, is a potential field of application of this method as has been mentioned in Section 1. Consider an autonomous system, for instance a house cleaning robot. The robot makes decisions to fulfill conflicting desires: it may want to minimize its energy consumption while maximizing the amount of dust collected. If the robot behaves according to a microeconomic model, its behavior may be hoped to be understandable: a utility maximizing robot should be easier to debug than a robot behaving according to a black box decision mechanism, especially under a social situation in which many robots interact.

An inevitable research direction is in how to set up the importance weights, as has been touched upon in Section 4. In weighted least squares the importance weights are set to be a^{-2} assuming $y = 0$ and $-a = b$, meaning that all independent outcomes are considered with equal importance after having each of them standardized to mean zero and standard deviation one. This is reasonable because in this case both the permissible interval and the numerical accuracy have the same representation, which is the standard deviation. In our case, however, the permissible interval and the numerical accuracy have separate representations, hence calls for a new method for importance weight setting.

An interesting way to set up the weights would be to “solve” (3) with respect to w_i , given instances of u_i , which may be computed, and U values, which may be

³ A natural response seems to be: “since the skilled workers are more efficient than the unskilled, forget about the unskilled: employ only the skilled”.

given subjectively. This problem is dual to (2) in the sense that both take the same form, hence no extra methodology or software will be necessary.

ACKNOWLEDGMENTS

The authors are grateful to an anonymous referee for valuable suggestions.

REFERENCES

- Kato, T., Omachi, S., Aso, H., 2002. *Asymmetric gaussian and its application to pattern recognition*. In: Joint IAPR International Workshops on Syntactical and Structural Pattern Recognition and Statistical Pattern Recognition (S+SSPR2002). Barcelona, pp. 405–413, www.iic.ecei.tohoku.ac.jp/~kato/papers/t.kato_spr2002a.pdf, viewed 2012-02-15.
- R Core Team, 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, www.R-project.org/
- Wichers, R., 1996. *A Theory of Individual Behavior*. Academic Press.
- Yoneda, K., 2008. A loss function for box-constrained inverse problems. *Decision Making in Manufacturing and Services* **2** (1–2), pp. 79–98, www.dmms.agh.edu.pl/Volume_2/DMMS_2_Yoneda.pdf