Journal of Uncertain Systems
Vol.9, No.2, pp.139-143, 2015
Online at: www.jus.org.uk

Why Convex Homotopy is Very Useful in Optimization: A Possible Theoretical Explanation

Nicholas Sun*

Rutgers, the State University of New Jersey, 96 Davidson Road, New Brunswick, New Jersey 08854, USA

Received 1 June 2014; Revised 15 June 2014

Abstract

One of the efficient methods of optimizing a complex function is a *homotopy method*, when we first optimize a simpler approximate objective function and then gradually adjust this solution by solving intermediate optimization problems obtained by an appropriate combination of the original and simplified objective function, until we reach the desired maximum of the original objective function. The success of this method depends on the selection of the appropriate combination function. Empirically, the most successful combination function is the *convex homotopy* function, i.e., a convex combination of the original and simplified objective functions. In this paper, we provide a possible theoretical explanation for this empirical success.

Keywords: homotopy method in optimization, convex homotopy function, theoretical explanation

1 Formulation of the Problem

Homotopy methods in optimization: brief reminder. In many practical problem, we want to find the best possible solution: the best possible design, the best possible control, the clustering which is most consistent with our intuition, etc.

To describe this problem in formal terms, we need to be able to gauge the quality of an alternative x by an appropriate number $F(x)$ describing how good is this alternative. Once the corresponding function $F(x)$ is defined, selecting the best alternative means finding an alternative x for which the value $F(x)$ is the largest possible.

In many cases, the function $F(x)$ (describing our preferences) is very complex, so it is difficult to optimize this function.

We can simplify the problem and find an approximate function $G(x)$ which is easier to optimize. The alternative x which optimizes the approximate function $G(x)$ can be viewed as an approximate solution to the original optimization problem $F(x) \to \max$. How can we use this *approximate* solution to find the *exact* (or, at least, more accurate) solution to the original optimization problem?

A widely used method for doing this is known as the *homotopy method*; see, e.g., [4]. This method is based on the fact that once we know a solution to an optimization problem, this usually helps us to find a solution to a "nearby" optimization problem as well.

To use this idea, we form a continuous family of functions $H(\lambda, x) = f(\lambda, F(x), G(x))$ with a parameter $\lambda \in [0, 1]$ that starts, for $\lambda = 0$, at the simplified objective function $H(0, x) = G(x)$ and ends up, for $\lambda = 1$, at the desired complex objective function $H(1, x) = F(x)$.

We then select a sequence of real numbers $\lambda_0 = 0 < \lambda_1 < \cdots < \lambda_k = 1$ for which all the differences $\lambda_{i+1} - \lambda_i$ are small, and sequentially solve the optimization problem $H(\lambda_i, x) \to \max$ for $i = 0, 1, \ldots, k$.

- We start by solving the optimization problem corresponding to $\lambda_0 = 0$ (i.e., to the objective function $G(x) = H(\lambda_0, x)$), and getting an alternative $x(\lambda_0)$ which optimizes this simpler-to-optimize objective function $G(x)$.

*Corresponding author.
Email: nicholas.sun@rutgers.edu (N. Sun).

- For each i, once we know the solution $x(\lambda_i)$ which optimizes the objective function $H(\lambda_i, x) \to \max$, we use this solution to find a solution $x(\alpha_{i+1})$ to the "close" optimization problem $H(\lambda_{i+1}, x) \to \max$.

Once we get to the value $i = k$, we thus have a solution $x(1)$ to the desired problem of maximizing the objective function $F(x) = H(\lambda_k, x) = H(1, x)$.

Convex homotopy. The success of a homotopy method depends on the proper selection of the combination function $f(\lambda, a, b)$ which is used to form the appropriate family. The most widely used function is a function

$$f(\lambda, a, b) = \lambda \cdot a + (1 - \lambda) \cdot b, \tag{1}$$

known as *convex homotopy* function.

Problem. In many cases, the convex homotopy function leads to a successful solution to the original optimization problem. However, while this success is an empirical fact, there seems to be no convincing theoretical explanation for this empirical success – and thus, it is not clear whether the convex homotopy function is indeed the best function or there may be other functions $f(\lambda, a, b)$ which may be even better (so that the convex homotopy function is just a first approximation to a – yet to be determined – better function).

What we do in this paper. In this paper, we provide a possible theoretical explanation for the empirical success of the convex homotopy function.

2 Analysis of the Problem

We need a family of homotopy functions. As we can see from our description of the homotopy method, the exact parametrization of different functions $f(\lambda, a, b)$ is not that important; what is important is that we have a family of functions $g(a, b) \stackrel{\text{def}}{=} f(\lambda, a, b)$ corresponding to different values of the parameter λ.

Reasonable properties of functions from the homotopy family must satisfy. What properties should these functions $g(a, b)$ satisfy?

First, in the case when the original objective function $F(x)$ is already simple, i.e., if $G(x) = F(x)$ for all x, then it is reasonable to require that the above procedure do not force us to perform any unnecessary job of optimizing any other objective function. In other words, for each of the homotopy functions $g(a, b)$, the resulting objective function $H(x) = g(F(x), G(x)) = g(F(x), F(x))$ should coincide with $F(x)$: $g(F(x), F(x)) = F(x)$).

This property should be satisfied for all possible numerical values of $F(x)$. Thus, we must have $g(a, a) = a$ for all real values s.

Another reasonable property is related to the known fact that smooth (differentiable) functions are easier to optimize than non-differentiable ones: starting from the gradient methods, many successful optimization techniques require differentiability; see, e.g., [1]. It is therefore reasonable to require that if $F(x)$ and $G(x)$ are differentiable functions, then the function $H(x) = g(F(x), G(x))$ should also be differentiable. In particular, for $F(x) = x_1$ and $G(x) = x_2$, this implies that the homotopy function $g(x_1, x_2)$ itself be differentiable.

Finally, let us go back to the need for gauging the user preferences. In modern decision theory (see, e.g., [3, 7, 9, 6]), a natural way to gauge user preferences is to use *utility functions*, and utility functions are determined modulo a linear transformation $u \to k \cdot u + m$ where $k > 0$; see Appendix for details.

This is a typical situation in measurement theory. For example, the numerical expression of the moment of time is also only determined modulo a linear transformation: it depends on the starting point (birth of Jesus Christ or the French Revolution) and on the measuring unit (year, second, etc.); see, e.g., [8].

The corresponding linear transformation does not change the meaning of the optimization problems. This is similar to the fact in financial problem, maximizing the profit in dollars $F(x)$ is exactly the same optimization problem as maximizing the profit in Euros $F'(x) \stackrel{\text{def}}{=} k \cdot F(x)$, where k is the current cost of 1 US dollar in Euros.

It is therefore reasonable to require that the transformation corresponding to each homotopy function should not depend on this choice of scale. In other words, if $H(x) = g(F(x), G(x))$, then for each $k > 0$ and m, for the re-scaled functions $F'(x) = k \cdot F(x) + m$, $G'(x) = k \cdot G(x) + m$, and $H'(x) = k \cdot H(x) + m$, we should have the same dependence: $H'(x) = g(F'(x), G'(x))$.

Now, we are ready to formulate our main result.

3 Definitions and the Main Result

Definition. *Let us call a differentiable function $g(a,b)$ of two variables a* reasonable homotopy function *if it is satisfies the following two properties:*

- *$g(a,a) = a$ for all real numbers a;*
- *for all real numbers F, G, H, $k > 0$, and m, if $H = g(F,G)$, $F' = k \cdot F + m$, $G' = k \cdot G + m$, and $H' = k \cdot H + m$, then $H' = g(F',G')$.*

Comment. One can easily check that for every real number λ, the function $g(a,b) = \lambda \cdot a + (1-\lambda) \cdot b$ is a reasonable homotopy function (in the sense of the above Definition). It turns out that these are the only reasonable homotopy functions.

Proposition. *Every reasonable homotopy function has the form $g(a,b) = \lambda \cdot a + (1-\lambda) \cdot b$ for some real number λ.*

Discussion. This result provided the desired theoretical justification for the convex homotopy function.

Comment. We are analyzing *generic* homotopy functions, i.e., homotopy functions which can be applied to all possible optimization problems. For *specific* optimization functions, different homotopy functions – which take the specific character of the corresponding optimization problem into account – are sometimes better than the convex one; see, e.g., [2, 5].

4 Proof

1°. By definition of a reasonable homotopy function $g(a,b)$, we have $g(a,b) = a$ when $a = b$. To complete our description of the function $g(a,b)$, we thus need to find the values $g(a,b)$ for the cases when $a < b$ and when $a > b$. Let us consider these two cases one by one.

2°. Let us first consider the case when $a < b$.

The simplest such case is when $a = 0$ and $b = 1$. Let us denote the corresponding value $g(0,1)$ by ν.

Let us now take any other pair (a,b) with $a < b$, and let us use the second property of a reasonable homotopy function to find $g(a,b)$. We will apply this property to values $F = 0$, $G = 1$, $H = \nu$, $m = a$, and $k = b - a > 0$. By definition of ν as $g(0,1)$, we have $H = g(F,G)$. Here, $F' = k \cdot F + m = m = a$,

$$G' = k \cdot G + m = (b-a) + a = b,$$

and

$$H' = k \cdot H + m = (b-a) \cdot \nu + a = a \cdot (1-\nu) + b \cdot \nu.$$

For $\mu \stackrel{\text{def}}{=} 1 - \nu$, we have $\nu = 1 - \mu$ and thus, $H' = \mu \cdot a + (1-\mu) \cdot b$. From the second property of a reasonable homotopy function, we conclude that $H' = g(F',G')$, i.e., that

$$g(a,b) = \mu \cdot a + (1-\mu) \cdot b. \quad (2)$$

3°. Let us now consider the remaining case when $a > b$.

The simplest such case is when $a = 1$ and $b = 0$. Let us denote the corresponding value $g(1,0)$ by λ.

Let us now take any other pair (a,b) with $a > b$, and let us use the second property of a reasonable homotopy function to find $g(a,b)$. We will apply this property to values $F = 1$, $G = 0$, $H = \lambda$, $m = b$, and $k = a - b > 0$. By definition of λ as $g(1,0)$, we have $H = g(F,G)$. Here,

$$F' = k \cdot F + m = (a-b) + b = a,$$

$G' = k \cdot G + m = m = b$, and

$$H' = k \cdot H + m = (a-b) \cdot \lambda + b = \lambda \cdot a + (1-\lambda) \cdot b.$$

From the second property of a reasonable homotopy function, we conclude that $H' = g(F', G')$, i.e., that

$$g(a, b) = \lambda \cdot a + (1 - \lambda) \cdot b. \tag{3}$$

4°. The function $g(a, b)$ should be differentiable, in particular, it should be differentiable when $a = b$, i.e., there should be a limit

$$\frac{\partial g}{\partial a} = \lim_{h \to 0} \frac{g(b + h, b) - g(b, b)}{h}.$$

Here:

- When $h < 0$, we have $a = b + h < b$ and thus, from the formula (2), we conclude that the limit is equal to μ.
- When $h > 0$, we have $a = b + h > b$ and thus, due to the formula (3), the limit is equal to λ.

Since we should have the same limit for $h < 0$ and for $h > 0$, this means that $\mu = \lambda$, and thus, the formula (3) describes the function $g(a, b)$ for all possible pairs (a, b).

The proposition is proven.

Acknowledgments

This work was performed when the authors was working at the the University of Texas at El Paso's National Center of Border Security and Immigration, a Center of Academic Excellence under the Department of Homeland Security.

The author would also like to thank Shahriar Hossain, Vladik Kreinovich, and Luc Longpré for their invaluable guidance.

Appendix: The Main Notions of Utility: Brief Reminder

What is utility. How can we describe the human preferences $A \prec B$ between alternatives A and B in numerical terms? A natural idea is to select two alternatives: a very bad alternative A_0 and a very good alternative A_1, so that every reasonable alternative is in between A_0 and A_1: $A_0 \prec A \prec A_1$. For example, in financial situations, A_0 can be "I lose everything", while A_1 can be "I gain billion dollars".

For each real number p from the interval $[0, 1]$, we can form a lottery in which we get A_1 with probability p and A_0 with the remaining probability $1 - p$. We will denote this lottery by $L(p)$.

When $p = 0$, the lottery $L(p)$ means that we get the bad alternative A_0. When $p = 1$, we get the good alternative A_1. The larger the probability p that we get A_1, the better: if $p < p'$ then $L(p) \prec L(p')$. Thus, we get a continuous scale ranging from the very bad alternative A_0 to the very good alternative A_1.

For every alternative A, we have $A_0 = L(0) \prec A$ and thus, $L(p) \prec A$ for small p. Similarly, we have $A \prec A_1 = L(1)$ and thus, $A \prec A(p)$ for $p \approx 1$. So,

- for small p, the alternative A is better than the lottery $L(p)$; while
- for large p, the lottery $L(p)$ is better than the alternative A.

As p increases, the lottery $L(p)$ becomes better and better; thus, there should be a point p_0 at which we switch from $L(p) \prec A$ to $A \prec L(p)$. This threshold value p_0 is called the *utility* of the alternative A and usually denoted by $u(A)$. We can say that the alternative A is neither better than $L(p_0)$ nor worse than this lottery, so A is *equivalent* to this lottery: $A \sim L(p_0) = L(u(A))$.

How the numerical value of utility depends on the choice of A_0 and A_1. In principle, we can choose a different pair (A'_0, A'_1) of a very bad and a very good alternatives. Based on this new pair, we will form different utility values $u'(A)$. What is then the relation between the old utility values $u(A)$ and the new utility values $u'(A)$?

Let us first consider the case when $A'_0 \prec A_0 \prec A_1 \prec A'_1$. In this case, since both A_0 and A_1 are in between A'_0 and A'_1, each of these alternatives is equivalent to the corresponding lottery:

- the alternative A_0 is equivalent to the lottery in which we get A'_1 with probability $u'(A_0)$ and A'_0 with the remaining probability $1 - u'(A_0)$; and
- the alternative A_1 is equivalent to the lottery in which we get A'_1 with probability $u'(A_1)$ and A'_0 with the remaining probability $1 - u'(A_1)$.

The alternative A is equivalent to a lottery in which we get A_1 with probability $u(A)$ and A_0 with the remaining probability $1 - u(A)$.

Thus, the original alternative A is equivalent to a complex lottery in which:

- first, we select A_1 with probability $u(A)$ and A_0 with probability $1 - u(A)$; and then,
- depending on the first choice of A_i, we select A'_1 with probability $u(A_i)$ and A'_0 with the remaining probability $1 - u(A_i)$.

As a result of this complex lottery, we get either A'_1 or A'_0. The probability $P(A'_1)$ of selecting A'_1 can be computed by using the complete probability formula as follows:

$$P(A'_1) = P(A'_1 \mid A_1) \cdot P(A_1) + P(A'_1 \mid A_0) \cdot P(A_0) = u'(A_1) \cdot u(A) + u'(A_0) \cdot (1 - u(A)).$$

This formula can be equivalently rewritten as

$$P(A'_1) = k \cdot u(A) + m,$$

where $k \stackrel{\text{def}}{=} u'(A_1) - u'(A_0)$ and $m \stackrel{\text{def}}{=} u'(A_0)$.

Thus, the alternative A is equivalent to a lottery in which we get A'_1 with probability $P(A'_1) = k \cdot u(A) + m$ and we get A'_0 with the remaining probability $1 - P(A'_1)$. By definition of utility, this means that in the new scale, the utility $u'(A)$ of the alternative A is equal to $P(A'_1)$, i.e., that

$$u'(A) = k \cdot u(A) + m$$

for some real values $k > 0$ and m which do not depend on A.

We have shown this relation for the case when $A'_0 \prec A_0 \prec A_1 \prec A'_1$. In all other cases, we can construct the alternatives A''_0 and A''_1 for which $A''_0 \preceq A_0 \prec A_1 \preceq A''_1$ and $A''_0 \preceq A'_0 \prec A'_1 \preceq A''_1$. For example:

- as A''_0, we can take the worst of the alternatives A_0 and A'_0; and
- as A''_1, we can take the best of the alternatives A_1 and A'_1.

In this case, from $A''_0 \preceq A_0 \prec A_1 \preceq A''_1$ and $A''_0 \preceq A'_0 \prec A'_1 \preceq A''_1$, we conclude that both utilities $u(A)$ and $u'(A)$ can be obtained from $u''(A)$ by an appropriate linear re-scaling – thus, they are linearly related to each other as well.

So, utilities $u(A)$ and $u'(A)$ corresponding to different scales are indeed connected with each other by a linear transformation $u'(A) = k \cdot u(A) + m$ for some $k > 0$ and m.

References

[1] Burden, R.L., and J.D. Faires, *Numerical Analysis*, Brooks/Cole, Cengage Learning, Boston, Massachusetts, 2011.

[2] Easterling, D.R., Hossain, M.S., Watson, L.T., and N. Ramakrishnan, Probability-one homotopy maps for tracking constrained clustering solutions, *Proceedings of the ACM High Performance Computing Symposium*, article no.17, 2013.

[3] Fishburn, P.C., *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

[4] Floudas, C.A., and P.M. Pardalos, *Encyclopedia of Optimization*, vol.2, Springer Verla, Berlin, Heidelberg, New York, 2008.

[5] Ji, S., Watson, L.T., and L. Carin, Semisupervised learning of Hidden Markov Models via a homotopy method, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.31, no.2, pp.275–287, 2009.

[6] Kreinovich, V., Decision making under interval uncertainty (and beyond), *Human-Centric Decision-Making Models for Social Sciences*, pp.163–193, 2014.

[7] Luce, R.D., and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[8] Rabinovich, S.G., *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.

[9] Raiffa, H., *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.