2019-01-01

Decision Making Under Uncertainty With Applications To Geosciences And Finance

Laxman Bokati
University of Texas at El Paso, laxmanbokati@gmail.com

DECISION MAKING UNDER UNCERTAINTY WITH APPLICATIONS

TO GEOSCIENCES AND FINANCE

LAXMAN BOKATI

Master's Program in Computational Science

APPROVED:

Vladik Kreinovich, Ph.D., Chair

Martine Ceberio, Ph.D.

Aaron Velasco, Ph.D.

Stephen L. Crites, Ph.D.
Dean of the Graduate School

DECISION MAKING UNDER UNCERTAINTY WITH APPLICATIONS TO GEOSCIENCES AND FINANCE

by

LAXMAN BOKATI

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Computational Science Program

THE UNIVERSITY OF TEXAS AT EL PASO

May 2019

Acknowledgements

Research cannot be done alone, it requires support of many people to perform research. Although it is not possible to name every individual that has directly or indirectly helped shape up this thesis, I have to thank certain people that were integral to completion of this thesis.

First and foremost, it would not have been possible for me to complete this thesis without remarkable support of my research advisor Prof. Vladik Kreinovich. Prof. Kreinovich is undoubtedly the best professor I have ever had in my academic career. His constant support and encouragement is the driving force behind success of this thesis work.

Besides my advisor, my sincere gratitude also goes to rest of my thesis committee: Prof. Aaron Velasco and Dr. Martine Ceberio, for their enlightening and insightful suggestions that helped finalize this thesis.

I would also like to thank US National Science Foundation for supporting the work done in this thesis via grant HRD-1242122 (Cyber-ShARE Center of Excellence).

I also have to thank Dr. Olga Kosheleva for all the help and suggestions without which this thesis would not have been complete.

Special thanks to Pawan Koirala, Kamal Nyaupane, Neelam Dumre, Bibek Aryal, and Subharaj Ranabhat for assisting me to organize my thesis presentation.

Finally, I must record a debt to my father Damber Dutta Bokati and mother Ishwori Devi Bokati as well as my brother Prakash Bokati for their constant inspiration, love and support without which I would not have been able to pursue my academic goals.

Abstract

In many practical situations, we need to make a decision. In engineering, we need to decide on the best design of a system, and, for existing systems – on the best control strategy. In financial applications, we need to decide what is the best way to invest money. In geosciences, we need to decide whether we should explore a possible mineral deposit – or whether we should perform more experiments and measurements (and what exactly).

In some cases, we can compute the exact consequences of each decision – e.g., if we are controlling a satellite. However, in many other cases, we do not know the exact consequences. In such situations, we need to make a decision under uncertainty.

In many application areas, uncertainty is small – and can be made even smaller by appropriate measurements. For example, when we control a self-driving car, if there is an uncertainty about the locations and speeds of other objects on the road, we can install more accurate sensors and get a better description of the driving environment.

However, there are applications when it is difficult or even impossible to decrease uncertainty. One such area is anything related to human activities. Humans make individual decisions based on their perceived value of different alternatives. The same alternative – be it a movie or a computer – have drastically different value to different people, so it is very difficult to predict their behavior. Such behavior affects economics and finance, so in decision making in economics and finance, it is very important to take such decision making under uncertainty into account.

Another area where it is difficult to decrease uncertainty is geosciences. In this thesis, we analyze the general problem of decision making under uncertainty and show how our results can be applied to finances and geosciences.

Table of Contents

Chapter 1

Introduction

Need for decision making. In many practical situations, we need to make a decision.

- In engineering, we need to decide on the best design of a system, and, for existing systems – on the best control strategy.
- In financial applications, we need to decide what is the best way to invest money.
- In geosciences, we need to decide whether we should explore a possible mineral deposit – or whether we should perform more experiments and measurements (and what exactly).

Need for decision making under uncertainty. In some cases, we can compute the exact consequences of each decision – e.g., if we are controlling a satellite. However, in many other cases, we do not know the exact consequences. In such situations, we need to make a decision under uncertainty.

Possible application areas. In many application areas, uncertainty is small – and can be made even smaller by appropriate measurements. For example, when we control a self-driving car, if there is an uncertainty about the locations and speeds of other objects on the road, we can install more accurate sensors and get a better description of the driving environment.

However, there are applications when it is difficult or even impossible to decrease uncertainty. One such area is anything related to human activities. Humans make individual decisions based on their perceived value of different alternatives. The same alternative – be it a movie or a computer – have drastically different value to different people, so it is

very difficult to predict their behavior. Such behavior affects economics and finance, so in decision making in economics and finance, it is very important to take such decision making under uncertainty into account.

Another area where it is difficult to decrease uncertainty is geosciences. In the cases of self-driving cars, we can decrease uncertainty by installing more accurate sensors, but to make a decision, e.g., on whether a certain underground area contains oil, there is only so much we can do with measurements. The only way to get a more accurate picture of what is going on beneath the earth surface is to dig a well and see – and the whole purpose of decision making is to decide whether such an expensive procedure is worth doing.

What we do in this thesis. In line with the above, in this thesis, we first, describe what is currently known about decision making under uncertainty. This will be discussed in Chapter 2.

In general, decision making under uncertainty requires that we fix some objective function that reflects the decision maker's interests. It is often difficult to find the expression that exactly captures these interests. In such cases, we have an approximate objective functions. It is well known that the quality of the resulting decisions and the computational complexity of computing this decision depend on the choice of approximating family. In Chapter 3, we formulate the problem of selecting an optimal approximating family as a precise mathematical problem, and we show that, under reasonable assumptions, the optimal approximating functions are polynomials.

Usually, decision makers do not just want to have a solution to their current problems, they also want to learn general techniques that will help them solve future problems. In particular, since it makes sense to consider polynomial objective functions, it is important to teach decision makers how to analyze such functions. A new way of doing this is presented in Chapter 4.

From the purely mathematical viewpoint, it is sufficient to formulate the practical problem in precise mathematical terms. However, from the practical viewpoint, even after this formulation, we still need to solve the corresponding computational problem. For simple

problems, the existing algorithms provide solutions in reasonable time. However, for more complex problems, the existing algorithms take too long a time. It is therefore necessary to explore the possibility of faster computations. One such possibility is the use of quantum computing. There are many existing algorithms for quantum computing, but, as we show in Chapter 5, they often do not provide an adequate representation of generic functions – and objective functions corresponding to decision making under uncertainty can be very complex. In this section, we show how to come up with more adequate representation of generic functions in quantum computing.

The practical usefulness of these theoretical developments is illustrated on the two above application examples: finances and geosciences.

We start with applications to finances. Of course, in finances, decision theory is already heavily used. What we are trying to do is illustrate the effectiveness of decision-making-under-uncertainty techniques in situations which are not well described by the usual decision theory approaches. Specifically, in Chapter 6, we explain why people sell and buy in the first place – it turns out that, from the usual economical viewpoint, it is still largely a mystery. In Chapter 7, we explain why for the same person, buy and sell prices are often different – another phenomenon that puzzles economists. These chapters describe the current decision making. If we take into account events in the past – or possible future consequences – then we need to describe how people perceive the corresponding time intervals. In Chapter 8, we explain that the seemingly counterintuitive empirical data about the human perception of time intervals can be naturally explained by the usual results about decision making under uncertainty.

Chapter 9 contains applications to geosciences. To illustrate the power of decision making techniques, we selected one of the most mysterious and challenging phenomenon – the Bhutan landscape anomaly.

Finally, in Chapter 10, we provide our future work plans.

Chapter 2

Decision Theory: A Brief Reminder

What is traditional decision theory. Traditional decision theory (see, e.g., [13, 22, 25, 30, 36]) describes preferences of rational agents, i.e., e.g., agents that when preferring A to B and B to C would always prefer A to C.

Comment. It is well known that real agents are not perfectly rational (see, e.g., [20, 24]), for the simple reason that our ability to process information and select an optimal decision is bounded. However, in many cases, traditional decision theory still provides a very good picture of human behavior.

The notion of utility. To describe the preferences of such an agent, we can select two alternatives:

- a very bad one A_- that is much worse than this agent will actually encounter, and
- a very good one A_+ that is much better than this agent will actually encounter.

For each value p from the interval $[0, 1]$, we can form a lottery $L(p)$ in which we get A_+ with probability p and A_- with the remaining probability $1 - p$.

When p is close to 1, this means that we are almost certainly getting a very good deal. So, for any realistic option A, the corresponding lottery $L(p)$ is better than A: $A < L(p)$. Similarly, when p is close to 0, this means that we are almost certainly getting a very bad deal, so $L(p) < A$. There should be a threshold u at which the preference $L(p) < A$ corresponding to smaller probabilities p is replaced by an opposite preference $A < L(p)$. In other words, we should have:

- $L(p) < A$ for all $p < u$ and

- $A < L(p)$ for all $p > u$.

This threshold value is called the *utility* of the alternative A; it is denoted by $u(A)$.

The above two conditions means that, in a certain reasonable sense, the original alternative A is equivalent to the lottery $L(u(A))$ corresponding to the probability $u(A)$: $A \equiv L(u(A))$.

A rational agent should maximize utility. Of course, the larger the probability of getting a very good outcome A_+, the better. Thus, among several lotteries $L(p)$, we should select the one for which the probability p of getting A_+ is the largest. Since each alternative A is equivalent to the corresponding lottery $L(u(A))$, this implies that we should select the alternative with the largest possible value of utility.

Main conclusion of traditional decision theory: a rational agent must maximize expected utility. In practice, we rarely know the consequences of each action. At best, we know possible outcomes $A_1, \ldots, A_n$, and their probabilities $p_1, \ldots, p_n$. Since each alternative A_i is equivalent to a lottery $L(u(A_i))$ in which we get A_+ with probability $u(A_i)$ and A_- with the remaining probability $1 - u(A_i)$, the whole action is equivalent to a two-stage lottery in which:

- first, we select one of the n alternatives A_i with probability p_i, and
- then, depending on which alternative A_i we selected on the first stage, we select A_+ with probability $u(A_i)$ and A_- with the remaining probability $1 - u(A_i)$.

As a result of this two-stage lottery, we get either A_+ or A_-. The probability u of getting A_+ can be computed by using the formula of complete probability, it is equal to

$$u = p_1 \cdot u(A_1) + \ldots + p_n \cdot u(A_n).$$

This is exactly the formula for the expected value of the utility $u(A_i)$. Thus, the utility of each action to a person is equal to the expected value of utility.

Thus, according to the traditional decision theory, rational agents should select the alternative with the largest possible value of expected utility.

Utility is defined modulo linear transformations. The numerical value of utility depends on the selection of the alternatives A_- and A_+. It can be shown that if we select a different pair (A'_-, A'_+), then the corresponding utility $u'(A)$ is related to the original utility by a linear transformation $u'(A) = a \cdot u(A) + b$ for some $a > 0$ and b; see, e.g., [22, 30].

How utility is related to money. The dependence of utility of money is non-linear: namely, utility u is proportional to the square root of the amount m of money $u = c \cdot \sqrt{m}$; see [20] and references therein.

Comment. This empirical fact can be explained. For example, the non-linear character of this dependence is explained, on a commonsense level, in [21], while the square root formula can also be explained – but it requires more mathematical analysis; see, e.g., [24]. In the current thesis, we simply take this fact as a given.

How to compare current and future gains: discounting. How can we compare current and future gains? If we have an amount m of money now, then we can place it in a bank and get the same amount plus interest, i.e., get the new amount $m' \stackrel{\text{def}}{=} (1 + i) \cdot m$ in a year, where i is the interest rate. Thus, the amount m' in a year is equivalent to the value $m = q \cdot m'$ now, where $q \stackrel{\text{def}}{=} 1/(1 + i)$. This reduction of future gains – to make them comparable to current gains – is known as *discounting*.

Discounting: a more detailed description. An event – e.g., a good dinner – a year in the past does not feel as pleasant to a person now as it may have felt a year ago. Similarly, a not-so-pleasant event in the past – e.g., a painful inoculation – does not feel as bad now if it felt a year ago, when it actually happened. Thus, the utility of an event changes with time: positive utility decreases, negative utility increases (i.e., gets closer to 0). If u is the utility of a current event, how can we describe the utility $f(u)$ of remembering the same event that happened 1 year ago?

We can normalize the utility values by assuming that the status quo situation has utility

0. Then the only remaining transformation is re-scaling $u' = a \cdot u$. Similarly to the case of interval uncertainty, it is reasonable to require that the function $f(u)$ is invariant with respect to such a transformation, i.e., that:

- if we have $v = f(u)$,
- then for each a, we should have $v' = f(u')$, where we denoted $v' = a \cdot v$ and $u' = a \cdot u$.

Substituting the expressions for v' and u' into the formula $v' = f(u')$, we conclude that $a \cdot v = f(a \cdot u)$, i.e., $a \cdot f(u) = f(a \cdot u)$. Substituting $u = 1$ into this formula, we conclude that $f(a) = q \cdot a$, where we denoted $q \stackrel{\text{def}}{=} f(1)$. Since $f(u) < u$ for $u > 0$, this would imply that $q < 1$.

So, an event with then-utility u that occurred 1 year ago has the utility $q \cdot u$ now. Similarly, an event with utility u that happened 2 years ago is equivalent to $q \cdot u$ a year ago, and thus, is equivalent to $q \cdot (q \cdot u) = q^2 \cdot u$ now. We can similarly conclude that an event with utility u that occurred t moments in the past is equivalent to utility $q^t \cdot u$ now.

Decision making under interval uncertainty. In real life, we rarely know the exact consequences of each action. As a result, for each alternative A, instead of the exact value of its utility, we often only know the bounds $\underline{u}(A)$ and $\overline{u}(A)$ on this unknown value. In other words, all we know is the interval $[\underline{u}(A), \overline{u}(A)]$ that contains the actual (unknown) value $u(A)$. How can we make a decision under this interval uncertainty?

In particular, for such an interval case, we need to be able to compare the interval-valued alternative with lotteries $L(p)$ for different values p. As a result of such comparison, we will come up with a utility of this interval. So, to make recommendations on decision under interval uncertainty, we need to be able to assign, to each interval $[\underline{u}, \overline{u}]$, a single utility value $u(\underline{u}, \overline{u})$ from this interval that describes this interval's utility.

Since utility is defined modulo a linear transformation $u \to u' = a \cdot u + b$, it is reasonable to require that the corresponding function $u(\underline{u}, \overline{u})$ should also be invariant under such transformations, i.e., that:

- if $u = u(\underline{u}, \overline{u})$,
- then $u' = u(\underline{u}', \overline{u}')$, where we denoted $u' = a \cdot u + b$, $\underline{u}' = a \cdot \underline{u} + b$, and $\overline{u}' = a \cdot \overline{u} + b$.

It turns out that this invariance requirement implies that

$$u(\underline{u}, \overline{u}) = \alpha_H \cdot \overline{u} + (1 - \alpha_H) \cdot \underline{u}$$

for some $\alpha_H \in [0, 1]$ [22, 30]. This formula was first proposed by a future Nobelist Leo Hurwicz and is, thus, known as the Hurwicz optimism-pessimism criterion [17, 25].

Theoretically, we can have values $\alpha_H = 0$ and $\alpha_H = 1$. However, in practice, such values do not happen:

- $\alpha_H = 1$ would correspond to a person who only takes into account the best possible outcome, completely ignoring the risk of possible worse situations;
- similarly, the value $\alpha_H = 0$ would correspond to a person who only takes into account the worst possible outcome, completely ignoring the possibility of better outcomes.

In real life, we thus always have $0 < \alpha_H < 1$.

Chapter 3

What Is the Optimal Approximating Family

Results from this chapter will appear in [32].

Need for approximations. In many practical situations, we need to find a good description of the observed data. In a computer, we can only store finitely many parameters. So, it is reasonable to consider finite-parametric approximation families, i.e., families that depend on finitely many parameters $C_1, \ldots, C_n$.

Need to consider families that linearly depend on the parameters. From the computational viewpoint, the easiest case is when the dependence on the parameters is linear, i.e., when the family has the form

$$f(x) = C_1 \cdot f_1(x) + \ldots + C_n \cdot f_n(x)$$

for some functions $f_1(x), \ldots, f_n(x)$. In this case, to find the values of the parameters C_i based on the known observations $x_1^{(k)}, \ldots, x_n^{(k)}, y^{(k)}$, it is sufficient to solve a system of linear equations

$$y^{(k)} = C_1 \cdot f_1(x^{(k)}) + \ldots + C_n \cdot f_n(x^{(k)}).$$

For solving systems of linear equations, there are efficient algorithms.

In principle, we can consider more complex dependencies – e.g., quadratic ones:

$$f(x) = \sum_i C_i \cdot f_i(x) + \sum_{i,j} C_i \cdot C_j \cdot f_{ij}(x).$$

However, in this case, to find the values of the corresponding parameters, we would need to solve systems of quadratic equations – and this is known to be NP-hard; see, e.g., [35].

Thus, unless P=NP (which most computer scientists believe to be impossible), no general feasible algorithm is possible for solving such systems.

Since we want to have efficient algorithms, it is reasonable to restrict ourselves to approximating families in which the dependence on the parameters is linear.

Observations are usually smooth. Sensors usually smooth the observed processes, so what we observe is usually smooth. So, we can safely assume that the corresponding functions $f_i(x)$ are smooth (differentiable).

Thus, we arrive at the following definition.

Definition 3.1. *Let n be a positive integer. By an* approximating family*, we mean a family of functions*

$$\{C_1 \cdot f_1(x) + \ldots + C_n \cdot f_n(x)\}_{C_1,\ldots,C_n}, \tag{3.1}$$

where the functions $f_1(x), \ldots, f_n(x)$ are fixed differentiable functions, and C_1, ..., C_n are arbitrary real numbers.

From this viewpoint, selecting a description means selecting n functions $f_1(x)$, ..., $f_n(x)$.

Towards the optimal description. Which description is the best? To answer this question, we need to be able to decide, for each two families of functions F and F', whether the first family is better (we will denote it by $F' < F$) or the second family is better ($F < F'$), or both families have the same quality (we will denote it by $F \sim F'$). Clearly, if F is worse than F' and F' is worse than F'', then F should be worse than F''. So, we arrive at the following definition.

Definition 3.2. *Let n be a positive number. By an* optimality criterion*, we mean the pair of relations $(<, \sim)$ on the set S of all possible n-dimensional approximating families that satisfies the following conditions:*

- *for every pair $F, F' \in S$, we have one and only one of the following options: either $F' < F$ or $F < F'$ or $F \sim F'$;*

- *for every F, F', and F'', if $F < F'$ and $F' < F''$, then $F < F''$;*
- *for every F, F', and F'', if $F < F'$ and $F' \sim F''$, then $F < F''$;*
- *for every F, F', and F'', if $F \sim F'$ and $F' < F''$, then $F < F''$;*
- *for every F, F', and F'', if $F \sim F'$ and $F' \sim F''$, then $F \sim F''$;*
- *for every F and F', if $F \sim F'$, then $F' \sim F$.*

Definition 3.3. *Let $(<, \sim)$ be an optimality criterion. We say that a family F is* optimal *with respect to this optimality criterion if for every other family F', we have either $F' < F$ or $F' \sim F$.*

We want to use an appropriate optimality criterion to select a family. If a criterion selected several different families as equally good, then we can use this non-uniqueness to optimize something else. For example, if we have several different families that provide an equally good approximation, then, from all these optimal families, we can select, e.g., the family which is the easiest to compute. This additional selection is, in effect, equivalent to replacing the original optimality criterion with the new one $<_{\text{new}}$, according to which $F <_{\text{new}} F'$ if:

- either $F < F'$ according to the original criterion,
- or $F \sim F'$ and F' is easier to compute (in some formal sense, e.g., in terms of the computation time).

If the new criterion still selects several families as equally optimal, we can again modify it, etc. – until we end up with a *final* criterion for which there is exactly one optimal family.

Definition 3.4. *We say that an optimality criterion is* final *if it has exactly one optimal family.*

As a starting point for measuring x, we can take different locations. If we select a different location which is x_0 units before the current one, then each new location x is identical to the old location $x' = x + x_0$. So, the same profile approximation that in the new units has the form $f(x)$ in the old units has the form $f(x + x_0)$. The relative quality of different profiles approximations should not change if we simply change the starting location. Thus, we arrive at the following definitions.

Definition 3.5. *For each family F as described by the formula* (3.1) *and for each x_0, by a* shift $S_{x_0}(F)$*, we mean a family*

$$\{C_1 \cdot (S_{x_0} f_1)(x) + \ldots + C_n \cdot (S_{x_0} f_n)(x)\},$$

where $(S_{x_0} f_i)(x) \stackrel{\text{def}}{=} f_i(x + x_0)$.

Definition 3.6. *We say that an optimality criterion is* shift-invariant *if for every F, F', and x_0, the following two properties hold:*

- *if $F < F'$, then $S_{x_0}(F) < S_{x_0}(F')$;*
- *if $F \sim F'$, then $S_{x_0}(F) \sim S_{x_0}(F')$.*

Similarly, nothing should change if we simply change the measuring unit for x – e.g., use miles instead of kilometers. If we replace the original measuring unit by a one which is λ times larger, then the new value x is identical to the old value $x' = \lambda \cdot x$. So, the same profile approximation that in the new units has the form $f(x)$ in the old units has the form $f(\lambda \cdot x)$. The relative quality of different profiles approximations should not change if we simply change the measuring unit. Thus, we arrive at the following definitions.

Definition 3.7. *For each family F as described by the formula* (3.1) *and for each $\lambda > 0$, by a* rescaling $R_\lambda(F)$*, we mean a family*

$$\{C_1 \cdot (R_\lambda f_1)(x) + \ldots + C_n \cdot (R_\lambda f_n)(x)\},$$

where $(R_\lambda f_i)(x) \stackrel{\text{def}}{=} f_i(\lambda \cdot x)$.

Definition 3.8. *We say that an optimality criterion is* scale-invariant *if for every F, F', and $\lambda > 0$, the following two properties hold:*

- *if $F < F'$, then $R_\lambda(F) < R_\lambda(F')$;*
- *if $F \sim F'$, then $R_\lambda(F) \sim R_\lambda(F')$.*

Proposition 3.1. *For every n and for every final shift- and scale-invariant optimality criterion, the optimal family F_{opt} consists of polynomials of order $\le n-1$.*

Comment. This result is similar to results from [31].

Proof.

1°. Let us first prove that the optimal family is shift- and scale-invariant, i.e., that $S_{x_0}(F_{\text{opt}}) = R_\lambda(F_{\text{opt}}) = F_{\text{opt}}$ for all x_0 and λ.

Let us first prove shift-invariance of F_{opt}. Since F_{opt} is optimal, for every family F, we have $F < F_{\text{opt}}$ or $F \sim F_{\text{opt}}$. In particular, this is true for the family $S_{-x_0}(F)$, i.e., either $S_{-x_0}(F) < F_{\text{opt}}$ or $S_{-x_0}(F) \sim F_{\text{opt}}$. Since the optimality criterion is shift-invariant, this implies that either $S_{x_0}(S_{-x_0}(F)) < S_{x_0}(F_{\text{opt}})$ or $S_{x_0}(S_{-x_0}(F)) \sim S_{x_0}(F_{\text{opt}})$. However, as one can easily check, we have $S_{x_0}(S_{-x_0}(F)) = F$. Thus, for every family F, we have either $F < S_{x_0}(F_{\text{opt}})$ or $F \sim S_{x_0}(F_{\text{opt}})$. By definition of optimality, this means that the family $S_{x_0}(F_{\text{opt}})$ is also optimal.

Since the optimality criterion is final, there is only one optimal family, hence $S_{x_0}(F_{\text{opt}}) = F_{\text{opt}}$. Shift-invariance is proven.

Scale-invariance is proven similarly, by taking into account that for every F and every λ, either $R_{1/\lambda}(F) < F_{\text{opt}}$ or $R_{1/\lambda}(F) \sim F_{\text{opt}}$. So, by applying the scaling R_λ to both sides of these relations, we conclude that $R_\lambda(F_{\text{opt}})$ is also optimal and thus, $R_\lambda(F_{\text{opt}}) = F_{\text{opt}}$.

2°. Shift-invariance means that every element of the family $S_{x_0}(F_{\text{opt}})$ also belongs to the same family F_{opt}. Let $f_i(x)$ denote the functions whose linear combinations (3.1) form the family F_{opt}. Then, in particular, invariance means that for every i, the shifted function

$f_i(x + x_0)$ is a linear combination of functions $f_j(x)$:

$$\begin{aligned} f_1(x + x_0) &= C_{11}(x_0) \cdot f_1(x) + \ldots + C_{1n}(x_0) \cdot f_n(x); \\ &\ldots \\ f_n(x + x_0) &= C_{n1}(x_0) \cdot f_1(x) + \ldots + C_{nn}(x_0) \cdot f_n(x), \end{aligned} \tag{3.2}$$

for some coefficients C_{ij} depending on x_0.

For each i, we can take n different values $x_1, \ldots, x_n$ of x and get a system of n linear equations with n unknowns $C_{i1}(x_0), \ldots, C_{in}(x_0)$:

$$\begin{aligned} f_i(x_1 + x_0) &= C_{i1}(x_0) \cdot f_1(x_1) + \ldots + C_{in}(x_0) \cdot f_n(x_1); \\ &\ldots \\ f_i(x_n + x_0) &= C_{i1}(x_0) \cdot f_1(x_n) + \ldots + C_{in}(x_0) \cdot f_n(x_n). \end{aligned}$$

By Cramer's rule, the solutions $C_{ij}(x_0)$ to this system can be represented as a ratio of two polynomials in terms of $f_i(\cdot)$. Since the functions $f_i(x)$ are smooth, this implies that the functions $C_{ij}(x_0)$ are also differentiable functions of x_0.

Thus, we can differentiate both sides of (3.2) by x_0 and take $x_0 = 0$. As a result, we get a system of linear differential equations with constant coefficients:

$$\begin{aligned} f_1'(x) &= c_{11} \cdot f_1(x) + \ldots + c_{1n} \cdot f_n(x); \\ &\ldots \\ f_n'(x) &= c_{n1} \cdot f_1(x) + \ldots + c_{nn} \cdot f_n(x), \end{aligned} \tag{3.3}$$

where we denoted $c_{ij} \stackrel{\text{def}}{=} C_{ij}'(0)$.

The general solution to such a system is well-known (see, e.g., [16, 31]): it is a linear combination of terms of the type $\exp(\lambda_i \cdot x)$ and $x^k \cdot \exp(\lambda_i \cdot x)$, where λ_i are eigenvalues of the matrix (c_{ij}), and $k \leq n - 1$ is a positive integer corresponding to the case when we have a multiple eigenvalue.

3°. Similarly, scale-invariance means that every element of the family $R_\lambda(F_{\text{opt}})$ also belongs to F_{opt}. In particular, this means that for every i, the re-scaled function $f_i(\lambda \cdot x)$ is a linear combination of functions $f_j(x)$:

$$\begin{aligned} f_1(\lambda \cdot x) &= C_{11}(\lambda) \cdot f_1(x) + \ldots + C_{1n}(\lambda) \cdot f_n(x); \\ &\ldots \\ f_n(\lambda \cdot x) &= C_{n1}(\lambda) \cdot f_1(x) + \ldots + C_{nn}(\lambda) \cdot f_n(x), \end{aligned} \tag{3.4}$$

for some coefficients C_{ij} depending on λ.

For each i, we can take n different values $x_1, \ldots, x_n$ of x and get a system of n linear equations with n unknowns $C_{i1}(\lambda), \ldots, C_{in}(\lambda)$:

$$\begin{aligned} f_i(\lambda \cdot x_1) &= C_{i1}(\lambda) \cdot f_1(x_1) + \ldots + C_{in}(\lambda) \cdot f_n(x_1); \\ &\ldots \\ f_i(\lambda \cdot x_n) &= C_{i1}(\lambda) \cdot f_1(x_n) + \ldots + C_{in}(\lambda) \cdot f_n(x_n). \end{aligned}$$

By Cramer's rule, the solutions $C_{ij}(\lambda)$ to this system can be represented as a ratio of two polynomials in terms of $f_i(\cdot)$. Since the functions $f_i(x)$ are smooth, this implies that the functions $C_{ij}(\lambda)$ are also differentiable functions of λ.

Thus, we can differentiate both sides of (3.4) by λ and take $\lambda = 1$. As a result, we get the following system of linear differential equations:

$$\begin{aligned} x \cdot f_1'(x) &= c_{11} \cdot f_1(x) + \ldots + c_{1n} \cdot f_n(x); \\ &\ldots \\ x \cdot f_n'(x) &= c_{n1} \cdot f_1(x) + \ldots + c_{nn} \cdot f_n(x), \end{aligned} \tag{3.5}$$

where we denoted $c_{ij} \stackrel{\text{def}}{=} C_{ij}'(1)$.

Here, for each i, we have

$$x \cdot f_i'(x) = x \cdot \frac{df_i}{dx} = \frac{df_i}{dx/x}.$$

Since $dx/x = d(\ln(x))$, we thus conclude that for the new variable $X = \ln(x)$ (for which $x = \exp(X)$) and for the corresponding functions $F_i(X) = f_i(\exp(X))$, we get the system of linear differential equations with constant coefficients:

$$F_1'(X) = c_{11} \cdot F_1(X) + \ldots + c_{1n} \cdot F_n(X);$$

$$\ldots \tag{3.6}$$

$$F_n'(x) = c_{n1} \cdot F_1(X) + \ldots + c_{nn} \cdot F_n(X).$$

Hence, similarly to the previous subsection, we conclude that each solution of this system is a linear combination of terms of the type $\exp(\lambda_i \cdot X)$ and

$$X^k \cdot \exp(\lambda_i \cdot X).$$

Substituting $X = \ln(x)$ into this formula, we conclude that each function $f_i(x) = F_i(\ln(x))$ is a linear combination of functions $\exp(\lambda_i \cdot \ln(x))$ and

$$\ln^k(x) \cdot \exp(\lambda_i \cdot \ln(x)).$$

Here, $\exp(\lambda_i \cdot \ln(x)) = (\exp(\ln(x))^{\lambda_i} = x^{\lambda_i}$.

Thus, each function $f_i(x)$ is a linear combination of functions x^{λ_i} and

$$\ln^k(x) \cdot x^{\lambda_i}.$$

4°. Our functions $f_i(x)$ are *both* shift-invariant and scale-invariant. Thus, each of them has to be both of form described at the end of Part 2 of this proof and of the form described at the end of Part 3. So, out of terms from Part 2, we cannot have exponential terms with non-zero λ_i – since these terms cannot be expressed in Part-3 form. Thus, the only possible terms are terms x^k with $k \le n - 1$.

So, each function $f_i(x)$ is a linear combination of such terms – and is, thus, a polynomial of order $\le n - 1$. The proposition is proven.

Chapter 4

Teaching Optimization: How to Generate "Nice" Cubic Polynomials

In the previous chapter, we have shown that it is reasonable to approximate functions by polynomials. In particular, it makes sense to consider polynomial objective functions. It is therefore important to teach decision makers how to analyze such functions.

In general, people feel more comfortable with rational numbers than with irrational ones. Thus, when teaching the beginning of calculus, it is desirable to have examples of simple problems for which both zeros and extrema points are rational. Recently, an algorithm was proposed for generating cubic polynomials with this property. However, from the computational viewpoint, the existing algorithm is not the most efficient one: in addition to applying explicit formulas, it also uses trial-and-error exhaustive search. In this chapter, we describe a new computationally efficient algorithm for generating all such polynomials: namely, an algorithm that uses only explicit formulas.

The abstracts related to this result appeared in [5] and [6]; the full paper will appear in [7].

4.1 Formulation of the Problem

Need for nice calculus-related examples. After students learn the basics of calculus, they practice in using the calculus tools to graph different functions $y = f(x)$. Specifically,

- they find the roots (zeros), i.e., the values where $f(x) = 0$,

- they find the extreme points, i.e., the values where the derivative is equal to 0,
- they find out whether the function is increasing or decreasing between different extreme points – by checking the signs of the corresponding derivatives,

and they use this information – plus the values of $f(x)$ at several points x – to graph the corresponding function.

For this practice, students need examples for which they can compute both the zeros and the extreme points.

Cubic polynomials: the simplest case when such an analysis makes sense. The simplest possible functions are polynomials. For linear functions, the derivative is constant, so there are no extreme point. For quadratic functions, there is an extreme point, but, after studying quadratic equations, students already know how to graph the corresponding function, when it decreases, when it increases. So, for quadratic polynomials, there is no need to use calculus.

The simplest case when calculus tools are needed is the case of cubic polynomials.

To make the materials simpler for students, it is desirable to limit ourselves to rational roots. Students are much more comfortable with rational numbers than with irrational ones. Thus, to make the corresponding example easier for students, it is desirable to start with examples in which all the coefficients, all the zeros, and all the extreme points of a cubic polynomial are rational.

Good news is that when we know that the roots are rational, it is (relatively) easy to find these roots. Indeed, to find rational roots, we can use the *Rational Root Theorem*, according to which for each rational root $x = p/q$ (where p and q do not have any common divisors) of a polynomial $a_n \cdot x^n + a_{n-1} \cdot x^{n-1} + \ldots + a_0$ with integer coefficients $a_0, \ldots, a_{n-1}, a_n$, the numerator p is a factor of a_0, and the denominator q is a factor of a_n; see, e.g., [28].

Thus, to find all the rational roots of a polynomial, it is sufficient:

- to list all factors p of the coefficient a_0,

- to list all factors q of the coefficient a_n, and then
- to check, for each pair (p, q) of the values from the two lists, whether the ratio p/q is a root.

How can we find polynomials for which both zeros and extreme points are rational?

What is known. An algorithm for generating such polynomials was proposed in [3, 4]. This algorithm, however, is not the most efficient one: for each tuple of the corresponding parameter values, it uses exhaustive trial-and-error search to produce the corresponding nice cubic polynomial.

What we do in this chapter. In this chapter, we produce an efficient algorithm for producing nice polynomials. Namely, we propose simple computational formulas with the following properties:

- for each tuple of the corresponding parameters, these formulas produce coefficients of a cubic polynomial for which all zeros and extreme points are rational, and
- every cubic polynomial with this property can be generated by applying these formulas to an appropriate tuple of parameters.

Thus, for each tuple of parameters, our algorithm requires the same constant number of elementary computational steps (i.e., elementary arithmetic operations) – in contrast with the existing algorithm, in which the number of elementary steps, in general, grows with the values of the parameters.

4.2 Analysis of the Problem

Let us first simplify the problem. A general cubic polynomial with rational coefficients has the form

$$a \cdot X^3 + b \cdot X^2 + c \cdot X + d. \tag{4.1}$$

We consider the case when this is a truly cubic polynomial, i.e., when $a \neq 0$.

Roots and extreme points of a function do not change if we simply divide all its values by the same constant a. Thus, instead of considering the original polynomial (4.1) with four parameters a, b, c, and d, it is sufficient to consider the following polynomial with only three parameters:

$$X^3 + p \cdot X^2 + q \cdot X + r, \tag{4.2}$$

where

$$p \stackrel{\text{def}}{=} \frac{b}{a}, \quad q \stackrel{\text{def}}{=} \frac{c}{a}, \quad r \stackrel{\text{def}}{=} \frac{d}{a}. \tag{4.3}$$

When the coefficients a, b, c, and d of the original polynomial (4.1) were rational, the coefficients of the new polynomial (4.2) are rational as well; vice versa, if we have a polynomial (4.2) with rational coefficients, then, for any rational a, we can have a polynomial (4.1) with rational coefficients $b = a \cdot p$, $c = a \cdot q$, and $d = a \cdot r$. Thus, to find cubic polynomials with rational coefficients, rational roots, and rational extreme points, it is sufficient to consider polynomials of type (4.2).

We can simplify the problem even further if we replace the original variable X with the new variable

$$x \stackrel{\text{def}}{=} X + \frac{p}{3} \tag{4.4}$$

for which

$$X = x - \frac{p}{3}. \tag{4.5}$$

Substituting this expression for X into the formula (4.2), we get

$$\left(x - \frac{p}{3}\right)^3 + p \cdot \left(x - \frac{p}{3}\right)^2 + q \cdot \left(x - \frac{p}{3}\right) + r =$$

$$x^3 - 3 \cdot \frac{p}{3} \cdot x^2 + 3 \cdot \left(\frac{p}{3}\right)^2 \cdot x - \left(\frac{p}{3}\right)^3 + p \cdot x^2 -$$

$$2 \cdot p \cdot \frac{p}{3} \cdot x + p \cdot \left(\frac{p}{3}\right)^2 + q \cdot x - q \cdot \frac{p}{3} + r =$$

$$x^3 + \alpha \cdot x + \beta, \tag{4.6}$$

where

$$\alpha = q - \frac{p^2}{3} \tag{4.7}$$

and

$$\beta = r - \frac{p \cdot q}{3} + \frac{2p^3}{27}. \tag{4.8}$$

The roots and extreme points of the new polynomial (4.6) are obtained from the roots and extremes of the original polynomial (4.2) by shifting by a rational number $p/3$, so they are all rational for the polynomial (4.6) if and only if they are rational for the polynomial (4.2).

Describing in terms of roots. Let r_1, r_2, and r_3 denote rational roots of the polynomial (4.6). Then, we have

$$x^3 + \alpha \cdot x + \beta = (x - r_1) \cdot (x - r_2) \cdot (x - r_3) =$$

$$x^3 - (r_1 + r_2 + r_3) \cdot x^2 + (r_1 \cdot r_2 + r_2 \cdot r_3 + r_1 \cdot r_3) \cdot x - r_1 \cdot r_2 \cdot r_3. \tag{4.9}$$

By equating the coefficients at x^2, x, and 1 at both sides, we conclude that

$$r_1 + r_2 + r_3 = 0, \tag{4.10}$$

$$\alpha = r_1 \cdot r_2 + r_2 \cdot r_3 + r_1 \cdot r_3, \tag{4.11}$$

and

$$\beta = -r_1 \cdot r_2 \cdot r_3. \tag{4.13}$$

From (4.10), we conclude that

$$r_3 = -(r_1 + r_2). \tag{4.14}$$

Substituting the expression (4.14) into the formulas (4.11) and (4.13), we conclude that

$$\alpha = r_1 \cdot r_2 - r_2 \cdot (r_1 + r_2) - r_1 \cdot (r_1 + r_2) = -(r_1^2 + r_1 \cdot r_2 + r_2^2) \tag{4.15}$$

and

$$\beta = r_1 \cdot r_2 \cdot (r_1 + r_2). \tag{4.16}$$

Now the polynomial (4.6) takes the following form:

$$x^3 - (r_1^2 + r_1 \cdot r_2 + r_2^2) \cdot x + r_1 \cdot r_2 \cdot (r_1 + r_2). \tag{4.17}$$

Using the fact that the extreme points should also be rational. Let us now use the fact that the extreme points should also be rational. Let x_0 denote an extreme point, i.e., a point at which the derivative of the polynomial (4.17) is equal to 0. Differentiating the expression (4.17) and equating the derivative to 0, we get

$$3x_0^2 - (r_1^2 + r_1 \cdot r_2 + r_2^2) = 0. \tag{4.18}$$

The expression in parentheses can be equivalently described as

$$\frac{3}{4} \cdot (r_1 + r_2)^2 + \frac{1}{4} \cdot (r_1 - r_2)^2 = 3y^2 + z^2, \tag{4.19}$$

where we denoted

$$y \stackrel{\text{def}}{=} \frac{r_1 + r_2}{2} \text{ and } z \stackrel{\text{def}}{=} \frac{r_1 - r_2}{2}. \tag{4.20}$$

Substituting this expression (4.20) into the formula (4.18), we arrive at the following homogeneous quadratic relation with integer coefficients between the rational numbers x_0, y, and z:

$$3x_0^2 - 3y^2 - z^2 = 0. \tag{4.21}$$

If we divide both sides of equation (4.21) by y^2, we get a new equation

$$3X_0^2 - 3 - Z^2 = 0, \tag{4.22}$$

where we denoted $X_0 \stackrel{\text{def}}{=} \frac{x_0}{y}$ and $Z \stackrel{\text{def}}{=} \frac{z}{y}$. When x_0, y, and z are rational, then X_0 and Z are also rational numbers. Vice versa, when X_0 and Z form a rational-valued solution of the equation (4.22), then, for any rational number y, by multiplying both sides of equation (4.22) by y^2, we can get a solution $x_0 = y \cdot X_0$, y, and $z = y \cdot Z$ of the equation (4.21). Thus, to find all rational solutions of the equation (4.21), it is sufficient to find al rational solutions of a simplified equation (4.22).

The simplest solution and the resulting "nice" polynomials. One of the solution of equation (4.22) is easy to find: namely, when $X_0 = -1$, the equation (4.22) takes the form $Z^2 = 0$, i.e., $Z = 0$.

This means that for every y, the values $x_0 = -y$, y and $z = 0$ solve the equation (4.21). The formulas (4.20) enable us to reconstruct r_1 and r_2 from y and z as

$$r_1 = y + z \text{ and } r_2 = y - z. \tag{4.23}$$

In our case, this means $r_1 = r_2 = y$. Thus, due to (4.15) and (4.16), we have a polynomial $x^3 + \alpha \cdot x + \beta$ with $\alpha = -3y^2$ and $\beta = 2y^3$.

By applying a shift by a rational number s, i.e., by replacing x with $x = X + s$, we transform a "nice" polynomial $x^3 + \alpha \cdot x + \beta$ into a new "nice" polynomial

$$(X + s)^3 + \alpha \cdot (X + s) + \beta = X^3 + 3s \cdot X^2 + (3s^2 + \alpha) \cdot X + (s^3 + \beta),$$

i.e., a polynomial (4.2) with $p = 3s$, $q = 3s^2 + \alpha$, and $r = s^3 + \beta$. Finally, by multiplying this polynomial by a rational number a, we get the following family of "nice" polynomials:

$$b = 3a \cdot s, \quad c = a \cdot (3s^2 + \alpha), \quad d = a \cdot (s^3 + \beta). \tag{4.24}$$

In our case, with $\alpha = -3y^2$ and $\beta = 2y^3$, we get

$$b = 3a \cdot s, \quad c = a \cdot (3s^2 - 3y^2), \quad d = a \cdot (s^3 + 2y^3). \tag{4.24a}$$

Using the general algorithm for finding all rational solutions to a quadratic equation. To find all rational solutions of the equation (4.21), we will use a general algorithm for finding all rational solutions of a homogeneous quadratic equation with integer coefficients; see, e.g., [38].

We have already found a solution of the equation (4.22) corresponding to $X_0 = -1$. For this value X_0, the equation (4.22) has only one solution $(-1, 0)$, for which $X_0 = -1$ and

$Z = 0$. Every other solution (X_0, Z) can be connected to this simple solution $(-1, 0)$ by a straight line. A general equation of a straight line passing through the point $(-1, 0)$ is

$$Z = t \cdot (X_0 + 1). \tag{4.25}$$

When X_0 and Z are rational, the ratio $t = \dfrac{Z}{X_0 + 1}$ is also rational.

Substituting the expression (4.25) into the equation (4.22), we get

$$3X_0^2 - 3 - t^2 \cdot (X_0 + 1)^2 = 0,$$

i.e.,

$$3 \cdot (X_0^2 - 1) - t^2 \cdot (X_0 + 1)^2 = 0. \tag{4.26}$$

Since we consider the case when $X_0 \neq -1$, we thus have $X_0 + 1 \neq 0$. So, we can divide both sides of the equation (4.26) by $X_0 + 1$ and thus, get the following equation:

$$3 \cdot (X_0 - 1) - t^2 \cdot (X_0 + 1) = 0.$$

From this equation, we can describe X_0 in terms of t: $(3 - t^2) \cdot X_0 = 3 + t^2$, hence

$$X_0 = \frac{3 + t^2}{3 - t^2}. \tag{4.27}$$

Substituting this expression for X_0 into the formula (4.25), we conclude that

$$Z = \frac{6t}{3 - t^2}. \tag{4.28}$$

Towards a general description of all "nice" polynomials. For every rational y, we can now take $x_0 = y \cdot X_0$, y, and

$$z = y \cdot Z = \frac{6yt}{3 - t^2}. \tag{4.29}$$

Based on y and z, we can compute r_1 and r_2 by using the formulas (4.23).

We can now use the values r_1 and r_2 from (4.23) and the formulas (4.15) and (4.16) to compute α and β. Since here, $r_1 + r_2 = 2y$, we get

$$\alpha = r_1 \cdot r_2 - (r_1 + r_2)^2 = (y + z) \cdot (y - z) - (2y)^2 = y^2 - z^2 - 4y^2 = -3y^2 - z^2 \tag{4.30}$$

and

$$\beta = r_1 \cdot r_2 \cdot (r_1 + r_2) = (y^2 - z^2) \cdot (2y) = 2y \cdot (y^2 - z^2). \tag{4.31}$$

Substituting these expressions for α and β into the formula (4.24), we get the formulas for computing the coefficients of the corresponding "nice" cubic polynomial:

$$b = 3a \cdot s; \tag{4.32}$$

$$c = a \cdot (3s^2 + \alpha) = a \cdot (3s^2 - 3y^2 - z^2); \tag{4.33}$$

$$d = a \cdot (s^3 + \beta) = a \cdot (s^3 + 2y \cdot (y^2 - z^2)). \tag{4.34}$$

Thus, we arrive at the following algorithm for computing all possible "nice" cubic polynomials.

4.3 Resulting Algorithm

Here is an algorithm for computing all "nice" cubic polynomials, i.e., all cubic polynomials with rational coefficients for which all three roots and both extreme points are rational.

In this algorithm, we use four arbitrary rational numbers t, y, s, and a. Based on these numbers, we first compute

$$z = \frac{6yt}{3 - t^2}. \tag{4.29a}$$

Then, we compute the coefficients b, c, and d of the resulting "nice" polynomial (the value a we already know):

$$b = 3a \cdot s; \tag{4.32}$$

$$c = a \cdot (3s^2 - 3y^2 - z^2); \tag{4.33a}$$

$$d = a \cdot (s^3 + 2y \cdot (y^2 - z^2)). \tag{4.34a}$$

These expressions cover almost all "nice" polynomials, with the exception of one family of such polynomials, which is described by the formula

$$b = 3a \cdot s, \quad c = a \cdot (3s^2 - 3y^2), \quad d = a \cdot (s^3 + 2y^3). \tag{4.24a}$$

Chapter 5

How to Speed Up Computations: Representing Functions in Quantum and Reversible Computing

From the purely mathematical viewpoint, it is sufficient to formulate the practical problem in precise mathematical terms. However, from the practical viewpoint, even after this formulation, we still need to solve the corresponding computational problem. For simple problems, the existing algorithms provide solutions in reasonable time. However, for more complex problems, the existing algorithms take too long a time. It is therefore necessary to explore the possibility of faster computations. One such possibility is the use of quantum computing. There are many existing algorithms for quantum computing, but, as we show in this chapter, they often do not provide an adequate representation of generic functions – and objective functions corresponding to decision making under uncertainty can be very complex. In this section, we show how to come up with more adequate representation of generic functions in quantum computing.

5.1 Formulation of the Problem

Need for faster computing. While computers are very fast, in many practical problems, we need even faster computations. For example, we can, in principle, with high accuracy predict in which direction a deadly tornado will turn in the next 15 minutes, but this computation requires hours even on the most efficient high performance computers – too

late for the resulting prediction to be of any use.

Faster computations means smaller processing units. One of the main limitations on physical processes is the fact that, according to modern physics, all processes cannot move faster than the speed of light. For a laptop of size ≈ 30 cm, this means that it takes at least 1 nanosecond (10^{-9} sec) for a signal to move from one side of the laptop to the other. During this time, even the cheapest laptops perform several operations. Thus, to speed up computations, we need to further decrease the size of the computer – and thus, further decrease the size of its memory units and processing units.

Need for quantum computing. Already the size of a memory cell in a computer is compatible with the size of a molecule. If we decrease the computer cells even more, they will consist of a few dozen molecules. Thus, to describe the behavior of such cells, we will need to take into account the physical laws that describe such micro-objects – i.e., the laws of quantum physics.

Quantum computing means reversible computing. For macro-objects, we can observe irreversible processes: e.g., if we drop a china cup on a hard floor, it will break into pieces, and no physical process can combine these pieces back into the original whole cup. However, on the micro-level, all the equations are reversible. This is true for Newton's equations that describe the non-quantum motion of particles and bodies, this is true for Schroedinger's equation that takes into account quantum effects that describes this notion; see, e.g., [12, 41].

Thus, in quantum computing, all elementary operations must be reversible.

Reversible computing beyond quantum. Reversible computing is also needed for a different reasons. Even at the present level of micro-miniaturization, theoretically, we could place more memory cells and processing cells into the same small volume if, instead of the current 2-D stacking of these cells into a planar chip, we could stack them in 3-D.

For example, if we have a Terabyte of memory, i.e., 10^{12} cells in a 2-D arrangement, this means $10^6 \times 10^6$. If we could get a third dimension, we would be able to place $10^6 \times$

$10^6 \times 10^6 = 10^{18}$ cells in the same volume – million times more than now.

The reason why we cannot do it is that already modern computers emit a large amount of heat. Even with an intensive inside-computer cooling, a working laptop warms up so much that it is possible to be burned if you keep it in your lap. If instead of a single 2-D layer, we have several 2-D layers forming a 3-D structure, the amount of heat will increase so much that the computer will simply melt.

What causes this heat? One of the reasons may be design flaws. Some part of this heat may be decreased by an appropriate engineering design. However, there is also a fundamental reason for this heat: Second Law of Thermodynamics, according to which, every time we have an irreversible process, heat is radiated, in the amount $T \cdot S$, where S is the entropy – i.e., in this case, the number of bits in information loss; see, e.g., [12, 41]. Basic logic operations (that underlie all computations) are irreversible. For example, when $a \,\&\, b$ is false, it could be that both a and b were false, it could be that one of them was false. Thus, the usual "and"-operation $(a, b) \to a \,\&\, b$ is not reversible.

So, to decrease the amount of heat, a natural idea is to use only reversible operations.

How operations are made reversible now? At present, in quantum (and reversible) computing, a bit-valued function $y = f(x_1, \ldots, x_n)$ is transformed into the following reversible operation:

$$T_f : (x_1, \ldots, x_n, x_0) \to (x_1, \ldots, x_n, x_0 \oplus f(x_1, \ldots, x_n)),$$

where x_0 is an auxiliary bit-valued variable, and $\oplus$ denotes "exclusive or", i.e., addition modulo 2; see, e.g., [33].

It is easy to see that the above operation is indeed reversible: indeed, if we apply it twice, we get the same input back:

$$T_f(T_f(x_1, \ldots, x_n, x_0)) = T_f(x_1, \ldots, x_n, x_0 \oplus f(x_1, \ldots, x_n)) =$$

$$(x_1, \ldots, x_n, x_0 \oplus f(x_1, \ldots, x_n) \oplus f(x_1, \ldots, x_n)).$$

For addition modulo 2, $a \oplus a = 0$ for all a, so indeed

$$x_0 \oplus f(x_1, \ldots, x_n) \oplus f(x_1, \ldots, x_n) = x_0 \oplus (f(x_1, \ldots, x_n) \oplus f(x_1, \ldots, x_n)) = x_0$$

and thus,

$$T_f(T_f(x_1, \ldots, x_n, x_0)) = (x_1, \ldots, x_n, x_0).$$

Limitations of the current reversible representation of functions. The main limitation of the above representation is related to the fact that we rarely write algorithms "from scratch", we usually use existing algorithms as building blocks.

For example, when we write a program for performing operations involving sines and cosines (e.g., a program for Fourier Transform), we do not write a new code for sines and cosines from scratch, we use standard algorithms for computing these trigonometric functions – algorithms contained in the corresponding compiler. Similarly, if in the process of solving a complex system of nonlinear equations, we need to solve an auxiliary system of linear equations, we usually do not write our own code for this task – we use existing efficient linear-system packages. In mathematical terms, we form the desired function as a composition of several existing functions.

From this viewpoint, if we want to make a complex algorithm – that consists of several moduli – reversible, it is desirable to be able to transform the reversible versions of these moduli into a reversible version of the whole algorithm. In other words, it is desirable to generate a reversible version of each function so that composition of functions would be transformed into composition. Unfortunately, this is not the case with with the existing scheme described above. Indeed, even in the simple case when we consider the composition $f(f(x_1))$ of the same function $f(x_1)$ of one variable, by applying the above transformation twice, we get – as we have shown – the same input x_1 back, and *not* the desired value $f(f(x_1))$.

Thus, if we use the currently used methodology to design a reversible version of a modularized algorithm, we cannot use the modular stricture, we have, in effect, to rewrite

the algorithm from scratch. This is not a very efficient idea.

Resulting challenge, and what we do in this chapter. The above limitation shows that there is a need to come up with a different way of making a function reversible, a way that would transform composition into composition. This way, we will have a more efficient way of making computations reversible.

This is exactly what we do in this chapter.

5.2 Analysis of the Problem and the Resulting Recommendation

Simplest case: description. Let us start with the simplest case of numerical algorithms, when we have a single real-valued input x and a single real-valued output y. Let us denote the corresponding transformation by $f(x)$.

In general, this transformation is not reversible. So, to make it reversible, we need to consider an auxiliary input variable u – and, correspondingly, an auxiliary output variable v which depends, in general, on x and u: $v = v_f(x, u)$. The resulting transformation $(x, u) \to (f(x), v_f(x, u))$ should be reversible.

How to make sure that composition is transformed into composition. Let us fix some value of the auxiliary variable u that we will use, e.g., the value $u = 0$. We want to make sure that when $x = 0$, then in the resulting pair (y, v), the second value v is also 0, i.e., that $v_f(x, 0) = 0$. This way, $(x, 0)$ is transformed into $(x', 0) = (f(x), 0)$. So, if after this, we apply the reversible analogue of $g(x)$, we get $(g(x'), 0) = (g(f(x)), 0)$.

What does "reversible" mean here? In the computer, real numbers are represented with some accuracy ε. Because of this, there are finitely many possible computer representations of real numbers.

Reversibility means that inputs and outputs are in 1-1 correspondence, and thus, for each 2-D region r, its image after the transformation $(x, u) \to (y, v)$ should contain exactly

as many pairs as the original region r.

Each pair (x, u) of computer-representable real numbers takes the area of ε^2 in the (x, u)-plane. In each region of this plane, the number of possible computer-representable numbers is therefore proportional of the area of this region. Thus, reversibility means that the transformation $(x, u) \to (f(x), v(x, u))$ should preserve the area.

From calculus, it is known that, in general, under a transformation

$$(x_1, \ldots, x_n) \to (f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n)),$$

the n-dimensional volume is multiplied by the determinant of the matrix with elements $\dfrac{\partial f_i}{\partial x_j}$. Thus, reversibility means that this determinant should be equal to 1.

Let us go back to our simple case. For the transformation $(x, u) \to (f(x), v_f(x, u))$, the matrix of the partial derivatives has the form

$$\begin{pmatrix} f'(x) & 0 \\ \dfrac{\partial v_f}{\partial x} & \dfrac{\partial v_f}{\partial u} \end{pmatrix},$$

where, as usual, $f'(x)$ denoted the derivative. Thus, equating the determinant of this matrix to 1 leads to the following formula

$$f'(x) \cdot \frac{\partial v_f}{\partial u} = 1,$$

from which we conclude that

$$\frac{\partial v_f}{\partial u} = \frac{1}{f'(x)}.$$

Thus,

$$v_f(x, U) = v_f(x, 0) + \int_0^U \frac{\partial v_f}{\partial u}\, du = v_f(x, 0) + \int_0^U \frac{1}{f'(x)}\, du =$$

$$v_f(x, 0) + \frac{U}{f'(x)}.$$

We know that $v_f(x, 0) = 0$, thus we have

$$v_f(x, u) = \frac{u}{f'(x)},$$

and the transformation takes the form

$$(x, u) \to \left(f(x), \frac{u}{f'(x)} \right).$$

Examples.

- For $f(x) = \exp(x)$, we have $f'(x) = \exp(x)$ and thus, the reversible analogue is $(x, u) \to (\exp(x), u \cdot \exp(-x))$.
- For $f(x) = \ln(x)$, we have $f'(x) = 1/x$ and thus, the reversible analogue is $(x, u) \to (x, u \cdot x)$.

Comment. The above formula cannot be directly applied when $f'(x) = 0$. In this case, since anyway, we consider all the numbers modulo the "machine zero" ε – the smallest positive number representable in a computer – we can replace $f'(x)$ with the machine zero.

General case. Similarly, if we have a general transformation

$$(x_1, \ldots, x_n) \to f(x_1, \ldots, x_n) \stackrel{\text{def}}{=} (f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n)),$$

we want to add an auxiliary variable u and consider a transformation

$$(x_1, \ldots, x_n, u) \to (f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n), v_f(x_1, \ldots, x_n, u)).$$

To make sure that composition is preserved, we should take $v_f(x_1, \ldots, x_n, 0) = 0$. Thus, from the requirement that the volume is preserved, we conclude that

$$v_f(x_1, \ldots, x_n, u) = \frac{u}{\det \left\| \dfrac{\partial f_i}{\partial x_j} \right\|}.$$

Resulting recommendation. To make the transformation

$$(x_1, \ldots, x_n) \to (f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n))$$

reversible, we should consider the the following mapping:

$$(x_1, \ldots, x_n, u) \to \left(f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n), \frac{u}{\det \left\| \dfrac{\partial f_i}{\partial x_j} \right\|} \right).$$

5.3 Discussion

Need to consider floating-point numbers. In the previous text, we considered only fixed-point real numbers, for which the approximation accuracy ε – the upper bound on the difference between the actual number and its computer representation – is the same for all possible values x_i.

In some computations, however, we need to use floating-point numbers, in which instead of directly representing each number as a binary fraction, we, crudely speaking, represent its logarithm: e.g., in the decimal case, 1 000 000 000 is represented as 10^9, where 9 is the decimal logarithm of the original number. In this case, we represent all these logarithms with the same accuracy ε. In this case, the volume should be preserved for the transformation of logarithms $\ln(x_i)$ into logarithms $\ln(f_j)$, for which

$$\frac{\partial \ln(f_i)}{\partial \ln(x_j)} = \frac{x_j}{f_i} \cdot \frac{\partial f_i}{\partial x_j}.$$

In this case, formulas similar to the 1-D case imply that the resulting reversible version has the form

$$(x_1, \ldots, x_n, u) \to \left(f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n), \frac{u}{\det \left\| \dfrac{x_j}{f_i} \cdot \dfrac{\partial f_i}{\partial x_j} \right\|} \right).$$

In some cases, the input is a fixed-point number while the output is a floating point number; this happens, e.g., for $f(x) = \exp(x)$ when the input x is sufficiently large. In this case, we need to consider the dependence of $\ln(f)$ of x.

Case of functions of two variables. If we are interested in a single function of two variables $f(x_1, x_2)$, then it makes sense not to add an extra input, only an extra output,

i.e., to consider a mapping $(x_1, x_2) \to (f(x_1, x_2), g(x_1, x_2))$, for an appropriate function $g(x_1, x_2)$.

The condition that the volume is preserved under this transformation means that

$$\frac{\partial f}{\partial x_1} \cdot \frac{\partial g}{\partial x_2} - \frac{\partial f}{\partial x_2} \cdot \frac{\partial g}{\partial x_1} = 1.$$

For example, for $f(x_1, x_2) = x_1 + x_2$, we get the condition

$$\frac{\partial g}{\partial x_2} - \frac{\partial g}{\partial x_1} = 1.$$

This expression can be simplified if, instead of the original variables x_1 and x_2, we use new variables $u_1 = x_1 - x_2$ and $u_2 = x_1 + x_2$ for which $x_1 = \dfrac{u_1 + u_2}{2}$ and $x_2 = \dfrac{u_2 - u_1}{2}$. In terms of the new variables, the original function $g(x_1, x_2)$ has the form

$$G(u_1, u_2) = f\left(\frac{u_1 + u_2}{2}, \frac{u_2 - u_1}{2}\right).$$

For this new function,

$$\frac{\partial G}{\partial u_1} = \frac{1}{2} \cdot \frac{\partial g}{\partial x_1} - \frac{1}{2} \cdot \frac{\partial g}{\partial x_2} = -\frac{1}{2}.$$

Thus,

$$G(u_1, u_2) = -\frac{1}{2} \cdot u_1 + C(u_2)$$

for some function $C(u_2)$, i.e., substituting the expressions for u_i,

$$g(x_1, x_2) = \frac{x_2 - x_1}{2} + C(x_1 + x_2).$$

So, to make addition reversible, we may want to have subtraction – the operation inverse to addition; this make intuitive sense.

Similarly, for $f(x_1, x_2) = x_1 \cdot x_2$, we get the condition

$$x_2 \cdot \frac{\partial g}{\partial x_2} - x_1 \cdot \frac{\partial g}{\partial x_1} = 1.$$

This expression can be simplified if we realize that $x_i \cdot \dfrac{\partial f}{\partial x_i} = \dfrac{\partial f}{\partial X_i}$, where we denoted $X_i \stackrel{\text{def}}{=} \ln(x_i)$. In these terms, we have

$$\frac{\partial g}{\partial X_2} - \frac{\partial g}{\partial X_1} = 1,$$

and thus, as in the sum example, we get

$$g(X_1, X_2) = \frac{X_2 - X_1}{2} + C(X_1 + X_2).$$

Thus, we get

$$g(x_1, x_2) = \frac{\ln(x_2) - \ln(x_1)}{2} + C(\ln(x_1) + \ln(x_2)),$$

i.e.,

$$f(x_1, x_2) = \frac{1}{2} \cdot \ln\left(\frac{x_2}{x_1}\right) + C(x_1 \cdot x_2).$$

So, to make multiplication reversible, we need to add a (function of) division – the operation inverse to multiplication. This also makes common sense.

Chapter 6

First Application to Finance: Is "No Trade Theorem" Really a Paradox

One of the challenges in foundations of finance is the so-called "no trade theorem" paradox: if an expert trader wants to sell some stock, that means that this trader believes that this stock will go down; however, the very fact that another expert trader is willing to buy it means that this other expert believes that the stock will go up. The fact that equally good experts have different beliefs should dissuade the first expert from selling – and thus, trades should be very rare. However, in reality, trades are ubiquitous. In this chapter, we show that a detailed application of decision theory solves this paradox and explains how a trade can be beneficial to both seller and buyer. This application also explains a known psychological fact – that depressed people are usually more risk-averse.

Results from this chapter first appeared in [8].

6.1 Formulation of the Problem

"No trade theorem" paradox. When a bank or a hedge fund wants to buy a stock, this means that professionals running this financial institution believe that, in the future, this stock will increase in price. This makes perfect sense until we realize that for this institution to be able to buy this stock, some other institution needs to be willing to sell it at this price – which means that professionals running that other instutition must believe that, in the future, this stock will decrease in price.

Stock market is not a game for amateurs, serious agents buying and selling stock are

smart experts who know what they are doing and who, in the past, have shown a good intuition about future stock values. So, even when such an expert initially thinks that this stock will increase in price, the very fact that this stock is available for sale means that another expert has an exactly opposite belief. This should, in many cases, dissuade the first expert from his or her original belief.

Similarly, an expert who is initially eager to sell, i.e., who initially believes that this stock will decrease in price, should be dissuaded by the presence of similarly qualified experts who are willing to buy, i.e., who believe that this stock will increase in price.

If we follow this logic, then very few agents will be trading stocks – but in reality, the trading volume is very high, every second, a huge amount of stocks change hands. This paradoxical behavior is known as a "no trade theorem"; see, e.g., [27, 40].

What we do in this chapter. In this chapter, we use decision making to show that in reality, trading makes perfect sense if we take into account different risks associated with different stocks.

Our explanation also explains another empirical phenomenon, a phenomenon from psychology – that depressed people are more risk-averse.

6.2 Analysis of the Problem and the Resulting Explanation of the "No Trade Theorem" Paradox

Towards formulation of the problem in precise terms. Let us assume that the person originally had the amount M of money. This person is thinking of possibly buying a stock which costs s.

Let us also assume an ideal situation, in which everyone has the same information about the future value of this stock, namely, everyone knows the probability distribution of its next year's gain. In particular, everyone knows the mean m' and the standard deviation σ' of this future gain. After discounting, we get the mean $m = q \cdot m'$ and the standard

deviation $\sigma = q \cdot \sigma'$ of the equivalent current gain.

Let us also make a realistic assumption that the price s, the mean m, and the standard deviation σ are much smaller than the current money amount M. In other words, we assume that we are talking about a usual trade, not about extreme situations in which a person gambles his or her whole fortune by investing it all in a seemingly attractive stock.

Analysis of the problem. Let us denote the difference between the actual (discounted) value v of the stock and its mean value m by $\Delta v \stackrel{\text{def}}{=} v - m$. By definition of the mean, we have $m = E[v]$, where $E[\cdot]$ denoted the mean value. Thus, the mean value of Δv is 0: $E[\Delta v] = 0$.

The mean value of $(\Delta v)^2$ is, by definition, equal to σ^2: $E\left[(\Delta v)^2\right] = \sigma^2$.

In these terms, the (discounted) future gain is equal to $v = m + \Delta v$. The discounted future amount of money can be obtained if we take the original amount M, subtract the cost s of the stock, and add the gained value $v = m + \Delta v$; as a result, we get the value $M - s + m + \Delta v$.

The utility is proportional to the square root of money. We can always select a unit of utility so that utility will be exactly equal to the square root of money. In this case, the original utility is $u = \sqrt{M}$, and the discounted future utility corresponding to buying a stock is equal to $u = \sqrt{M - s + m + \Delta v}$.

As we have mentioned in the previous section, a rational person should select the alternative with the largest possible value of expected utility $E[u]$. Thus, for the agent, it makes sense to buy the stock if

$$E[u] = E\left[\sqrt{M - s + m + \Delta v}\right] > \sqrt{M}.$$

If we have a reverse inequality, then, as one can easily see, it is beneficial for this person to sell this stock. So, to decide whether it is beneficial for a person to sell or buy the stock, we need to estimate the value $E[u] = E\left[\sqrt{M - s + m + \Delta v}\right]$ of the expected utility.

Estimating the value of the expected utility. We assumed that the values m, s, and σ are much smaller than M. Thus, the corresponding random value Δv is also much smaller

than M. So, we can expand the expression $\sqrt{M-s+m+\Delta v}$ in Taylor series in terms of s, m, and Δv, and keep only linear and quadratic terms in this expansion. As a result, we get the following expression:

$$\sqrt{M-s+m+\Delta v}=\sqrt{M}+\frac{1}{2\sqrt{M}}\cdot(-s+m+\Delta v)-\frac{1}{4\cdot M^{3/2}}\cdot(-s+m+\Delta v)^2.$$

If we open the parentheses and take into account that $E[\Delta v]=0$ and $E\left[(\Delta v)^2\right]=\sigma^2$, we conclude that the expected utility of buying the stock is equal to

$$E[u]=\sqrt{M}+\frac{1}{2\sqrt{M}}\cdot(m-s)-\frac{1}{4\cdot M^{3/2}}\cdot((m-s)^2+\sigma^2).$$

Thus, this value is larger than the original utility $\sqrt{M}$ if and only if $E[u]-\sqrt{M}>0$, i.e., if and only if

$$\frac{1}{2\sqrt{M}}\cdot(m-s)-\frac{1}{4\cdot M^{3/2}}\cdot((m-s)^2+\sigma^2)>0.$$

Multiplying both sides by $4\cdot M^{3/2}$, we get an equivalent inequality

$$2M\cdot(m-s)-((m-s)^2+\sigma^2)>0,$$

i.e., equivalently, $2M\cdot(m-s)>((m-s)^2+\sigma^2)$ and

$$M>M_0\stackrel{\text{def}}{=}\frac{(m-s)^2+\sigma^2}{2(m-s)}. \tag{6.1}$$

This explains the "no trade theorem" paradox. For the same stock with the same information about its future gains, whether it is beneficial to buy it or sell it depends on the initial amount of money that a trader has:

- if the trader has a large amount of money M, then buying a stock whose expected benefits m exceed the buying cost s makes perfect sense, even when the risk σ is reasonably high;
- on the other hand, if the trader has a not so large amount of money and the stock is risky, then for this trader, it makes sense to sell this stock.

For this stock, for almost all traders (with a rare exception of a trader whose current amount is exactly M_0), it is either beneficial to buy (if $M > M_0$) or to sell (if $M < M_0$).

Thus, for the same stock, with the same information, we always have many traders for whom it is beneficial to buy, and we have many traders for whom it is beneficial to sell. This explains the ubiquity of trading.

6.3 Auxiliary Result: Decision Theory Explains Why Depressed People Are More Risk-Averse

Empirical fact. It has been observed that depressed people are more risk-averse, i.e., they are less willing to make decisions involving risks; see, e.g., [23, 29].

Our explanation. Each risky decision is described by the same formulas as a particular case of buying-a-stock risky decision: we may gain something, we may lose something, all we know is the probability distribution of the corresponding gains and losses.

Thus, to decide when it is beneficial to participate in a risky activity, we can use the same formulas as above – the only difference is that instead of just money amount M and the corresponding initial utility $u_0 = \sqrt{M}$, we can take into account different things that affect the person's utility. In terms of utility u_0, the inequality (6.1) – that describes when it is beneficial for a person to engage in a risky behavior – takes the form

$$u_0^2 > M_0 \stackrel{\text{def}}{=} \frac{(m-s)^2 + \sigma^2}{2(m-s)}. \tag{6.2}$$

This formula says that when the initial value of the utility u_0 is small, risky behavior – with large σ – is not beneficial. And this is exactly what depression means in decision-theoretic terms: that a person is not very happy, i.e., that the corresponding utility value u_0 is small.

Thus, our decision-theoretic analysis explains the above-mentioned psychological phenomenon.

Chapter 7

Second Application to Finance: Decision Theory Can Explain Why Buying and Selling Prices Are Different

7.1 Buying and Selling Prices Are Different: a Phenomenon and Its Current Quantitative Explanations

Buying and selling prices are different: a phenomenon. According to the naive understanding of economic behavior, we should decide, for ourselves, how much each object is worth to us. This worth amount should be the largest amount that we should be willing to pay if we are buying this object, and this same amount should be the smallest amount for which we should agree to sell this objects.

However, in many experiments, the price participants are willing to pay to buy a certain item is different from the price they are willing to accept to part with this item. For example, students are willing to pay $3 for a mug but require to be paid at least $7 to sell it back. In other words, people estimate the consequences of losing an object differently than the consequences of gaining the same object; see, e.g., [20, 40] and references therein.

Current explanations of this phenomenon. The current explanation of this phenomenon is based on the fact that people are not clear on the value of each object. Instead of the exact monetary amount, at best, they have a range $[\underline{u}, \overline{u}]$ of possible values of this object's worth; see, e.g., [14, 15].

Need for a more detailed analysis. While [14, 15] provide a qualitative explanation for the loss aversion phenomenon, it is desirable to extend this to a quantitative analysis, an analysis that takes into account known results about rational decision making under interval uncertainty. This is what we do in this chapter.

Results from this chapter will appear in [10].

7.2 Quantitative Explanation

Decision making under interval uncertainty: case of monetary values. How can we make decision if, instead of the exact value of an object, we only know the interval $[\underline{u}, \overline{u}]$ of possible values? In other words, what is the value $u(\underline{u}, \overline{u})$ that we are willing to pay for this object?

Clearly, since we know that the object is worth at least $\underline{u}$ and at most $\overline{u}$, this means that the price $u(\underline{u}, \overline{u})$ that we are willing to pay should also be at least $\underline{u}$ and at most $\overline{u}$:

$$\underline{u} \le u(\underline{u}, \overline{u}) \le \overline{u}. \tag{7.1}$$

This property is known as *boundedness*.

Another reasonable requirement is that if we have two different objects, with values in $[\underline{u}, \overline{u}]$ and $[\underline{v}, \overline{v}]$, then the price that we are willing to pay to buy both should be equal to the prices that we pay for each of them. Let us describe this second requirement in precise terms.

When we get two objects together, the smallest possible value of our purchase is when both objects are worth their smallest amounts $\underline{u}$ and $\underline{v}$. In this case, the overall worth of both objects is equal to the sum $\underline{u} + \underline{v}$. Similarly, the largest possible value of our purchase

is when both objects are worth their largest amounts $\overline{u}$ and $\overline{v}$. In this case, the overall worth of both objects is equal to the sum $\overline{u} + \overline{v}$. Thus, for two objects sold together the interval of possible worth values is $[\underline{u} + \underline{v}, \overline{u}, \overline{u} + \overline{v}]$. So, the second requirement takes the following form:

$$u(\underline{u} + \underline{v}, \overline{u} + \overline{v}) = u(\underline{u}, \overline{u}) + u(\underline{v}, \overline{v}). \tag{7.2}$$

This property is known as *additivity*.

It turns out (see, e.g., [22]) that the only functions that satisfy both requirements (7.1) and (7.2) are functions of the type

$$u(\underline{u}, \overline{u}) = \alpha_H \cdot \overline{u} + (1 - \alpha_H) \cdot \underline{u}, \tag{7.3}$$

for some $\alpha_H \in [0, 1]$. This fact easily follows from the fact that all bounded additive functions are linear; see, e.g., [1].

As we have mentioned in Chapter 2, the formula (7.3) was first proposed – for the case of utilities – by a future Nobelist Leo Hurwicz and is thus known as Hurwicz optimism-pessimism criterion [17, 25]. We have just shown that a similar formula can be used to estimate monetary value under interval uncertainty.

Hurwicz criterion explains the difference between buy and sell prices. When we buy an object whose worth is between $\underline{u}$ and $\overline{u}$, the best possible gain is $\overline{u}$ and the worst possible gain is $\underline{u}$. Thus, according to the Hurwicz criterion, we should be willing to pay the amount u_b (b for *buy*) which is equal to

$$u_b = \alpha_H \cdot \overline{u} + (1 - \alpha_H) \cdot \underline{u}. \tag{7.4}$$

On the other hand, if we already own this object and we sell it, then our loss is between $-\overline{u}$ and $-\underline{u}$. The most optimistic estimate for our resulting state is $-\underline{u}$ and the most pessimistic estimate is $-\overline{u}$. In this case, according to the Hurwicz criterion, this is equivalent to the value of

$$\alpha_H \cdot (-\underline{u}) + (1 - \alpha_H) \cdot (-\overline{u}). \tag{7.5}$$

Thus, to compensate for this loss, we need to get the amount u_s (s for *sell*) that, when added to the value (7.5), will result in 0, i.e., the value

$$u_s = \alpha_H \cdot \underline{u} + (1 - \alpha_H) \cdot \overline{u}. \tag{7.6}$$

We can see that, in general, the expressions for the buy u_b and sell u_s prices are different. Indeed, the only time when the prices are equal, i.e., when $u_b = u_s$, is when

$$\alpha_H \cdot \underline{u} + (1 - \alpha_H) \cdot \overline{u} = \alpha_H \cdot \overline{u} + (1 - \alpha_H) \cdot \underline{u}.$$

Moving all the terms to the left-hand side and adding resulting coefficients at $\overline{u}$ and $\underline{u}$, we conclude that

$$(2\alpha_H - 1) \cdot \underline{u} - (2\alpha_H - 1) \cdot \overline{u} = 0,$$

i.e., $(2\alpha_H - 1) \cdot (\underline{u} - \overline{u}) = 0$. Since we consider the case when we have uncertainty, i.e., when $\underline{u} \neq \overline{u}$, we thus conclude that $2\alpha_H - 1 = 0$, i.e., that $\alpha_H = 0.5$.

So, only people with $\alpha_H = 0.5$ buy and sell at exactly the same price. For everyone else – who is even slightly more optimistic or even slightly less optimistic than $\alpha_H = 0.5$ – the buy and sell prices are different, and this is exactly what we observe.

Chapter 8

Decision Theory Explains "Telescoping Effect" – That Our Time Perception Is Biased

The two previous chapters describe the current decision making. If we take into account events in the past – or possible future consequences – then we need to describe how people perceive the corresponding time intervals.

People usually underestimate time passed since distant events, and overestimate time passed since recent events. There are several explanations for this "telescoping effect", but most current explanations utilize specific features of human memory and/or human perception. We show that the telescoping effect can be explained on a much basic level of decision theory, without the need to invoke any specific ways we perceive and process time.

Results from this chapter will appear in [9].

8.1 Formulation of the Problem

Telescoping effect. It is known that when people estimate how long ago past events happened, their estimates are usually biased (see, e.g., [11, 18, 39]):

- for recent events, people usually *overestimate* how much time has passed since this event;
- on the other hand, for events in the more distant past, people usually *underestimate*

how much time has passed since the event.

This phenomenon is called *telescoping effect* since the bias in perceiving long-ago past events is similar to what happens when we look at the celestial objects via a telescope: all the objects appear closer than when you look at them with a naked eye.

How can this effect be explained. There are many explanations for the telescope effect [11, 18, 39], but most current explanations utilize specific features of human memory and/or human perception.

What we do in this chapter. In this chapter, we show that the telescoping effect can be explained on a much basic level of decision theory, without the need to invoke any specific ways we perceive and process time.

8.2 How Decision Theory Can Explain the Telescoping Effect

People's perceptions are imprecise. In the ideal situation, an event of utility u_0 that occurred t moments in the past should be equivalent to exactly the utility $u = q^t \cdot u_0$ now. In practice, however, people's perceptions are imprecise.

Let us describe this imprecision: first approximation. Let us denote by ε the accuracy of people's perception. Then, for an event with actual utility u, the perceived utility can differ by ε, i.e., it can take any value from the corresponding interval $[u-\varepsilon, u+\varepsilon]$. In particular, our perceived utility u of the past event can take any value from the interval $[q^t \cdot u_0 - \varepsilon, q^t \cdot u_0 + \varepsilon]$.

How we perceive events form the distant past. The above interval can be somewhat narrowed down if we take into account that for a positive event, with utility $u_0 > 0$, the perception cannot be negative, while the value $q^t \cdot u_0 - \varepsilon$ is negative for large t. Thus, when $q^t \cdot u_0 - \varepsilon < 0$, i.e., when $t > T_0 \stackrel{\text{def}}{=} \dfrac{\ln(u_0/\varepsilon)}{|\ln(q)|}$, the lower bound of the interval is 0, and thus,

the interval has the form

$$[\underline{u}, \overline{u}] = [0, q^t \cdot u_0 + \varepsilon].$$

Based on Hurwicz's optimism-pessimism criterion, this interval is equivalent to the value $\alpha_H \cdot (q^t \cdot u_0 + \varepsilon)$. How does this translate into a perceived time? For any time t_p, the utility of the event t_p moments in the past is equal to $q^{t_p} \cdot u_0$. Thus, the perceived time t_p can be found from the condition that the utility $\alpha_H \cdot (q^t + \varepsilon)$ is equal to $q^{t_p} \cdot u_0$. This equality $\alpha_H \cdot (q^t \cdot u_0 + \varepsilon) = q^{t_p} \cdot u_0$ implies that

$$t_p = \frac{\ln((\alpha_H \cdot (q^t \cdot u_0 + \varepsilon))/u_0)}{\ln(q)}.$$

In particular, when t tends to infinity, we have $q^t \to 0$ and thus, the perceived time tends to a finite constant

$$\frac{\ln((\alpha_H \cdot \varepsilon)/u_0)}{\ln(q)}.$$

Thus, for large t we indeed have $t_p \ll t$, which is exactly what we observe in the telescoping effect for events from the distant past.

How we perceive very recent events. For recent events, the interval

$$[q^t \cdot u_0 - \varepsilon, q^t \cdot u_0 + \varepsilon]$$

can also be somewhat narrowed down if we take into account that the perceived utility of a past event cannot exceed its utility now, i.e., the value u_0. Thus, when $q^t \cdot u_0 + \varepsilon > u_0$, i.e., when $q^t > 1 - \varepsilon/u_0$ and thus, $t < t_0 \stackrel{\text{def}}{=} \dfrac{\ln(1 - u_0/\varepsilon)}{\ln(q)}$, the upper bound of the interval is u_0, and thus, the interval has the form

$$[\underline{u}, \overline{u}] = [q^t \cdot u_0 - \varepsilon, u_0].$$

Based on Hurwicz's optimism-pessimism criterion, this interval is equivalent to the value $\alpha_H \cdot u_0 + (1 - \alpha_H) \cdot (q^t \cdot u_0 - \varepsilon)$. Similarly to the distant-past case, the perceived time t_p can be found from the condition that the above value is equal to $q^{t_p} \cdot u_0$, i.e., that

$$\alpha_H \cdot u_0 + (1 - \alpha_H) \cdot (q^t - \varepsilon) = q^{t_p} \cdot u_0.$$

This implies that

$$t_p = \frac{\ln(\alpha_H + (1-\alpha_H)\cdot(q^t - \varepsilon/u_0))}{\ln(q)}.$$

In particular, when t tends to 0, we have $q^t \to 1$ and thus, the perceived time t_p tends to a finite positive constant

$$\frac{\ln(\alpha_H + (1-\alpha_H)\cdot(1 - \varepsilon/u_0))}{\ln(q)}.$$

Thus, for small t, we indeed have $t_p \gg t$, which is exactly what we observe in the telescoping effect for recent events.

Chapter 9

Application to Geosciences: Bhutan Landscape Anomaly Explained

Economies of countries located in seismic zones are strongly effected by this seismicity. If we underestimate the seismic activity, then a reasonably routine earthquake can severely damage the existing structures and thus, lead to huge economic losses. On the other hand, if we overestimate the seismic activity, we waste a lot of resources on unnecessarily fortifying all the buildings – and this too harms the economies. From this viewpoint, it is desirable to have estimations of regional seismic activities which are as accurate as possible. Current predictions are mostly based on the standard geophysical understanding of earthquakes as being largely caused by the movement of tectonic plates and terranes. This understanding works in most areas, but in Bhutan area of the Himalayas region, there seems to be a landscape anomaly. As a result, for this region, we have less confidence in the accuracy of seismic predictions based on the standard understanding and thus, have to use higher seismic thresholds in construction. In this chapter, we find the optimal description of landscape-describing elevation profiles, and we use this description to show that the seeming anomaly is actually in perfect agreement with the standard understanding of the seismic activity. Our conclusion is that it is safe to apply, in this region, estimates based on the standard understanding and thus, avoid unnecessary expenses caused by an increased threshold.

Results from this chapter will appear in [32].

9.1 Formulation of the Problem

Seismicity affects economy. In highly seismic areas like the Himalayas, economy is affected by our knowledge of possible seismicity.

Protection against possible earthquakes is very costly. If we have only a vague idea about possible seismic events – i.e., if we can potentially expect high-energy earthquakes at all possible locations – then, every time we build a house or a factory, we need to spend a lot of money on making it protected against such events – with little money left for any other development project.

On the other hand, if we can reasonably accurately localize potential hazards, then we can concentrate our building efforts mostly in safer zones. This will require less investment in earthquake protection and thus, leave more money for other development projects.

Thus, the economy of a highly seismic zone is directly affected by our understanding of the corresponding seismic processes.

Bhutan landscape anomaly. In general, modern geophysics has a reasonably good understanding of seismic processes and seismic zones. Specifically, the current understanding is that seismicity is usually caused by mutual movement of tectonic plates and their parts (terranes), and it is mostly concentrated on the borderline between two or more such plates or terranes. In general, while we still cannot predict the exact timing of earthquakes, geoscientists can reasonably well predict the size of a future earthquake based on the corresponding geophysical models.

Researchers and practitioners are reasonably confident in these predictions – at least for locations whose geophysics is well understood by the traditional geophysical models.

However, there are locations where observed phenomena are different from what we usually expect. In such cases, there are reasonable doubts in seismicity estimates produced by the traditional techniques – and thus, it is reasonable to be cautious and use higher strengths of potential earthquakes when building in these locations, which invokes significant additional expenses. For such domains, it is therefore desirable to come up with a

better understanding of the observed geophysical phenomena – thus hopefully allowing us to make more accurate predictions and hence, save money (which is now wasted on possibly too-heavy earthquake protection) for other important activities.

One such areas in the vicinity of the Himalayan country of Bhutan, where the landscape profile is drastically different from the profiles of other Himalayan areas such as areas of Nepal. In general, a landscape can be described in numerical terms if we take a line orthogonal to the prevailing rivers (which are usually the lowest points on the landscape) and plot the elevation as a function of the distance from the corresponding river. The shape of the landscape (elevation) profile in Bhutan is visually drastically different from the landscape profile in Nepal; see, e.g., [2]. Namely, in most of the Himalayas – and, in general, in the most of the world – the corresponding curve is first convex (corresponding to the river valley), and then becomes concave – which corresponds to the mountain peaks. In contrast, in Bhutan, the profile turns concave very fast, way before we reach the mountain peaks area.

As of now, there are no good well-accepted explanations for this phenomenon – which makes it an anomaly. To be more precise, we know that the geophysics of the Bhutan area is somewhat different: in Nepal (like in most areas in the world), the advancing tectonic plate in orthogonal to the border of the mountain range, while in Bhutan, the plate pushes the range at an angle. However, it is not clear how this can explain the above phenomenon. This leads us to the following questions.

Questions. The first question is: can we explain the Bhutan anomaly within the existing geophysical paradigm? If we can do, this would mean that this anomaly is not an obstacle to applying this paradigm, and thus, that the estimates of future seismic activity obtained within this paradigm can be safely applied – without the need to make expensive extra precautions.

A related question is related to the fact that while we use convexity and concavity to describe elevation profiles, the only reason for using these two properties is because these are the basic properties that we learn in math. Is there any geophysical meaning in

convexity vs. concavity?

What we do in this chapter. In this chapter, we provide answers to both questions: we explain why convexity and concavity are adequate ways to describe elevation profiles, and we explain how the at-an-angle pressure in the Bhutan area leads to the observed convex-followed-by-concave phenomenon.

To answer these questions, we first formulate the problem of adequately describing elevation profiles as an optimization problem. Then, we solve this problem, and use the solution to answer the above two questions.

9.2 What Is the Optimal Description of Elevation Profiles

How can we describe elevation profiles? An elevation profile results from the joint effect of many different physical processes, from movement of tectonic plates to erosion. These process are largely independent from each other: e.g., erosion works the same way whether we have the landscape on the sea level or the same landscape which the geological processes raised to some elevation. Because of this independence, the observed profile $f(x)$ can be reasonably well represented as the sum of profiles corresponding to different processes:

$$f(x) = f_1(x) + \ldots + f_n(x).$$

Different profile-changing processes may have different intensity. So, to describe the effect of the i-th process, instead of a fixed function $f_i(x)$, it is more appropriate to use the correspondingly re-scaled term $C_i \cdot f_i(x)$, where the coefficients C_i describe the intensity of the i-th process, so that

$$f(x) = C_1 \cdot f_1(x) + \ldots + C_n \cdot f_n(x).$$

Due to erosion, discontinuities in the elevation profiles are usually smoothed out, so we can safely assume that the corresponding functions $f_i(x)$ are smooth (differentiable).

For such families, the problem of selecting the optimal description was formulated and solved in Chapter 3. The result was that the optimal approximating family is the family of polynomials.

9.3 Why Convexity and Concavity Are Important in Elevation Profiles: An Explanation Based on the Optimality Result

Discussion. The above result provides us, for different n, with families of approximations to the elevation profiles. Let us start with the simplest possible approximation.

For $n = 1$, we get the class of constant functions – no landscape at all. For $n = 2$, we get a class of linear functions – no mountains, no ravines, just a flat inclined surface. So, the only non-trivial description of a landscape starts with $n = 3$, i.e., with quadratic functions.

We want to provide a qualitative classification of all such possible elevation functions. It is reasonable to say that the two elevation functions are equivalent if they differ only by re-scaling and shift of x and y:

Definition 9.1. *We say that two quadratic functions $f(x)$ and $g(x)$ are* equivalent *if for some values $\lambda_x > 0$, $\lambda_y > 0$, x_0, and y_0, we have*

$$g(x) = \lambda_y \cdot f(\lambda_x \cdot x + x_0) + y_0$$

for all x.

Proposition 9.1. *Every non-linear quadratic function is equivalent either to x^2 or to $-x^2$.*

Discussion. Thus, in this approximation, we have, in effect, two shapes: the shape corresponding to x^2 (convex) and the shape corresponding to $-x^2$ (concave). This result

explains why our visual classification into convex and concave shapes makes perfect sense.

Proof. Every non-linear quadratic function $g(x)$ has the form

$$g(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2,$$

for some $a_2 \neq 0$.

If $a_2 > 0$, then this function can be represented as

$$a_2 \cdot \left(x + \frac{a_1}{2a_2} \right)^2 + \left(a_0 - \frac{a_1^2}{4a_2} \right),$$

i.e., can be represented in the desired form, with $f(x) = x^2$, $\lambda_x = 1$, $\lambda_y = a_2$, $x_0 = \dfrac{a_1}{2a_2}$, and $y_0 = a_0 - \dfrac{a_1^2}{4a_2}$.

If $a_2 < 0$, then this function can be represented as

$$|a_2| \cdot \left(- \left(x + \frac{a_1}{2a_2} \right)^2 \right) + \left(a_0 - \frac{a_1^2}{4a_2} \right),$$

i.e., can be represented in the desired form, with $f(x) = -x^2$, $\lambda_x = 1$, $\lambda_y = |a_2|$, $x_0 = \dfrac{a_1}{2a_2}$, and $y_0 = a_0 - \dfrac{a_1^2}{4a_2}$.

The proposition is proven.

9.4 Bhutan Anomaly Explained

Discussion. In the previous text, we have shown that the optimal description of an elevation profiles is by polynomials of a fixed degree.

In the first approximation, a landscape profile can be described by a quadratic function. To get a more accurate description, let us also consider cubic terms, i.e., let us consider profiles of the type

$$f(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3. \tag{9.1}$$

As a starting point $x = 0$ for the elevation profile, it makes sense to select the lowest (or the highest) point. In both cases, according to calculus, the first derivative of the elevation profile is equal to 0 at this point: $f'(0) = 0$. Substituting the above expression for $f(x)$ into this formula, we conclude that $a_1 = 0$ and thus,

$$f(x) = a_0 + a_2 \cdot x^2 + a_3 \cdot x^3. \tag{9.2}$$

Let us analyze how this approximation works for the above two cases: the case of Nepal and the case of Bhutan.

Case of Nepal. In the case of Nepal, the forces compressing the upper plate are orthogonal to the line of contact. This means that in this case, the forces do not change if we change left to right and right to left.

Since the whole mountain range was created by this force, it is reasonable to conclude that the corresponding elevation profile is also invariant with respect to swapping left and right, i.e., with respect to the transformation $x \to -x$:

$$f(x) = f(-x). \tag{9.3}$$

Substituting the cubic expression (9.2) for the profile $f(x)$ into this formula, we conclude that $a_3 = 0$. Thus, in this case, the elevation profile is quadratic even in this next approximation – and is, therefore, either convex or concave.

Case of Bhutan. In the case of Bhutan, the force is applied at an angle. Here, there is no symmetry with respect to $x \to -x$, so, in general, we have $a_3 \neq 0$. Thus, the second derivative – that describes whether a function is locally convex (when this second derivative is positive) or locally concave (when the derivative is negative) – becomes a linear function $6a_3 \cdot x + 2a_2$, with $a_3 \neq 0$.

A non-constant linear function always changes signs – this explains why in the case of Bhutan, convexity follows by concavity.

9.5 Auxiliary Question: How to Best Locate an Inflection Point

Practical problem. Many geophysical ideas are applicable only to valley-type convex domains or only to mountain-type concave domains. So, to apply these ideas to a real-life landscape, it is necessary to divide the whole landscape into convex and concave zones. What is the best way to do it? In other words, what is the best way to locate an *inflection point*, i.e., the point at which local convexity changes to local concavity?

First idea: a straightforward least squares approach. The first natural idea – motivated by the above analysis – is to approximate the actual elevation profile by a cubic function (9.1). The corresponding coefficients c_0, c_1, c_2, and c_3 can be obtained, e.g., by applying the least squares method to the corresponding system of linear equations

$$y_i \approx c_0 + c_1 \cdot x_i + c_2 \cdot x_i^2 + c_3 \cdot x_i^3,$$

where x_i is the i-th location and y_i is the i-th elevation.

The least squares method minimizes the sum

$$\sum_i (y_i - (c_0 + c_1 \cdot x_i + c_2 \cdot x_i^2 + c_3 \cdot x_i^3))^2.$$

Differentiating this expression with respect to each of the unknowns c_j and equating all four derivatives to 0, we get an easy-to-solve system of four linear equations with four unknowns.

Once we find the characteristics, we then estimate the location of the inflection point as the value at which the second derivative is equal to 0, i.e., the value $x_{\text{infl}} = -\dfrac{c_2}{3c_3}$.

Second idea: a model-free least squares approach. Instead of restricting ourselves to a cubic approximation, we can consider general convex functions. For a function $f(x)$ defined by its values $y_1 = f(x_1)$, $y_2 = f(x_2)$, ..., on a equally spaced grid

$$x_1, x_2 = x_1 + \Delta x, x_3 = x_1 + 2\Delta x, \ldots, x_N,$$

convexity is equivalent to the sequence of inequalities

$$y_i \le \frac{y_{i-1} + y_{i+1}}{2}. \tag{9.4}$$

For each set of actual profile points $\widetilde{y}_i$, we can therefore find the closest convex profile by looking for the values y_i that minimize the mean square error (MSE)

$$\frac{1}{N} \cdot \sum_i (\widetilde{y}_i - y_i)^2$$

under the constraints (9.4). The minimized expression is a convex function of the unknowns y_i, and each constraint – and thus, their intersection – defines a convex set. Thus, we can find the corresponding minimum by using a known algorithm for convex optimization (= minimizing a convex function on a convex domain); see, e.g., [34, 35, 37].

By applying this algorithm to actually convex profiles, we can find the largest and thus, the corresponding MSE. Let us denote the largest of such values by M. Then, to find an inflection point, we can consider larger and larger fragments of the original series $f(x_1)$, $f(x_2)$, ..., until we reach a point at which the corresponding MSE exceeds M. This is the desired inflection point.

We can speed up this algorithm if instead of slowly increasing the size of the still-convex fragment, we use bisection. Specifically, we always keep two values $\underline{p}$ and $\overline{p}$ such that the fragment until $\underline{p}$ is convex (within accuracy M), while the fragment up to the point $\overline{p}$ is not convex within the given accuracy.

In the beginning, we first apply our criterion to the whole list of N values. If the result is M-close to convex, we consider the profile convex – no inflection point here. If the result is not M-convex, then we take $\underline{p} = 1$ and $\overline{p} = N$.

Once we have two values $\underline{p} < \overline{p}$, we then take a midpoint $m \stackrel{\text{def}}{=} \dfrac{\underline{p} + \overline{p}}{2}$. If the segment up to this midpoint is M-convex, then we replace $\underline{p}$ with m. If this segment is not M-convex, we replace $\overline{p}$ with m.

In both case, we get a new interval $[\underline{p}, \overline{p}]$ whose width decreased by a factor of two. We started with width N. Thus, in $\log_2(N)$ steps, this size decreases to $N/2^{\log_2(N)} = N/N = 1$, i.e., we get the exact location of the inflection point.

Comment. Other algorithms for detecting inflection points are described, e.g., in [19, 26].

Chapter 10

Future Work Plans: In Brief

This thesis is devoted to a practically important problem of decision making under uncertainty and its applications to finances and geosciences. In this thesis, we provide our preliminary results and first applications. We plan to continue this work, both by expanding our theoretical analysis and by coming up with more examples of practical applications.

In terms of theoretical analysis, what we have done so far is based on *deterministic* decision making, when a decision maker consistently selects the same "best" alternative. In practice, human decision making is often probabilistic: in the same situation, we may select different alternatives, with different probabilities. This situation has been analyzed in decision theory by a Nobelist D. McFadden. However, his analysis assumes that we know the exact gains related to different alternatives. In practice, we usually know the expected gains only with some uncertainty. So, our main theoretical research direction would be to extend McFadden's analysis to the case of decision making under uncertainty.

This planned research is closely related to another related topic that we mentioned in this thesis: the need to speed up computations. At present, in many problems, deep machine learning is effectively used to extract dependencies from data. Interestingly, one of the important stages of deep learning – softmax – uses the exact same formulas as McFadden's description of human decision making. As a result, is suffers from the same limitation as McFadden's analysis – that it assumed that we know the exact expected gains of different possible decisions. We hope that extending softmax to a more realistic case of decision making under uncertainty will help further speed up deep learning-related computations.

We also plan to work on applications. Just like in this thesis, we plan to both work

on explaining the existing phenomena – which would provide a good validity test for our models – and on producing new recommendations. In terms of explanation, there are several directions that we plan to pursue:

- First, in addition to seemingly counterintuitive aspects of human behavior that have been explained in the past and we explain in this thesis, there are other phenomena that still need explaining. For example, an often cited phrase that giving is better than receiving definitely reflects some aspects of human behavior, but seems to be inconsistent with the usual utilitarian models of this behavior.

- Second, the Hurwicz-style analysis of decision making under uncertainty – that we use in this thesis – explains different types of people's behavior by assigning, to each person, a numerical degree of this person's optimism-pessimism. However, this usual analysis does not explain why some people are more optimistic and some are more pessimistic. It is therefore desirable to try to understand this by analyzing which type of behavior works best in different situations.

- Finally, it is desirable not just to look at the results of human decision making, but also at procedures that human use to teach their results. For example, as part of these procedures, humans perform some non-traditional approximate computations. We plan to analyze how these unusual procedures can be explained in the general framework of decision making under uncertainty.

Finally, we plan to work on applications:

- First, we plan to analyze how all this can be applied to an area with which we are all very familiar: grading of student papers. As of now, deciding which problems and which tests are worth how many points is more of an art. This problem is difficult to solve in precise terms because clearly we need to make decisions under uncertainty. We plan to see what recommendations can be extracted if we apply the general techniques of decision making under uncertainty to this problem.

- Second, we plan to apply these general techniques to robotics, to see if we can use at least of these techniques to make robots behavior more human-like and thus, more acceptable to potential users.
- Finally, we will continue to look for possible challenges and applications in geosciences.

References

[1] J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 2008.

[2] B. A. Adams, K. X. Whipple, K. V. Hodges, and A. M. Heimsath, "In situ development of high-elevation, low-relief landscapes via duplex deformation in the Eastern Himalayan hinterland, Bhutan", *Journal of Geophysical Research: Earth Surface*, 2016, Vol. 121, pp. 294–319.

[3] C. L. Adams, "Introducing roots and extrema in calculus: use cubic polynomial functions before increasing the difficulty with irrational values", *Mathematics Teacher*, 2018, Vol. 112, No. 2, pp. 132–135.

[4] C. L. Adams and J. Board, "Conditions on a coefficients of a reduced cubic polynomial such that it and its derivative are factorable over the rational numbers", *Electronic Proceedings of the 28th Annual International Conference on Technology in Collegiate Mathematics*, Atlanta, Georgia, March 10–13, 2016, pp. 33–45; available from http://archives.mat.utk.edu/ICTCM/i/28/A003.html

[5] L. Bokati, O. Kosheleva, and V. Kreinovich, "How to generate 'nice' cubic polynomials – with rational coefficients, rational zeros and rational extrema: a fast algorithm", *Abstracts of the 11th International Conference of Nepalese Student Association NeSA'11*, Las Cruces, New Mexico, March 23, 2019.

[6] L. Bokati, O. Kosheleva, and V. Kreinovich, "How to generate 'nice' cubic polynomials – with rational coefficients, rational zeros and rational extrema: a fast algorithm", *Abstracts of the NMSU/UTEP Workshop on Mathematics, Computer Science, and Computational Science*, Las Cruces, New Mexico, April 6, 2019.

[7] L. Bokati, O. Kosheleva, and V. Kreinovich, "How to generate 'nice' cubic polynomials – with rational coefficients, rational zeros and rational extrema: a fast algorithm", *Journal of Uncertain Systems*, to appear.

[8] L. Bokati and V. Kreinovich, "Is 'No Trade Theorem' Really a Paradox: Analysis Based on Decision Theory", *Applied Mathematical Sciences*, 2019, Vol. 13, No. 8, pp. 397–404.

[9] L. Bokati and V. Kreinovich, "Laxman Bokati and Vladik Kreinovich, "Decision Theory Explains 'Telescoping Effect' – That Our Time Perception Is Biased", *Journal of Uncertain Systems*, to appear.

[10] L. Bokati and V. Kreinovich,"Decision Theory Can Explain Why Buying and Selling Prices Are Different", *Journal of Uncertain Systems*, 2020, to appear.

[11] S. E. Crawley and L. Pring, "When did Mrs. Thatcher resign? The effects of ageing on the dating of public events", *Memory*, 2000, Vol. 8, No. 2, pp. 111–121.

[12] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

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

[14] D. Gal, "Selling behavioral economics", *New York Times*, October 10, 2018, International Edition, p. 10.

[15] D. Gal and D. R. Rucker, "Loss Aversion, Intellectual Inertia, and a Call for a More Contrarian Science: A Reply to Simonson & Kivetz and Higgins & Liberman", *Journal of Consumer Psychology*, 2018, to appear, available at https://ssrn.com/abstract=3127716

[16] K. B. Howell, *Ordinary Differential Equations: An Introduction to the Fundamentals*, CRC Press, Boca Raton, Florida, 2016.

[17] L. Hurwicz, *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.

[18] S. M. J. Janssen, A. G. Chessa, and J. M. J. Murre, "Memory for time: how people date events", *Memory and Cognition*, 2006, Vol. 34, No. 1, pp. 138–147.

[19] N. K. Kachouie and A. Schwartzman, "Non-parametric estimation of a single inflection point in noisy observed signal", *Journal of Electrical and Electronic Systems*, 2013, Vol. 2, No. 2, Paper 1000108.

[20] D. Kahneman, *Thinking, Fast and Slow*, Farrar, Straus, and Giroux, New York, 2011.

[21] O. Kosheleva, M. Afravi, and V. Kreinovich, "Why utility non-linearly depends on money: a commonsense explanation", *Proceedings of the 4th International Conference on Mathematical and Computer Modeling*, Omsk, Russia, November 11, 2016, pp. 13–18.

[22] V. Kreinovich, "Decision making under interval uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.

[23] Y. Leykin, C. S. Roberts, and R. J. DeRubeis, "Decision-making and depressive symptomatology", *Cognitive Therapy and Research*, 2011, Vol. 35, pp. 333–341.

[24] J. Lorkowski and V. Kreinovich, *Bounded Rationality in Decision Making Under Uncertainty: Towards Optimal Granularity*, Springer Verlag, Cham, Switzerland, 2018.

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

[26] D. Manocha and J. F. Canny, "Detecting cusps and inflection points in curves", *Computer Aided Geometric Design*, 1992, Vol. 9, No. 1, pp. 1–24.

[27] P. Milgrom and N. Stokey, "Information, trade and common knowledge", *Journal of Economic Theory*, 1982, Vol. 26, No. 1, pp. 17–27.

[28] C. D. Miller, M. L. Lial, and D. I. Schneider, *Fundamentals of College Algebra*, Scott and Foresman/Little and Brown Higher Education, 1990.

[29] F. C. Murphy, J. S. Rubinsztein, A. Michael, R. D. Rogers, T. W. Robbins, E. S. Paykel, and B. J. Sahakian, "Decision-making cognition in mania and depression", *Psychological Medicine*, 2001, Vol. 31, pp. 679–693.

[30] H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.

[31] H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.

[32] Thach Ngoc Nguyen, L. Bokati, A. Velasco, and V. Kreinovich, "Bhutan Landscape Anomaly: Possible Effect on Himalayan Economy (In View of Optimal Description of Elevation Profiles)", *Thai Journal of Mathematics*, to appear.

[33] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.

[34] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, 2006.

[35] P. Pardalos, *Complexity in Numerical Optimization*, World Scientific, Singapore, 1993.

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

[37] R. T. Rockafeller, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1997.

[38] N. P. Smart, *The Algorithmic Resolution of Diophantine Equations*, Cambridge University Press, Cambridge, UK, 1998.

[39] Wikipedia, *Telescoping effect*, https://en.wikipedia.org/wiki/Telescoping_effect, downloaded on December 29, 2018

[40] R. H. Thaler, *Misbehaving: The Making of Behavioral Economy*, W. W. Norton & Co., New York, 2015.

[41] K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.

Curriculum Vitae

Laxman Bokati was born on April 8, 1989 in Mahendranagar, a small town in far western part of Nepal. He entered Institute of Engineering Pulchowk Campus (Lalitpur, Nepal) in November 2008, and graduated four years later with a Bachelors degree in Electrical Engineering.

He came to The University of Texas at El Paso (UTEP) to pursue PhD in Computational Science in Fall of 2017. He has been working as an Teaching Assistant at The University of Texas at El Paso since September 2017.

His work titled "Is 'no trade theorem' really a paradox: analysis based on decision theory" has been published in *Applied Mathematical Sciences*, Vol. 13, 2019.

He has presented his research work in various conferences such as NMSU/UTEP Workshop on Mathematics, Computer Science, and Computational Science (Las Cruces, New Mexico), 11th International Conference of Nepalese Student Association NeSA'11 (Las Cruces, New Mexico), and Southwest Local Algebra Meeting (El Paso, Texas).

Permanent address: Katan-18, Mahendranagar, Kanchanpur, Nepal.